A Proportional Odds Model with Complex Sampling Design to Identify Key Determinants of Malnutrition of Children under Five Years in Rwanda

Faustin Habyarimana

*School of Mathematics, Statistics and Computer Sciences,
University of KwaZulu-Natal, South Africa
Email: singizwabb@gmail.com*

Temesgen Zewotir

*School of Mathematics, Statistics and Computer Sciences,
University of KwaZulu-Natal, South Africa
Email: zewotir@ukzn.ac.za*

Shaun Ramroop

*School of Mathematics, Statistics and Computer Sciences,
University of KwaZulu-Natal, South Africa
Email: RamroopS@ukzn.ac.za*

Doi:10.5901/mjss.2014.v5n23p1642

Abstract

The main objective of this study is to identify the key determinants of malnutrition of children under five years in Rwanda. The Rwanda demographic health survey (2010) data was used as application. The anthropometric indicator for underweight (weight-for-age) was considered and categorized as severely undernourished when z-score <-3.0, moderately undernourished when -3.0 ≤ z-score <-2.0 and nourished when z-score ≥-2.0. The score test and Brant test were used to test the proportional odds assumption and it was satisfied. As the data was collected using multistage sampling, this study extended the classical proportional odds model to include the complex sampling design. This research revealed that birth order, mother's education, gender of the child, birth weight of the child, marital status of the mother, body mass index, Anemia, multiple birth and whether or not the child had fever before the survey were found to be determinants of malnutrition of children under five years in Rwanda. The influence of these actors can be used to develop the strategies of reducing child malnutrition in Rwanda. When a complex design sample is used to draw a sample from finite population, the sample design should be incorporated in the analysis of the survey data in order to make statistically valid inferences for the finite population.

Keywords: *Child malnutrition; complex sampling design; POM; PPOM; Rwanda; underweight*

1. Introduction

Malnutrition is a serious problem in developing countries. Children are more prone to suffer from malnutrition deficiencies than adults because they are in a physiologically less stable situation. Child malnutrition is a clinical sign of nutrient deficiency manifested as stunting, underweight and wasting. These manifestations are often measured using biomedical or anthropometric indicators. However, anthropometric indicators are frequently used for its cheapness and relation availability. Commonly used anthropometric indicators of child malnutrition under the age of five years (WHO, 1995) are: 1) Height-for-age, known as stunting which is an indicator of child's long-term or chronic nutritional status and is also affected by recurrent or chronic illness. 2) Wasting is weight-for-height index which measures body mass in relation to body height and describes current nutritional status of the child. Wasting represents the failure to receive adequate nutrition in the period immediately preceding the survey and may be the result of inadequate food intake or a recent episode of illness causing loss of weight and the onset of malnutrition. 3) Weight-for-age (underweight) which is a composite index given by weight-for-height and height-for-age. Depending on the purpose of assessment and the nature of intervention, the above three indices can either be used separately or together. When anthropometric measurements are taken regularly over time, they could provide information on how the health status of the population is changing and provide a timely warning on the food supply and poverty status of a given area. If the purpose is to obtain a quick picture of a community or large body of population to understand the extent of the problem, the measurement of wasting alone would provide sufficient information. However, if the purpose is to obtain information to decide what types of programs are needed in the specific area, the study involves all three indexes of anthropometric measurements. The child malnutrition still remains a public health problem mostly in developing countries including Rwanda. But there is a strong commitment from the government of Rwanda, its development partners and educational institutions to find solutions (NISR, ICF, & MoH, 2012). In Rwanda, as in other developing countries, malnutrition is a serious problem where 44 % of children under five years old are stunted and 17 % are severely stunted (NISR et al., 2012); 11 % of children under five years old are underweight and 2 % are severely underweight; whereas 3 % of children under five years old are wasted and 1 % are severely wasted. The demographic health survey data are collected using complex survey designs such as stratification, multi-stage clustering and unequal sampling weights. Most of the studies done using binary or ordinal logistic regression did not consider the complex sampling design. In this study we try to find the determinants of malnutrition (underweight) of children under five years old in Rwanda using ordinal logistic regression with complex sampling design.

1.1 Ordinal logistic regression

In binary logistic regression, the response variable is dichotomous, where level one is experiencing the events and level two is not experiencing the events. There are some other situations that lead to a response variable with more than two categories. This response variable may be ordinal when considering ordered categorical response or multinomial when non ordered categorical response is considered. The principal logistic regression models (with ordinal and non-ordinal categories) are binary model (BM), proportion odds models (POM), partial proportional odds model without restriction (PPOM-UR) and with restriction (PPOM-R), continuation ratio model (CRM), multinomial model and stereotype model(SM) (Abreu, 2009; Siqueira, Cardoso &Caiaffa, 2008; Agresti, 2002) Mathematically these models are formulated as (Abreu, 2009; Abreu et al., 2008; Ananth & Kleinbaum, 1997; Liu, 2009)

$$\eta(X) = \ln\left[\frac{\Pr(Y=1|x_1,x_2,\dots,x_p)}{\Pr(Y=0|x_1,x_2,\dots,x_p)}\right] = \gamma + \beta_1 x_1 + \cdots + \beta_p x_p \quad (1)$$

$$\eta_j(X) = \ln\left[\frac{\Pr(Y\le j|x_1,x_2,\dots,x_p)}{\Pr(Y>j|x_1,x_2,\dots,x_p)}\right] = \gamma_j + \beta_1 x_1 + \cdots + \beta_p x_p \quad (2)$$

$$\eta_j(X) = \ln\left[\frac{\Pr(Y\le j|x_1,x_2,\dots,x_p)}{\Pr(Y>j|x_1,x_2,\dots,x_p)}\right] = \gamma_j + (\beta_1+\alpha_{j1})x_1 + \cdots + (\beta_q+\alpha_{jq})x_q + \beta_{q+1}x_{q+1} + \cdots + \beta_p x_p \quad (3)$$

$$\eta_j(X) = \ln\left[\frac{\Pr(Y\le j|x_1,x_2,\dots,x_p)}{\Pr(Y>j|x_1,x_2,\dots,x_p)}\right] = \gamma_j + \omega_j[(\beta_1+\alpha_{j1})x_1 + \cdots + (\beta_q+\alpha_{jq})x_q] + \beta_{q+1}x_{q+1} + \cdots + \beta_p x_p \quad (4)$$

$$\eta_j(X) = \ln\left[\frac{\Pr(Y=j|x_1,x_2,\dots,x_p)}{\Pr(Y>j|x_1,x_2,\dots,x_p)}\right] = \gamma_j + \beta_{j1} x_1 + \beta_{j2}x_2 + \cdots + \beta_{jp}x_p \quad (5)$$

$$\eta_j(X) = \ln\left[\frac{\Pr(Y=j|x_1,x_2,\dots,x_p)}{\Pr(Y=0|x_1,x_2,\dots,x_p)}\right] = \gamma_j + \beta_{j1} x_1 + \beta_{j2}x_2 + \cdots + \beta_{jp}x_p \quad (6)$$

$$\eta_j(X) = \ln\left[\frac{\Pr(Y=j|x_1,x_2,\dots,x_p)}{\Pr(Y=0|x_1,x_2,\dots,x_p)}\right] = \gamma_j + \tau_j(\beta_1 x_1 + \beta_2x_2 + \cdots + \beta_p x_p) \quad (7)$$

where Y is the response variable, $X = (x_1 \dots, x_p)$ is the vector of explanatory variables, $\Pr(Y \le j|X) = \Pr(Y \le j|x_1, \dots, x_p)$, from equation (2) to (4) and equation (5) to (7), $j = 1, \dots, J-1$ and $j = 1, \dots, J$ respectively, γ_j are intercepts, $\beta_1, \dots, \beta_p$ are logit coefficients, equations (1), (2), (3), (4), (5), (6) and (7) are binary model, POM, PPOM-UR, PPOM-R, CRM, multinomial and stereotype models respectively. Equation (1) is valid when the response variable is dichotomous, equation (2) is valid when originally continuous response variable was subsequently grouped and the proportional odds assumption is satisfied, equation (3) is valid when the proportional odds assumption is not valid, equation (4) is used when the proportional odds assumption is not satisfied and linear relationship for odds ratio between covariate and the response variable are not valid, equation (5) is used when the interest is in a specific category of the outcome variable (Abreu et al., 2008), equation (6) is valid when response variable is nominal with three or more categories without ordering (Abreu et al., 2008) and equation (7) is used when the response variable is intrinsically ordinal and not a discrete version of some continuous variable (Abreu et al., 2008). Ordinal data are widely available in social sciences, for example, in a clinical trial on painkiller the degree of pain control may be described as totally ineffective, poor, moderate or good. These data are commonly modeled by proportional odds model (POM) also known as cumulative odds model (Agresti, 2002, 2007; Collett, 2002; Hosmer, Lemeshow, & Sturdivant, 2000; McCullagh, 1980; McCullagh & Nelder, 1989). The POM as defined by (McCullagh, 1980) is the most popular model for ordinal logistic regression (Bender & Grouven, 1998) because of its wide applicability and intuitive interpretation (Gameroff, 2005). It estimates the cumulative probabilities of being at or below a specific level of outcome variable given a set of explanatory variables; for instance if there are *j* levels of ordinal outcomes, this model will make *j-1* predictions, each estimating the cumulative probabilities at or below the j^{th} level of the response variable (Agresti, 2002; Collett, 2002; Hosmer et al., 2000; McCullagh, 1980; O'Connell & Liu, 2011). However, when the interest of study is on a particular category not at or below that category, and that an individual must pass through a lower category level before achieving a higher category level, the continuation ratio model is more appropriate than proportional odds model (Abreu et al., 2008; Agresti, 2002, 2007; Ananth & Kleinbaum, 1997; Hardin, Hilbe, & Hilbe, 2007; Hosmer et al., 2000; Lee & Forthofer, 2006). In addition, when the need is to compare each response category to the next larger, in this case the adjacent category logistic regression model is better than POM (Agresti, 2002; Hosmer et al., 2000). The ordinal logistic regression considers that the data are collected using simple random sampling where each sampling unit has the same probability of being chosen from the population. However, if the data is collected using complex survey sampling designs, where stratified sampling, clustered sampling are used, ordinal logistic regression may not be appropriate (Anthony, 2002; Liu & Koirala, 2013). Therefore fitting such data without considering the survey sampling design may lead to biased estimates of parameters and incorrect variance estimates (Anthony, 2002; Liu & Koirala, 2013). The model formulation and interpretations will still be the same, the only difference comes when analyzing. Two techniques are widely used for unbiased variance estimation in complex sampling survey designs, including linearization and replicated sampling method (Lee & Forthofer, 2006; Levy & Lemeshow, 2011, 2013; Lohr, 2009). The linearization method known as Taylor series approximation is the default variance estimator in many statistical software. This method linearizes complex nonlinear statistics into linear functions of samples totals. The variances of these linearized approximations of the original nonlinear statistics can then be calculated by using simpler known formulas for the variance and covariance of sample totals within strata. The replicated methods estimate variance by first breaking the sample into subsamples. The desired estimate is then computed for each subsample, and the variance is calculated among the subsample estimates. The replicated methods, also referred to as resampling methods, include the balanced repeated replication (BRR), the Jackknife repeated replication and the bootstrap method (Lee & Forthofer, 2006; Levy & Lemeshow, 2011, 2013; Rao & Shao, 1996; Rust & Rao, 1996). Like linearization method, replication applies to linear estimates as well as nonlinear combinations of linear estimates. Linearization and Jackknife repeated replication are generally more accurate and more stable when the survey logistics represent functions of means; however when estimating the variance of a quintile balanced repeated replication is better than Taylor series linearization and Jackknife repeated replication (Kreuter & Valliant, 2007; West, 2008). The stratification in general decreases the variance while clustering and weighting increase the variance. This is a better scenario in practice than underestimating the variance or overstating the precision of the survey estimates. As a result, the methods tend to provide conservative estimates of variances for the survey statistics. It protects the survey analyst against the risk of overestimating the precision of the survey estimates (West, 2008). There are some studies done on malnutrition of children under five years old using either geo-additive semi-parametric mixed model, binary logistic regression or ordinal logistic regression (Das & Rahman, 2011; Kandala, Madungu, Emina, Nzita, & Cappuccio, 2011). In these studies, only geo-additive semi-parametric mixed model took into account complex sampling design but to our knowledge no study fitted the DHS data using binary logistic or ordinal logistic regression incorporated complex sampling design and yet these data are from multistage sampling. Therefore fitting such data without considering the survey sampling design may lead to biased estimates of parameters and incorrect variance estimate (Anthony, 2002; Liu & Koirala, 2013). For this reason, the current study extends the work of (Das & Rahman, 2011) to account for complex sampling design. The variance was estimated using replication method (Jackknife). We compared the parameter estimates, standard error estimates and the statistical significance from both models.

2. Method

2.1 Source of data and measurements of child malnutrition

The Rwanda Demographic Health Survey (2010) was used in the current research where its sampling technique was done in two-stage stratified sampling. In the first stage 492 villages (known as clusters or enumeration areas) were selected with probability proportional to the number of households residing in that village, 12540 households, where 2009 and 10531 households were urban and rural respectively. Secondly, systematic sampling was used to all households existing in the selected villages. RDHS (2010) collected information on women aged 15-49 and 4,356 children under five years old, on height in cm, weight in kg and age in months, birth-weight, anemia levels, gender of the child, multiple birth, region, body mass index, wealth quintile, birth order, parent's education, nutritional knowledge, type of residence (urban or rural), types of housing and toilet, sickness such as cold, cough, diarrhea, or others during the last two weeks of the study, marital status of the mother, and child's caretaker, are used as explanatory variables on malnutrition; however for brevity, only significant variables are reported. The RDHS (2010) provides the data on three anthropometric indices known as weight-for-age, height-for-age and weight-for-height. In this research, we considered weight-for-age known as underweight. We have categorized the children's nutrition status into nourished *(z-score ≥ -2.0),* moderately undernourished *(-3.0 ≤ z-score<-2.0)* and severely undernourished *(z-score <-3.0),* which made our response variable to be ordinal from a continuous variable data.

2.2 Model fitting

Nutrition status in this research is an ordinal response variable obtained from grouped continuous variables. Therefore, it lends itself to use ordinal logistic regression models (Agresti, 2002, 2007; Ananth & Kleinbaum, 1997; Collett, 2002; Das & Rahman, 2011; Hosmer et al., 2000; McCullagh, 1980). However due to the nature of sampling technique used in the demographic health survey, we have also used the complex survey design for ordinal logistic regression(Anthony, 2002; Liu & Koirala, 2013) where the variance was estimated by replicated sampling methods (Jackknife). For a POM to be valid, the assumption that all the logit surfaces are parallel or proportional odds assumption must be tested (Ananth & Kleinbaum, 1997). A nonsignificant test is taken as evidence that the logit surfaces are parallel and that the odds ratios can be interpreted as constant across all possible cut points of the outcome. If this assumption is violated it may lead to wrong interpretations(Ananth & Kleinbaum, 1997). In such case, the alternative way is to fit the data with partial proportional odds model (Ananth & Kleinbaum, 1997; Koch, Amara, & Singer, 1985; Peterson & Harrell Jr, 1990) which is available in SAS PROC GENMOD (Gameroff, 2005) and Gologit2_autofit in Stata (Williams, 2006) . The PPOM relaxes the parallel lines assumption and allows the covariates which failed to satisfy the parallel line assumption to differ across the cut-off points (Fullerton & Xu, 2012) and let constrained to parallel line assumption other covariates which satisfied the proportion odds assumption. Another alternative is to dichotomize the ordinal outcome variable by means of several cut-off points and then use separate binary logistic regression model for each dichotomous outcome variable (Bender & Grouven, 1998). However, it is suggested that the separate binary logistic regression model should be not used if possible because of the loss in statistical power and reduced generality of analytical solution (Gameroff, 2005). We have used SAS 9.3 with PROC LOGISTIC procedure to find the score test for underweight. However, this test is nonconservative (that is, it rejects the assumption very often) (Bender & Grouven, 1998; Peterson & Harrell Jr, 1990). It is convenient to use other techniques. We used Brant test command of Stata SPost package (Freese & Long J S, 2006) to find the single score test for each explanatory variable; this test can show which variable violated or did not violate the proportional odds assumption. PROC SURVEYLOGISTIC was used to fit ordinal logistic regression with sampling design. The Jackknife method was used as variance estimators. The results from PROC LOGISTIC and PROC SURVEYLOGISTIC were compared.

3. Results and Interpretation

The score test of proportional odds assumption is found not significant at 5 % level of significance (p-value=0.6421) Table 2; this means that the proportional odds assumption is satisfied. The single score test for each explanatory variable is also not significant at 5 % level of significance which also confirmed the validity of proportional odds model (Table 1). Therefore, the results revealed that the children born at 2-3, 4-5 and 6+ birth order were found 2.183 (p-value <.0001), 2.235 (p=0.0002) and 3.062 (p-value <.0001) times more likely to be in worse nutrition status respectively as compared to children born at first order Table 2. The risk of having worse nutrition status were 12.247 (p-value <.0001) and 10.555 (p-value <.0001) times higher for children born to mother without education and mother with primary education respectively as compared to children born to mother with secondary or higher education Table 2. It was found that female children were 0.687 (p-value <.0001) times less likely to be in worse nutrition status as compared to male children Table 2. The risk of having worse underweight status was 3.192 (p-value =0.0033) times higher among children born with lower weight (<2500g) as compared to children born with higher weight (≥ 2500g) Table 2. The children born at first multiple (twin) were 3.574 (p-value =0.0020) times more likely to be in worse nutrition status as compared to singleton child at birth and the effect of second multiple birth was not significant (p-value =0.1302). The risk of having worse underweight was 1.403 (p-value =0.0045) times higher among the children born to anemic mother, when compared to children born to non-anemic mother Table 2. A child born to married mother or mother living with a partner was 0.577 (p-value =0.0166) times less likely to be in worse nutrition status as compared to child born to divorced or separated mother; however, the effect of child born to widower or mother who had never been in union was not significant as compared to child born to divorced or separated mother Table2. A child born to thin mother (BMI<18.5) was 2.601 (p-value =0.0002) times more likely to be in worse nutrition status as compared to a child born to normal or obese mother (BMI ≥ 18.5) Table 2. Children without fever in two weeks before the survey were 0.705 (p=0.0283) times less to be in worse nutrition status as compared to children who had fever in two weeks before the survey Table 2.

Table 1. Parameter estimate from POM without sampling design

Indicator	Estimate	SE	P-value	OR	Single P-value
Intercept1	-5.554	0.5391	<.0001		
Intercept2	-3.6564	0.5258	<.0001		
Birth order (First =reference)					0.133
2-3	0.7505	0.1788	<.0001	2.118	
4-5	0.7320	0.1980	0.0002	2.079	
6+	1.0510	0.1964	<.0001	2.861	
Mother's education (Secondary & higher=reference)					0.413
Primary	1.9191	0.4535	<.0001	6.815	
No education	2.1339	0.4650	<.0001	8.448	
Gender of the child (Male=reference)					0.996
Female	-0.4013	0.1191	0.0008	0.669	
Knowledge on nutrition (No=reference)					0.4171
Yes	-0.2768	0.1287	0.0315	0.758	
Birth weights ($\geq$ 2500g=reference)					0.837
$\leq$ 2500g	1.1736	0.2462	<.0001	3.234	
Multiple birth (Singleton=reference)					0.539
First multiple	1.3445	0.4198	0.0014	3.836	
Second multiple and more	0.7221	0.3980	0.0696	2.059	
Anemia (No anemic=reference)					0.492
Anemic	0.3327	0.1179	0.0048	1.395	
Marital status (Divorced/separated=reference)					0.757
Never in union	-0.2268	0.3258	0.4864	0.797	
Married/partner	-0.6268	0.2186	0.0041	0.534	
Widowed	-0.4261	0.4132	0.3024	0.653	
BMI (BMI $\geq$ 18.5=reference)					0.538
BMI <18.5	0.9272	0.2156	<.0001	2.527	
Had fever (Yes=reference)					0.282
No	-0.3842	0.1440	0.0076	0.681	
Score test for proportional odds assumption	$\chi^2 = 14.868$	Df=16	p-value=0.5343		
Goodness of fit(likelihood ratio)	$\chi^2 = 162.334$	Df=16	p-value<.0001		

Table 2. Comparison of the POM without and with complex survey design

Indicator	Estimate	**SE**	P-value	Estimate	OR	SE	P-value
Intercept1	-5.554	0.5391	<.0001	-6.1520		0.6432	<.0001
Intercept2	-3.6564	0.5258	<.0001	-4.2422		0.6318	<.0001
Birth order (First =reference)							
2-3	0.7505	0.1788	<.0001	0.7807	2.183	0.1717	<.0001
4-5	0.7320	0.1980	0.0002	0.8043	2.235	0.2190	0.0002
6+	1.0510	0.1964	<.0001	1.1190	3.062	0.2050	<.0001
Mother's education (Secondary & higher=reference)							
Primary	1.9191	0.4535	<.0001	2.3566	10.556	0.5649	<.0001
No education	2.1339	0.4650	<.0001	2.5053	12.247	0.5637	<.0001
Gender of the child (Male=reference)							
Female	-0.4013	0.1191	0.0008	-0.3753	0.687	0.1276	0.0033
Knowledge on nutrition (No=reference)							
Yes	-0.2768	0.1287	0.0315	-0.2806	0.765	0.1380	0.0523
Birth weights ($\geq$ 2500g=reference)							

$\leq$ 2500g	1.1736	0.2462	<.0001	1.1607	3.192	0.2563	<.0001
Multiple birth (Singleton=reference)							
First multiple	1.3445	0.4198	0.0014	1.2737	3.574	0.4129	0.0020
Second multiple and more	0.7221	0.3980	0.0696	0.6080	1.837	0.4018	0.1302
Anemia (No anemic=reference)							
Anemic	0.3327	0.1179	0.0048	0.3389	1.403	0.1194	0.0045
Marital status (Divorced/separated=reference)							
Never in union	-0.2268	0.3258	0.4864	-0.0568	0.945	0.3435	0.8687
Married/partner	-0.6268	0.2186	0.0041	-0.5494	0.577	0.2293	0.0166
Widowed	-0.4261	0.4132	0.3024	-0.2960	0.744	0.5003	0.5541
BMI (BMI $\geq$ 18.5=reference)							
BMI <18.5	0.9272	0.2156	<.0001	0.9559	2.601	0.2563	0.0002
Had fever (yes=reference)							
No	-0.3842	0.1440	0.0076	-0.3497	0.705	0.1595	0.0283
Score test for proportional odds assumption	$\chi^2 = 14.8680$	Df=16	p-value=0.5343	$\chi^2 = 13.4160$	Df=16	p-value=0.6421	
Goodness of fit (likelihood ratio)	$\chi^2 = 162.334$	Df=16	p-value<.0001	$\chi^2 = 166.633$	Df=16	p- value<.0001	

4. Discussion

When comparing the results from classical proportional odds model and proportional odds model with sampling design, we found that the model with sampling weight had higher standard deviation as compared to model without sampling weights (Table 2); this is in line with (Liu & Koirala, 2013). It means that the proportional odds model without sampling design may underestimate the standard error. As a result some covariates may be significant and yet when sampling design is included are not significant; for instance in our case knowledge on nutrition was found significantly affecting malnutrition when proportional odds model used (p-value=0.0320) Table 1 but is not significant (p-value=0.0523) when proportional odds model with sampling design is used Table 2. When the proportional odds model without sampling design was considered, we found that birth order, mother's education, gender of the child, knowledge on nutrition, birth weights, multiple birth, body mass index, anemia, marital status and whether the child had or not fever in two weeks before the survey were determinants of malnutrition of children under five years in Rwanda. However, in the case of proportional odds model with sampling design, we found almost the same determinants of malnutrition of children under five years except knowledge on nutrition which is not significantly affecting malnutrition. When the data is from simple random sampling, the ordinal logistic model and ordinal logistic model with sampling design are identical. We were also interested to use interaction effects, unfortunately we did not find any significant interaction effect.

5. Conclusion and Suggestions

This study used ordinal logistic regression (classical proportional odds model) and ordinal logistic regression with sampling design (proportional odds model with complex sampling design) to identify the key determinants of malnutrition of children under five years in Rwanda. We first fitted the proportional odds model without considering the sampling design and thereafter proportional odds model with sampling design was fitted. All results of these two models were compared. From these results, we found that it is better to include sampling design when the data was obtained from multistage sampling in order to make statistically valid inferences from the finite population. Birth order, mother's education, gender of the child, knowledge on nutrition, birth weights, multiple birth, body mass index, anemia, marital status and whether the child had fever or not in two weeks before the survey were determinants of malnutrition of children under five years in Rwanda. The influence of these factors can be used to develop the strategies of reducing child malnutrition in Rwanda.

6. Acknowledgements

The authors acknowledge National Institute of Statistics of Rwanda (NISR) [Rwanda], Ministry of Health (MOH) [Rwanda], and ICF International for data.

References

Abreu, M. N. S. (2009). Ordinal logistic regression in epidemiological studies *Revista de Saude publica, 43*(1), 183-194.

Abreu, M. N. S., Siqueira, A. L., Cardoso, C. S., & Caiaffa, W. T. (2008). Ordinal logistic regression models: application in quality of life studies. *Cadernos de Saúde Pública, 24*, 581-591.

Agresti, A. (2002). *Categorical Data Analysis*: John.

Agresti, A. (2007). An introduction to categorical data analysis. Hoboken: NJ: Wiley.

Ananth, C. V., & Kleinbaum, D. G. (1997). Regression models for ordinal responses: a review of methods and applications. *International journal of epidemiology, 26*(6), 1323-1333.

Anthony, B. A. (2002). *Performing logistic regression on survey data with the new SURVEYLOGISTIC procedure.* Paper presented at the Proceedings of the twenty-seventh annual SAS® users group international conference.

Bender, R., & Grouven, U. (1998). Using binary logistic regression models for ordinal data with non-proportional odds. *Journal of clinical epidemiology, 51*(10), 809-816.

Collett, D. (2002). *Modelling binary data*: CRC press.

Das, S., & Rahman, R. M. (2011). Application of ordinal logistic regression analysis in determining risk factors of child malnutrition in Bangladesh. *Nutrition journal, 10*(1), 124.

Freese , J., & Long J S, J. S. (2006). *Regression models for categorical dependent variables using stata*: Stata Press.

Fullerton, A. S., & Xu, J. (2012). The proportional odds with partial proportionality constraints model for ordinal response variables. *Social science research, 41*(1), 182-198.

Gameroff, M. (2005). *Using the proportional odds model for health-related outcomes: Why, when, and how with various SAS® procedures.* Paper presented at the SUGI.

Hardin, J. W., Hilbe, J. M., & Hilbe, J. (2007). *Generalized linear models and extensions*: Stata Press.

Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2000). *Introduction to the logistic regression model*: Wiley Online Library.

Kandala, N.-B., Madungu, T. P., Emina, J. B., Nzita, K. P., & Cappuccio, F. P. (2011). Malnutrition among children under the age of five in the Democratic Republic of Congo (DRC): does geographic location matter? *BMC public health, 11*(1), 261.

Koch, G. G., Amara, I. A., & Singer, J. M. (1985). A two-stage procedure for the analysis of ordinal categorical data. *Biostatistics: Statistics in Biomedical, Public Health and Environmental Sciences, ed. PK Sen. North Holland: Elsevier Science*, 357-387.

Kreuter, F., & Valliant, R. (2007). A survey on survey statistics: What is done and can be done in Stata. *Stata Journal, 7*(1), 1-21.

Lee, E. S., & Forthofer, R. N. (2006). *Analyzing complex survey data*: Sage.

Levy, P. S., & Lemeshow, S. (2011). *Sampling of populations: Methods and applications* (Vol. 543): John Wiley & Sons.

Levy, P. S., & Lemeshow, S. (2013). *Sampling of populations: methods and applications*: John Wiley & Sons.

Liu, X. (2009). Ordinal regression analysis: Fitting the proportional odds model using Stata, SAS and SPSS. *Journal of Modern Applied Statistical Methods, 8*(2), 30.

Liu, X., & Koirala, H. (2013). Fitting Proportional Odds Models to Educational Data with Complex Sampling Designs in Ordinal Logistic Regression. *Journal of Modern Applied Statistical Methods, 12*(1), 26.

Lohr, S. (2009). *Sampling: design and analysis*: Cengage Learning.

McCullagh, P. (1980). Regression models for ordinal data. *Journal of the royal statistical society. Series B (Methodological)*, 109-142.

McCullagh, P., & Nelder, J. A. (1989). Generalized linear models.

NISR, ICF, & MoH. (2012). Rwanda Demographic and Health Survey 2010.

O'Connell, A. A., & Liu, X. (2011). Model Diagnostics for proportional and partial proportional odds models. *Journal of Modern Applied Statistical Methods, 10*(1), 15.

Peterson, B., & Harrell Jr, F. E. (1990). Partial proportional odds models for ordinal response variables. *Applied Statistics*, 205-217.

Rao, J., & Shao, J. (1996). On balanced half-sample variance estimation in stratified random sampling. *Journal of the American Statistical Association, 91*(433), 343-348.

Rust, K. F., & Rao, J. (1996). Variance estimation for complex surveys using replication techniques. *Statistical methods in medical research, 5*(3), 283-310.

West, B. T. (2008). Statistical and methodological issues in the analysis of complex sample survey data: practical guidance for trauma researchers. *Journal of traumatic stress, 21*(5), 440-447.

WHO. (1995). Pysical Status: the Use and Interpretation of Anthropometry:Report of a WHO Expert Committe.

Williams, R. (2006). Generalized ordered logit/partial proportional odds models for ordinal dependent variables. *Stata Journal, 6*(1), 58-82.