Evaluating Likelihood Estimation Methods in Multilevel Analysis of Modern Contraceptive Use in Nigeria

by

BAKRE, BABATUNDE BOWALE
MATRIC NO: 133749

A DISSERTATION SUBMITTED TO THE DEPARTMENT OF EPIDEMIOLOGY AND MEDICAL STATISTICS, FACULTY OF PUBLIC HEALTH, COLLEGE OF MEDICINE, UNIVERSITY OF IBADAN, IN PARTIAL FULFILLMENT OF THE REQUIREMENT FOR THE AWARD OF A MASTER OF SCIENCE (M.SC.)DEGREE IN BIOSTATISTICS.

FEBRUARY, 2015

Certification

I certify that this project work was carried out under my supervision by. Bakre, Babatunde Bowale, of the Department of Epidemiology and Medical Statistics, Faculty of Public Health, College of Medicine, University of Ibadan.

03/12/14

....

Supervisor

Dr. A.F. Fagbamigbe

B.Sc(Ilorin) M.Sc. (Lancaster) Ph.D (Ibadan)

Department of Epidemiology and Medical

Statistics,

Faculty of Public Health. University of Ibadan.

Dedication

I dedicate this dissertation to the God Almighty through whom all things are possible and to my parent Mr. and Mrs. Bakre, for their love and support throughout my stay in the University.

Acknowledgement

I am grateful to God the Alpha and omega of my life and the provider of all that I needed for the success of this course of study.

My unreserved gratitude goes to my able supervisor, Dr. A.F. Fagbamigbe, who despite his busy schedule and program, found time to read the work and offered ideas to improve its content. I appreciate his efficiency and constructive criticisms to bring the best out of me.

My sincere gratitude goes to all my lecturers in the Department of Epidemiology and Medical Statistics, College of Medicine , University of Ibadan, for being helpful especially in the impartation of Knowledge in one way or the other . They include: Prof. Olushola Ayeni, Prof. Afolabi E. Bamigboye, Dr. Adeniyi F. Fagbamige, Dr (Mrs.) Bidemi O.Yusuf, Dr. Onoja O. Akpa, Dr. Joshua O. Akinyemi, Dr. Ayo S. Adebowale Dr. Babatunde O. Adedokun, Dr. David M. Dairo, Dr (Mrs.)Ibitola O. Fawole, Dr (Mrs.)Oyeneye I. Ajayi , Dr (Mrs.) Adejoke I. Adeoye, Mr.Rotimi F. Afolabi, Mr.Babatunde M., Gbadebo, Mr. Nathaniel Afolabi and Mr. Aduroja.

I express my profound gratitude to every member of my family especially my parents; Mr and Mrs Bakre and my Brothers and Sisters and also to Oladoyin family

I am grateful to the Federal Government of Nigeria and Federal Ministry of Health for allowing access to the 2012 National AIDS and Reproductive Health Survey data(NARHS).

Words are inadequate to appreciate my friends; Alo Oluwafemi David, Oyedele Oyewole, Gabriel Obokon, Owolabi Boluwaji, Motunrayo Shodimu, Mr. Adetunji, Omisile Gbenga, Idowu Philip, Ugalahi Linda and other classmate that I didn't mention.

Finally my heartfelt gratitude goes to people who are found of asking me "Tunde, how far with your project?"

TABLE OF CONTENTS

List of Tables

List of Tables

List of Tables

List of Figure

Abbreviations

AGQ	Adaptive Gaussian Quadrature
CUMC	Currently Using Modern Contraception
FMOH	Federal Ministry of Health
MGQ	Marginal Quasi Likelihood
MLE	Maximum Likelihood Estimation
MSE	Mean Square Error
NAGQ	Non- Adaptive Gaussian Quadrature
NARHS	National AIDS and Reproductive Health Survey
NDHS	Nigeria Demography Health Survey
POR	Place of Resident
PQL	Penalized Quasi Likelihood
PSU	Primary Sampling Unit
SE	Standard Error
TFR	Total Fertility Rate

Abstract

INTRODUCTION: Likelihood plays an important role in parameter estimation. It is one of the tools used in estimating parameters of multilevel models, including multilevel binary logistic models. Cluster sampling scheme often introduces multilevel dependency among clustered observations whereby samples from same cluster tends to have related characteristics but different from samples from other clusters. This dependency may render single-level statistical models inefficient in the process parameter estimation. Despite the inadequacy of single-level estimates in the cluster data, public health researchers, lay little emphasis on estimation technique. This has hitherto led to improper inferences. The aim of this research is to evaluate different multilevel likelihood analysis estimation procedures including the traditional methods and to identify the best parameter estimation method in clustered data.

METHODOLOGY: This study utilized the 2012 National AIDS and Reproductive Health Survey (NARHS), a multistage stratified cluster dataset. The nationally representative survey used semi structured questionnaire to obtain information on reproductive behavior of women aged 15-49 years, The use of modern contraceptive was used as dependent variable while ages of respondents, place of residence, wealth status, religion, education among others were the independent variables. The standard binary logistic regression was first compared with multilevel binary logistic regression to obtain the percentage relative bias, then comparison of the performance of Penalized Quasi-Likelihood (PQL), Non-Adaptive Gaussian Quadrature(NAGQ) and Adaptive Gaussian Quadrature (AGQ) using XTMELOGIT and GLLAMM syntax in estimating parameters for multi-level logistic regression models were carried out. The comparisons were in terms of bias, numerical convergence, best fitted model and computational time. STATA version 12 and SPSS version 20 were used for data analysis at 5% significant level.

RESULT: Using -2logL, AIC and BIC, as yardstick to determine the fitness of the models from the different likelihood estimation method. AGQ had highest values and lowest standard error and was considered the best model for both two and three levels logistic regression. PQL was less biased compared to the other multilevel maximum likelihood methods, the conventional logistic model has overestimated the parameters by about 2%, 19% and 20% compared to multilevel model using by the corresponding methods PQL, NAGQ and AGQ respectively. AGQ using XTMELOGIT syntax gave the largest ICC result (ICC=0.201) which means 20% of the total variance is explained by the variance within the cluster. The PQL method generate the smallest intral cluster correlation coefficient(ICC=0.052). Also current age of the respondents, their wealth index, place of residence, Education , religion and their cluster have significant contribution to modern contraceptive use.

CONCLUSION: The adaptive Gaussian quadrature (AGQ) performed better than the Laplacian approximation (NAGQ) and penalized quasi likelihood (PQL) when considering two and three levels. but PQL performed relatively well in term of unbias estimate. In terms of computational time AGQ with XTMELOGIT syntax were adequate for two-level models while AGQ using GLLAMM syntax was adequate for three levels. Multilevel analysis should be encouraged in analyzing cluster data rather than the traditional individual level analysis.

Key words: Cluster survey, Likelihood, Adaptive Gaussian Quadrature, Penalized quasi likelihood, Laplacian approximation, Akaike's information criteria and Bayesian information criteria.

CHAPTER ONE

1.1 Introduction

Multilevel analysis also known as hierarchical modeling has been used in the fields of education(Bryk & Raudenbush et al 2001), demography (James, 2003, Hermalin 1986 and Mason 1983), and sociology(Guang & Hongxin 2000, DiPrete & Forristal 1994) to describe an analytical approach that allows the simultaneous examination of the effects of group-level and individual-level variables on individual-level outcomes. Over the years, interest in the use of multilevel analysis to investigate public health problems (Diez-Roux 2000 and Duncan et al. 1998) has grown. This growth has been stimulated in part by a resurgence of interest in the potential ecological macro or group-level determinants of health and the notion that variables referring to groups or to how individuals are related to each other within groups may be relevant to understanding the distribution of health outcomes (Diez-Roux 2000, Duncan et al.1996, Susser . 1994, Von Korff et al 1992). A second driving force in the use of multilevel methods has been the accelerated development of the statistical methods themselves (as well as the accompanying software) and the recognition that they are applicable in a broad range of circumstances involving nested data structures.

The availability of these complex statistical methods challenges public health researchers to articulate theories of the causes of disease that bring together factors at different levels. This will ensure that the method does not become an end in itself, but rather serves as a tool to investigate more sophisticated and hopefully more realistic models of disease causation.

1.2 An Overview of Multilevel Model

Multilevel model is a statistical model of parameter that varies at more than one level (Bryk et al 2002), this model can be seen as generalization of linear model, although they can also extend to nonlinear models. Multilevel model are particularly appropriate for research design where data for participant are organized at more than one level (i.e nested data). The unit of analysis is usually individual (at a lower level) who are nested in within aggregate unit (at high level) (Kim and Kawachi 2007). While the lowest level of the data in multilevel model is usually individual, repeated measurement of individual may be examined. Multilevel model provide an alternative type of analysis for univariate or multivariate analysis of repeated measures.

1.3 Advantages of Multilevel Analysis

Multilevel modeling offers several advantages. Some public health work was conceptualized as multilevel analysis but analyzed by traditional model, however, traditional linear or nonlinear models (that is single level model) do not enjoy all the advantages that will be described.

- A multilevel model provides a convenient framework for studying multilevel data. Such framework encourages a systematic analysis of how covariates measured at various levels of a hierarchical structure affect the outcome variable and how the interactions among covariate measured at different level affect the outcome variable. One of the frequently examined cross-level interaction effect is how the group context affect the impact of a covariate at the individual level. For example, (Entwisle et al 1986) tested the idea that the strength of the effect of maternal education on fertility depends on the characteristics of a country such as gross national product (GNP) and the intensity of family planning efforts.
- Multilevel modeling correct for the biases in parameter estimates resulting from clustering. In contrast to the popular belief, ignoring multilevel structure can result in biases in parameter estimates as well as biases in their standard errors. The more highly correlated the observations are within clusters, the more likely that ignoring clustering would result in biases in parameter estimates.(Goldstein 1995)
- Multilevel modeling provide correct standard error and thus produce correct confidence intervals and significance tests. When observations are clustered into higher-level units, the observations are no longer independent. Independent is one of the most basic assumption's underlying traditional linear and binary logistic regression models. When the clustering structure in a hierarchical data is ignored and the independent assumption is violated, the traditional linear and binary model tends to underestimate the standard errors (Diez-Roux 2000). The following is an intuitive argument for this statement, the observation in the same cluster tends to be more similar in their outcome measures. Similarity within a cluster implies that one can, to some extent, predict the outcome of an observation if the outcome of another observation in the cluster is known. This suggested that not every observation provide an independent piece of information and that the total amount of information contained in a sample with clustering is less than that in a sample without clustering.
- Estimate of the variances and covariance of random effect at various levels enable investigator to decompose the total variance in the outcome variable into portions associated with each level. (Guang and Hongxin 2000)

1.4 Contraceptive Use in Nigeria

The Federal Government of Nigeria adopted the National Policy on Population for Development, Unity, Progress and Self-Reliance in 1988. A revised policy in 2004 has included the aim of reduction of maternal deaths by 75% in 2015 in accordance with the Millennium Development Goal Number 5. The National Policy on Population back in 1988 encouraged open discussion and promotion of family planning. The goals of the policy were to improve the standard of living of Nigerians, promote health and welfare of the people through the reduction of deaths and disease among women and children, achieve a lower population growth rate through voluntary fertility regulation, and stem the population drift to urban areas.

An evaluation of the policy and the specific targets of the Nigerian Population Policy (NPP) by Adẹkunle et al (2000) indicated a total failure of all set targets for the year 2000. The population has continued to grow at an annual rate of approximate 3.0% and it is estimated to be about 148 million. The contraceptive prevalence rate, currently at 11%–13%, is far from the estimated 84% expected in 2005. The total fertility rate, although decreased from 6.2 in the earlier half of the decade, is still far from the targeted 4.0. The reasons for the policy's failure are an underestimation of the huge financial resources required for its implementation, the lack of political will, poor and uncoordinated organizational strategies, "gender-divide"(reducing women's fertility to four children, while leaving men free to have as many as children as they wish), and Nigeria's prolonged political instability with frequent policy changes.

In addition, the public sector and clinic-based, physician-controlled family planning programs carried out by the NPP cannot provide the needed coverage to satisfy the large unmet demand for family planning services, which currently stand at over 28%, involving over 4.76 million women, especially in the rural areas and northern part of Nigeria. Emmanuel Monjok et al (2010). The level of contraception among sexually active young women is particularly low, with a reported prevalence of 7.3% (Oye-Adeniran *et al.*, 2004) and 10% of modern contraception (NARHS 2013). This contributes to the high level of unwanted pregnancy, unsafe abortions and maternal mortality. (Ankomah *et al.* 2013) identified substantial geographical variations and a decline trend (between 2003 and 2007) in use of modern contraceptive methods in Nigeria. This is worrisome and calls for review of strategies to enhance improved use of modern contraceptive methods.

1.5 Empirical Applications of Multilevel Analysis in Public Health

Multilevel analysis is applicable to the study of a broad range of studies, the vast majority of applications in the health field have focused on geographically defined contexts, such as countries,(Chung H. et al(2007), states, (Kim and Kawachi 2007) and most commonly ''neighbourhoods'' defined in various ways, and also by smaller administrative areas.(Chaix B. et al 2007, Rundle A et al 2007) . The types of group-level constructs investigated have included, for example, income inequality,(Subramanian SV, et al.2006) social capital,(kim and Kawachi 2007) residential segregation, women's status, and neighbourhood characteristics such as neighbourhood disadvantage or other measures of neighbourhood social and physical environments. Most studies have used multilevel analysis to isolate associations of group-level factors with individual-level health outcomes after accounting for individual-level confounders (i.e. individual level variables associated with the health outcomes and with group membership, and, therefore, with group characteristics). A smaller number have focused on the complementary objective of decomposing variance into between and within-group components.

1.6 Problem Statement

The results of multilevel analyses published to date are not consistent with main effects of a variety of group-level variables on individual-level outcomes that persist after controlling for individual- level variables. The strength of this main effect has varied substantially depending on the study and the research question investigated. Detection of the group effects often generate a very distal relationship to the health outcomes being studied(Ana & Diez 2000), misspecification of groups and group-level variables, and the often extensive adjustment for much better measured individual-level variables, many of which are mediators rather than true confounders of the group-level effects.

Generally, the percent of total variance in the individual-level health outcome that is between groups (as compared to within groups) has been small. However, this result must be viewed in light of the fact that the relevant ''groups'' are generally grossly misspecified, that partitioning variance is complex for health outcomes that are not continuous variables, and that even well-established individual-level risk factors often explain only a very small amount of the observed inter-individual variability. So far the methods of parameter estimation have led to several problems in the realm of multilevel analysis. Where some leads to under estimation of parameters and some are biased (Guang & Hongxin 2000). In this study some important methods of estimating multilevel binary logistic model parameters will be deal with and the best method will be determined, perhaps even more important than the specific empirical results obtained to date, multilevel analysis has stimulated and promoted multilevel thinking generally within epidemiology, challenging researchers to begin to think more specifically.

1.7 Research Justification

Cluster sampling scheme often introduces multilevel dependency or correlation among the observations that can have implications for model parameter estimates. For multistage-clustered samples, the dependence among observations often comes from several levels of the hierarchy. The problem of dependencies between individual observations also occurs in survey research, where the sample is not taken randomly but cluster sampling from geographical areas is used instead. In this case, the use of single-level statistical models is no longer valid and reasonable. Hence, in order to draw appropriate inferences and conclusions from multistage stratified clustered survey data, it is very objective to use tricky and complicated modeling techniques like multilevel modeling; Traditional logistic regression (which, in multilevel analysis terms, is a single-level) requires :

(a) Independence of the observations conditional on the explanatory variables and

(b) Uncorrelated residual errors.

These assumptions are not always met when analyzing nested data, hence the option of the multilevel logistic regression analysis. It considers the variations due to hierarchy structure in the data. It allows the simultaneous examination of the effects of group level (cluster and Region) and individual level variables on individual level outcomes while accounting for the non-independence of observations within groups.

The number of groups, the group sizes, the variance of the random effects and the size of the correlation between random effects may be influential factors affecting the performance of the analysis method. Some methods of estimations were biased in this case, there is need to investigate the best method of parameter estimation.

Gidado 2013 has worked on community and individual factors influencing modern contraceptive use among women in Northern and Southern part of Nigeria. Since modern contraception is use by both men and women, there is need to include men and also to consider geo political zones , clusters and individual level factors on the uses of modern contraception. And to investigate if geopolitical zone can stand as a level.

This study aimed to:

(a) Describes the likelihood methods involved in estimation of multilevel parameters and how they are compare with traditional methods.

(b) To verify the best likelihood method of parameter estimation in multilevel analysis.

(c) To apply both multilevel and traditional logistic regression on cluster data using a nationally representative data collected from a multistage sampling procedure.

1.9 Objectives

- The main objective of this study is to evaluate the performance of different likelihood estimation procedures (MLE,PQL, AGQ and NAGQ) to determine on factors affecting modern contraceptive use in Nigeria.

The specific objectives are to:

1. Determine the best multilevel analysis method of estimation among PQL, AGQ and NAGQ.
2. Examine the effects of group-level and individual level predictors on modern contraceptive use.
3. Examine the inter individual and inter group variation as well as the contributions of individual level and group-level variables to these variations.
4. Examine the impact of some socio-demographic and socioeconomic factors on modern contraceptive use.

1.8 Research Hypotheses

The hypotheses in this study are.

1. Adaptive Gussian Quadrature (AGQ) will be unbiased and most efficient when there are large number of clusters, but these properties may not hold when there are fewer clusters. In particular, variance estimates may be biased when the number of clusters is large and the number of fixed effects is small.
2. Penalized quasi likelihood (PQL) estimates will be attenuated, especially when the variances of the random effects are large and when the cluster sizes are large.
3.
4. Penalized quasi likelihood (PQL) is not better than Adaptive Gussian Quadrature (AGQ) when using a multistage cluster survey for binary outcomes.

CHAPTER TWO

2.1 Literature Review

Several works had been done on multilevel analysis, in this chapter previous research on multilevel modeling will be reviewed. Foremost, handling multilevel analysis will be considered.

2.1.1 Handling Multilevel Analysis

Many years back, several authors have warned against the rapid incorporation of complex multilevel models before their performance is adequately understood, evaluated and especially when it is done with little regard to the adequacy of the data and the inferences that can be drawn from it.

Dipiete and Forristal, 1994 and Morris C. 1995 described the advantages of multilevel models over traditional methods as expense of greater model complexity. Models that are more complicated may be closer to reality but testing model fit and examination of model assumptions is more difficult. The authors maintained that if the model is true, multilevel estimates are less biased and more efficient than those obtained using other methods; however, models are less parsimonious and need larger data sets, and estimation becomes complicated. According to (Cohen, 1998), sample size and power calculations for multilevel hypotheses testing are particularly complex. Power, for example, depends both on the number of groups and on the number of individuals per group. Kreft, 1995 and Morris, C. 1995 said the centering of explanatory variables also raises more complicated issues than it does in traditional regression models estimation of variance explained at different levels and by different variables, particularly for models with many random coefficients and for nonlinear models. Ana V. Diez-Roux (2000), described random effect model as the model that can be reduced to standard regression model including both individual level and group level independent variables. He said the model with no random effect in which all regression coefficients are modeled as fixed with no random component at all group level is known as standard regression model. The persistence of significant variation in intercepts or slopes after inclusion of group-level variables suggests that other group-level factors possibly responsible for this variation may need to be explored. Empirical Bayes estimates of regression coefficients have been used to obtain improved estimates of associations in studies investigating the role of multiple exposures. He concluded that multilevel can also work like other statistical method to describe, summarize, and quantify patterns present in the data. but it will not explain these patterns; explanation will emerge from the reciprocal interplay between theory formulation and empirical testing.

2.1.2 Steps in Understanding the Multilevel Determinant of Health

A.V. Diez Roux (2008) examined the needs of any research by stating the impact of moving beyond neighborhoods to investigate other contexts. He said application of multilevel analysis in epidemiology has become almost synonymous with investigation of neighbourhood health effects. An important need is, therefore, to expand research into health effects of other well-defined contexts (e.g. countries or other policy-relevant units, schools, workplaces) with modifiable features likely to be related to health. Another research needs is the improved measurement of group level constructs at different levels of organization. In the absence of adequate measurement, not even the most sophisticated analytical techniques will allow convincing causal inference. Acyclic graphs, propensity score matching, instrumental variables, and marginal structural models were used to improve multilevel causal inference, the evaluation of their impact on results in real-life scenarios. Recent interest in applications of systems approaches to health has highlighted the potential utility of methods such as agent-based models or dynamic systems models in understanding the determinants of health. Applications of these approaches to multilevel questions, and research that contrasts the insights obtained from multilevel models and systems models would be important contributions to the field. Uthman (2007) make use of both descriptive statistics and multilevel modeling. The estimates of unadjusted effects of household wealth status on stunting, underweight, and wasting indicated that the household wealth status had strong negative effects on both stunting and underweight, but not on wasting. The effect was stronger on stunting than on underweight. the random effects model shows that there are significant variation in the log odds of stunting and underweight across the communities. The multilevel framework used in his study has shown that both individual-level and community-level characteristics are important predictors of childhood malnutrition in Nigeria, and demonstrates significant neighborhood variation in chronic childhood malnutrition.

2.1.3 Robustness, Power and Sample Size Selection in Multilevel Modeling

The robustness issue and the choice of sample size and power in multilevel modeling for both categorical and continuous dependent variables has been studied by several authors (Snijders and Bosker 1993, Raudenbush and Liu 2001, Hox 2002, Bingenheimer and Raudenbush 2004 and Maas and Hox 2005). Austin 2005 , used Monte Carlo simulation to assess the impact of misspecification of the distribution of random effects on estimation of and inference about both the fixed effects and the random effects in multilevel logistic regression models. He concluded that estimation and inference concerning the fixed effects were insensitive to misspecification of the distribution of the random effects, but estimation and inferences concerning the random effects were affected by model misspecification. Simulation studies indicate that a larger number of groups is more important than a larger number of individuals per group [Maas and Hox 2004 and Hox 2002]. The overall conclusion from these studies is that the estimates of the regression coefficients are unbiased, but the standard errors and the variance components tend to be biased downward (underestimated) when the number of level 2 units is small (e.g. less than 30) [Maas and Hox 2005].

2.1.4 Challenges in the use of Multilevel Analysis

The structure of clustered survey data is hierarchical, and a sample from such a population can be viewed as a multistage sample in multilevel analysis. There are some challenges people encountered when multilevel is used and without the use of multilevel. according to Blalock 1984 who work on Contextual-effects models: theoretical and methodological issues. His 1984 review, described many of the theoretical and methodological challenges facing contextual analysis. Despite the methodological sophistication of multilevel models, many of these challenges are still valid today. Perhaps chief among these is the need to develop theories that specify how group-level and individual-level factors may jointly shape the distribution of health and disease, theories that can be operationalised and tested. An example of the use of multilevel analysis in the context of a theoretical model that specifies how neighborhood attributes may be related to violent crime is provided by (Sampson et al 1997). Based on their underlying model, Sampson and collaborators conceptualized the relevant neighborhood-level attributes and developed operational measures of them. Multilevel analysis was then used as a statistical tool to examine aspects of the model in different ways. An important challenge to public health researchers is to develop substantive explanations and move beyond the use of multilevel analysis simply to document and statistically explain residual variability across groups after accounting for individual-level variables. In the absence of this, multilevel analysis runs the risk of being reduced to a method that examines variation across meaningless groups or associations with meaningless group-level variables and of either not finding much or finding patterns that are difficult to understand.

Early methodology work on multilevel logit model includes (wong & Mason 1985, Anderson & Aitkin 1985 and Goldstein 1991). Using data from fifteen world fertility survey (WFS), Entwisle et al (1986) studied contraceptive behavior of couples as a function of socioeconomics origins at the individual, of the gross national product per capital (GNP) and of the family planning effort at the country level.

2.1.5 The Consistency of Parameters in Multilevel Models.

Neil H. Spencer (2003) considered method of estimation of parameter of lagged multilevel model. This type of model is used for data collected over time where changes in test result over time can be modeled.

Simulations are used to demonstrate their success in obtaining consistent parameter estimates. He said maximum likelihood method can be used to estimate the multilevel parameter by using Iterative Generalized Least Squares (IGLS) method for the parameters in the random part of the model (the variances and co variances). These estimates are then used to obtain estimate of parameters from the fixed part of the model (the intercept and coefficients for the regressors) using Least Squares methods. He discovered that the estimation algorithms incorporate Least Squares methods, and it is this fact that leads to the problem of inconsistency for the model.

2.1.6 Comparison of Multilevel Methods of Estimation

Box and Tiao 1973 reviewed results of Klotz et al. (1969) and Portnoy (1971) which contrast the mean squared error (MSE) behavior of the following estimators of σ_u^2: the classical unbiased estimator based on mean squares, the ML estimator, and the mean and mode of the marginal posterior distribution for σ_u^2 with several choices of relatively diffuse priors. They found, over all values of the intra-class (intra-cluster) correlation $\rho = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2}$ they examined, that (a) the MSEs of the ML and posterior-mode estimators are comparable and much smaller than that of the unbiased estimator, and (b) the posterior mean is, by a substantial margin, the worst estimator on MSE grounds. Box and Tiao criticized MSE as an arbitrary criterion for performance assessment, and resisted the distillation of an entire posterior distribution down to a single point estimate. We are sympathetic with their position from the Bayesian viewpoint of the choice of posterior summaries should ideally be based on decision criteria arising from possible actions when using models like ; $\ln[\frac{p_{ijk}}{1-p_{ijk}}] = \beta_0 + \beta_1 x_{ijk} + u_{1jk}x_{ijk} + v_{0k} + u_{0jk}$ to solve real-world problems, but we nevertheless find it relevant, particularly in the context of general-purpose multilevel modeling software (where the eventual use of the output is far from clear), to examine operating characteristics such as bias and interval coverage. Rodriguez and Goldman (1995) used the structure of the Guatemalan child health data to examine how well quasi-likelihood methods compare with fitting a standard logistic regression model and ignoring the multilevel structure. As noted in Section 1.2, their approach involved creating simulated data sets based on the original structure but with known true values for the fixed effects (the β_i in model (2)) and variance parameters. They considered the MQL method and showed that estimates of the fixed effects produced by MQL were even worse, in terms of bias, than estimates produced by standard logistic regression disregarding the hierarchical nature of the data. The compared four approximation estimation procedures (first-order MQL or MQL-1, second-order MQL or MQL-2, first-order PQL or PQL-1, and second-order PQL or PQL-2) with the maximum likelihood achieved through high-dimensional numerical integration and the method of Gibbs sampling. They concluded that all approximation methods underestimate the random as well as fixed effects and that the underestimations of all except PQL-2 are severe. They preferred PQL-2 to all other methods as it has been found least biased. In the context of this research work, PQL and the full likelihood estimation methods and their percentage relative bias will be considered. Rountree and land (1996) reported distinctive differences between a general perceived risk of crime and a burglary- specific fear. They based their analysis on a victimization survey collected in Seattle, Washington in 1990. In the data set, more than 5000 individual are clustered into about 300 city-block, which are in turn clustered into 100 census track. In an effort to explain the southern migrant advantage in family stability, which refers to more stability among black southern family that migrated to northern cities. He realized that considering cluster the estimate of the parameters were not underestimated compare to without involving cluster. This study looked into the effect of the cluster and geopolitical zone on the use of modern contraceptive in Nigeria. Also to determine the rate at which the levels overestimate or underestimate the factors that was considered in the research work. Marc Callens et al 2003, who worked on performance of likelihood-based estimation methods for multilevel binary regression models. focused on two different likelihood-based estimation procedures frequently used in the applied multilevel-modeling literature: Non-Adaptive Gaussian Quadrature (NGQ) and Adaptive Gaussian Quadrature (AGQ). Their computing algorithms were standard implemented in the SAS macro GLIMMIX and procedure NLMIXED. The two estimation methods are evaluated at four performance dimensions: numerical convergence, bias, mean squared error and computation time. Numerical convergence is measured by the convergence rate. This convergence rate is based on the indicator variables produced by the macro GLIMMIX and PROC NLMIXED to confirm whether numerical convergence has been reached or not. He discovered that there is problem of convergence among the two method of estimation. Comparing the quadrature methods yields close results with respect to Bias and Mean Squared Error, but the Non-Adaptive version was by far the slowest. Hence, it is confirmed that AGQ is to be preferred above NGQ. hereby confirming previous studies that mainly focused on the bias. However, AGQ gave the most precise estimates, as measured by the MSE.

Ecevit Eyduran (2008) used penalized maximum likelihood estimation method as an alternative to maximum likelihood estimation methods in medical research. He generated four unreal small sample dataset which were in 2 x 2 contingency table form, due to separation problem PMLE was used to reduce the biased estimate in traditional logistic regression, but the level are not consider . MLE and PMLE methods were applied and compared for separation case, including biased estimation in the logistic regression. The parameter and their standard error estimates showed clearly that MLE's are bias estimate and PMLE was unbiased estimate, it is also show clearly that the standard error for PMLE was reduced compare to MLE. He concluded that PMLE performed unbiased (reliable) . Adam C Carie 2009, who worked on fitting multilevel models in complex survey data with design weights, used data from 2005-2006 National survey of children with special Health Care needs and fit a series of multilevel linear models using Mplus , MLwiN and GLLAMM, he compare and contrast the estimates and there standard error across the program and scaling methods. He examined the effect of the software by using continuous outcome variable and categorical variable. And also fitted six models , across the model each software program converge on nearly identical result with regard to the weighted analysis across the fixed and random effect model the program achieved nearly identical weighted result, he discovered that MLwiN consistently estimated a marginal larger variance in the intercept across state but the different methods of parameter estimation was not examine. GLLAMM gave a precise estimate, in other to get large variance, weighting may not be needed in the case, though weight leads to more representative population estimate but the failure to include them did not bias the inference decisions.

2.1.7 Multilevel Model with Ordinal Outcome Variable

Daniel J. and Sonya K. (2011), fitted multilevel model with ordinal outcome variable, they checked if fitting multilevel linear model to ordinal outcome is justified. Also, compared Adaptive Guassian quadrature and penalized quasi likelihood across variation in sample size, they also checked the magnitude of variance component and the distribution shape . They considered two levels which are individual (level one) and cluster (level two). Using data generated by SAS version 9.1, they fitted multilevel cumulative logit model by PQL using SAS version 9.1 and the adaptive Gauss –Hermit Quadrature was done using Mplus version 5. Comparison were done in three scenario (that is, when the clusters are small, middle and high) PQL performed best when the random effects were small and the cluster sizes were large. In addition, a new result of the study is that the performance of PQL greatly improves with the number of categories for the outcome. The AGQ estimator also behaved as expected. Consistent with asymptotic theory, AGQ was least biased and most efficient for data with 100 or 200 clusters. With 25 or 50 clusters, however, AGQ estimates were more variable and often had higher MSE than PQL estimates.

There are still compelling reasons to compare the ML(AGQ) and PQL estimators for the multilevel binary logit model. First, although ML(AGQ) is an asymptotically unbiased estimator, it suffers from small sample bias (Demidenko, 2004, Raudenbush & Bryk, 2002, p. 53).

When the number of clusters is small, ML(AGQ) produces negatively biased variance estimates for the random effects. Additionally, this small-sample bias increases with the number of fixed effects. Second, Bellamy (2005) showed analytically and empirically that when there are small numbers of clusters, as often occurs in group-randomized trials, the efficiency of PQL estimates can equal or exceed the efficiency of ML estimates. Third, as discussed above, AGQ may compare more favorably to PQL when the data are Binary rather than ordinal.

2.2 Modern Contraceptive Use

Oyefara,2013 define Contraception as a means of controlling fertility by using various methods that inhibit contraception which can be traditional or modern method. In his review of 1990 Nigeria Demographic and health Survey which showed that knowledge of family planning method witnessed a remarkable improvement between 1981 and 1990 because about 46% of all women aged 15-49 involve in the study knew at least one method of family planning, with about 44% Identifying modern methods of which the pill, injection, condom, IUCD and female sterilization were most commonly known. The 2008 NDHS results also confirmed this. According to his analysis between age at first birth and Contraceptive use among women of child bearing age in Osun, the study revealed that older mothers had relatively better knowledge about contraceptives than adolescent. 94.0% of older mothers against 83.2% among adolescent mothers had knowledge about contraceptive use.

2.3 Geographical Regions (Zones) in Nigeria

The main regions in Nigeria are the North central, North East, North West, South East, South South and South West regions. The prevalence of use of modern contraceptive vary across all these regions, which leads to different fertility rate across the nation. According to Mberu and Reed 2008, the North central region has a Total Fertility Rate (TFR) that is lower than the core Northern regions, but which is still on the average , one child more than the TFR of the Southern regions. The total fertility in the North has been over two children per woman higher than that of the South in both 2003 and 2008. Also, the mean number of children ever born (a measure of past fertility) was 3.1 in 2003 for Nigeria as a whole but a difference of over one child per woman was observed between the North and the South. In the Study by Mberu and Reed 2008, it was showed that adolescent motherhood and pregnancy are lowest in the South

West and South East regions in 2003. In Contrast, 38% and 37% of adolescents aged 15-19 in the North East and North West were mothers in 2008, the.highest level in the country.

2.4 Education

The education variable was divided into four categories—Qur'anic only, primary, secondary and higher based on the highest grade of schooling the respondent had attended. Attending school rather than obtaining a degree was used as a metric of education so that teenage women attending school or who might soon attend secondary school would not be excluded from those over age 18, who have had a chance to complete their degree.

2.5 Wealth Index

The economic state of a woman as measured by wealth index has impact on her reproductive status. Research done by Gidado 2013 has shown that women from higher wealth quintile are more likely to be better educated than those from lower wealth quintile. Also, wealth quintile has been found to be positively associated with contraceptive use and age at first sexual intercourse. Consequently, It is tempting to argue that wealth quintile have influence on modern contraceptive use in Nigeria.

2.6 Religion

Religion, according to Christiano et al(2008).is the opium of the masses. Religion is believed to play a part in shaping the views, norms, belief, attitudes and practices of the people which in turn affects the reproductive behaviour. In Nigeria, religion has a great effect on the pattern of childbearing. According to Blom and Reddy, 1986 Studies done in India indicate that Hindus marry and bear children at younger ages than non-hindus . In Tanzania, religion influences childbearing ages. Religions such as Islam that places absolute emphasis on pre-marital chastity, this will ultimately result in early marriage and as a result early pre-disposition to sexual intercourse, thereby leading to early childbearing ages, (Ngalinda.I. 1998). In his study, Muslims showed a lower mean age of child bearing of 18years than other religious affiliate.

Srikanthan and Reid(2008) work on religious and cultural influences on contraception, they have said that the perception and behavior related to reproduction are strongly determined by prevailing cultural and religious values.

2.7 Age at First Sexual Intercourse

The age at first sexual intercourse is the age when sexual initiation begins. It is an important indicator that will notify people when to start using contraceptive. In cases where the use of the most effective contraceptive method is absent, unwanted pregnancy will be increased . In a study by Uthman (2008), North West and North East had the highest proportion of women who had reported early sexual debut.

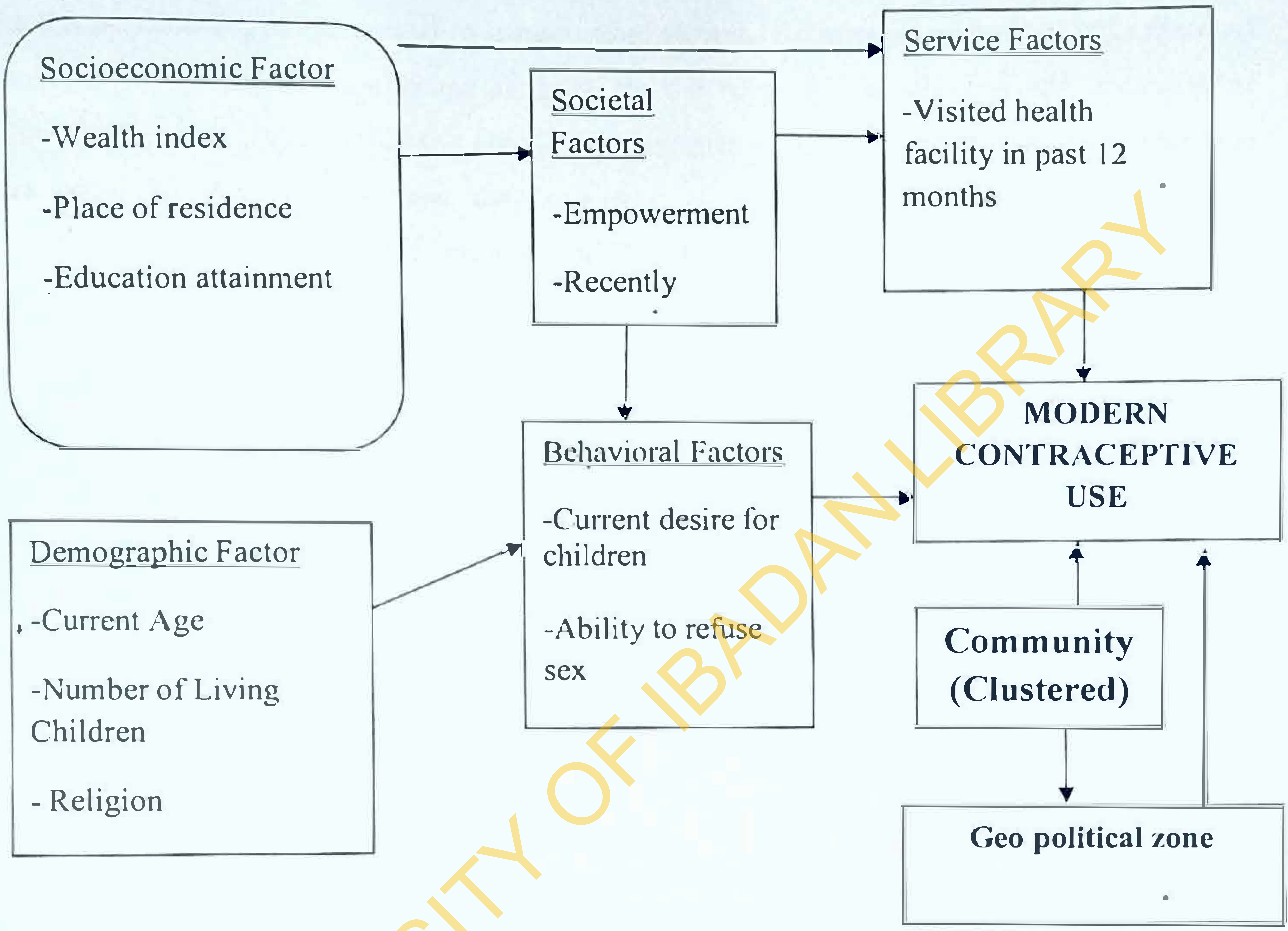
2.8 Place of Residence

So many studies done in the past have observed that women living in urban areas start bearing child early enough than their counterparts in the rural setting. Adebimpe et al, (2011) in Osun found that the mean age at first birth among rural respondents was 20.8±3.7 years and 23.2±5.1 years among urban respondents: also the mean number of births per woman was 3.4±1.8 births per woman in rural, and 2.9±1.5 births per urban woman. They concluded that there was a significant association between locations of residence and modern contraceptive use.

Cohen (1993), argued that women living in urban areas are assumed to have better knowledge of contraception and access that affect their reproductive outlook.

2.9 Determinant of Modern Contraceptive Use

Conceptual Framework



Conceptual framework for determinants of modern contraceptive use

Many researchers have examined the determinants of contraceptive use, from both the providers' and clients' perspectives (Cleland et al. 2006). The customized conceptual framework builds on existing knowledge to analyze the socio-economic and demographic factors associated with contraceptive use among young married women compared with older women in Nigeria. While the framework used is generalized for both the young women and older women, I hypothesize that the factors associated with contraceptive use may operate differently within each age group due to differences such as empowerment, education, and desire for children. This hypothesis is premised on the fact that, as in many of the least-developed countries, health services and policies in Nigeria are not clearly streamlined to consider the special needs of young women.

CHAPTER THREE

Methodology

3.1 Multilevel Analysis for Multistage Clustered Data

In multilevel research, the structure of data in the population is hierarchical, and a sample from such a population can be viewed as a multistage sample. Because of cost, time and efficiency considerations, stratified multistage samples are the norm for sociological and demographic surveys (Hongxin Zhao and Guang Guo 2000). For such samples the clustering of the data is in the phase of data analysis and data reporting, a nuisance which should be taken into consideration. However, these samples, while efficient for estimation of the descriptive population quantities, pose many challenges for model-based statistical inference.

This cluster sampling scheme often introduces multilevel dependency or correlation among the observations that can have implications for model parameter estimates. For multistage clustered samples, the dependence among observations often comes from several levels of the hierarchy. The problem of dependencies between individual observations also occurs in survey research, where the sample is not taken randomly but cluster sampling from geographical areas is used instead. In this case, the use of single-level statistical models is no longer valid and reasonable.) (Hasinur et al 2011). Hence, in order to draw appropriate inferences and conclusions from multistage stratified clustered survey data one may require tricky and complicated modeling techniques like multilevel modeling, and very often the computation required for this is not straightforward and is not very time consuming depends on software used for the model. There are numbers of software packages.

3.2 Multilevel Analysis Software

Multilevel models can be formulated in two ways: (1) by presenting separate equations for each of the levels, and (2) by combining all equations by substitution into a single model-equation. The software HLM (Raudenbush et al., 2000) requires specification of the separate equations at each available level. Most other software (e.g., MLwiN; Raudenbush et al., 2005), SAS Proc Mixed (Marc Callens et al 2003) uses the single equation representation. Both representations have their advantages and disadvantages. The separate-equation representation has the advantage that it is always clear how the model is built up. The disadvantage is that it hides from view that modeling regression slopes by other variables results in adding an interaction to the model.

3.3 Brief Description of the Study Area

Nigeria came into existence as a nation – state in 1914 through the amalgamation of the Northern and Southern protectorates. Prior to that time, there were various separate cultural, ethnic and linguistic groups. The British established a crown colony type of government after the amalgamation. The affairs of the colonial administration were conducted by the British until 1942, when a few Nigerians became involved Nigeria is in the administration of the country. In the early 1950's, Nigeria achieved partial self government with a legislature in which the majority of the members were elected into an executive council of which most were Nigerians. Nigerians became fully independent in October 1960 as a federation of three regions (Northern, Western, and Eastern) under a constitution that provided for a parliamentary system of governance. The Lagos area became the federal capital territory.

Nigeria is in the West African sub region, lying between latitudes 4°16′ and 13°53′ north and longitudes 2°40′ and 14°41′ east. It is bordered by Niger in the North, Chad in the North east, Cameroon in the East, and Benin in the west. To the south, Nigeria is bordered by Approximately 850 kilometres of the Atlantic Ocean, stretching from Badagry in the West to the Rio del Rey in the east, with a total land area of 923,768 square kilometres. Nigeria is the fourteenth largest country in Africa.

Presently, Nigeria is made up of thirty-six states and a federal capital territory (FCT), grouped into six geopolitical Zones namely: North Central, North East, North West, South East, South-South, and South West with about 774 constitutionally recognized local government areas. She has two predominant religions namely, Islam and Christianity.

3.4 Data

The dataset used in this study has been taken from the 2012 National HIV and AIDS and Reproductive Health Survey (NARHS Plus II). which was a nationally representative survey carried out to provide information on key HIV & AIDS and reproductive health knowledge and behaviour related issues. Data collection took place between September and December 2012 from 31,235 individual respondents interviewed in NARHS Plus II; consisting of 15,596 males and 15,639 females showed a response rate of 88%. The 2012 NARHS Plus II had valid responses of 10733 currently -married women and 1519 sexually active unmarried female aged 15–49.

The 2012 NARHS+ data set used for this study was collected using a multistage stratified cluster sampling. The appropriate approach to analyze contraceptive data from this survey is therefore based on nested sources of variability. Here the units at lower level (level-1) are individuals (ever and never -married women aged 15−49 and male aged 15-64) who are nested within units at higher level (clusters: level-2 in which they were community from either rural or urban area) and the clusters are again nested within units at the next higher level (region: level-3 which is the six geopolitical zone in Nigeria). Clusters are primary sampling units (PSU) defined by the National Census of 1991, and correspond approximately to rural and urban areas. All clusters are approximately of equal size in terms of area. On the other hand, Regions are administrative areas each of which consists of a number of sub-administrative areas. Due to this nested structure, the odds of women and men experiencing the outcome of interest are not independent, because individual from the same cluster may share common exposure to community characteristics.

	LEVEL3 (ZONES)	LEVEL 2 (CLUSTERS)	LEVEL 1 (INDIVIDUAL)
2012 NARHS+ DATA	North Central	210	6008
	North East	180	4875
	North West	210	6152
	South East	150	4282
	South South	180	4939
	South west	180	4979
	6 units	1110 units	31235 units

Figure 1: Hierarchical structure of the 2012 NARHS data

3.5 Data Collection

Data were collected by using two structured and semi-structured questionnaires – one each for individuals and households (Federal Ministry of Health 2013). And these data was pretested in two states (kogi and cross river) using one urban and one rural cluster in each of the state, which assisted in identifying gaps that could have arisen during the actual exercise. The survey personnel were also trained at two level (central and state level).

3.6 Operationalization of Key Variables

3.6.1 Dependent Variable

The response variable in this study is "currently using modern contraception" (CUMU)Q1212 which is binary. For the study purpose the response variable was recorded as follows: those women currently using the methods are coded as 1 and those not currently using the method are coded as 0.

3.6.2 Independent Variable

The primary choice of explanatory variables for this study was based on previous studies on factors influencing contraceptive prevalence rate (S. Amin *et al.*, 2002; Kalam and Khan, 2002). It consists of socio-demographic and socio-economic variables. Current age (CAge)(Q103), education (Educ)(Q106), Religion (Q111) place of residence (POR)(003 locality) and wealth index (wealth quintile), are the explanatory variables that was used in the course of this research..

Socio-Demographic variables

Age: This was measured by the current age of the respondents as at the time of survey and was Recoded into 15-19,20-24, 25-29, 30-39, 40-49 and 50-64

Socio-economic variables

Education: This was grouped into No Formal Education, Quranic only, Primary, Secondary and Higher education categories.

Religion: The religions of the respondents were recoded into Islam, Non Catholic Christian, Catholic, Traditional, No religion and others.

Wealth Index: The wealth index was grouped into Poorest, Poorer, Average, Wealthier and Wealthiest.

Cluster level variable

The cluster level variable in this study is place of residence (POR) in which people are clustered based on their community in either urban or rural area in the country, for the purpose of this study I assumed that it is fixed.

Zonal level variable

The variable used is geo-political zone which is divided into six include; North Central, North East, North West, South East , South South and South west.

This study used multilevel analysis in which current age, education, religion and the wealth index measured on individuals were level one variables, cluster(community) and place of residence were level two variables and geo political zone was level three variable.

3.7 Methods

In this research work, the methods for estimating multilevel binary logistic models are based on likelihood. In statistics, likelihood function is a function of parameters of a statistical model. The likelihood is a set of parameter value θ given outcome y is equal to the probability of those observed outcome given those parameter values, that is $l(\theta / y) = p(y/\theta)$. Likelihood play an important role in the method of parameter estimation and it is synonymous with probability. Likelihood is used when describing the function of parameter given an outcome.

Among the likelihood methods, Marginal Quasi Likelihood (MQL) (Goldstein, 1991; Goldstein and Rasbash, 1996) and Penalized Quasi Likelihood (PQL) (Breslow and Clayton, 1993, Daniel and sonya 2011) are the two most used approximation procedures. In this study, penalized quasi likelihood was considered as the quasi likelihood. After applying this quasi likelihood methods, the model is then estimated using iterative generalized least squares (IGLS) (Goldstein, 2003), which is full maximum likelihood estimation procedure (that is, Adaptive Gaussians Quadrature and Laplacian approximation) to estimate the random intercept and fixed effect model.

The multilevel process was stepwise.

Steps

In this research work, the following steps will be considered in other to achieve the objectives stated in chapter one, the steps are to:

1. Fit a simple model with no pedictors i.e an intercept-only model that predicts the probability of contraceptive use. The functional form of the model is $\beta_{00k} = \beta_{000} + u_{0jk} + \tau_{00k}$ The estimates of parameters and standard errors will be determined. Using PQL and ML(AGQ and NAGQ)
2. Presents a random intercept and a fixed slope for the variable by using multilevel univariate analysis.
3. Assess all significant factors found in previous univariate analysis that affect contra captive use.
4. Fit random effects univariate model for the covariate that is significant.
5. Compare the three methods of parameter estimation by checking with the smallest standard error.

This allows the effect of the explanatory variable to vary from zone to zone and from cluster to cluster. Also this analysis allows the examination of both between group and within group variability as well as how group level and individual level variables are related to variability at both levels and also the performance of likelihood estimation method was examined.

3.8 The Multilevel Linear Model

Multilevel linear model which focus on a few specific multilevel model that statistician are likely to estimate. For description of general form of multilevel linear model. I first consider a simple two- level model with a single explanatory variable.

$$y_{ij} = \beta_{0j} + \beta_{1j}\,\mathrm{x}_{1ij} + \mathrm{e}_{ij} \qquad \text{level 1}$$

$$\beta_{0j} = \beta_{00} + \beta_{01}w_{1j} + u_{0j}$$
$$\beta_{1j} = \beta_{10} + u_{1j} \qquad \text{level 2}$$

$$y_{ij} = \beta_{00} + \beta_{01}w_{1j} + \beta_{10}x_{1ij} + u_{1j}x_{1ij} + u_{0j} + e_{ij} \qquad \text{(combined model)}$$

Where y_{ij} is the outcome variable for the ith unit at level one and the jth unit at level two, β_{00} is the intercept, x_{1ij} is the explanatory variable in level one while β_{10} is its effect and w_{1j} is the explanatory variable in level two while β_{01} is its effect, u_{0j} and u_{1j} are random effect accounting for the random variation at level two, and e_{ij} is the level one random effect. The parameter for the random effects are $E[u_{0j}] = E[e_{ij}] = 0$, $\mathrm{var}(u_{0j}) = \sigma^2$, $\mathrm{var}(e_{ij}) = \sigma_e^2$, $\mathrm{cov}(u_{0j}, e_{ij}) = 0$ for $j \neq i$. The within cluster or intraclass correlation after controlling for the explanatory variable can be obtained from $\rho = \frac{\sigma_u^2}{(\sigma_u^2 + \sigma_e^2)}$. Equation (1) can also be considered as a random effect model for panel data. We next extend the simple two level model to a three- level model with random coefficients.

$$y_{ijk} = \beta_{000} + \beta_{001}z_{1k} + \beta_{110}w_{1jk} + \tau_{01k}w_{1jk} + u_{1jk}x_{1ijk} + \beta_{100}x_{1ijk} + \tau_{01k}x_{1ijk} + u_{0jk} + \tau_{00k} \qquad \text{.....(2)}$$

where z_{1k}, w_{1jk} and x_{1ijk} are third second and first level explanatory variable respectively and β_{001}, β_{110} and β_{100} are those levels fixed effect, β_{000} is the intercept. τ_{00k} and u_{0jk} are the random intercept for level three and level two, respectively, and u_{1jk} is x_{ijk}'s random effect at level two, τ_{01k} is x_{1ijk}'s random effect at level three .

Other parameter of the model include E[v_{ok}] =E[U_{0jk}] = E[e_{0ijk}] = 0, var(τ_{00k}) = σ^2_{ro}, var(u_{ojk}) = σ^2_{uo}, var(u_{1jk}) = σ^2_{u1}, var (e_{0ijk}) = σ^2_{eo}, and cov ($u_{ojk,}$ u_{1jk}) = σ^2_{u01}. The model assumes that the random effect across different level and the random effects across different cluster in the same level are uncorrelated by adding more observed variable to equation (2) more complex model will be constructed which will allow cross-level interaction.

3.9 Multilevel Modeling for Binary Data

By considering two level model for binary outcome with a single explanatory variable. Intuitively, this model is equivalent to model (1) except for the outcome variable. Since NARHS data consisting of individual, (level one) grouped into clusters (level two). Then y_{ij} is a binary response for women i using modern contraceptive in cluster j and x_{ij} is an explanatory variable at individual level.

The probability of using modern contraceptive is equal to one that is p_{ij} = $pr(y_{ij}=1)$ while probability of not using equal to zero[p_{ij} = $pr(y_{ij}=0)$] and let p_{ij} be modeled using a logit link function. The standard assumption is that y_{ij} has a Bernoulli distribution. Then the two- level model can be written as

$$\ln\left[\frac{p_{ij}}{1-p_{ij}}\right] = \beta_{00} + \beta_{01}w_{1j} + \beta_{10}x_{1ij} + u_{1j}x_{1ij} + u_{0j} \text{(combine model)}$$
....(3)

Where u_{0j} and u_{1j} are the random effect at level two. Without random effect, equation(3) would be a standard logistic regression model. In the case of multilevel u_{0j} and u_{1j} are assumed to be independent , u_j is also assumed to be normally distributed with mean 0 and variance σ^2_u, model (3) is often described alternatively in the literature on multilevel by the next equation.

$$\ln\left[\frac{p_{ij}}{(1-p_{ij})}\right] = \beta_{00} + \beta_{01}w_{1j} + \beta_{10}x_{1ij} \qquad \text{(level 1 model)} \qquad (4)$$

Where,

$$\begin{aligned} \beta_{0j} &= \beta_{00} + \beta_{01}w_{1j} + u_j \\ \beta_{1j} &= \beta_{10} \end{aligned} \qquad \text{(level 2 model)} \qquad (5)$$

Relative to equation (4) and (5) equation (3) is called combine model.

The multilevel model for binary outcome can also be derived through a latent variable concept. We assume that there exist a latent continuous variable y_{ij}^{*} underline y_{ij}. by observing only the binary response variable y_{ij} directly, but not y_{ij}^{*}. However, $y_{ij}^{*} > 0$ if $y_{ij} = 1$ and $y_{ij}^{*} \leq 0$ if $y_{ij} = 1$. A multilevel model for y_{ij}^{*} equivalent to (3) can be written as

$$y_{ij}^{*} = \beta_{00} + \beta_{01} w_{1j} + \beta_{10} x_{1ij} + u_{1j} x_{1ij} + u_{0j} + e_{ij} \qquad (6)$$

Condition on the random effect u_j's at level two, either a logit multilevel model such as (3) or a probit multilevel model can be derived from equation (6) depending on whether e_{ij} in (6) is assumed to be standard logistic distribution or normal distribution. This concept illustrate the close connection between the multilevel models for linear data and those for binary data. The result of the connections will be used to calculate the intra-cluster correlation for binary data.

By assuming u_j, the conditional density function for cluster j for model (3) was identical to that of the logistic regression.

$$f(y_j \setminus x_j, u_j) = \prod_{i=1}^{n_j} \frac{\exp[(\beta_{00} + \beta_{01} w_{1j} + \beta_{10} x_{1ij} + u_j)]}{1 + \exp(\beta_{00} + \beta_{01} w_{1j} + \beta_{10} x_{1ij} + u_j)} \qquad (7)$$

Where y_j and x_{1ij}, respectively denote the responses and the explanatory variable in cluster j. The standard for estimating the model parameter is to assume that u_j, is normally distributed and to integrate out the unobserved random effect u_j.

$$f(y_i \setminus x_j) = \int f(y_j \setminus x_j, u_j) g(u_j) du_j, \qquad (8)$$

Where g(.) represents the normal density function. The unconditional density $f(y_i \setminus x_j)$ does not have a closed expression, therefore, maximum likelihood estimation has to resort to approximation procedure such as numerical integration.

Model (3) is almost the simplest possible multilevel model for binary data. Greater challenges arise in estimation of general model with multiple random effects. The next equation describe a three – level model with a single explanatory variable that has both fixed effect and random effect,

$$\ln[\frac{p_{ijk}}{1 - p_{ijk}}] = \beta_{000} + \beta_{001} z_{1k} + \beta_{110} w_{1jk} + \tau_{01k} w_{1jk} + u_{1jk} x_{1ijk} + \beta_{100} x_{1ijk} + \tau_{01k} x_{1ijk} + u_{0jk} + \tau_{00k} \quad \text{(combined model)(9)}$$

3.9.1 Three- Level Model With Predictors at All Levels with Random Intercept And Random Slopes.

$$\ln\left(\frac{p_{ijk}}{1-p_{ijk}}\right) = \beta_{0jk}cons + \beta_{1jk}cage_{ijk} + \beta_{2jk}wi_{ijk} + \beta_{3j}Edu_{ijk} + \beta_{4jk}\,\mathrm{Re}l_{ijk} + \beta_{5jk}POR_{ijk} \quad \text{(level one)(i)}$$

Here five explanatory variables at individual level and one variable each at level two and level three that contribute to the odd of using modern contraceptive among both male and female in reproductive age range in Nigeria. The fixed effect will be randomized

$$\begin{aligned}
\beta_{0jk} &= \beta_{00k} + \beta_{01k}w_{1jk} + u_{0jk}\\
\beta_{1jk} &= \beta_{10k} + \beta_{11k}w_{1jk} + u_{1jk}\\
\beta_{2jk} &= \beta_{20k} + \beta_{21k}w_{1jk} + u_{2jk}\\
\beta_{3jk} &= \beta_{30k} + \beta_{31k}w_{1jk} + u_{3jk}\\
\beta_{4jk} &= \beta_{40k} + \beta_{41k}w_{1jk} + u_{4jk}\\
\beta_{5jk} &= \beta_{50k} + \beta_{51k}w_{1jk} + u_{5jk}
\end{aligned} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(ii)}$$

By substituting equation (ii) in (i) I have level two multilevel logistic model , which consist of individual level effect, cross level effect and the random effect at level two.

$$\ln\left[\frac{p_{ijk}}{1-p_{ijk}}\right] = \beta_{00k} + \beta_{01k}W_{1jk} + u_{0jk} + \beta_{10k}X_{1ijk} + \beta_{11k}X_{1ijk}W_{1jk} + u_{1jk}X_{1ijk} + \beta_{20k}X_{2ijk} + \beta_{21k}X_{2ijk}W_{1jk} + u_{2jk}X_{2ijk} +$$

$$\beta_{30k}X_{3ijk} + \beta_{31k}X_{3ijk}W_{1jk} + u_{3jk}X_{3ijk} + \beta_{40k}X_{4ijk} + \beta_{41k}X_{4ijk}W_{1jk} + u_{4jk}X_{4ijk} + \beta_{50k}X_{5ijk} + \beta_{51k}X_{5ijk}W_{1jk} + u_{5jk}X_{5ijk} \quad \ldots\text{(iii)}$$

(Level Two)

The above fixed effect at level two will generate another random model by considering level three variable (Regions) .

$$
\begin{aligned}
\beta_{00k} &= \beta_{000} + \beta_{001} z_{1k} + \tau_{00k} \\
\beta_{01k} &= \beta_{010} + \beta_{011} z_{1k} + \tau_{01k} \\
\beta_{11k} &= \beta_{110} + \beta_{111} z_{1k} + \tau_{11k} \\
\beta_{20k} &= \beta_{200} + \beta_{201} z_{1k} + \tau_{20k} \\
\beta_{21k} &= \beta_{210} + \beta_{211} z_{1k} + \tau_{21k} \\
\beta_{30k} &= \beta_{300} + \beta_{301} z_{1k} + \tau_{30k} \qquad \text{-- (iv)} \\
\beta_{31k} &= \beta_{310} + \beta_{311} z_{1k} + \tau_{31k} \\
\beta_{40k} &= \beta_{400} + \beta_{401} z_{1k} + \tau_{40k} \\
\beta_{41k} &= \beta_{410} + \beta_{411} z_{1k} + \tau_{41k} \\
\beta_{50k} &= \beta_{500} + \beta_{501} z_{1k} + \tau_{50k} \\
\beta_{51k} &= \beta_{510} + \beta_{511} z_{1k} + \tau_{51k}
\end{aligned}
$$

By substituting (iv) in (iii) we have combined multilevel logistic regression

$$
\begin{aligned}
\ln\left[\frac{p_{ijk}}{1-p_{ijk}}\right] &= \beta_{000} + \beta_{001} z_{1k} + \tau_{00k} + \left(\beta_{010} + \beta_{011} z_{1k} + \tau_{01k}\right) w_{1jk} + u_{0jk} + \\
&\left(\beta_{100} + \beta_{101} z_{1k} + \tau_{10k}\right) x_{1ijk} + \left(\beta_{110} + \beta_{111} z_{1k} + \tau_{11k}\right) x_{1ijk} w_{1jk} + \\
&u_{1jk} x_{1ijk} + \left(\beta_{200} + \beta_{201} z_{1k} + \tau_{20k}\right) x_{2ijk} + \left(\beta_{210} + \beta_{211} z_{1k} + \tau_{21k}\right) x_{2ijk} w_{1jk} + \\
&u_{2jk} x_{2ijk} + \left(\beta_{300} + \beta_{301} z_{1k} + \tau_{30k}\right) x_{3ijk} + \left(\beta_{310} + \beta_{311} z_{1k} + \tau_{31k}\right) x_{3ijk} w_{1jk} + \\
&u_{3jk} \cdot x_{3ijk} + \left(\beta_{400} + \beta_{401} z_{1k} + \tau_{40k}\right) x_{4ijk} + \left(\beta_{410} + \beta_{411} z_{1k} + \tau_{41k}\right) x_{4ijk} w_{1jk} + \\
&u_{4jk} x_{4ijk} + \left(\beta_{500} + \beta_{501} z_{1k} + \tau_{50k}\right) x_{5ijk} + \left(\beta_{510} + \beta_{511} z_{1k} + \tau_{51k}\right) x_{5ijk} w_{1jk} + u_{5jk} x_{5ijk}
\end{aligned}
$$

...(level three)

where,

$\left(\dfrac{p_{ijk}}{1-p_{ijk}}\right)$ = Odd *of using modern contraceptive*

$x_{1ijk} = CurrentAge,$

$x_{2ijk} = wealthindex$

$x_{3ijk} = Education\ attainment,$

$x_{4ijk} = \mathrm{Re}\,ligion$

$x_{5ijk} = POR$

$\mathrm{w}_{1jk} = cluster(\text{level }2)$

$z_{1k} = zone(\text{level }3)$

$\beta's = Fixed\ effect\ parameters$

$u's = Random\ effect\ parameter$ of the cluster *at level two*

$\tau's = Random\ effect\ parameter$ of region *at level* three

Equation (v) above is the three levels five predictors logistic regression with fixed effect, random intercept and random slope.

3.9.2 Model for Three Levels Five Predictors Logistic Regression With Random Intercept and Fixed Slope

The most basic expansion of a fixed-effects regression model to a multilevel model is to allow the intercept term to vary randomly over groups. This parameterization implies that the regression slopes remain fixed (i.e., are invariant over groups), but the intercept term does not fixed. The Level-1 model is given as

$$\ln\left(\frac{p_{ijk}}{1-p_{ijk}}\right) = \beta_{0jk} + \beta_{1jk}x_{1ijk} + \beta_{2jk}x_{2ijk} + \beta_{3j}x_{3ijk} + \beta_{4jk}x_{4ijk} + \beta_{5jk}x_{5ijk} \quad \text{................(vi)}$$

By randomizing β_{0jk} we have,

$$\beta_{0jk} = \beta_{00k} + u_{0jk} \ \text{(level two)}\text{..(vii)}$$

$$\beta_{00k} = \beta_{000} + \tau_{00k} \ \text{(level three)}\text{...(viii)}$$

The reduced form is derived by the simple substitution of Equation (viii) into Equation (vii), which results in

$$\beta_{0jk} = \beta_{000} + \tau_{00k} + u_{0jk} \ \text{...(ix)}$$

By substituting equation (ix) in (vi) the model result in

$$\ln\left(\frac{p_{ijk}}{1-p_{ijk}}\right) = \beta_{000} + \tau_{00k} + u_{0jk} + \beta_{1jk}x_{1ijk} + \beta_{2jk}x_{2ijk} + \beta_{3j}x_{3ijk} + \beta_{4jk}x_{4ijk} + \beta_{5jk}x_{5ijk}$$

..........................(x)

The random intercept (denoted β_{0jk}) is thus expressed as an additive function of a grand mean (β_{000}) and a group-levels deviation from this mean are u_{0jk} and τ_{00k}.

The random effect are assumed to be normally distributed that is;

$$\begin{bmatrix} u_{0jk} \\ \tau_{00k} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}\begin{bmatrix} \sigma^2_{u0} \\ \sigma^2_{\tau 0} \end{bmatrix}\right)$$

The parameter of the above equation (fixed effect, random effect, variance of the random effect and residual variance) are simultaneously estimated using iterative method (Bryk and Raudenbush 1992, Kreft and deleeuw 1998, Goldstein 2003, Hasinur et al 2011)

3.10 Marginal Likelihood

In order to estimate the parameters of hierarchical generalized linear model (HGLM) one usually makes use of marginal maximum likelihood estimation. In this method, the marginal likelihood of the observed data, obtained by integrating out the distribution of the random effects, marginal likelihood L(y) (conditional on the covariates), can be written as ;

$$L(y) = \int \pi_{j=1}^{N} \pi_{i=1}^{n_j} f_{y_{ij}/u_j}(y_{ij}/u_j) f_{u_j}(u_j)du_j = \pi_{j=1}^{N} \pi_{i=1}^{n_j} f_{y_{ij}/u}(y_{ij}/u) f_u(u)du \text{ ....(11)}$$

where $L(y)$ depends on unknown parameters $\gamma_0, \gamma_1, \gamma_2, \gamma_3, \sigma_0, \sigma_1 . \sigma_{01}$ which are the random effect parameters. The likelihood (11), can thus be considered as a product of independent contributions from each of N clusters. In general the integral (11) has no closed form and needs to be evaluated numerically. Maximization of the likelihood proceeds then by standard methods such as the EM algorithm. Since the likelihood needs to be evaluated many times during the iterative maximization procedure, fast but reliable approximations to (11) are needed.

3.11 Gaussian Quadrature Methods

An alternative approach is to approximate the integral (11) by numerical integration and then to maximize the likelihood with approximate values for the integrals. Numerical integration proceeds by Gauss-Hermite quadrature formula:

$$\int_{-\infty}^{\infty} h(v)e^{-v^2}\,dv = \sum_{q=1}^{d} h(x_q)\,w_q \;\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(12)$$

where h is a smooth function. Here $x_1,\ldots,x_d$ are the quadrature points, and $w_1,\ldots,w_d$ are the associated weights summing to one. The larger d (the number of quadrature points), the better the approximation in (12). For a given d, quadrature points and weights are tabulated. Note that, since the distribution of the random effects is supposed to be normal, the integrals appearing in (11) are of the above form. The estimator obtained by maximizing the likelihood approximated in this way is called the Non adaptive Gaussian quadrature (NAGQ) estimator. Gauss-Hermitian quadrature can be poor for functions that are not properly centered or non-smooth (McCulloch and Searle, 2001). A likely reason for this is that in these conditions, the cluster-specific integrands have very sharp peaks that may be located between adjacent quadrature points (Lesaffre and Spiessens, 2001). The performance of Gaussian quadrature can be improved by integration methods that are called adaptive in the sense that they take into account the properties of the integrand. Such methods scale and translate the quadrature locations to place them under the peak of the integrand. In this way, the position of the quadrature points may vary from cluster to cluster. For more detail, we refer to (Pinheiro and Bates 1995) who developed such an improvement over non adaptive Gaussian quadrature in the context of two-level random coefficient models. Since the quadrature points need to be scaled and translated, computing the approximations of the integral will be more time consuming for a fixed value. But, since the quadrature points will now be placed much more central in the region of interest, the approximation will be much more accurate, allowing for a smaller number of quadrature points. The resulting estimator will be called the Adaptive Gaussian quadrature estimator (AGQ). Details of the quasi likelihood methods are given below.

3.12 Relationship between Marginal and Penalized Quasi-Likelihood (MQL and PQL)

Maximum likelihood (ML) and Restricted Maximum Likelihood (REML) are relevant to linear multilevel models with Gaussian outcomes; different likelihood based methods are needed with models for dichotomous outcomes, such as (9). Following Goldstein (1995), in the simpler case of a two-level structure a reasonably general multilevel model for the binary outcome yij has the form (yij / pij) ~Bernoulli(pij) with

$$p_{ij} = f(x_{ij}\beta + z^1_{ij}e_{ij} + z^2_{ij}u_j) \;\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(11)$$

let $l = x_{ij}\beta + z_{ij}^{1}e_{ij} + z_{ij}^{2}u_{j}$

where f(l) has a nonlinear character such as $\text{logit}^{-1} = \frac{1}{(1+e^{-l})}$ One approach to the fitting of (11) is through quasi-likelihood methods, which proceed (e.g., Breslow and Clayton 1993) by linearizing the model via Taylor series expansion; for instance, with H_t as a suitably chosen value around which to expand, the f(l) expression in (11) for the ijth unit at iteration (t + 1) may be approximated by

$$f(H_t) + x_{ij}(\beta_{t+1} - \beta_t)f'(H_t) + (z_{ij}^{(1)}e_{ij} + z_{ij}^{(2)}u_j)f'(H_t) + \frac{1}{2}(z_{ij}^{(1)}e_{ij} + z_{ij}^{(2)}u_j)^2 f''(H_t) \text{..........}(12)$$

in terms of parameter values estimated at iteration t. The simplest choice, $H_t = x_{ij}\beta_t$, the fixed part predicted value of the argument of f in (5), yields the marginal quasi-likelihood (MQL) algorithm.

This can be improved upon by expanding around the entire current predicted value for the ijth unit, $H_t = x_{ij}\beta_t + z_{ij}^{(1)}e_{ij} + z_{ij}^{(2)}u_j$. where $\hat{e}_{ij}$ and $\hat{u}_j$ are the current estimated random effects; when this is combined with an improved approximation obtained by replacing the third and forth term in (12) with

$$[z_{ij}^{(1)}(e_{ij} - \hat{e}_{ij}) + z_{ij}^{(2)}(u_j + \hat{u}_j)]f'(H_t) + \frac{1}{2}[z_{ij}^{(1)}(e_{ij} - \hat{e}_{ij}) + z_{ij}^{(2)}(u_j + \hat{u}_j)]^2 f''(H_t) \text{.........}(13)$$

the result is the penalized or predictive quasi-likelihood (PQL) algorithm. The order of an MQL or PQL algorithm refers to how many terms are used in the Taylor expansion underlying the linearization; for example, equation (12) is based on expansion up to second order and leads to MQL2 and PQL2 estimates. Estimated asymptotic standard errors for MQL/PQL estimates typically derive from a version of observed Fisher information based on the quasi-likelihood function underlying the estimation process Breslow and Clayton (1993).

3.13 Penalized Quasi-Likelihood

The PQL estimation procedure is described here for three level logistic regression models. Consider a level-1 outcome Y_{ijk} taking on a value of 1 with conditional probability pij. Then the logit model or the generalized linear model is,

$$\ln\left[\frac{p_{ijk}}{1-p_{ijk}}\right] = n_{ijk} = \gamma x_{ijk} + z_{ijk}u_{jk} \text{.................................} (14)$$

for level-1 unit i nested within level-2 unit j which is also nested in level- 3 unit k. At level 1, one can assume Y_{ijk} conditionally distributed as Bernoulli, while the random effects vector u_{jk} is distributed as $N(0, \sigma^2 u)$ across the level-2 units. Let variance σ_u^2 be T throughout this PQL estimation procedure. The PQL approach can be derived as a nonlinear regression model. In the case of binary outcomes with logit link, we start with the level-1 model

$$Yij = p_{ij} + e_{ij}, \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (15)$$

where $E(e_{ij}) = 0$ and $Var(e_{ij}) = p_{ij}(1 - p_{ij})$. This is a nonlinear model which we linearize by means of the first-order Taylor series expansion. At this iteration, we have

$$p_{ijk} \approx p_{ijk}^{(s)} + \frac{dp_{ijk}}{dn_{ijk}}(\mathrm{n}_{ijk} - \mathrm{n}_{ijk}^{(s)}) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (16)$$

And evaluate the derivative

$$\frac{dp_{ijk}}{dn_{ijk}} = p_{ijk}(1 - \mathrm{p}_{ijk}) = \omega_{ijk} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (17)$$

At $p_{ijk}^{(s)}$. Substituting the linear approximation for p_{ijk} in equation (15) yields

$$y_{ijk} = p_{ijk}^{(s)} + \omega_{ijk}^{(s)}(\mathrm{n}_{ijk} - \mathrm{n}_{ijk}^{(s)}) + \mathrm{eij}. \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (18)$$

Algebraically, rearrange this equation so that all known quantities are on the left- hand side of the equation produces

$$\frac{Y_{ijk} - p_{ijk}^{(s)}}{\omega_{ijk}^{(s)}} + n_{ijk}^{(s)} = n_{ijk} + \frac{e_{ijk}}{\omega_{ijk}^{(s)}} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (19)$$

This equation has the form of the familiar three- level hierarchical linear model

$$Y_{ijk}^{(s)} = X_{ijk}^{T}\gamma + Z_{ijk}^{T}u_j + \varepsilon_{ijk} \ldots\ldots\ldots\ldots\ldots\ldots\ldots (20)$$

Which gives a straightforward updating scheme. This is known as penalized quasi- likelihood (involving only 1st and 2nd ordered derivatives) with a penalty term on the random effect. Here ,

$$Y_{ijk}^{(s)} = \frac{(Y_{ijk} - P_{ijk}^{(s)})}{\omega_{ijk}^{(s)} + n_{ijk}^{(s)}}$$

$$\varepsilon_{ijk} = \frac{e_{ijk}}{\omega_{ijk}^{(s)}} \sim N(0,T). \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots(21)$$

The estimate of $n_{ijk}^{(s)}$ can be written as below

$$n_{ijk} = \gamma^{(s)} x_{ijk}^{T} + z_{ijk}^{T} u_{j}^{(s)} + v_{0k} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots(22)$$

3.14 Performance Measures

I examined both the bias and efficiency of the estimates. Bias indicates whether a parameter tends to be over- or underestimated, and is computed as the difference between the mean of the estimates (across samples) and the true value, or

$$B = E\left(\hat{\theta}_r\right) - \theta$$

where θ is the parameter of interest, $\hat{\theta}_r$ is the estimate of θ for replication r, and $E(\hat{\theta}_r)$ is the mean estimate across replications. A good estimator should have bias values near zero, indicating that the sample estimates average out to equal the population value. Bias of 5-10% is often considered tolerable (Kaplan, 1989). The accuracy of the parameter estimates is also quantified by percentage relative bias for parameter $(\theta) = \left(\frac{\hat{\theta}_r - \theta}{\theta}\right) * 100$. (Maas and Hox 2005)

Likewise, to evaluate efficiency, one can examine the variance of the estimates,

$$v = E\left|\left(\hat{\theta} - E\left(\hat{\theta}_r\right)^2\right)\right|$$

A good estimator will have less variance than other estimators, indicating more precision and, typically, higher power for inferential tests. Bias and variance should be considered simultaneously when judging an estimator. For instance, an unbiased estimator with high variance is not very useful, since the estimate obtained in any single sample is likely to be quite far from the population value. Another estimator may be more biased, but have low variance, so that any given estimate is usually not too far from the population value. An index which combines both bias and variance is the Mean Squared Error (MSE), which is computed as the average squared difference between the estimate and the true parameter value across samples (Daniel and Sonya 2011). -2loglikelihood, Akaike's information criteria and Bayesian information criteria was also used to detect the best model obtained from the three method of parameter estimation.

3.15 Residual Intra class Correlation Coefficient

In a multilevel model, the sources of variation could be within –group and between groups. In this study, the total variation in individual outcomes can be partitioned into two variance component: within the group variance (that is, variance among individual in the same cluster group and in the same geo-political zone) and between the group variance (that is, variance between individual in different cluster and cluster in different geo-political zone). Thus, when individual within group are very similar to each other, less information is obtained compared to when the same number of individual is obtained compared to when the same number of individuals is obtained in an un clustered sample (that is , by simple random sample). The amount of variation in the use of modern contraceptive explained by the cluster variable and geo political zone variable is known as Intra class correlation coefficient(ICC). It is a measure that describe the dependencies in the data and it measure the extent to which individuals within the same group are more similar to each other than they are to individual in different groups. It is a population estimate of the variance explained by the grouping structure, which is equal to the estimated proportion of group level variance compared to the estimated total variance. For binary responses, the ICC is often expressed in term of the correlation between the latent responses. The logistic distribution for the level one residual e_{ij} implies a variance of $\Pi^2/3=3.29$. This implies that for three level logistic random intercept model the ICC level three is $\rho = \dfrac{\sigma_{r0}^2}{\sigma_{r0}^2 + \sigma_{u0}^2 + \Pi^2/3}$

Where σ_{r0}^2 is the level three constant variance

σ_{u0}^2 is the level two constant variance

ICC for level two is $\rho = \dfrac{\sigma_{u0}^2 + \sigma_{r0}^2}{\sigma_{r0}^2 + \sigma_{u0}^2 + \Pi^2/3}$

3.16 Akaike's Information Criteria(AIC) and Bayesian Information Criteria(BIC)

AIC and BIC are both penalized-likelihood criteria. They are sometimes used for choosing best predictor subsets in regression and often used for comparing non nested models, which ordinary statistical tests cannot do. The AIC or BIC for a model is usually written in the form [$-2\log L + kp$], where L is the likelihood function, p is the number of parameters in the model, and k is 2 for AIC and $\log(n)$ for BIC.AIC is an estimate of a constant plus the relative distance between the unknown true likelihood function of the data and the fitted likelihood function of the model, so that a lower AIC means a model is considered to be closer to the truth. BIC is an estimate of a function of the posterior probability of a model being true, under a certain Bayesian setup, so that a lower BIC means that a model is considered to be more likely to be the true model (John dziak et al 2012).Both criteria are based on various assumptions and asymptotic approximations. Each, despite its heuristic usefulness, has therefore been criticized as having questionable validity for real world data. But despite various subtle theoretical differences, their only difference in practice is the size of the penalty; BIC penalizes model complexity more heavily. The only way they should disagree is when AIC chooses a larger model than BIC.

3.17 Data Management and Analysis

SPSS version 20 was used for data cleaning and also for fitting mixed effect model which represent penalized quasi likelihood method which is the only estimator currently used in SPSS (Daniel and Sonya 2011) and STATA 12 was used for estimating single level fixed effect and multilevel fixed and random effect parameter, by using both Adaptive Gauss-Hermit quadrature and non Adaptive Gaussian quadrature (Laplacian approximation).

GENLINMIXED Syntax was written for penalized quasi likelihood method on SPSS while GLLAMM was downloaded on STATA version 12 which was used for adaptive Gaussian quadrature with 15 integration point and XTMELOGIT is default on STATA software. However, XTMELOGIT syntax was used for Laplacian approximation which does not use any quadrature point (that is, integration point) . Microsoft excel 2007 was used for all mathematical calculation (that is, for calculating the estimates that are over estimated or underestimated) , stop watch was also used for obtaining the computational time for the four syntax (GENLINMIXED, GLLAMM, XTMELOGIT for Laplacian approximation and for adaptive Gaussian quadrature.

CHAPTER FOUR

4.0 Result

In this chapter equation (X) in previous chapter which is three levels five predictors logistic regression with random intercept and fixed slope was considered. The three multilevel methods of parameter estimation applied in this study were (PQL NAGQ and AGQ) using GENLINMIXED, XTMELOGIT and GLLAMM syntax and maximum likelihood method was used for single level binary logistic regression which is the standard logistic regression method.

4.1 Three Level Intercept Only Multilevel Logistic Model

From Table 1 below, the fixed and random intercept for three level in all the methods are significant except the random effect at level three. And the standard logistic regression which is single level model overestimate the parameter compare to the multilevel methods and also the random effect for the third level using XTMELOGIT syntax was approximately zero in the Laplacian approximation (NAGQ) and adaptive Gaussian quadrature methods in which their intral class correlation (that is within the zone correlation) is zero. This implies that using geo-political zone or region as a level is not reliable. Table 2 shows level three model comparison using -2log-likelihood, Akaike's information criteria and Bayesian information criteria, it was discovered that among the intercept only model for three levels using the quasi and the full maximum likelihood methods with different syntax(GENLINMIXED, XTMELOGIT and GLLAMM) , Adaptive Gaussian quadrature using GLLAMM syntax have the smallest -2logL(21191.626), AIC (21197.626) and BIC(21222.673) and from table-1 AGQ with GLLAMM syntax have the smallest standard error for both fixed and random effect except for level three which is the regional level. which implies that AGQ with GLLAMM syntax is the best for fitting three levels model.

4.1.1 Intra Class Correlation for Three Levels Intercept Only Model

From Table-1 below, the intra cluster correlation coefficient for multilevel methods reduced from AGQ using GLLAMM (32%) to PQL(14%). in which AGQ (XTMELOGIT) have 31% of the total variance that was explained by the variance within the cluster, while AGQ(GLLAMM), NAGQ(XTMELOGIT) and PQL has 32%, 30%and 14% of the total variance that was explained by variance in the cluster respectively. And for the geo political zone AGQ (GLLAMM) have 11% of the total variation that was explained by the variance within the zone while AGQ(XTMELOGIT) and NAGQ (XTMELOGIT) has zero percent of total variance that was explain by variance across the zone and PQL has 10% of the total variance that is explained by the variance within the zone.

Table3 and 4 below is for two level random intercept model.

Table 1: Three-level estimates of multilevel analysis using an intercept only single level and multilevel logistic model to predict modern contraceptive use

Model Effect	Standard logistic	PQL	NAGQ	AGQ XTMELOGIT	AGQ GLLAMM
Fixed effect Intercept	-1.95** (0.171)	-1.987** (0.180)	-2.428** (0.046)	-2.430** (0.047)	-2.425** (0.030)
σ^2_{u0} (BCV)		0.172 (0.245)	1.411717 (.100)	1.457 (.104)	1.012 (.076)
σ^2_{z0} (BZV)		0.395 (0.170)	5.17e-07 (.001)	2.12e-08 (0.001)	0.5458 (.0211)
Intra CCC		0.147	0.300	0.307	0.321
Intra ZCC		0.102	1.10e-7	4.47e-9	0.115
-2logL	23418.350	152294.570	21503.090	21490	21191.626
AIC	23427.350	152300.570	21509.090	21496	21197.626
Iteration	1	3	13	11	3
Computation	30sec	1min,30sec	3mins,3secs	6 Mins	3hrs,5mins and16secs
N	31235	31235	31235	31235	31235

Note: BCV Between the Cluster Variance, BZC Between the Zone Variance, CCC Cluster Correlation Coefficient, ZCC Zonal correlation coefficient. The symbol ** and * indicate that the estimate is significant at 0.01 and 0.05 respectively. Standard error in parenthesis.

Table 2: Comparison of Different Multilevel Methods using Three-Levels Intercept only Model

	-2LOGL	AIC	BIC
PQL	162294.571	162296.571	162304.920
AGQ(XTMELOGIT)	21496	21496.133	21521.181
AGQ(GLLAMM)	21191.626	21197.626	21222.673
NAGQ	21503.090	21509.086	21534.134

4.2 Two Level Intercept Only Multilevel Logistic Model

Here a simple model was fitted with no predictors for level two (i.e. an intercept-only model) that predicts the probability of modern contraceptive use. The estimates of parameters and standard errors are presented in Table 3 below. The maximum likelihood (ML) estimate from the standard logistic model of the ratio of modern contraceptive user to Modern contraceptive nonuser is $e^{(-1.955)} = 0.142$, which is the same as the sample ratio of 3874 modern contraceptive users to 28000 nonusers. It is the odds-ratio when no predictors have been considered in the model. In comparison, the same ratio is estimated to be $e^{(-1.987)}=0.137$, $e^{(-2.428)}=0.088$ and $e^{(-2.430)}=0.088$ from the multilevel model by the PQL, NAGQ and AGQ methods respectively.

A crude (each of mean and median is a measure of central tendency) comparison has been made to understand the multilevel effects. Compared to the odds-ratios obtained by all multilevel estimation methods, the standard logistic model odds-ratio has overestimated. It is observed that there is a significant difference between the standard logistic estimate and the multilevel logistic estimate. Therefore, by failing to take into account the clusters (level 2), the standard logistic model has overestimated the odds-ratio by about 2% [((-1.954)-(-1.987))*100/(-1.987)], 19% and 20% compared to multilevel model using by the corresponding methods PQL, NAGQ and AGQ (Table3). Since bias of 5-10% is often considered tolerable (Kaplan, 1989) then PQL give better estimate in term of bias. The random quantity at cluster level is under estimated for PQL compared to full likelihood method. However, the full likelihood have the smallest standard error.

4.2.1 Convergence of the Estimation Methods in Two Levels

In Table 3 below, Adaptive Gaussian Quadrature (AGQ) with XTMELOGIT syntax converge at iteration eleven after six minutes of computation, Adaptive Gaussian quadrature (AGQ) with GLLAMM syntax converge at iteration three after three hours, five minutes and sixteen seconds of computation, laplacian approximation(NAGQ) with XTMELOGIT syntax converges at iteration thirteen after three minutes and three seconds and also the penalized quasi likelihood(PQL) converges after third iteration where the estimate converge after one minutes and thirty seconds. Table 4 also shows that AGQ method using GLLAMM and XTMELOGIT have the smallest -2logL(21490.132), AIC (21496.132) and BIC(21510.831)among the multilevel methods even when considering the standard logistic method of estimation. AGQ method with XTMELOGIT to all other methods is the best for two level when the log likelihood estimate, Akaike information criteria and Bayesian information criteria was considered.

4.2.2 Random Effect of Two Levels Intercept Only Model

The parameters under random effect in Table 3 was the estimated variances of the random intercepts at level 2 for fitting a two-level intercept-only model. To understand the random effect in this two-level intercept-only model, one can imagine a unique effect for each cluster (level 2) in addition to the fixed intercept of -2.430 (AGQ estimate with XTMELOGIT), -2.425(AGQ estimate with GLLAMM), -2.428 (NAGQ estimate with XTMELOGIT) and -1.874 (PQL estimate with GENLINMIXED) which is the average of modern contraceptive use in all cluster. The addition of the cluster specific effects makes the model more accurate than the fixed intercept only model. In the random effect model, the cluster effects are assumed to be distributed normally for the purpose of estimation. In Table 3 the estimate of the random effect at levels two does increase from PQL to NAGQ and even to AGQ. And the standard error of the random effect in Adaptive Gaussian quadrature using XTMELOGIT is the smallest which implies that AGQ using XTMELOGIT is more efficient.

4.2.3 The Predicted Probability of Modern Contraceptive use

When the multilevel AGQ method with XTMELOGIT syntax is applied, the expected log-odds of modern contraceptive use is -2.430, corresponding to an odds of $e^{(-2.430)}=0.088$.

This corresponds to a predicted probability of $1/(1 + e^{(-2.430)}) = 0.919$.

For AGQ with GLLAMM: the expected log-odds of contraceptive use is -2.430.

This corresponds to a predicted probability of $1/(1 + e^{(-2.430)}) =0.919$

For NAGQ: the expected log-odds of contraceptive use is -2.428. This corresponds to a predicted probability of $1/(1 + e^{(-2.428)}) = 0.919$.

For PQL: the expected log-odds of contraceptive use is -1.987. This also corresponds to a predicted probability of $1/(1 + e^{(-1.874)}) = 0.879$.

For the standard logistic model which is single level model, the predicted probability is

$1/(1 + e^{(-1.955)}) = 0.876$. based on the estimate for the predicted probability, multilevel estimation methods provide an estimate that have higher prediction compared to that of standard logistic estimation method but estimated value for PQL is very close to standard logistic regression.

Table 3: Two-level estimates of multilevel analysis using an intercept only single level and multilevel logistic model to predict modern contraceptive use

Model Effect	Standard logistic	PQL	NAGQ	AGQ XTMELOGIT	AGQ GLLAMM
Fixed effect Intercept	-1.955** (0.017)	-1.874** (0.294)	-2.428** (.046)	-2.430** (0.046)	-2.430** (0.047)
σ^2_{uo} (BCV)		0.172 (0.245)	1.412 (0.100)	1.207 (.0.043)	1.456 (0.104)
Intra CCC		0.0497 (5%)	0.300 (30%)	0.268 (27%)	0.307 (31)
-2logL	23418.35	158912.752	21503.090	21490.132	21490.132
AIC	23422.35	158916.752	21507.090	21494.132	21494.132
Iteration	1	6	3	2	3
Computation	30sec	45seconds	54seconds	1 MINS and 24 secs	3MINS And 16secs
Number of observation	31235	31235	31235	31235	31235
Number of group	1076	1076	1076	1076	1076

Note: BCV Between the Cluster Variance, CCC Cluster Correlation Coefficient. The symbol ** and * indicate that the estimate is significant at 0.01 and 0.05 respectively.

For NAGQ: the expected log-odds of contraceptive use is -2.428. This corresponds to a predicted probability of $1/(1 + e^{(-2.428)}) = 0.919$.

For PQL: the expected log-odds of contraceptive use is -1.987. This also corresponds to a predicted probability of $1/(1 + e^{(-1.874)}) = 0.879$.

For the standard logistic model which is single level model, the predicted probability is

$1/(1 + e^{(-1.955)}) = 0.876$. based on the estimate for the predicted probability, multilevel estimation methods provide an estimate that have higher prediction compared to that of standard logistic estimation method but estimated value for PQL is very close to standard logistic regression.

Table 3: Two-level estimates of multilevel analysis using an intercept only single level and multilevel logistic model to predict modern contraceptive use

Model Effect	Standard logistic	PQL	NAGQ	AGQ XTMELOGIT	AGQ GLLAMM
Fixed effect Intercept	-1.955** (0.017)	-1.874** (0.294)	-2.428** (.046)	-2.430** (0.046)	-2.430** (0.047)
$\hat{\sigma}^2_{u0}$ (BCV)		0.172 (0.245)	1.412 (0.100)	1.207 (.0.043)	1.456 (0.104)
Intra CCC		0.0497 (5%)	0.300 (30%)	0.268 (27%)	0.307 (31)
-2logL	23418.35	158912.752	21503.090	21490.132	21490.132
AIC	23422.35	158916.752	21507.090	21494.132	21494.132
Iteration	1	6	3	2	3
Computation	30sec	45seconds	54seconds	1 MINS and 24 secs	3MINS And 16secs
Number of observation	31235	31235	31235	31235	31235
Number of group	1076	1076	1076	1076	1076

Note: BCV Between the Cluster Variance, CCC Cluster Correlation Coefficient. The symbol ** and * indicate that the estimate is significant at 0.01 and 0.05 respectively.

Table 4:Comparison of different multilevel methods using two-level intercept only model

	-2LOGL	AIC	BIC
PQL	158912.752	158914.752	158912.752
AGQ (XTMELOGIT)	21490.132	21494.133	21510.831
AGQ(GLLAMM)	21490.132	21494.133	21510.831
NAGQ	21503.090	21507.086	21523.785

4.3 Multilevel Univariate Logistic Model

4.3.1 Comparison Between Single Level and Multilevel Estimates

In the multilevel univariate analysis represented in Table5,6,7 and 8 below, each of the models presents a random intercept and a fixed slope for the variable. Column two of the tables are the effects of individual predictor ($\hat{\beta}$) obtained from the standard (single level) logistic regression, where the 3rd column of the Tables represents odds ratios ($\dot{\psi}$) of the standard logistic model. In standard logistic regression, the odds of outcome for a non reference case in a predictor variable divided by the odds of outcome for a reference case for the same predictor variable does not depend on the level. Thus, although odds ratios ($\dot{\psi}$) can be calculated from the effect of those predictors ($\hat{\beta}$). To correctly interpret the parameter estimates related to predictors in a multilevel model, it is more meaningful to state that the individual estimates increase or decrease the contribution of the explanatory variables on the outcome. Column five and six in Table 5,6,7 and 8 presented the percentage increase or decrease of the estimate. β coefficients was presented (for notational convenience, β_s for single level and β_m for multi-level; in Table-5, 6,7 and 8) for the four type of models.

It was observed that there exist significant differences between the β coefficients (that is, β_s for single level and β_m for multi-level) of these four models for each of the explanatory variables. Also the β coefficients of primary predictors (that is, reproductive age group) in the standard (single level) logistic model have been underestimated in comparison with the multilevel models while for other covariate some were either overestimated or underestimated. The difference between β coefficients estimated of the multilevel models and standard model arises because of the addition of the random effects which is the cluster level effect in the multilevel, this implies that using single level model for modern contraceptive use in clustered survey data is not appropriate. It also implies that it is not only the fixed variables that contribute to the use of modern contraceptive use but the Table-5,6,7 and 8 shows that cluster level have a significant effect on the use of modern contraception.

4.3.2 Significance of the Estimates

Table5,6,7 and 8 below compared the level at which the multilevel logistic regression and standard logistic regression were significant. From the analysis, age group of the respondent was found significantly associated with modern contraceptive use in the three methods of multilevel binary logistic regression at one percent level of significant (p-value <0.001) including standard logistic regression. The wealth index was also found to be significantly associated with modern contraceptive use at one percent level of significant (p-value <0.001) in all the methods, among the education category, Qur'anic only was not significant while others were significant at one percent level of significant in both standard and multilevel logistic regression (p-value <0.001), in religion categories, traditional religion which was significant at one percent level of significant (That is. p-value <0.001) , it is significant at five percent level of significant (that is, p-value <0.05) in penalized quasi likelihood(PQL) and not significant in all the full likelihood methods. For place of residence (POR), the significant occur at one percent level of significant in standard logistic regression method and it was significant at five percent level of significant in all the multilevel logistic regression methods.

Table-5,6,7 and 8 also show that the -2log likelihood and Akaike's information criteria estimate for multilevel model is less than that of the standard logistic regression which even shows that the model obtain with the multilevel is better than model of standard logistic regression though multilevel methods have a longer computational time than standard logistic regression.

Table 5: Two-level estimates of univariate single-level and multilevel logistic model predicting the probability of contraceptive use with random intercept and fixed effect using PQL method.

	SINGLE LEVEL		MULTILEVEL	
	$\hat{\beta}_s$	$\hat{\psi}$	$\hat{\beta}_m$	Under estimated(%)
Constant	-4.789 (0.107)**	0.008	-4.883(0.135)**	2
Age				0
20-24	1.095(0.075)**	2.988	1.094(0.085)**	0
25-29	1.372(0.074)**	3.944	1.372(0.084)**	0
30-39	1.249(0.071)**	3.488	1.249(0.081)**	0
40-49	1.063 (0.077)**	2.896	1.063(0.087)**	0
50-64	0.5788 (0.010)**	1.784	0.579(0.150)**	0
Wealth Index				
Poorer	0.408(0.076)**	1.503	0.408(0.086)**	0
Average	0.581(0.074)**	1.789	0.584(0.085)**	0
Wealthier	0.629(0.077)**	1.876	0.634(0.097)**	1
Wealthiest	0.682(0.081)**	1.977	0.688(0.091)**	0
Education				
Quranic only	-0.204(0.155)	0.815	-0.205(0.175)	0
Primary	0.945(0.080)**	2.572	0.945(0.090)**	0
Secondary	1.241(0.076)**	3.460	1.242(0.086)**	1
Higher Education	1.507(0.083)**	4.514	1.508(0.093)**	1
Religion				
Non catholic Xtian	0.655(0.047)**	1.925	0.653(0.067)**	0
Catholic	0.708(0.060)**	2.031	0.705(0.079)**	0
Traditional	0.472(0.2380)*	1.603	0.470(0.288)*	0
No Religion	0.875(0.275)**	2.398	0.875(0.295)**	0
Others	-0.184(0.405)	0.832	- 0.185(0.455)	0
POR	-0.182(0.042)**	0.834	-0.091(1.544)*	100
σ^2 intercept			0.180(0.427)	
Intral CCC			0.0519	
-2logl	20760.79		158912.752	
AIC	20977.79		158914.752	
Iteration	5		20	
Computation	36 seconds		39seconds	
N	31135		31135	

Note: BCV Between the Cluster Variance, CCC Cluster Correlation Coefficient The symbol ** and * indicate that the estimate is significant at 0.01 and 0.05 respectively.

Reference categories are: '15-19' for Age, 'urban' for POR, 'No formal education' for Education, 'Islam' for Religion, and 'Poorest' for WI Standard errors are placed in parentheses.

Table 6:Two-level estimates of univariate single-level and multilevel logistic model predicting the probability of contraceptive use with random intercept and fixed effect using AGQ(GLLAMM syntax) method.

	Single Level		Multilevel		
	$\hat{\beta}_s$	$\hat{\psi}$	$\hat{\beta}_m$	Underestimate(%)	Over estimate(%)
Constant	-4.789(0.107)**	0.008	-5.094(0.130)**	6	
Age					
20-24	1.095(0.075)**	2.988	1.204 (0.080)**	9	
25-29	1.372(0.074)**	3.944	1.5142(0.079)**	9	
30-39	1.249(0.071)**	3.488	1.408(0.076)**	11	
40-49	1.063 (0.077)**	2.896	1.178(0.081)**	10	
50-64	0.5788(0.010)**	1.784	0.706(0.105)**	18	
Wealth Index					
Poorer	0.408(0.076)**	1.503	0.409(0.084)**	0	
Average	0.581(0.074)**	1.789	0.614(0.087)**	5	
Wealthier	0.629(0.077)**	1.876	0.738(0.091)**	14	
Wealthiest	0.682(0.081)**	1.977	0.883(0.097)**	23	
Education					
Quranic Only	-0.204(0.155)	0.815	-0.108(0.161)	89	
Primary	0.945(0.080)**	2.572	0.8020(0.084)**		18
Secondary	1.241(0.076)**	3.460	1.139(0.080)**	9	
Higher Education	1.507(0.083)**	4.514	1.464(0.088)**		3
Religion					
Non Catholic Xtian	0.655(0.047)**	1.925	0.504(0.059)**		30
Catholic Christian	0.708(0.060)**	2.031	0.606(0.075)**		17
Traditional	0.472(0.2380)*	1.603	0.407(0.257)		16
No Religion	0.875(0.275)**	2.398	0.972(0.293)**	10	
Others	-0.184(0.405)	0.832	-0.279(0.425)		33
POR	-0.182(0.042)**	0.834	-0.166(0.072)*	10	
σ^2_{uo} i(BCV)			0.681(0.056)		
IntralCCC			0.177		
-2logl	20760.79		19900.46		
AIC	20977.79		19944.46		
Iteration	5		6		

Computation	36 seconds		3hr:23mins			
N	31135		31135			

Note: Note: BCV Between the Cluster Variance CCC Cluster Correlation Coefficient. The symbol ** and * indicate that the estimate is significant at 0.01 and 0.05 respectively.

Reference categories are: '15-19' for Age, 'urban' for POR, 'No formal education' for Education, 'Islam' for Religion, and 'Poorest' for WI Standard errors are placed in parentheses.

Table 7: Two-level estimates of univariate single-level and multilevel logistic model predicting the probability of contraceptive use with random intercept and fixed effect using Laplacian approximation method.

	Single Level		Multilevel		
	$\hat{\beta}_s$	$\hat{\psi}$	$\hat{\beta}_m$	Underestimate(%)	Over Estimate(%)
Constant	-4.789 (0.107)**	0.008		6	
Age					
20-24	1.095(0.075)**	2.988	1.202(0.079)**	9	
25-29	1.372(0.074)**	3.944	1.512(0.079)**	9	
30-39	1.249(0.071)**	3.488	1.406(0.076)**	11	
40-49	1.063 (0.077)**	2.896	1.177(0.081)**	10	
50-64	0.5788(0.010)**	1.784	0.705(0.105)**	18	
Wealth Index					
Poorer	0.408(0.076)**	1.503	0.409(0.084)**	0	
Average	0.581(0.074)**	1.789	0.614(0.086)**	5	
Wealthier	0.629(0.077)**	1.876	0.738(0.091)**	15	
Wealthiest	0.682(0.081)**	1.977	0.882(0.969)**	23	
Education					
Quranic Only	-0.204(0.155)	0.815	-0.108(0.161)	88	
Primary	0.945(0.080)**	2.572	0.802(0.084)**		18
Secondary	1.241(0.076)**	3.460	1.139(0.080)**		9
Higher	1.507(0.083)**	4.514	1.464(0.088)**		3
Religion					
Non Catholic	0.655(0.047)**	1.925	0.307(0.039)**		133
Catholic	0.708(0.060)**	2.031	0.6088(0.0747)*		16
Traditional	0.472(0.2380)*	1.603	0.409(0.237)		15
No Religion	0.875(0.275)**	2.398	0.974(0.293)**	10	
Others	-0.184(0.405)	0.832	-0.276(0.425)		33
Por	-0.182(0.042)**	0.834	-0.166(0.072)*	9	
σ^2_{u0} I(BCV)			0.818(0.033)		
IntralCCC			0.199		
-2logl	20760.79		199492.3		
AIC	20977.79		199506.3		
Iteration	5		6		
Computation	36 Seconds		12 Minutes		
N	31135		31135		

Note: BCV Between the Cluster Variance CCC, Cluster Correlation Coefficient. The symbol ** and * indicate that the estimate is significant at 0.01 and 0.05 respectively.

Reference categories are: '15-19' for Age, 'urban' for POR, 'No formal education' for Education, 'Islam' for Religion, and 'Poorest' for WI Standard errors are placed in parentheses.

Table 8: Two-level estimates of univariate single-level and multilevel logistic model predicting the probability of contraceptive use with random intercept and fixed effects using AGQ method (XTMELOGIT syntax) .

	Single Level		Multilevel		
	$\hat{\beta}_s$	$\hat{\psi}$	$\hat{\beta}_m$	Underestimate(%)	Over Estimate(%)
Constant	-4.789	0.008	-5.094(0.130)**	6	
Age					
20-24	1.095	2.988	1.204(0.079)**	9	
25-29	1.372	3.944	1.514(0.079)**	9	
30-39	1.249	3.488	1.408(0.081)**	11	
40-49	1.063	2.896	1.178(0.081)**	10	
50-64	0.5788587	1.784	0.706(0.105)**	18	
Wealth Index					
Poorer	0.408	1.503	0.409(0.084)**	0	
Average	0.581	1.789	0.614(0.087)**	5	
Wealthier	0.629	1.876	0.738(0.091)**	15	
Wealthiest	0.682	1.977	0.883(0.097)**	23	
Education					
Quranic Only	-0.204	0.815	-0.108(0.161)	89	
Primary	0.945	2.572	0.802(0.084)**		18
Secondary	1.241	3.460	1.139(0.080)**	9	
Higher	1.507	4.514	1.4643(0.088)**	3	
Religion					
Non Catholic	0.655	1.925	0.504(0.059)**		30
Catholic	0.708	2.031	0.606(0.075)**	17	
Traditional	0.472	1.603	0.407(0.257)		16
No Religion	0.875	2.398	0.972(0.293)**	10	
Others	-0.184	0.832	-0.279(0.425)		34
Por	-0.182	0.834	-0.166(0.425)*		10
σ^2_{u0} I(BCV)			0.825		
IntralCCC			0.201		
-2logl	20760.79		19948.830		
AIC	20977.79		19996.830		
Iteration	5		6		
Computation	36 seconds		15 minutes		
N	31135		31135		

Note: BCV Between the Cluster Variance CCC, Cluster Correlation Coefficient. The symbol ** and * indicate that the estimate is significant at 0.01 and 0.05 respectively.

Reference categories are: '15-19' for Age, 'urban' for POR, 'No formal education' for Education, 'Islam' for Religion, and 'Poorest' for WI Standard errors are placed in parentheses.

4.3.3 Model Comparison for Three Level Methods of Estimation

Table 9 below shows the effect of geo-political zone on some factors that contributed to modern contraception using the three methods (PQL, NAGQ and AGQ), I discovered that the computational time for all the methods were longer than if I did not include the third level like the one in table-11 below , the fixed effect were significant at the same error rate except the quranic education and other types of religion that are not included in this study, but the fixed effect and the random intercept estimate for AGQ using XTMELOGIT was different from estimate of AGQ using GLLAMM despite the fact that both syntax are on the same quadrature point (that is, integration point (15)). Standard error for the estimate increase from PQL to AGQ using GLLAMM syntax. XTMELOGIT syntax for NAGQ and AGQ show that the random intercept for geo-political zone is zero which implies that geo political zone can not be level because the intral geo-political zone correlation coefficient is zero. GLLAMM syntax random intercept estimate for geo-political zone is 0.002(ICC=0.0003) which is also approximately zero. There is no different between the estimates obtained using only cluster as level and using cluster and zone as levels.

Also the -2logL, Akaike's information criteria (AIC) and Bayesian information criteria (BIC) in Table-10 above shows that adaptive Gaussian quadrature using GLLAMM have the smallest estimate compared to other multilevel syntax, which implies that when three levels is involve Adaptive Gaussian Quadrature using GLLAMM is the best.

Table 9:Three-level estimates of univariate multilevel quasi likelihood (PQL) and full maximum likelihood (NAGQ, AGQ with XTMELOGIT and AGQ with GLLAMM) methods.

	PQL	NAGQ	AGQ (XTMELOGIT)	AGQ(GLLAMM)
Constant	-4.844(0.189)**	-5.226(0.117)**	-5.226(0 .117)**	-5.050(0.117)**
Age				
20-24	1.164(0.086)**	1.204(0 .080)**	1.2052(0.080)**	1.2084(0.080)**
25-29	1.450(0.084)**	1.517(0.079)**	1.518(0 .079)**	1.519(0.093)**
30-39	1.289(0.079)**	1.412(0.077)**	1.414(0 .077)**	1.415(0.078)**
40-49	1.067(0.093)**	1.184(0.084)**	1.186(0 .084)**	1.184(0.084)**
50-64	0.589(0.121)**	0.704(0.105)**	0.705(0 .105)**	0.703(0.105)**
Wealth Index				
Poorer	0.396(0.087)**	0.421(0 .084)**	0.421(0 .084)**	0.360(0.085)**
Average	0.556(0.087)**	0.644(0.085)**	0.644(0 .086)**	0.538(0.087)**
Wealthier	0.617(0.093)**	0.786(0.089)**	0.786(0 .089)**	0.657(0.091)**
Wealthiest	0.696(0.097)**	0.946(0.093)**	0.946(0 .093)**	0.810(0.096)**
Education				
Quranic Only	-0.184(0.167)	-0.117(0.161)	-0.116(0 .161)	-0.073(0.163)
Primary	0.904(0.091)**	0.805(0.084)**	0.805(0 .084)**	0.749(0.085)**
Secondary	1.207(0.088)**	1.140(0.081)**	1.1400(0.082) **	1.086(0.082)**
Higher	1.475(0.096)**	1.468(.091)**	1.468(0.091)**	1.427(0.091)**
Religion				
Non-Catholic	0.548(0.065)**	0.493(0.059)**	0.491(0.059)**	0.347(0.064)**
Catholic	0.710(0.086)**	0.586(0 .074)**	0.584(0.074)**	0.472(0.080)**
Traditional	0.340(0.260)*	0.393(0.257)*	0.391(0.257)*	0.262(0.258)*
No Religion	0.631(0.299)**	0.954(0.293)**	0.953(0 .294)**	0.753(0.295)**
Others	- 0.353(0.438)	-0.286(0.425)	-0.289(0 .425)	-0.433(0.425)
Por	-0.522(0.086)	-0.003(0.010)	-0.003(0.010	-0.041(0.010)
σ^2_{uo} (BCV)	0.180(0.067)	0.673(0.055)	0.685(0.056)	0.684(0.057)

σ^2_{ro}(BZV)	0.055(0.066)	3.18e-06(.003)	2.12e-08(0.002)	0.002(0.064)
-2logl	169799.647	19957.610	19954.040	19900.460*
AIC	169791.647	20001.610	19998.040	19944.460
Iteration	20	6	6	6
Computation	2mins 23seconds	67mins	10hrs:21 mins	37hr:23mins
IntraICCC	0.067	0.183	0.172	0.172
IntraZCC	0.016 (2%)	0	0	0.00039
N	31135	31135	31135	31135

Note: BCV Between the Cluster Variance CCC Cluster Correlation Coefficient ZCC Zonal correlation coefficient BZC Between the Zone Variance. The symbol ** and * indicate that the estimate is significant at 0.01 and 0.05 respectively.

Reference categories are: '15-19' for Age, 'urban' for POR, 'No formal education' for Education, 'Islam' for Religion, and 'Poorest' for WI Standard errors are placed in parentheses.

Table 10: Comparison of different multilevel methods using three-level random intercept and fixed effect model.

	-2LOGL	AIC	BIC
PQL	169799.647	169791.647	169799.991
AGQ(XTMELOGIT)	19954.040	19998.040	20129.040
AGQ(GLLAMM)	19900.460	19944.460	20075.460
NAGQ	19957.610	20001.610	20132.610

4.3.4 Convergence of the Estimation Methods when Fixed Effects for two Level Models are included.

When the fixed effects were included in the two level model in the Table-11 below, there is change in computational time, convergence rate and the estimate of the intercept. Adaptive Gaussian quadrature (AGQ) with GLLAMM syntax) converge at iteration three after three hours, twenty-three minutes and six seconds of computation, . Adaptive Gaussian quadrature(AGQ) with XTMELOGIT syntax converge at iteration two after thirty minutes and twelve seconds of computation, Laplacian approximation(NAGQ) converges at iteration three after twelve minutes and twelve seconds and also the penalized quasi likelihood(PQL) converges after iteration twenty where the estimate converge after thirty-nine seconds. Table-12 above also shows that adaptive Gaussian quadrature (AGQ) have the smallest -2logL (19948.825)(19948.939), AIC(19992.825)(19992.939) and BIC (20176.439)(20176.553) estimates which is approximately the same for both XTMELOGIT AND GLLAMM syntax respectively among the multilevel methods even when considering the standard method of estimation. AGQ method to all other methods is the best when the log likelihood, AIC and BIC estimate were considered. Although it has longest computational time the fixed effect and the standard error for the AGQ with XTMELOGIT syntax and AGQ with GLLAMM syntax were equal and XTMELOGIT have shorter computational time for the convergence rate. XTMELOGIT syntax is therefore preferable in two level than GLLAMM syntax.

4.3.5 Variance Component for Two-Level Model.

The random effect of the two levels model in Table-11 below shows that the variance between the cluster in AGQ using XTMELOGIT is more than variance obtained from all the methods of parameter estimation,. Comparing variance of XTMELOGIT adaptive Gaussian quadratue to variance of GLLAMM adaptive Gaussian quadrature, the standard error for XTMELOGIT is smaller than that of every other methods which means the estimate obtained using XTMELOGIT syntax is better than estimate of every other methods when considering two levels binary logistic regression. Penalized quasi likelihood have the smallest random intercept with largest standard error and it also have the largest -2logL which minimized the reliability of the method.

4.3.6 Intra Cluster Correlation Coefficient

Among the intra cluster correlation coefficient of the three methods (AGQ, NAGQ and PQL) obtained from table-11, it was discovered that adaptive Gaussian quadrature using XTMELOGIT syntax gave the largest ICC result (ICC=0.201) which means 20% of the total variance is explained by the variance within the cluster. The penalized quasi likelihood method generate the smallest intral cluster correlation coefficient(ICC=0.052) which is also mean 5% of total variance is explain by the variance within the cluster.

Table 11: Two-level estimates of univariate multilevel quasi likelihood (PQL) and full maximum likelihood (NAGQ, AGQ with XTMELOGIT and AGQ with GLLAMM) methods.

	PQL	NAGQ	AGQ (XTMELOGIT)	AGQ(GLLAMM)
Constant	-4.883(0.135)**	-5.092(0.130)**	-5.094(0.130)**	-5.094(0.130)**
Age				
20-24	1.094(0.085)**	1.2020(0.079)**	1.2036(0.079)**	1.204(0.080)**
25-29	1.372(0.084)**	1.512(0.079)**	1.514(0.079)**	1.5142(0.079)**
30-39	1.249(0.081)**	1.406(0.076)**	1.408(0.081)**	1.408(0.076)**
40-49	1.063(0.087)**	1.177(0.0811)**	1.178(0.081)**	1.178(0.081)**
50-64	0.579(0.150)**	0.705(0.105)**	0.706(0.105)**	0.706(0.105)**
Wealth Index				
Poorer	0.408(0.086)**	0.409(0.0840)**	0.409(0.084)**	0.409(0.084)**
Average	0.584(0.085)**	0.614(0.086)**	0.614(0.087)**	0.614(0.087)**
Wealthier	0.634(0.097)**	0.738(0.091)**	0.738(0.091)**	0.738(0.091)**
Wealthiest	0.688(0.091)**	0.882(0.969)**	0.883(0.097)**	0.883(0.097)**
Education				
Quranic Only	-0.205(0.175)	-0.108(0.161)	-0.108(0.161)	-0.108(0.161)
Primary	0.945(0.090)**	0.802(0.084)**	0.802(0.084)**	0.802(0.084)**
Secondary	1.242(0.086)**	1.139(0.080)**	1.139(0.080)**	1.139(0.080)**
Higher	1.508(0.093)**	1.464(0.088)**	1.464(0.088)**	1.464(0.088)**
Religion				
Non-Catholic	0.653(0.067)**	0.307(0.039)**	0.504(0.059)**	0.504(0.059)**
Catholic	0.705(0.079)**	0.609(0.075)**	0.606(0.075)**	0.606(0.075)**
Traditional	0.470(0.288)*	0.409(0.237)	0.407(0.257)	0.407(0.257)
No Religion	0.875(0.295)**	0.974(0.293)**	0.972(0.293)**	0.972(0.293)**
Others	-[illegible].185(0.455)	-0.276(0.425)	-0.279(0.425)	-0.279(0.425)
Por	-0.091(1.544)*	-0.166(0.072)*	-0.166(0.072)*	-0.166(0.072)*
σ^2_{uo} (BCV)	0.180(0.427)	0.818(0.033)	0.825(0.034)	0.681(0.056)
Intral CCC	0.0519	0.199	0.201	0.177

-2logl	158912.752	199952.313	19948.83	19948.83
AIC	158914.752	19996.313	19962.83	19992.825
Iteration	20	3	2	3
Computation	39seconds	12 minutes and 6 seconds	30 minutes and 12 seconds	3hr:23mins and 6seconds
N	31135	31135	31135	31135

Note: BCV Between the Cluster Variance CCC Cluster Correlation Coefficient. The symbol ** and * indicate that the estimate is significant at 0.01 and 0.05 respectively.

Reference categories are: '15-19' for Age, 'urban' for POR, 'No formal education' for Education, 'Islam' for Religion, and 'Poorest' for WI Standard errors are placed in parentheses.

Table 12:Comparison of different multilevel methods using two-level random intercept and fixed effect model.

	-2LOGL	AIC	BIC
PQL	158912.752	158914.752	158923.101
AGQ(XTMELOGIT)	19948.825	19992.825	20176.439
AGQ(GLLAMM)	19948.939	19992.939	20176.553
NAGQ	19952.313	19996.313	20179.927

4.3.7 Comparing Quasi Likelihood with Full Maximum Likelihood

Table-13,14 and 15 below, show the increase and decrease in estimate of full likelihood methods compared to quasi likelihood method, penalized quasi likelihood method under estimate the primary predictor while other covariance effect were either under estimated or over estimated. The deviance between quasi and full likelihood methods are much in qur'anic education and the random effect , Traditional religion was significant at five percent level of significant in penalized quasi likelihood method and it was not significant in the quadrature methods that is, full maximum likelihood methods, also penalized quasi likelihood (PQL) method have the largest log likelihood estimate which implies that it is not the best method of parameter estimation for multilevel binary logistic regression though it converges earlier than every other methods. Base on the standard error of the estimate, it was discovered that penalized quasi likelihood have the largest standard error compared to the full maximum likelihood, this implies that the full likelihood methods were more précised compare to the quasi likelihood that was considered in this study.

Table 13: Two-level estimates of univariate multilevel quasi likelihood (PQL) and full maximum likelihood (NAGQ with XTMELOGIT) methods.

	PQL	XTMELOGIT (NAGQ)		
	$\hat{\beta}_m$	$\hat{\beta}_m$	Under estimated(%)	Overestimated(%)
Constant	-4.883(0.135)**	-5.092(0.130)**	4	
Age				
20-24	1.094(0.085)**	1.202(0.079)**	10	
25-29	1.372(0.084)**	1.512(0.079)**	10	
30-39	1.249(0.081)**	1.406(0.076)**	13	
40-49	1.063(0.087)**	1.177(0.081)**	11	
50-64	0.579(0.150)**	0.705(0.105)**	22	
Wealth Index				
Poorer	0.408(0.086)**	0.409(0.084)**	0	
Average	0.584(0.085)**	0.614(0.086)**	5	
Wealthier	0.634(0.097)**	0.738(0.091)**	16	
Wealthiest	0.688(0.091)**	0.882(0.969)**	28	
Education				
Quranic only	-0.205(0.175)	-0.108(0.161)		47
Primary	0.945(0.090)**	0.802(0.084)**		15
Secondary	1.242(0.086)**	1.139(0.080)**		8
Higher	1.508(0.093)**	1.464(0.088)**		3
RELIGION				
Non catholic	0.653(0.067)**	0.307(0.039)**		53
Catholic	0.705(0.079)**	0.609(0.075)**		14

Traditional	0.470(0.288)*	0.409(0.237)		13
No Religion	0.875(0.295)**	0.974(0.293)**	11	
Others	- 0.185(0.455)	-0.276(0.425)	49	
POR	-0.091(1.544)*	-0.166(0.072)*	83	
σ^2_{uo} (BCV)	0.180(0.427)	0.818(0.033)	355	
Intral CCC	0.0519	0.199		
-2logl	158912.752	19952.313		
AIC	158914.752	19996.313		
Iteration	20	3		
Computation	39seconds	12 minutes and 6 seconds		
N	31135	31135		

Note: BCV Between The Cluster Variance, CCC Class Correlation Coefficient. The symbol **and *indicate that the estimate is significant at 0.01 and 0.05 respectively.

Reference categories are: '15-19' for Age, 'urban' for POR, 'No formal education' for Education, 'Islam' for Religion, and 'Poorest' for WI. Standard errors are placed in parentheses

Table 14: Two-level estimates of univariate multilevel quasi likelihood (PQL) and full maximum likelihood (AGQ with XTMELOGIT) methods.

	PQL	XTMELOGIT (AGQ)		
	$\hat{\beta}_m$	$\hat{\beta}_m$	Under estimated(%)	Over estimated(%)
Constant	-4.883(0.135)**	-5.094(0.130)**	4	
Age				
20-24	1.094(0.085)**	1.204(0.079)**	9	
25-29	1.372(0.084)**	1.514(0.079)**	9	
30-39	1.249(0.081)**	1.408(0.081)**	11	
40-49	1.063(0.087)**	1.178(0.081)**	10	
50-64	0.579(0.150)**	0.706(0.105)**	18	
Wealth Index				
Poorer	0.408(0.086)**	0.409(0.084)**	0	
Average	0.584(0.085)**	0.614(0.087)**	5	
Wealthier	0.634(0.097)**	0.738(0.091)**	14	
Wealthiest	0.688(0.091)**	0.883(0.097)**	22	
Education				
Quranic only	-0.205(0.175)	-0.108(0.161)		90
Primary	0.945(0.090)**	0.802(0.084)**		18
Secondary	1.242(0.086)**	1.139(0.080)**		9
Higher	1.508(0.093)**	1.464(0.088)**		3
RELIGION				
Non catholic	0.653(0.067)**	0.504(0.059)**		30
Catholic Christian	0.705(0.079)**	0.606(0.075)**		16
Traditional	0.470(0.288)*	0.407(0.257)		15
No Religion	0.875(0.295)**	0.972(0.293)**	10	
Others	- 0.185(0.455)	-0.279(0.425)	34	
POR	-0.091(1.544)*	-0.166(0.425)*	448	

σ^2_{uo} (BCV	0.180(0.427)	0.825(0.338)	78	
Intral CCC	0.0519	0.201		
-2logl	158912.752	19948.830		
AIC	158914.752	19992.830		
Iteration	20	2		
Computation	39seconds	30 minutes and 12 seconds		
N	31135	31135		

Note: BCV Between The Cluster Variance, CCC Class Correlation Coefficient. The symbol **and *indicate that the estimate is significant at 0.01 and 0.05 respectively.

Reference categories are: '15-19' for Age, 'urban' for POR, 'No formal education' for Education, 'Islam' for Reli, and 'Poorest' for WI. Standard errors are placed in parentheses

Table 15: Two-level estimates of univariate multilevel quasi likelihood (PQL) and full maximum likelihood (AGQ with GLLAMM) methods.

	PQL	GLLAMM (AGQ)		
	$\hat{\beta}$(S.E)	$\hat{\beta}$(S.E)	Under estimated(%)	Overestimated(%)
Constant	-4.883(0.135)**	-5.094(0.130)**	4	
Age				
20-24	1.094(0.085)**	1.204(0.080)**	9	
25-29	1.372(0.084)**	1.514(0.079)**	9	
30-39	1.249(0.081)**	1.408(0.076)**	11	
40-49	1.063(0.087)**	1.178(0.081)**	10	
50-64	0.579(0.150)**	0.706(0.105)**	18	
Wealth Index				
Poorer	0.408(0.086)**	0.409(0.084)**	0	
Average	0.584(0.085)**	0.614(0.087)**	5	
Wealthier	0.634(0.097)**	0.738(0.091)**	14	
Wealthiest	0.688(0.091)**	0.883(0.097)**	22	
Education				
Quranic only	-0.205(0.175)	-0.108(0.161)		90
Primary	0.945(0.090)**	0.802(0.084)**		18
Secondary	1.242(0.086)**	1.139(0.080)**		9
Higher	1.508(0.093)**	1.464(0.088)**		3
Religion				
Non catholic	0.653(0.067)**	0.504(0.059)**		30
Catholic	0.705(0.079)**	0.606(0.075)**		16
Traditional	0.470(0.288)*	0.407(0.257)		15
No Religion	0.875(0.295)**	0.972(0.293)**	10	
Others	- 0.185(0.455)	-0.2786(0.425)	34	

POR	-0.091(1.544)*	-0.166(0.072)*	45	
σ^2_{u0} (BCV	0.180(0.427)	0.681(0.056)	74	
Intral CCC	0.0519	0.177		
-2logL	158912.752	19948.939		
AIC	158914.752	19992.939		
Iteration	20			
Computation	39seconds	3hr:23mins		
N	31135	31135		

Note: BCV Between The Cluster Variance, CCC Class Correlation Coefficient. The symbol **and *indicate that the estimate is significant at 0.01 and 0.05 respectively.

Reference categories are: '15-19' for Age, 'urban' for POR, 'No formal education' for Education, 'Islam' for Reli, and 'Poorest' for WI. Standard errors are placed in parentheses.

Table-16 above, shows the estimate from the Laplacian approximation (NAGQ) and the Adaptive Gaussian Quadrature(AGQ) method are not too different , most of the two methods estimate are equal except the estimate in non catholic religion, Laplacian approximation (NAGQ) have the largest -2logL (19952.313),AIC(21507.086)and BIC(21523.785) which implies that Laplacian approximation is not the best method of parameter estimation for fitting multilevel binary logistic regression though it converges earlier than Adaptive Gaussian Quadrature methods .

Table 16: Two-level estimates of univariate multilevel quasi likelihood (PQL) and full maximum likelihood (AGQ with XTMELOGIT) methods.

	XTMELOGIT (NAGQ)	XTMELOGIT (AGQ)		
	$\hat{\beta}_m$	$\hat{\beta}_m$	Under estimated(%)	Overestimated(%)
Constant	-5.092(0.130)**	-5.094(0.130)**	0	
Age				
20-24	1.202(0.079)**	1.204(0.079)**	0	
25-29	1.512(0.079)**	1.514(0.079)**	0	
30-39	1.406(0.076)**	1.408(0.081)**	0	
40-49	1.177(0.081)**	1.178(0.081)**	0	
50-64	0.705(0.105)**	0.706(0.105)**	0	
Wealth Index				
Poorer	0.409(0.084)**	0.409(0.084)**	0	
Average	0.614(0.086)**	0.614(0.087)**	0	
Wealthier	0.738(0.091)**	0.738(0.091)**	0	
Wealthiest	0.882(0.969)**	0.883(0.097)**	0	
Education				
Quranic only	-0.108(0.161)	-0.108(0.161)	0	
Primary	0.802(0.084)**	0.802(0.084)**	0	
Secondary	1.139(0.080)**	1.139(0.080)**	0	
Higher	1.464(0.088)**	1.464(0.088)**	0	
Religion				
Non catholic	0.307(0.039)**	0.504(0.059)**	39	
Catholic	0.609(0.075)**	0.606(0.075)**		1
Traditional	0.409(0.237)	0.407(0.257)	0	
No Religion	0.974(0.293)**	0.972(0.293)**	0	

Others	-0.276(0.425)	-0.279(0.425)	1	
POR	-0.1664(0.716)*	-0.166(0.4251)*	0	
σ^2_{uo} (BCV)	0.818(0.033)	0.825(0.338)	1	
Intral CCC	0.199	0.201		
-2logl	19952.313	19948.830		
AIC	19996.313	19992.830		
Iteration	3	2		
Computation	12 minutes and 6 seconds	30 minutes and 12 seconds		
N	31135	31135		

Note: BCV Between The Cluster Variance CCC Class Correlation Coefficient. The symbol **and *indicate that the estimate is significant at 0.01 and 0.05 respectively.

Reference categories are: '15-19' for Age, 'urban' for POR, 'No formal education' for Education, 'Islam' for Religion, and 'Poorest' for WI. Standard errors are placed in parentheses

From the Figure 4 below, penalized quasi likelihood was under estimated among age-group, wealth index and some path of religion categories. and it was over estimated in education category compare to full likelihood methods. Also, the quadrature part are almost equal in the primary predictor (current age- group), while the AGQ estimate from GLLAMM syntax was equal to AGQ estimate from XTMELOGIT syntax. But from table 2 and table 6, the -2logL,AIC and BICestimate shows that multilevel model from AGQ method is the best (that is, intercept model without explanatory variable in all levels and with explanatory variables). Because Adaptive Gaussian quadrature using XTMELOGIT syntax is the fastest and have the smallest -2logL AIC and BIC, odd ratio and their confidence interval was reported in the table 13 below for variables that were considered in this analysis .

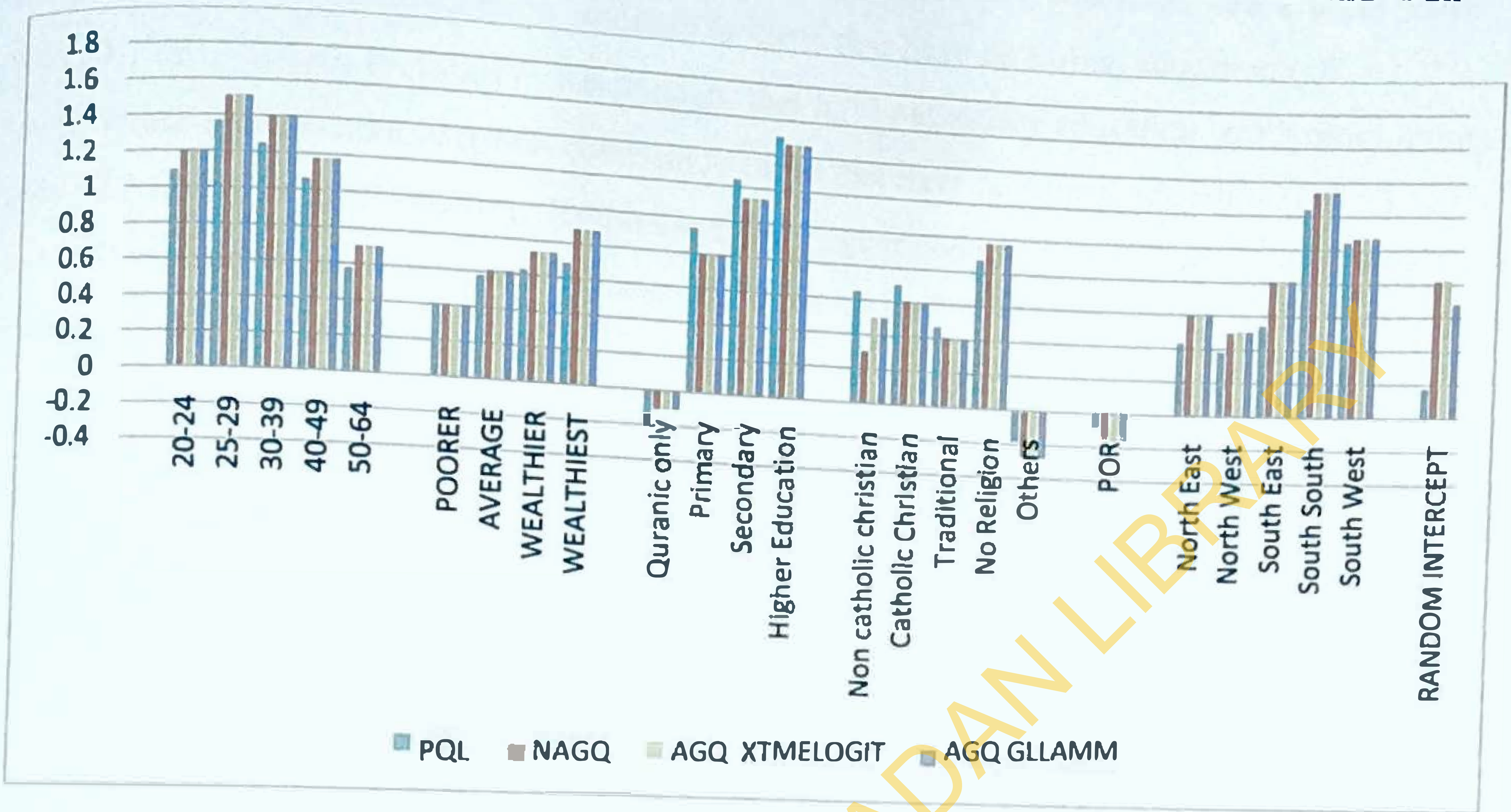
Others	-0.276(0.425)	-0.279(0.425)	1	
POR	-0.1664(0.716)*	-0.166(0.4251)*	0	
σ^2_{u0} (BCV)	0.818(0.033)	0.825(0.338)	1	
Intral CCC	0.199	0.201		
-2logl	19952.313	19948.830		
AIC	19996.313	19992.830		
Iteration	3	2		
Computation	12 minutes and 6 seconds	30 minutes and 12 seconds		
N	31135	31135		

Note: BCV Between The Cluster Variance CCC Class Correlation Coefficient. The symbol **and *indicate that the estimate is significant at 0.01 and 0.05 respectively.

Reference categories are: '15-19' for Age, 'urban' for POR, 'No formal education' for Education, 'Islam' for Religion, and 'Poorest' for WI. Standard errors are placed in parentheses

From the Figure 4 below, penalized quasi likelihood was under estimated among age-group, wealth index and some path of religion categories. and it was over estimated in education category compare to full likelihood methods. Also, the quadrature part are almost equal in the primary predictor (current age- group), while the AGQ estimate from GLLAMM syntax was equal to AGQ estimate from XTMELOGIT syntax. But from table 2 and table 6, the -2logL,AIC and BICestimate shows that multilevel model from AGQ method is the best (that is, intercept model without explanatory variable in all levels and with explanatory variables). Because Adaptive Gaussian quadrature using XTMELOGIT syntax is the fastest and have the smallest -2logL AIC and BIC, odd ratio and their confidence interval was reported in the table 13 below for variables that were considered in this analysis .

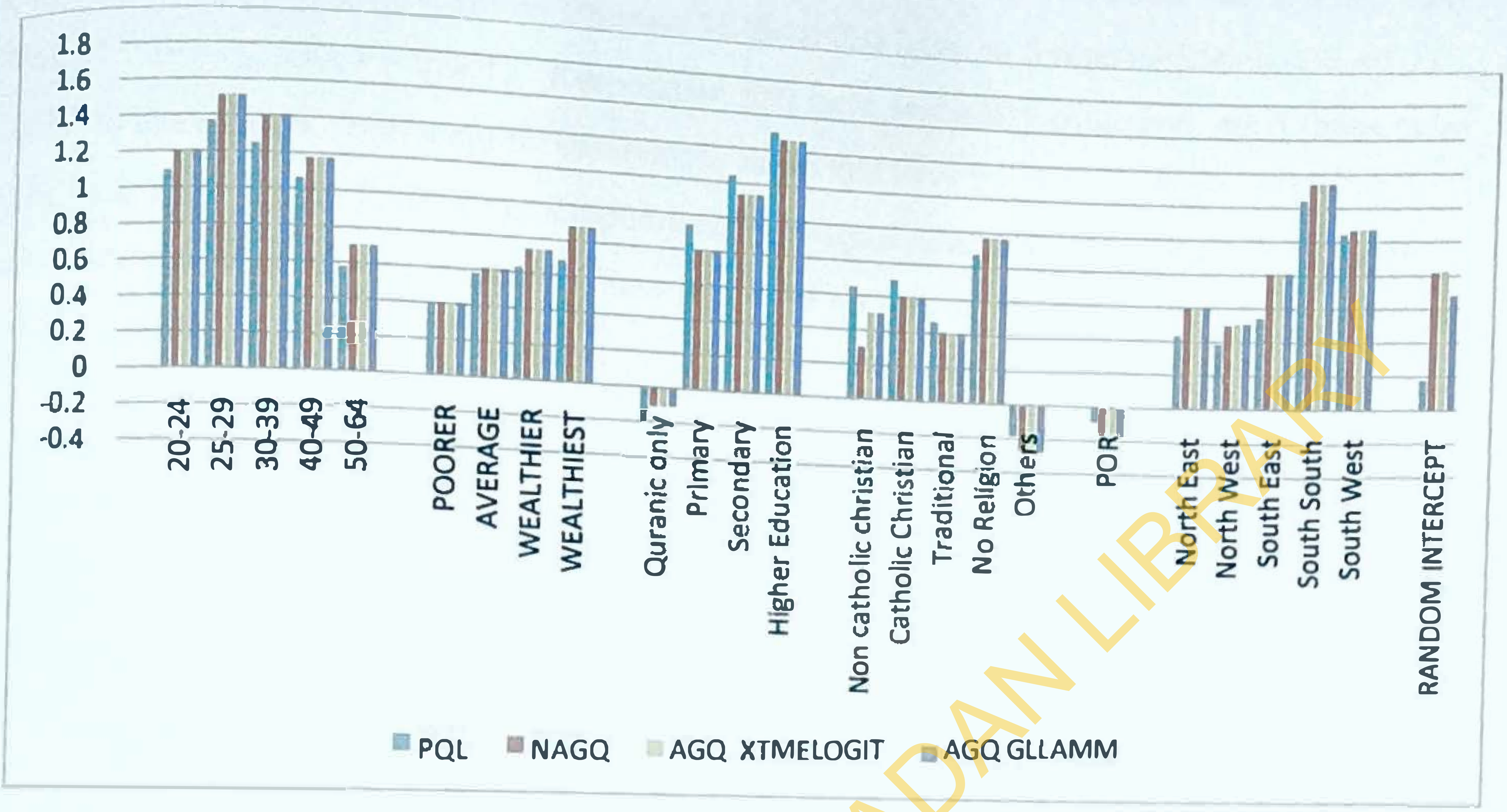
Figure 2: Multiple Bar Chart Showing the Contribution of Factors Affecting the Modern Contraceptive Use from Four Different Syntax for Both Quasi Likelihood and Full Maximum Likelihood.



From the multilevel binary logistic regression in Table 17 below, respondent aged between 20-24 years are 3 times more likely to use modern contraception compared to those of age 15 – 19 years (OR = 3.332, 95% CI:2.852 – 3.894, P<0.001),Respondent aged between 25-29 years are 5 times more likely to use modern contraception compared to those of age 15 - 19 (OR= 4.546, 95% CI : 3.897 -5.302, P <0.001), Respondent aged between 30 -39 years are 4 times more likely to use modern contraception compared to those of age 15 - 19 (OR= 4.088, 95% CI 3.524 -4.744, P <0.001). Respondent aged between 40 -49 years are 3 times more likely to use modern contraception compared to those of age 15 - 19 (OR= 3.248, 95% CI : 2.771 -3.808, P < 0.001), Respondent aged between 50-64 years are 2 times more likely to use modern contraception compared to those of age 15 - 19 (OR= 2.026, 95% CI : 1.649 -2.488, P < 0.001).

Respondent that were poorer are 2 times more likely to use modern contraception compare to those that are poorest (OR = 1.506, 95% CI: 1.277 -1.776, P <0.001). Respondent that were average are 2 times more likely to use modern contraception compared to those of poorest (OR= 1.848, 95% CI : 1.560 – 2.189, P < 0.001),). Respondent that were wealthier are 2 times more likely to use modern contraception compared to those of poorest (OR= 2.092, 95% CI : 1.750 – 2.502, P < 0.001). Respondent that were wealthiest are 2 times more likely to use

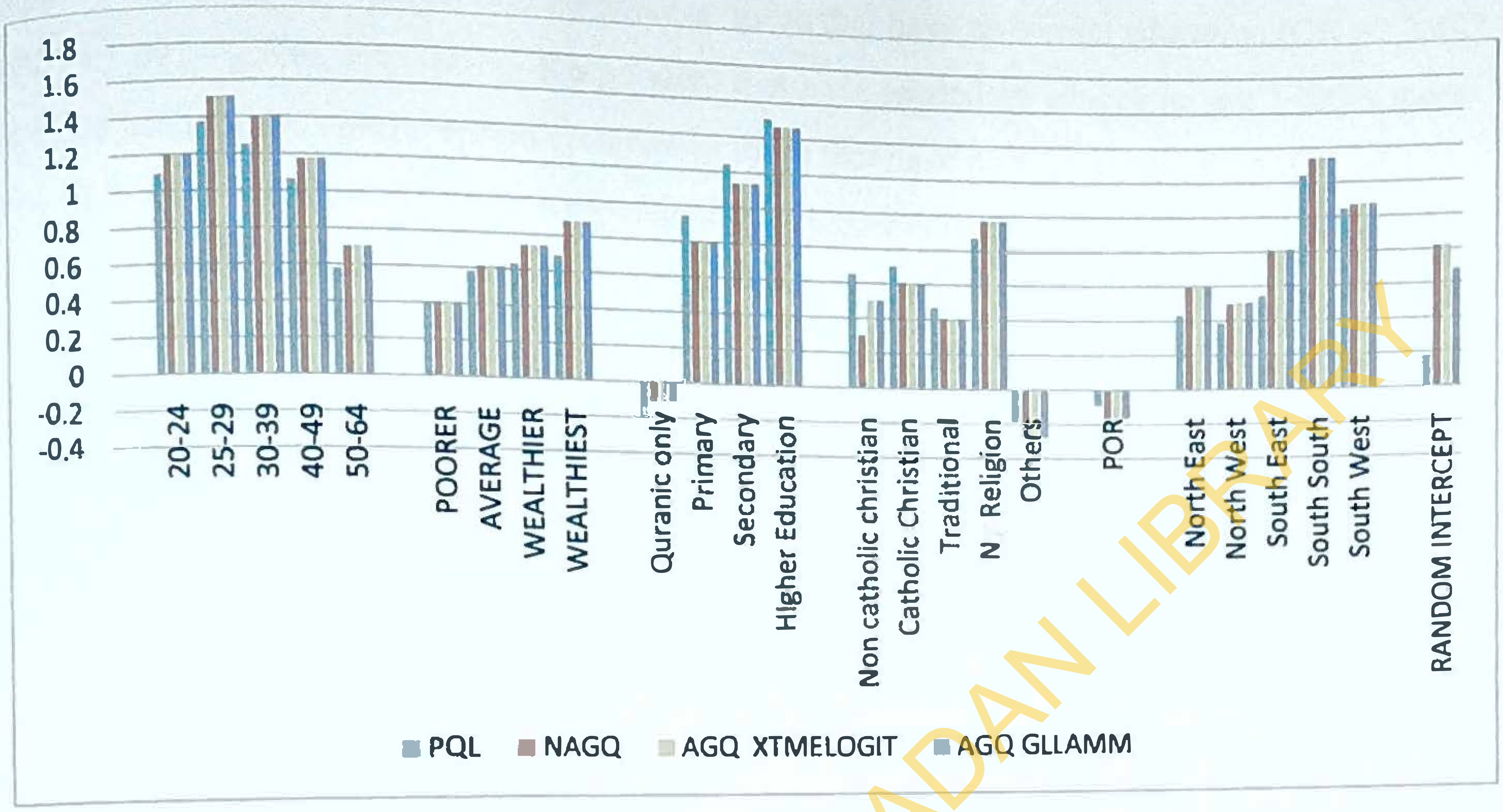
Figure 2: Multiple Bar Chart Showing the Contribution of Factors Affecting the Modern Contraceptive Use from Four Different Syntax for Both Quasi Likelihood and Full Maximum Likelihood.



From the multilevel binary logistic regression in Table 17 below, respondent aged between 20-24 years are 3 times more likely to use modern contraception compared to those of age 15 – 19 years (OR = 3.332, 95% CI:2.852 – 3.894, P<0.001),Respondent aged between 25-29 years are 5 times more likely to use modern contraception compared to those of age 15 - 19 (OR= 4.546, 95% CI : 3.897 -5.302, P <0.001), Respondent aged between 30 -39 years are 4 times more likely to use modern contraception compared to those of age 15 - 19 (OR= 4.088, 95% CI 3.524 -4.744, P <0.001), Respondent aged between 40 -49 years are 3 times more likely to use modern contraception compared to those of age 15 - 19 (OR= 3.248, 95% CI : 2.771 -3.808, P < 0.001), Respondent aged between 50-64 years are 2 times more likely to use modern contraception compared to those of age 15 - 19 (OR= 2.026, 95% CI : 1.649 -2.488, P < 0.001).

Respondent that were poorer are 2 times more likely to use modern contraception compare to those that are poorest (OR = 1.506, 95% CI: 1.277 -1.776, P <0.001), Respondent that were average are 2 times more likely to use modern contraception compared to those of poorest (OR= 1.848, 95% CI : 1.560 – 2.189, P < 0.001),), Respondent that were wealthier are 2 times more likely to use modern contraception compared to those of poorest (OR= 2.092, 95% CI : 1.750 – 2.502, P < 0.001). Respondent that were wealthiest are 2 times more likely to use

Figure 2: Multiple Bar Chart Showing the Contribution of Factors Affecting the Modern Contraceptive Use from Four Different Syntax for Both Quasi Likelihood and Full Maximum Likelihood.



From the multilevel binary logistic regression in Table 17 below, respondent aged between 20-24 years are 3 times more likely to use modern contraception compared to those of age 15 – 19 years (OR = 3.332, 95% CI:2.852 – 3.894, P<0.001),Respondent aged between 25-29 years are 5 times more likely to use modern contraception compared to those of age 15 - 19 (OR= 4.546, 95% CI : 3.897 -5.302, P <0.001), Respondent aged between 30 -39 years are 4 times more likely to use modern contraception compared to those of age 15 - 19 (OR= 4.088, 95% CI 3.524 -4.744, P <0.001), Respondent aged between 40 -49 years are 3 times more likely to use modern contraception compared to those of age 15 - 19 (OR= 3.248, 95% CI : 2.771 -3.808, P < 0.001), Respondent aged between 50-64 years are 2 times more likely to use modern contraception compared to those of age 15 - 19 (OR= 2.026, 95% CI : 1.649 -2.488, P < 0.001).

Respondent that were poorer are 2 times more likely to use modern contraception compare to those that are poorest (OR = 1.506, 95% CI: 1.277 -1.776, P <0.001), Respondent that were average are 2 times more likely to use modern contraception compared to those of poorest (OR= 1.848, 95% CI : 1.560 – 2.189, P < 0.001),), Respondent that were wealthier are 2 times more likely to use modern contraception compared to those of poorest (OR= 2.092, 95% CI : 1.750 – 2.502, P < 0.001). Respondent that were wealthiest are 2 times more likely to use modern contraception compared to those of poorest (OR= 2.418, 95% CI : 1.999 – 2.925, P < 0.001).

Among the Education categories, Respondent that have primary education are 2 times more likely to use modern contraception compare to those that have no formal education,(OR =2.230, 95% CI:1.891 - 2.630, P< 0.001), Respondent that have secondary education are 3 times more likely to use modern contraception compare to those that have no formal education,(OR =3.123, 95% CI:2.670 – 3.652, P <0.001), Respondent that Higher education are 4 times more likely to use modern contraceptive compare to those that have no formal education,(OR =4.324, 95% CI: 3.639 - 5.140, P < 0.001),

Among the Religion categories, Respondent that were Non catholic Christian are 2 times more likely to use modern contraception compare to those that were Islamic religion, (OR =1.655, 95% CI:1.474 – 1.859, P < 0.001), Respondent that were catholic Christian are 2 times more likely to use modern contraception compare to those that were Islamic religion,(OR =1.832, 95% CI:1.582 – 2.122, P < 0.001), Respondent that have no religion are 3 times more likely to use modern contraception compare to those that are Islamic religion,(OR =2.644, 95% CI:1.488 – 4.698, P < 0.001),

Respondent that were in rural area are 1.2 times less likely to use modern contraception compare to those that were in urban area,(OR =0.847, 95% CI: 0.736 – 0.800, P <0.05),

Among the geo political zone categories, Respondent that were in North East are 2 times less likely to use modern contraception compare to those that were in North Central, (OR =0.603, 95% CI: 0.470– 0.772, P < 0.001), Respondent that were in North west are 2 times less likely to use modern contraception compare to those that were in North Central,(OR = 0.506, 95% CI: 0.393 – 0.651, P < 0.001), Respondent that were in South South are 1.3 times more likely to use modern contraception compare to those that were in North Central,(OR = 1.344, 95% CI: 1.087 – 1.662, P < 0.001), The last row gives me the random effect estimates. This represents the estimated standard deviation in the intercept on the logit scale that is, the effect of the cluster on the use of contraception among the respondent. The total variance in the use of modern contraception among the respondent base on their clustering (β=0.827, 95% CI: 0.762 - 0. 897)

Table 17: AGQ with XTMELOGIT syntax estimates of the significant factors that contribute to modern contraceptive use, using two-level random intercept and fixed effects model.

CUMC	Odds Ratio	Std. Err.	P>|z|	[95% Conf. Interval]
Constant	0.008	0.001	0.000	(0.006 0.011)
Age				
20-24	3.332	0.265	0.000	(2.852 3.894)
25-29	4.546	0.357	0.000	(3.897 5.302)
30-39	4.088	0.310	0.000	(3.524 4.744)
40-49	3.248	0.264	0.000	(2.771 3.808)
50-64	2.026	0.212	0.000	(1.649 2.488)
Wealth Index				
Poorer	1.506	0.127	0.000	(1.277 1.776)
Average	1.848	0.160	0.000	(1.560 2.189)
Wealthier	2.092	0.191	0.000	(1.750 2.502)
Wealthiest	2.418	0.235	0.000	(1.999 2.925)
Education				
Quranic only	0.898	0.144	0.502	(0 .655 1.230)
Primary	2.230	0.188	0.000	(1.891 2.630)
Secondary	3.123	0.249	0.000	(2.670 3.652)
Higher	4.324	0.381	0.000	(3.639 5.140)

Religion				
Non catholic	1.655	0.098	0.000	(1.474 1.859)
Catholic	1.832	0.137	0.000	(1.583 2.122)
Traditional	1.502	0.386	0.114	(0.908 2.487)
No Religion	2.644	0.776	0.001	(1.488 4.698)
Others	0.757	0.322	0.512	(0.329 1.741)
POR	0.847	0.061	0.021	(0.736 0.975)
Region				
North East	0.603	0.076	0.000	(0.470 0.772)
North West	0.506	0.065	0.000	(0.393 0 .651)
South East	0.811	0.095	0.075	(0.644 1.021)
South South	1.344	0.146	0.006	(1.087 1.662)
South West	1.071	0.119	0.534	(0.862 1.331)
RANDOM EFFECT	Estimate	Std. Err.	P>|z|	[95% Conf. Interval]
INTERCEPT	0. 827	0.056		(0.762 0 .897)
ICC	0.201			

Note: ICC Intra Class Correlation Coefficient. The symbol ** and * indicate that the estimate is significant at 0.01 and 0.05 respectively.

Reference categories are: '15-19' for Age, 'Poorest' for WI, 'urban' for POR, 'No formal education' for Education, 'Islam' for Reli, and 'North Central' for geo political zone.

Religion				
Non catholic	1.655	0.098	0.000	(1.474 1.859)
Catholic	1.832	0.137	0.000	(1.583 2.122)
Traditional	1.502	0.386	0.114	(0.908 2.487)
No Religion	2.644	0.776	0.001	(1.488 4.698)
Others	0.757	0.322	0.512	(0.329 1.741)
POR	0.847	0.061	0.021	(0.736 0.975)
Region				
North East	0.603	0.076	0.000	(0.470 0.772)
North West	0.506	0.065	0.000	(0.393 0 .651)
South East	0.811	0.095	0.075	(0.644 1.021)
South South	1.344	0.146	0.006	(1.087 1.662)
South West	1.071	0.119	0.534	(0.862 1.331)
RANDOM EFFECT	Estimate	Std. Err.	P>|z|	[95% Conf. Interval]
INTERCEPT	0. 827	0.056		(0.762 0 .897)
ICC	0.201			

Note: ICC Intra Class Correlation Coefficient. The symbol ** and * indicate that the estimate is significant at 0.01 and 0.05 respectively.

Reference categories are: '15-19' for Age, 'Poorest' for WI, 'urban' for POR, 'No formal education' for Education, 'Islam' for Reli, and 'North Central' for geo political zone.

Figure 3: Estimated mean contribution for child bearing age group of male and female on modern contraceptive use in Nigeria.

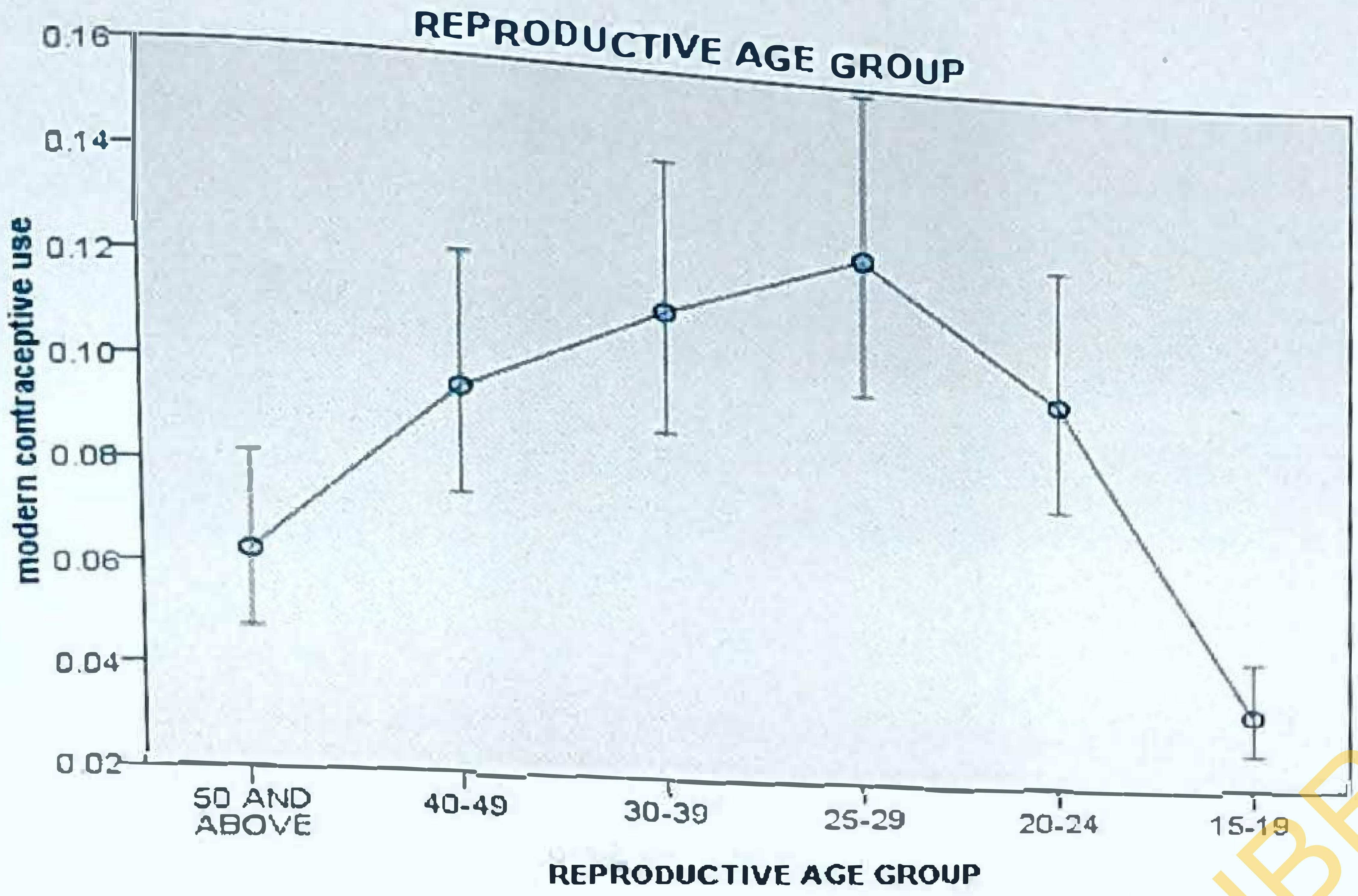


Figure 4: Estimated mean contribution of wealth index on modern contraceptive use in Nigeria

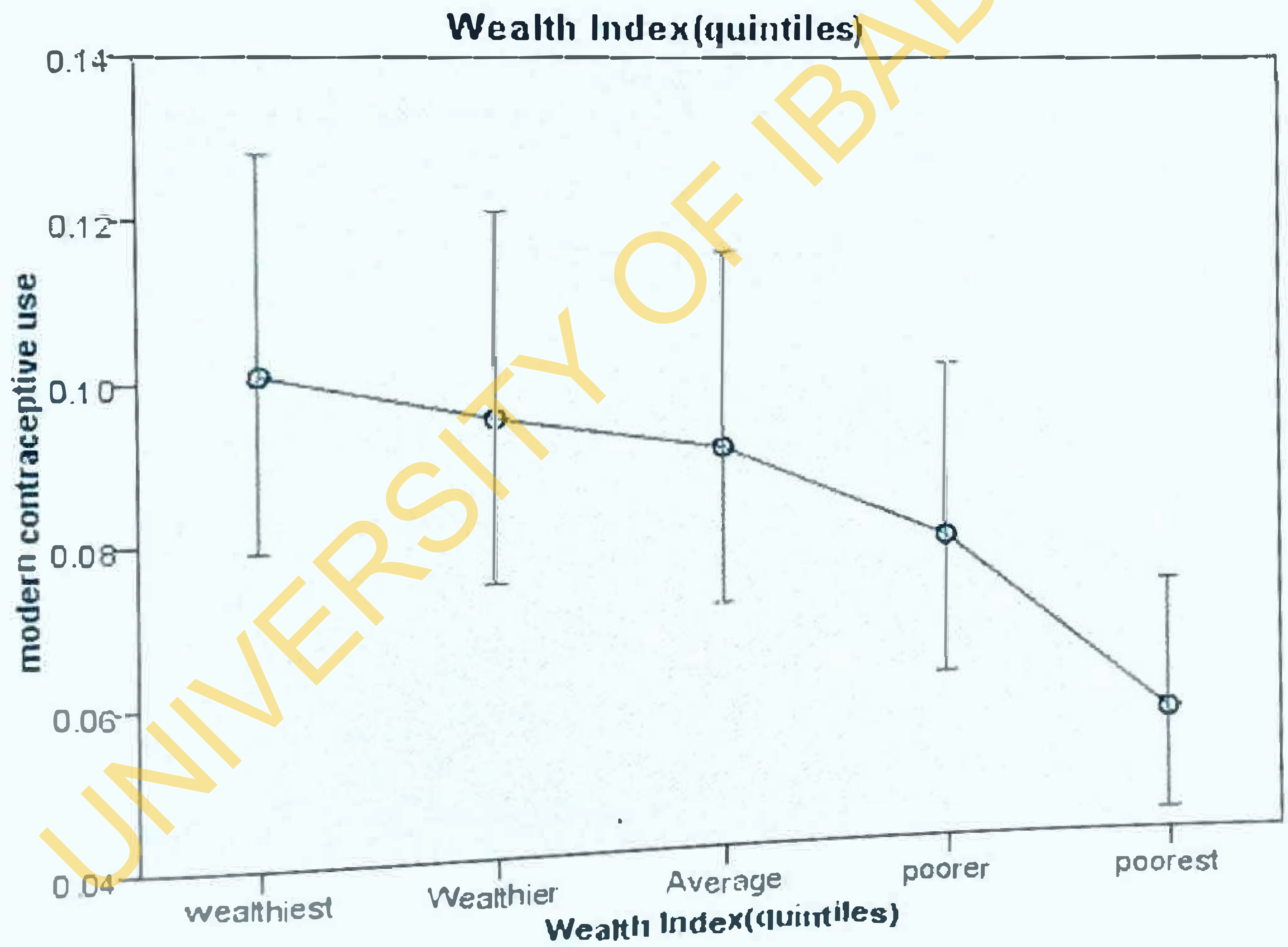


Figure 3: Estimated mean contribution for child bearing age group of male and female on modern contraceptive use in Nigeria.

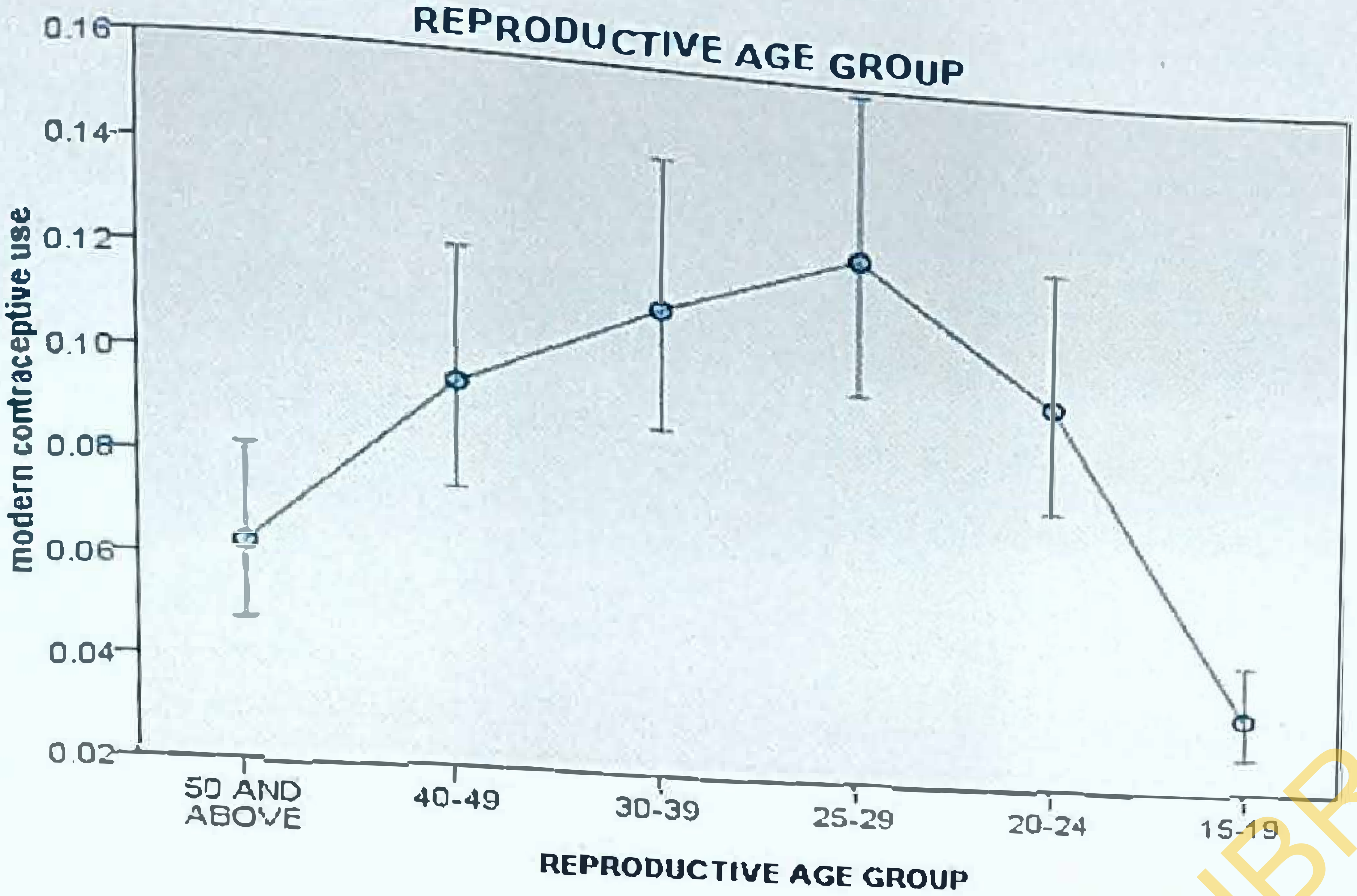


Figure 4: Estimated mean contribution of wealth index on modern contraceptive use in Nigeria

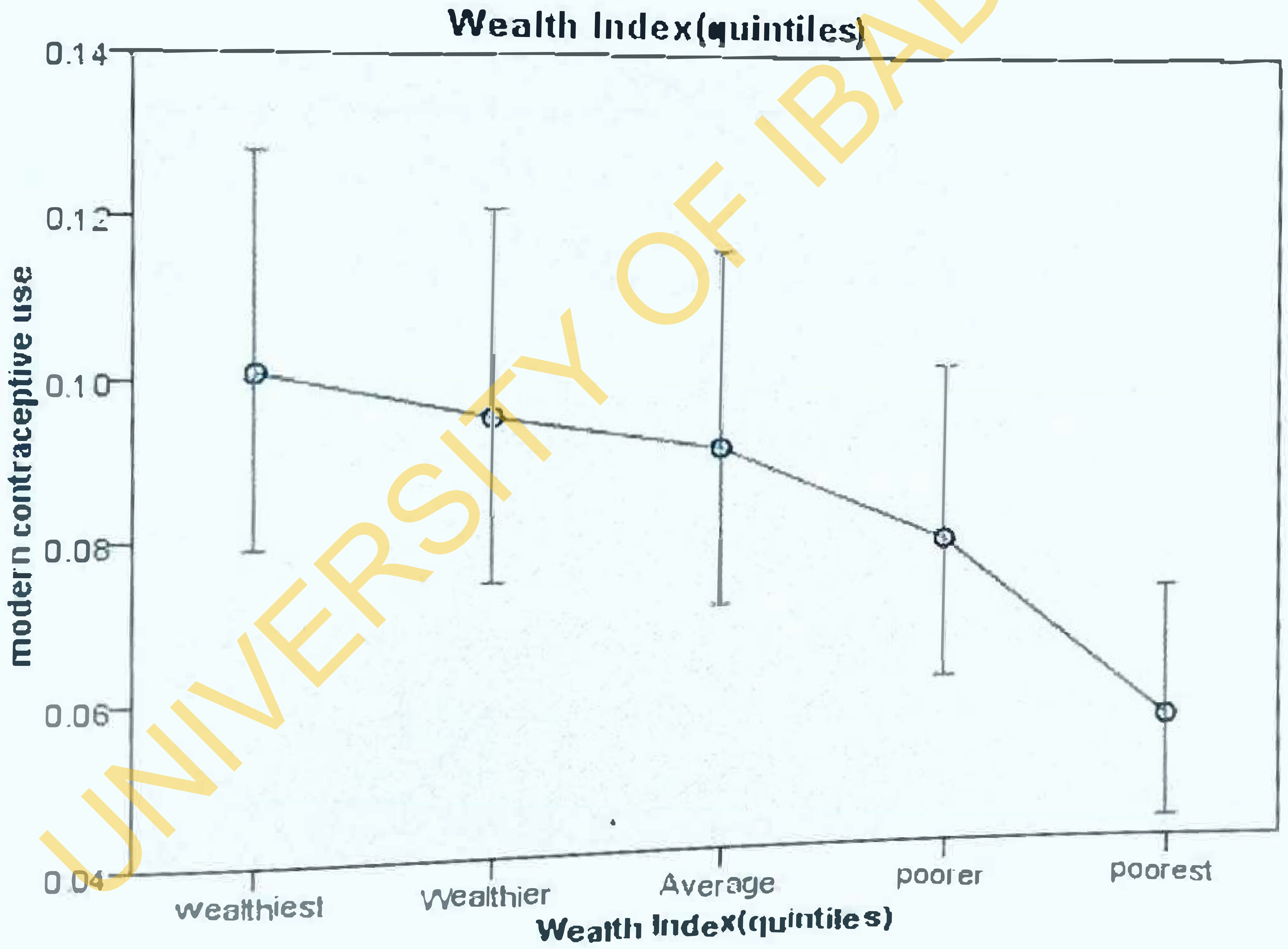


Figure 5: Estimated mean contributions of education categories on modern contraceptive use in Nigeria

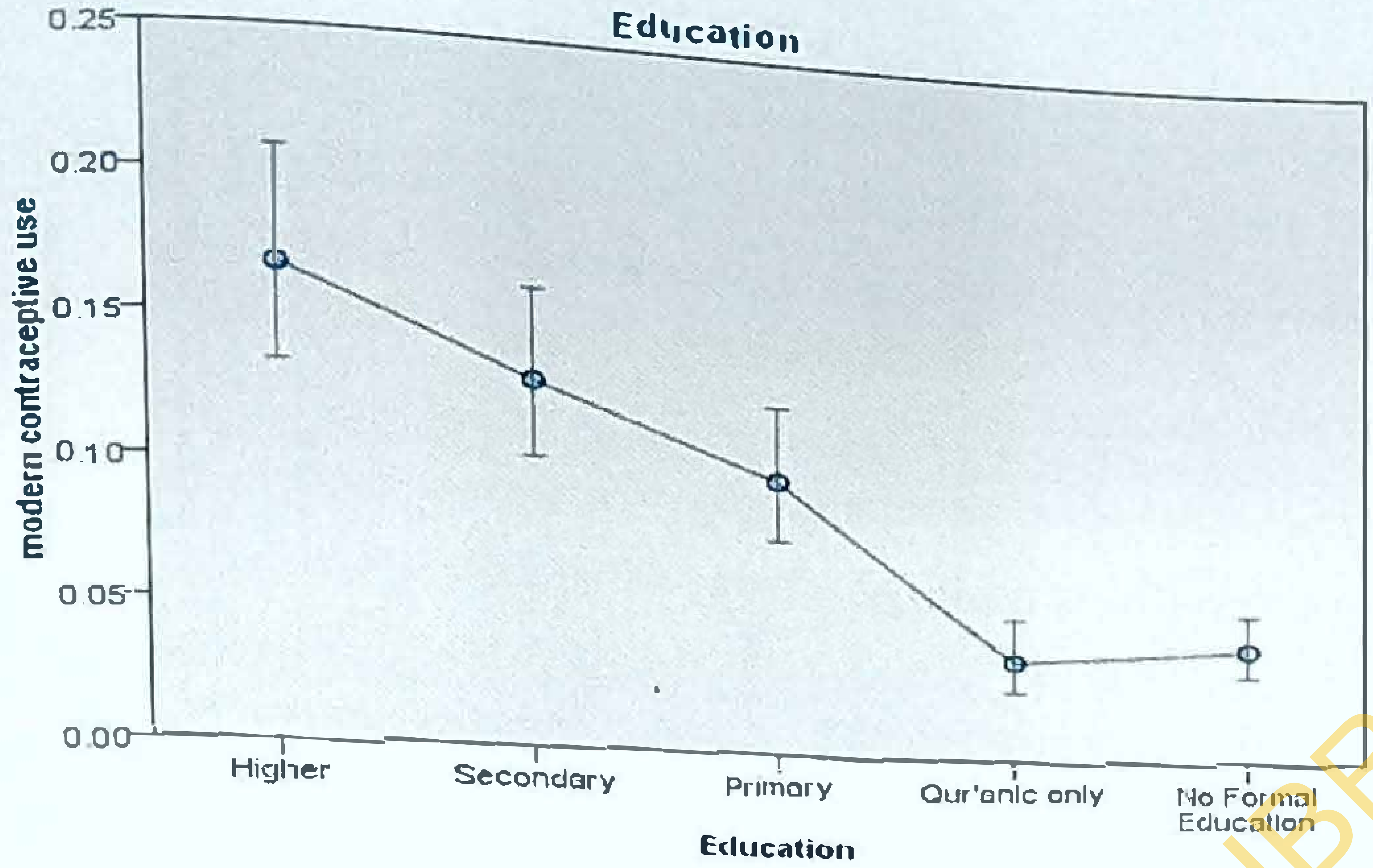


Figure 6: Estimated mean contributions of religion on modern contraceptive use in Nigeria

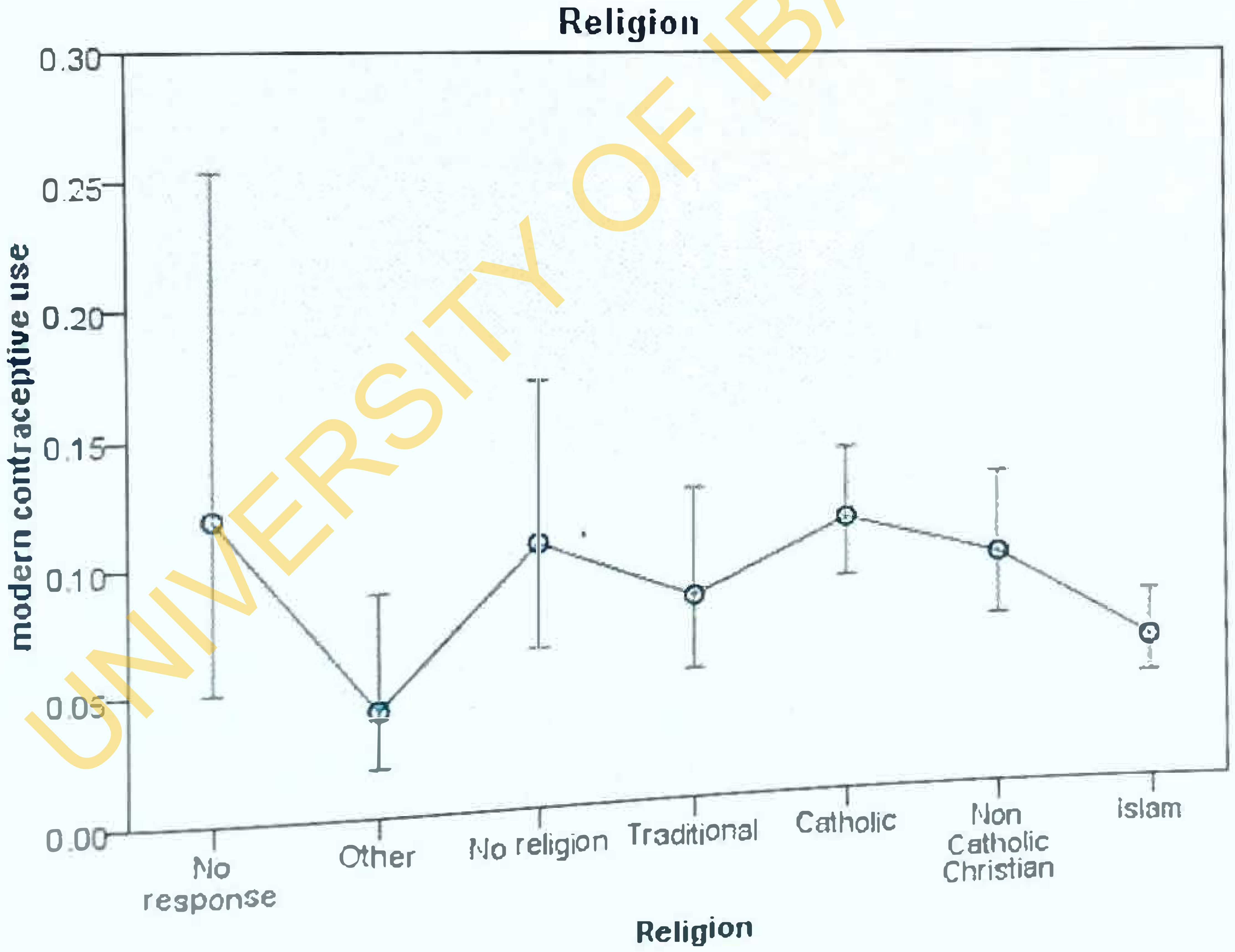
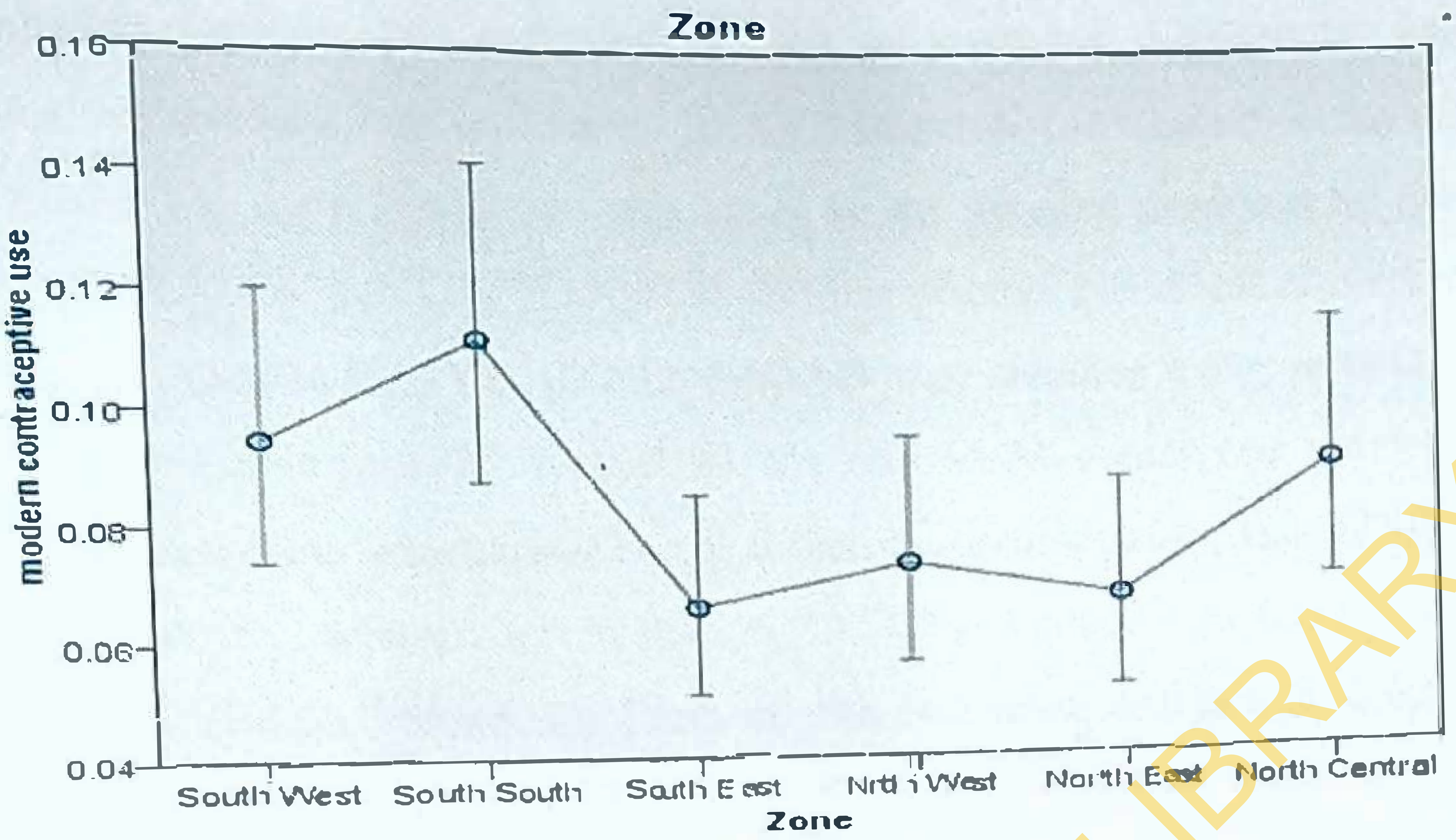


Figure 7: Estimated mean contribution of the respondent in each geo political zone on modern contraceptive use in Nigeria.



CHAPTER FIVE

5.0 Discussion

The three estimation methods are evaluated at four performance dimensions: numerical convergence, bias, computation time and model fitting. Numerical convergence is measured by the convergence rate. The convergence rate was based on the iteration produced by the macro GENLINMIXED, GLLAMM and XTMELOGIT to confirm whether numerical convergence has been reached or not. Output from the GENLINMIXED was obtained using penalized quasi likelihood, standard available in SPSS version 20, the GLLAMM output was obtained using Adaptive Gaussian Quadrature which make use of fifteen quadrature point , and XTMELOGIT syntax allow estimation via Laplacian approximation(NAGQ) and adaptive Gaussian quadrature which is available in STATA version 12. From all the multilevel analysis in chapter four, GLLAMM syntax for adaptive Gaussian quadrature(AGQ) has the smallest standard error, -2logl, AIC and BIC if three level is consider followed by XTMELOGIT for adaptive Gaussian quadrature. But XTMELOGIT syntax for AGQ has the smallest standard error, -2logl, AIC and BIC if two level is consider.

Also, comparison between single level and multilevel models were made and it turn out that the effect of the primary predictor in the standard logistic regression model have been underestimated in comparison with multilevel models and for other covariate, some are either over estimated or underestimated. This implies that the difference in β coefficients estimated from the multilevel models and standard model arises because of the addition of the random effects. Therefore, using single level model to predict the future value of modern contraceptive use in cluster survey is inappropriate. This is in line with the study done by Hasinur et al 2011.

5.1 Best Method In Term Of -2LOGL, AIC and BIC

For two level binary logistic regression done in Table -12 Adaptive Gaussian quadrature have the smallest -2logL, AIC and BIC in which the estimate for XTMELOGIT and GLLAMM for AGQ are equal. XTMELOGIT syntax have the smallest computational time and smallest standard error for the random effect, this implies that among the three method of parameter estimation (PQL, NAGQ and AGQ) the adaptive Gaussian quadrature is the best for three level binary logistic regression though it has a longest computational time this is in line with the study done by Adam C Carie 2009 . Also Marc Callen et al 2003concluded that Adaptive Gaussian quadrature performance was the best but his conclusion was based on small generated sample size. However, in this research work, AGQ is the best when we compared NAGQ with AGQ although AGQ is the slowest in term of computational time which disagree with Marc Callens conclusion based on time of computation.

5.2 Best Method in Term of Bias

This study has further demonstrated the tendency for the standard logistic model to seriously bias the parameter estimates of observed covariates when analyzing multilevel data. However, the estimated bias generally differs depending on the estimation procedure used for the multilevel logistic model. The differences between estimates obtained using PQL and NAGQ as well as between NAGQ and AGQ were minimal as obtained in the analysis. This is consistent with the observation in Goldstein and Rasbash (1996) and Hasinur and Ewart (2011) that in the more common case where variances in a multilevel logistic model do not exceed about 0.5, the PQL model can be expected to perform well in term of bias. That is, SPSS software's PQL are likely to be adequate for producing nearly unbiased estimates. PQL was also preferred in term of bias in the work done by Rodriguez and Goldman 1995.

5.3 Effect of Level Misspecification

By using geo political zone as level three, It was realized that the fixed effect estimate of adaptive Gaussian quadrature using XTMELOGIT and GLLAMM in STATA are different despite the fact that they are in the same integration point (that is, fifteen integration point) But when the level three which is geo-political zone was excluded in table 2 and table 8, the fixed effect estimate for XTMELOGIT and GLLAMM for Adaptive Gaussian quadrature were equal but the random effect are different. Also the -2logL, AIC and BIC were also equal. It was discovered that the fixed effect obtained from all the methods when only the cluster was involve is not different from the estimate obtained when the geopolitical zone was added to the model. The intral-class correlation coefficient for the geopolitical zone was approximately zero for all the methods which implies that there is no agreement between the data obtained from all the geopolitical zone and also the variation in modern contraceptive use among the respondents was not explained by the geopolitical zone, which implies that using geo political zone as level is not reliable and also not valid.

5.4 Problem Encountered with XTMELOGIT Syntax

The problem encountered with XTMELOGIT syntax during the analysis is that it crashed as soon as it begin initial parameter selection and it returns the following error: "Initial value not feasible". With the help of the internet, some solution were found, which advocates using the from() option to load up the model initial values. The option suggested that there is need to run a non – hierarchical logistic model, in other to extract the coefficient that was not feasible, before multilevel logistic model can be run. It was discovered that if there is K level, STATA will be expecting K additional coefficient but apparently it can handle this automatically. The only solution that was found was to generate a vector matrix of model dimension, in other to know the dimension that was needed for the model, a simpler multi-level model for the initial vector matrix that was created was ran, if it conform with the dimension needed for the model then it generates result else it specify the dimension that was needed and then extract the values from it.

5.5 Contributions of Some Socio-Demographic and Socio Economic Factors to Modern Contraceptive Use in Nigeria.

The 2012 National AIDS and Reproductive health survey (NARHS) data was based on multistage stratified cluster sampling. This study found that for hierarchical structured data the multilevel effects are significant and have to be taken into consideration in logistic regression model, in order to avoid overestimation or under estimation that may occur in single level logistic model, one has to use the best multilevel method of estimation. From the result of the analysis, it was discovered that AGQ using XTMELOGIT syntax is the best method for fitting two- level binary logistic regression model which was done in Table-17.

From table-17, respondents between age 25 and 29 have the highest rate of modern contraceptive use while those respondents between 15 and 19 are the least categories that use modern contraception in Nigeria. For wealth index categories, the rate at which they use contraception are in levels, those in wealthiest class have the highest contribution to modern contraceptive use followed by wealthier class to the last class which is poorest. Among the education categories, those with higher education use modern contraception than any other categories, while the respondents that do not have any form of education have lowest contribution to modern contraceptive use .

In the religion categories, respondents with no religion mostly use modern contraception while the least categories is Islamic religion. Also respondents in urban area are more likely to use modern contraception compared to those in rural area. Findings obtained are similar to the result of Gidado Omolola 2013 and also similar to findings of Adebimpe et al 2011 who concluded that there was a significant association between locations of residence on modern contraceptive use. Among the geopolitical zone, respondents in south south part of the country mostly use modern contraception compare to other geo political zone followed by respondents in south west while the least users of modern contraception are those in south east.

These were also shown on the chart in figures 4 to 7,that is, the log of odd of some significant factors on modern contraceptive use in Nigeria. The Community factor which is the cluster was significantly associated with the use of modern contraception. Interventions aimed at promoting the use of contraception among Nigerians should not only be implemented at the individual level but also tailored to the community (cluster) level as interventions conceived without consideration for cluster context are likely to have limited impact.

5.6 Conclusion

This project evaluated the performance of three estimation methods for multilevel binary logistic regression models: Penalised Quasi-Likelihood (PQL), Non-Adaptive Gaussian Quadrature (NGQ) that is Laplacian approximation and Adaptive Gaussian Quadrature (AGQ). These likelihood- based methods are frequently used in the applied multilevel-modeling literature to estimate multilevel binary logistic regression.

Large cluster schemes were used in this study. The information on performance of the estimators were under two different circumstances (that is, when considering the intercept only model and when including the explanatory variable in level one with random intercept for both level two and level three). Also, the multilevel binary logistic regression was used to quantify the effect of different syntax model parameters on the performance of the estimators. Bias, computing time, best fitted model and convergence of the estimation routine were considered as performance measures. In this study, AGQ had better performance than PQL and NAGQ due to the smallest -2logL, AIC and BIC. Comparison PQL with full maximum likelihood method showed that the bias was larger for full likelihood,. However, AGQ gave the most precise estimates,. STATA version 12 has two syntaxes for estimation of adaptive Gaussian quadrature. These include; XTMELOGIT and GLLAMM, based on the result obtained for three level binary logistic regression GLLAMM had the smallest -2logL,AIC and BIC which implies that GLLAMM AGQ is the best for three level model. While considering two level both XTMELOGIT and GLLAMM syntaxes for AGQ are good but in term of computational time, XTMELOGIT for AGQ was the fastest. These conclusions hold for multilevel binary logistic regression (logit link) when the number of cluster is large.

5.7 Recommendation

1. Fitting two levels binary logistic regression model, Adaptive Gaussian quadrature using XTMELOGIT syntax is better than other methods of estimation, though it gives the same result with Adaptive Gaussian quadrature using GLLAMM syntax but XTMELOGIT syntax has a shorter convergence time than GLLAMM.
2. Adaptive Gaussian quadrature using GLLAMM syntax is robust for fitting three level binary logistic regression model on STATA software .
3. Level specification is important in multilevel cluster survey analysis. Researchers should ensure that there is dependency between the levels and it should be investigated using random effect and Intra class correlation coefficient before it is use.
4. Interventions aimed at promoting the use of contraception among Nigerians should not only be implemented at the individual level but tailored to the community (that is, cluster) level, as interventions conceived without consideration for cluster context are likely to have limited impact.

References

Adekunle A.O, Arowojolu A.O, Adedimeji A.A, and Okunlola M.A. (2000). Emergency Contraception: Survey of knowledge, Attitudes and Practice of Healthcare Professionals in Ibadan. *Journal of Obstetrics and Gynecology*(2000). 20(3) pp. 284–289.

Amin, S., Basu, A. M. and Stephenson, R. (2002). Spatial variation in contraceptive use in Bangladesh: Looking beyond the borders. *Demography* **39**, pp.251-267.

Ankomah A., Anyanti J., Adebayo S. and Giwa A.(2013). Barriers to Contraceptive Use among Married Young Adult in Nigeria: A Qualitative Study. *International Journal of Tropical Disease and Health* vol.3, pp 267-282.

Austin P.C. (2005). Bias in Penalized Quasi-Likelihood Estimation in Random Effects Logistic Regression Models When the Random Effects are Not Normally Distributed. *Communications in Statistics: Simulation & Computation*, 34 pp. 549-565.

Blalock H. M. 1984. Contextual-Effects Models: Theoretical and Methodological Issues. *Annu. Rev. Sociol*1984.. 10:353–72

Blessing U. Mberu Holly E. Reed (2014). Understanding Subgroup Fertility Differentials in Nigeria. *Population Review* 2014. 53(2),pp.23-46.

Box G.E.P and Tiao G.C (1973). Bayesian Inference in Statistical Analysis. New York: Wiley.

Breslow N.E and Clayton D.G (1993). Approximate Inference in Generalized Linear Mixed Models. *Journal of The American Statistical Association*, 88, pp.9-25.

Bingenheimer J.B. and Raudenbush S.W. (2004). Statistical and Substantive Inferences in Public Health : Issues in the Application of Multilevel Models. *Annual Review of Public Health*, 25 pp.53-77.

Bryk A.S, and Raudenbush S.W. (1992) Hierarchical Linear Models: Applications and Data Analysis Methods. *Newbury Park, CA: Sage*. Pp.265

Bryk A.S, and Raudenbush S.W. (2002) Hierarchical Linear Models: Applications and Data Analysis Methods Second Edition. University of Chicago. *SAGE Publication, Inc*. Vol. 1 Pp.512

Chaix B. Rosvall M. and Merlo J. (2007) Neighborhood Socioeconomic Deprivation and Residential Instability: Effects On Incidence Of Ischemic Heart Disease And Survival After Myocardial Infarction Epidemiology 18:104–11.

Christiano, Swatos, Kivisto, Kevin, William, Peter (2008). Sociology of Religion. *New York: Rowman & Littlefield Publishers Inc.* p. 137. ISBN9780742561113.

Cleland J. S. Bernstein A. Ezeh, A. Faundes, and A. Glasier. (2006). Family Planning: The Unfinished Agenda. *Lancet* 2006 368(9549) pp.1810–1827.

Cohen B. (1993). Fertility Levels: Differentials and Trends in K.A Feet. Population Division Department of Economic and Social Affairs. UN Secretariat, New York.

Cohen M.P. (1998).Sample Size in Hierarchical Linear Models. *Journal Of Statistics1998.* 14 pp. 267– 276

Chung H. and Muntaner C. (2007). Welfare State Matters: A Typological Multilevel Analysis of Wealthy Countries. *Health Policy.* 80 pp.328–39.

Cleland, J., S. Bernstein, A. Ezeh, A. Faundes, and A. Glasier. 2006. Family Planning: The Unfinished Agenda. Lancet 368(9549): 1810–1827.

Daniel, J. and Sonya, K. (2011). Fitting Multilevel Models with Ordinal Outcomes: Performance of Alternative Specifications and Methods of Estimation. *NIH Public Access Psychol Methods* 2011. 16(4) pp.373–390.

DeLeeuw J. Kreft I.G. (1995). Questioning multilevel Models. Journal of Educational. *Behavioral Statistics* 1995. 20:171–89

Diez Roux, A. 2009. The Steps in Understanding the Multilevel Determinant of Health. *Journal of Epidemiol Community Health* 2008. 62 pp.957-959

Diez-Roux A V. (2000). Multilevel Analysis in Public Health Research. *Annu. Rev. Public Health* 2000. 21 pp.171–92

Diez-Roux A.V. (1998). Bringing Context Back into Epidemiology: Variables and Fallacies in Multilevel Analysis. *American Journal. Public Health* 1998.88 pp.216–22

DiPrete T.A. and Forristal J.D. (1994). Multilevel Models: Methods and Substance. *Annu. Rev Sociol.* 20 pp.331–57

Duncan C. Jones K and Moon G. (1996). Health Related Behaviour in Context. A Multilevel Modeling Approach. *Social Science. Med.* 42 pp.817–30

Duncan C. Jones K. and Moon G. (1998). Context, Composition, and Heterogeneity: Using Multilevel Models in Health Research. *Social Science. Med.* 46 pp.97–117

Ecevit Eyduran, 2008. Usage of Penalized Maximum Likelihood Estimation Method in Medical Research: An Alternative to Maximum Likelihood Estimation Method. *Journal of Research in Medical Sciences* 2008; 13(6) pp.325-330

Emmanuel M., Andrea S., John E.and James E. (2010). Contraceptive Practices in Nigeria: Literature Review and Recommendations for Future Policy Decision, Department of Obstetrics and gynecology, University of Calabar, Nigeria. *Open Access Journal of Contraception* Volume 1,pp 9-22

Federal Ministry of Health;2013. National Policy on population for Development, Unity, Progress and Self-reliance. Lagos, Nigeria: Department of Population Activities, [http://www.fmh.gov.ng/index]

Gidado Omolola (2013). Community and Individual Factors Influencing Modern Contraceptive Use Among Married Women in Northern and Southern Nigeria. [http://www.xcdsystem.com]

Goldstein, H. (1991). Nonlinear multilevel models with an application to discrete response data. *Biometrika* 78, 45-51.

Goldstein H. (1999). Multilevel Statistical Model, New York: Halsted. pp178. [http://www.arnoldpublisher.com/support/Goldstein.htm]

Goldstein, H. (2003). *Multilevel Statistical Models* 3rd ed.. Oxford University Press.

Goldstein, H. and Rasbash, J. (1996). Improved Approximations for Multilevel Models with Binary Responses. *Journal of Royal Statistical Society*. 159, pp.505-513.

Guang, G. and Hongxin, (2000) Multilevel Modeling for Binary Data . *Annu Rev. Sociol* 2000, 26 pp.441-62.

Hasinur R.K. Ewart J.and Shaw H. (2011). Multilevel Logistic Regression Analysis Applied to Binary Contraceptive Prevalence Data. *Journal of Data Science2011*. 9, pp.93-110

Hermalin A. (1986). The Multilevel Approach: Theory and Concepts. The Methodology for Measuring The Impact of Family Planning Programs on Fertility. *Addendum Manual IX Population Study*(1986). 66 pp.15–31

Hox J.J. (2002). Multilevel Analysis: Techniques and Applications. *Mahwah, N.J., Lawrence Erlbaum Publishers*. [joophox.net/publist/pubenjh.htm]

John J.D. Donna L.C Stephanie T.L, and Runze L. (2012). Sensitivity and Specificity of Information Criteria. *The Methodology Centre and Department of Statistics, Pennsylvania State University,*

Kalam, E. M. N. and Khan, H. T. A. (2002). Modeling contraceptive prevalence in Bangladesh: A hierarchical approach. *Asian Meta Centre, Research Paper Series*, 21, N0. 5.

Kaplan D. (1989). A Study of The Sampling Variability and Z-Values of Parameter Estimates from Misspecified Structural Equation Models.*Multivariate Behavioral Research.*;24(10):41–57.

Kreft I. and DeLeeuw J. (1998). Introducing Multilevel Modeling. *London: Sage*. 149 pp.

Kreft I. DeLeeuw J. and Aiken L.S. 1995. The Effect of Different Forms of Centering in Hierarchical Linear Models. *Multivariate Behavoural Research* 1995. 30. pp.1–20

Klotz J.H., Milton R.C. and Zacks S. (1969). Mean Square Efficiency of Estimators of Variance Components. *Journal of the American Statistical Association* 1969, 64, pp.1383-1394.

Kim D. and Kawachi I. (2007). U.S. state-level social capital and health-related quality of life: Multilevel Evidence of Main, Mediating and Modifying Effects. *Ann Epidemiol* 2007; 17:258–69.

Lesaffre E.and Spiessens B. (2001). The Effect of the Number of Quadrature Points in a Logistic Random-Effects Model: An Example, *Applied Statistics*, 50, pp. 325-335.

Marc Callens et al (2003). Performance of Likelihood-Based Estimation Methods for Multilevel Binary Regression Models. Dept. of Applied Economics, Katholieke Universiteit Leuven, Naamsestraat 69, B-3000 Leuven

Mason, W., Wong, G. and Entwisle, B. (1983). Contextual Analysis Through the Multilevel Linear Model. In Sociological Methodology 1983–1984, ed. S Leinhardt, pp. 72–103.

Maas C.J.M. and Hox J.J. (2004). Robustness issues in multilevel regression analysis. *Statistica Neerlandica*, 58 pp.127-137.

Maas, C.J.M and Hox, J.J (2005). Sufficient Sample Sizes for Multilevel Modeling. *European Journal of Research Methods for the Behavioral and Social Sciences* 2005, **1**:85-91.

McCulloch C.E. and Searle S.R. (2001), Generalized, Linear and Mixed Models, *New York: Wiley*.

Morris C. (1995). Hierarchical Models in Educational Data: An Overview. *Journal of Educational. Behaviour Statistics*. 20:190–200.

Ngalinda, I. (1998). Age at first Birth, Fertility and Contraception in Tanzania. PhD Thesis. Department of Demography, Humboldt University of Berlin.

Oyefara, J. L. (2012). Maternal Age at First Birth and Childhood Mortality in Yoruba Society: The Case of Osun State, Nigeria. IISTE. *Journal of Sociological Research* 2012. 3:1.

Oye-Adeniran, B.A., Adewole, I.F., Umoh, A.V. (2004). Community-based survey of unwanted pregnancy in Southwestern Nigeria. *African Journal of Reproductive Health*. 8(3):103–115.

Patrick J and Daniel J. (2007). Building Path Diagrams for Multilevel models. *American Psychological Association* 12(3), PP 283 – 297.

Pinheiro J. and Bates D. (1995). Approximations to the Log-likelihood Function in The Nonlinear Mixed-Effects Model, *Journal of Computational and Graphical Statistics*, 4(1), pp. 12-35.

Portnoy S. (1971). Formal Bayes Estimation with Application to a Random-Effects Model. *Annals of Mathematical Statistics* 1971. 42, 1379-1388.

Raudenbush S.W. and Liu X.F (2001). Effects of study duration, frequency of observation, and sample size on power in studies of group differences in polynomial change. *Psychological Methods*. **2001** 6:387-401

Rodriguez G and Goldman N. (1995). An Assessment of Estimation Procedures for Multilevel Models with Binary Responses. *Journal of the Royal Statistical Society, Series A* 1995, 158, 73-89.

Rountree P.W. and Land K.C. (1996). Perceived Risk Versus Fear of Crime: Empirical Evidence of Conceptually Distinct Reactions in Survey Data. *Social Forces* 74: 1353-76

Sanagou M., Wolfe R., Forbes A. and Reid C.M. (2012). Hospital-level associations with 30-day patient mortality after cardiac surgery: a tutorial on the application and interpretation of marginal and multilevel logistic regression. *BMC Medical Research Methodology* 2012, 12(28).

Sampson R. Raudenbush S. and Earls F. (1997).Neighborhoods and Violent Crime: A Multilevel Study of Collective Efficacy.*Science*277:918–24

Saei, A. and MacGilchris C.A. (1998). Longitudinal Threshold Models with Random Components. The Statistical Journal1998. 47:365–375.

Snijders T.A.B. and Bosker R.J.(1993). Standard Errors and Sample Sizes for 2-Level Research. *Journal of Educational Statistics* 1993, **18**:237-259.

Srikanthan A., Reid R.L.(2008). Religious and Cultural influences on contraception. Journal of Obstetric and Gynaecol . 30(2) pp129 - 37.

Susser M. (1994). The Logic in Ecological analysis. *American. Journal of Public Health*1994. 84 pp.825–29

Uthman O (2007). Multilevel Analysis of the Effect of Household Wealth Inequality on under-Five Child Under-Nutrition: Evidence from the 2003 Nigeria Demographic and Health Survey. *The Internet Journal of Nutrition and Wellness* 2007. 6 (2).

Von Korff M. (1992). Multilevel Research in Epidemiologic Research on Health Behaviors and Outcomes. *America Journal. Epidemiol* 1992. 135 pp.1077–82

Appendices

The syntax that was used for solving "initial value not feasible " problem

xi: logit nQ1212 i.NEWAGEGROUP i.Wealthquin i.Education i.Q111 i.H004_LOCATION

mata=e(b)

mat a1=(a,0,0)

xi: xtmelogit nQ1212 i.NEWAGEGROUP i.Wealthquin i.Education i.Q111 i.H004_LOCATION, || H00B_CLUSTER:,covariance (independent) intpoints(15)variance from(a1, copy).

STATA XTMELOGIT syntax for three levels

xtmelogit Q1212n .NEWAGEGROUP i.Wealthquin i.Education i.Q111 i.H004_LOCATION, || H001_ZONE

SPSS Penalized Quasi Likelihood Syntax for Two Levels

GENLINMIXED

/DATA_STRUCTURE SUBJECTS=CLUSTER

/FIELDS TARGET= new_Q1212

/TARGET_OPTIONS DISTRIBUTION=BINOMIAL LINK=LOGIT

/FIXED EFFECTS= NEWAGEGROUP Wealthquin Education Q111 H001_ZONE USE_INTERCEPT=TRUE

/RANDOM USE_INTERCEPT=TRUE SUBJECTS=CLUSTER

COVARIANCE_TYPE=VARIANCE_COMPONENTS

/BUILD_OPTIONS TARGET_CATEGORY_ORDER=DESCENDING

INPUTS_CATEGORY_ORDER=DESCENDING

MAX_ITERATIONS=1500 CONFIDENCE_LEVEL=95 DF_METHOD=SATTERTHWAITE.

The "TARGET" is the outcome and the "INPUTS" are the predictors. The SUBJECTS variable is the level designation.

ORDER=DESCENDING is used to specify that the 0 level is used as the comparison (typically what is desired) for the dependent or the independent variable. If omitted, the 1 level is used as the default.

STATA GLLAMM Syntax for Three Level

xi:gllamm nQ1212 i.NEWAGEGROUP i.Wealthquin i.Education i.Q111, i(H00B_CLUSTER H001_ZONE) family (binomial) link (logit) nip(15) adapt.