Establishing a robust technique for monitoring and early warning of food insecurity in post-conflict Southern Sudan using Ordinal Logistic Regression

LB. LOKOSANG[1], S. RAMROOP AND SL HENDRIKS

The lack of a "gold standard" to determine and predict household food insecurity is well documented. While a considerable volume of research continues to explore universally applicable measurement approaches, robust statistical techniques have not been applied in food security monitoring and early warning systems, especially in countries where food insecurity is chronic. This study explored the application of various Ordinal Logistic Regression techniques in the analysis of national data from Southern Sudan. Five Link Functions of the Ordinal Regression model were tested. Of these techniques, the *Probit* Model was found to be the most efficient for predicting food security using ordered categorical outcomes (Food Consumption Scores). The study presents the first rigorous analysis of national food security levels in post-conflict Southern Sudan and shows the power of the model in identifying significant predictors of food insecurity, surveillance, monitoring and early warning.

Key Words: Ordinal Logistic Regression, Proportional Odds Model, Probit Model, Generalised Linear Regression, Link Function, Food Insecurity, Food Consumption Scores/Groups.

1. INTRODUCTION

Food security programmers and policy makers seem to have drawn a foregone conclusion that food insecurity in Southern Sudan is a function of war and displacement, combined with unfavourable environmental conditions such as drought and floods. For this reason, current food insecurity interventions in the Intergovernmental Authority on Development (IGAD) region seem to be geared towards food aid and short-term relief, rather than chronic food insecurity.

[1] Respectively: PhD candidate, University of KwaZulu-Natal; Senior Lecturer in Statistics, University of KwaZulu-Natal; and Associate Professor in Food Security, University of Pretoria

Although some smaller community and household data collection surveys were conducted in the region over the last few years, they were conducted in an isolated and uncoordinated fashion. Analysis of data from these surveys is often limited to satisfying the purposes of the initiating organisation. The Food and Agricultural Organisation (FAO, 2008) of the United Nations' Sudan Information for Food Security in Action Programme notes the gaps and limitations facing food security information systems as: reliance on out-of-date census and baseline data, lack of standard data collection methodologies, and lack of coordination among stakeholders. Often events overtake the outcomes of the surveys, rendering the information too late to inform action. Sometimes the commissioning institutions lack analytical capacity.

This study sets out to forecast possible occurrence of food insecurity based on a set of determined factors using Ordinal Logistic Regression. An analysis of possible causes of household food insecurity was conducted by examining a subset of data collected in a major household survey – the Sudan Household Health Survey (2006).

Most food insecurity measures are calculated from household surveys that yield qualitative or categorical responses. The response (or dependent) variable is often in the form of a score. Examples of the measures currently in use are Food Consumptions Scores, dietary diversity indexes, the Household Food Insecurity Access Scale and others. To determine the level of vulnerability, the scores are classified into ordered categories ranging from, say, 'extremely vulnerable' to 'better off'.

Since food security surveys are often conducted for vulnerability assessment and intervention programmes to reduce hunger through provision of food aid, analysis often stops at determining the percentage of households falling into each ordered category. Knowing what determines the level and magnitude of the relationship and predicting the probability of occurrence of food insecurity are often not the concerns of food security programmers. However, emerging food security policies, such as those posited by the Framework for Africa's Food Security (FAFS) of the African Union's New Partnerships for African Development (NEPAD) Comprehensive Africa Agriculture Development Programme (CAADP), categorically recommend implementing measures to predict risk to food security and livelihoods.

Ordinal logistic regression procedures - featuring the *Probit* Model - avail the opportunity to predict the probability of occurrence of food insecurity by modelling the relationships between the ordered categories of a response variable and a set of predictor (or explanatory) variables. Ordinal logistic regression enables determination of the magnitude and significance of each relationship and identifies the set of predictor variables that can recurrently be used in food insecurity surveillance and monitoring. This paper demonstrates the potential of the *Probit* model as a practical tool for not only predicting the risk of food insecurity for early warning and contingency planning but also for surveillance and monitoring. The tool offers opportunities for national food security information systems to measure and monitor food insecurity and reduce risk as set out in the CAADP's FAFS (The African Union/NEPAD, 2009).

2. CURRENT CHALLENGES IN MEASURING FOOD INSECURITY

The importance of measuring household food insecurity cannot be over-emphasised (Bently & Pelto, 1991). National sampling is usually based on census data sampling frames and use households as the base unit of analysis. Households are the social institutions through which individuals access food (FANTA, 2003; Maxwell *et al*., 2003). Bickel *et al*. (2000) affirm that food security is an essential, universal dimension of household and personal well-being and food security and hunger are precursors to nutrition, health, and developmental problems. Hoddinott (1999) reported on the value of household-based measurements to identify food insecurity, assess the severity of food insufficiency, and characterise the nature of the insecurity.

Food security measurement and monitoring is carried out for various purposes. Riely *et al.* (1999) underscore the importance of food security information that goes beyond programme monitoring and impact evaluation, to the design of relief and development interventions. Human nutrition practitioners monitor food intakes. Public health practitioners and leading international health organisations, such as the United Nations Children Fund (UNICEF) and the World Health Organisation (WHO), are interested in information on mortality and morbidity, especially of mothers and children under five years.

Webb *et al*. (2006) observe that a number of agencies lack methodologies that distinguish household food security levels to target and evaluate programmes. De Haen (FAO/FIVIMS, 2002), speaking at the closure of the International Scientific Symposium on Measurement and Assessment of Food Deprivation and Under-nutrition, points out that analysis of food insecurity lacks a perfect single measure that captures all aspects of food insecurity. The complexity of food security, as a crosscutting issue, has complicated the challenge of finding a summative (or 'gold standard') measure of household food insecurity (Coates *et al*., 2003). For example, the Food and Nutrition Technical Assistance (FANTA) Project (2003) includes 33 recommended indicators for measuring food insecurity access alone.

A number of food insecurity indicators are outlined below in five categories, namely: food sufficiency, food access, food utilisation, vulnerability, and resilience to shocks and stresses. Table 1 summarises the available measures of household food insecurity. The table provides ample evidence to confirm that the diverse measures of household food insecurity differ in purpose, scope, scale and efficacy. They also vary in type of data collection instruments, methodology and analysis approaches. Vulnerability and food access dominate the literature on food security measurement.

Table 1 Common household food insecurity measurement approaches, uses and main purposes

	Approach	**Main Aim(s) or Focus**	**Main Uses**
1	Household Economy Analysis/ Approach (Save the Children Fund-UK, 2000)	– Helps in operational decision-making (Boudreau, 1998) – Quantitatively predicts that an event (e.g. crop failure or price change) is likely to affect people's ability to access food (SMART‡, 2006)	– Assessing relief needs, rationalising the use of food aid and early warning of food crises (Boudreau, 1998) – Understanding how poor people make ends meet and the reasons for rural-urban migration (Boudreau, 1998) – Developing policies against chronic hunger (Boudreau, 1998)

	Approach	Main Aim(s) or Focus	Main Uses
2	Household Food Insecurity Access Scale	– Prevalence of household food insecurity access – Monitoring of food insecurity access over time (Coates *et al.*, 2007)	– Monitoring of food assistance programmes (Coates *et al.*, 2007) – Assessment of programme impact (Coates *et al.*, 2007)
3	Food Access Survey Tool (FAST)	Food security-related programming and assessment for operational purposes (Coates *et al.*, 2003)	Guiding, monitoring and evaluating food security access operational interventions (Coates *et al.*, 2003) Assessing poor people's perceptions of food insecurity and measure the experience of hunger (Coates *et al.*, 2003)
4	Malnutrition Surveillance Measurement (Anthropometry)	– Highlights the need for a special food insecurity-related intervention and the target population (Setboonsarng, 2005)	– Assessment of magnitude, distribution, and severity of a nutrition problem (Setboonsarng, 2005) – Proxy measure of a household's socio-economic level (Swindale & Bilinsky, 2006)
5	Dietary Consumption/ Food Intake	– Impact of household food Access programmes – Dietary Diversity Score is a proxy measure of food access (Swindale & Bilinsky, 2006)	– Improved household food consumption (Swindale & Bilinsky, 2006) – Quality of diet (Setboonsarng, 2005) – A proxy for socio-economic level of the household (Swindale & Bilinsky, 2006)

‡ Measuring Mortality, Nutritional Status, and Food Security in Crisis Situations

Access indicators based on quantitative data enable predictive analysis and require expert, rigorous, and time-consuming data collection and analysis. In a crisis situation which demands timely decision-making and immediate intervention, these tools are not helpful.

The strengths and weaknesses of seven widely used indicators are compared (Table 2). Despite their dependence on qualitative data, the Coping Strategy Index, Dietary Diversity Score and Household Food Insecurity Score, are rated as easy-to-use and readily estimated.

Table 2 Strength and weaknesses of seven major indicators used in measuring and monitoring household food insecurity

	Indicator	**Strengths**	**Weaknesses**
1	Coping Strategy Index (CSI)	– Rapid measure of short-term household food insecurity (Maxwell *et al.*, 2003) – Easy to implement and directly captures perceptions of availability and vulnerability (Hoddinott, 1999a) – Good proxies for food intake etc. (Christiansen & Boivert, 2000; Hendriks, 2005; Maxwell *et al.*, 2003) – Ability to identify changes in household conditions as a result of emergency food aid operations (Hendriks, 2005)	– The assessment cannot be repeated for the same community as respondents may alter their responses to the coping strategy behaviour questions in subsequent rounds (Hendriks, 2005) – Caution needs to be taken in the interpretation of results as some coping strategies are reversible while others are not (Gillespie et al., 2001; Loevinsohn & Gillespie, 2003) – High susceptibility to misreporting (Hoddinott, 1999a)
2	Dietary Diversity Score (DDS)	– Easy to use and straightforward taking less than 10 minutes to complete a questionnaire (Hoddinott, 1999a; Swindale & Bilinsky, 2006) – Tracks seasonal changes in	– If responses are not weighted, the method does not record quantities (Hoddinott, 1999a) – It is not possible to estimate the extent to which diets are inadequate in terms of caloric availability, unless the frequency

	Indicator	Strengths	Weaknesses
		food security (Hoddinott, 1999a) – Enables examination of food insecurity at household and intra-household levels (Swindale & Bilinsky, 2006)	of consumption of particular diets is probed (Hoddinott, 1999a)
3	Child malnutrition	– Relatively easy data collection (Nandy *et al*., 2003) – Based on quantitative data and therefore more objective – Can lead to additional summary descriptive statistics – Powerful for evaluation of household, sub-national and national food insecurity status and programmes	– Requires a large sample to arrive at efficient statistics for concluding the nutritional level of an area – Results not easy to interpret *at any one time* i.e. requires time series (Setboonsarng, 2005)
4	Household Food Insecurity Access Scale (HFIAS) Anthropometry	– Method has been tested and validated in some developing countries and generated required indicators (Coates *et al*., 2003; Frongillo & Nanama,2006; Webb & Rogers, 2003) – Proven sensitivity to different cultural contexts (Coates *et al*., 2007)	– Heavily depends on individual perceptions of food security access aspects and thus cannot be standardised with regard to different cultural contexts – requires adaptations to local settings and compromises (do you mean ‘compromises’?) comparability of information across countries or regions

	Indicator	Strengths	Weaknesses
5	Dietary energy consumption	– Can produce more accurate measures of individual energy intake if measured correctly (Hoddinott, 1999b) – Can indicate within household disparities of food insecurity status (Hoddinott, 1999b)	– Too cumbersome to be used for targeting food aid (Chung *et al*., 1997) – Questionable reliability of data sources (Boudreau, 1998) – Reliance on expert data collectors and analysts (Hoddinott, 1999b) – Requires repeated measurements (Hoddinott, 1999b)
6	Household Economy Analysis	– Provides decision makers with an understanding of the picture of a rural economy, thus helping in food aid programming (Boudreau, 1998)	– Reliance on expert analysis (own observation) – Requires a high degree of training, well educated, committed and enthusiastic staff (Boudreau, 1998) – Non-verifiability of results, costly, time consuming and impracticable (Boudreau, 1998)
7	Sustainability Assessment-based Resilience	– Quick, does not require rigorous calculation and prediction of impending vulnerability – Can be used for surveillance of food insecurity – Instantaneous results; readily shared with the respondent and easily summarised	– Untested. Therefore, no universal agreement on method – Weighting of scores is arbitrary i.e. susceptible to prolonged debate – How to adjust scores on the percentile scale threshold could be challenging

Note: Where no citation is provided, the postulations are those of the authors.

Hoddinott (1999a) paints a clearer picture of the advantages and disadvantages of four of the above indicators as presented in Table 3 in terms of some intrinsic qualities. Table 3 shows that the Dietary Diversity Scores and the Coping Strategy Index are relatively stronger measures for rapid household surveys.

Table 3 Comparison of methods for monitoring household food insecurity (Hoddinott, 1999a)

	Food Intake	**Household Energy Acquisition**	**Dietary Diversity**	**Coping Strategies**
Data collection costs	High	Moderate	Low	Low
Time required for analysis	High	Moderate	Low	Low
Skill level required	High	Moderately High	Moderately low	Low
Susceptibility to misreporting	Low	Moderate	Low	High

3. DATASET AND METHODOLOGY

3.1 Raw data and sample

The raw data used for the study were obtained from the 2006 Sudan Household Health Survey, which was modelled on Multiple Indicator Cluster Survey (MICS) methodologies (United Nations Children Fund, 2008). There was a separate instrument for collecting household food security and livelihoods data. The Food Security Questionnaire covered seven sections or modules, namely: household circumstances; household belongings, including ownership of land

and livestock; livelihoods and agricultural production; household expenditures; food consumption and sources; shocks, coping mechanisms and food aid.

A total of 9,557 households were enumerated in Southern Sudan, representing a response rate of 95.6 *per cent*. However, cases with missing entries were removed and the resultant working sample size was 9,220 households. The raw dataset was extracted from the Survey Data Processing (CSPro) version 3.0 package (United States Census Bureau, 2006)) and analysed in SPSS version 15 (SPSS, 2006).

3.2 The main response variable

The main response variable under study was the Food Consumption Score. This score was based on dietary diversity, food frequency and the relative weight of food with nutritional importance consumed in a defined period by a household or in a geographical area (World Food Programme (WFP), 2008). The score indicates the availability and consumption of specific food groups to determine the extent of vulnerabilities *vis-à-vis* the level of food insecurity in an area. According to Swindale and Bilinsky (2006), the Household Dietary Diversity Score measures food access as a proxy of socio-economic status. Twelve food groups were identified: cereals; roots and tubers; vegetables; fruits; meat, poultry and offal; eggs; fish and seafood; pulses, legumes or nuts; milk and milk products; oil/fats; sugar or honey and miscellaneous food items.

A number of steps led to the calculation of the Food Consumption Score (FCS), which was calculated for each household based on the standards set out in Table 4 below. The number of food groups consumed by members of the same household was aggregated. The number of times a food item was consumed in a week, or the frequency of food consumption and the standard weighting of the food group, provided the basis for calculation of the FCS.

Table 4 **Standard food groups and standard weights for calculation of the Food Consumption Score (World Food Programme (WFP) Vulnerability Assessment and Mapping Unit, 2008)**

	Food consumption group	**Food group**	**Weight (definitive)**
1	Maize , maize porridge, rice, sorghum, millet pasta, bread and other cereals	Main staples	2
	Cassava, potatoes and sweet potatoes, other tubers, plantains		
2	Beans. Peas, groundnuts and cashew nuts	Pulses	3
3	Vegetables, leaves	Vegetables	1
4	Fruits	Fruit	1
5	Beef, goat, poultry, pork, eggs and fish	Meat and fish	4
5	Milk yogurt and other dairy	Milk	4
6	Sugar and sugar products, honey	Sugar	0.5
7	Oils, fats and butter	Oil	0.5
8	Spices, tea, coffee, salt, fish power and small amounts of milk for tea.	Condiments	0

Food consumption groups (FCGs) were derived by aggregating individual frequencies of food items in the same group. The Food Consumption Scores were then calculated for each household by summing the products of the frequencies of the FCG multiplied by the corresponding weight. The thresholds for the FCS were based on the criteria shown in Table 5 below.

Table 5 Profiling of food consumption behaviour based on the Food Consumption Score (World Food Programme, 2008)

Food consumption score	Food security level
≤ 28	Poor
28.1 – 42	Borderline
42.1 – 105	Good

3.3 The set of predictor variables

The dataset included a number of explanatory variables that were assumed to be associated with the Food Consumption Scores (FCS). Nineteen predictors were included in an Ordinal Logistic Regression Model. The selected set of predictors included three factors with quantitative (or ratio scale) values (Table 6), otherwise referred to as covariates in this text. Some variables were excluded during refinement of the model to avoid multicolinearity, non-additivity and heteroscedasticity (Hosmer & Lemeshow, 2000; Menard, 2002). These variables included: meals eaten per day by adults in normal and hunger periods (in the week preceding the survey), meals eaten per day by children in normal and hunger periods, the number of days in the past week each food item was eaten, the number of under 5 children, and the number of women in the household.

Table 6 List of independent variables included in the first model

	Variable description	Reason for Inclusion
1	State	States are units of monitoring and focus of economic policy-making. Sharp disparities in food consumption could arouse concerns and instigate further investigation and action by relevant federal and state authorities.
2	Household type	Households might differ in their consumption levels according to whether they are displaced or not.
3	Number of household members	It is possible that food consumption is strained by large household sizes. It is also possible that a large household size could present an opportunity in that there could be more working adults and hence more sources of food. Either way, food consumption might be affected.
4	Gender of household head	Are male-headed household better off in food consumption levels than female-headed, or vice-versa?
5	Level of education of household head	Are households headed by persons with higher education levels significantly better off than those of lower education?
6	Ownership of land for agricultural purposes (farming, grazing or fishing)	It is possible that households with land tenure are presented with opportunities to have better income and food sources than those without tenure.
7	Use of land for farming	Households using land for farming are likely to produce food and can turn farm produce into income and improve access to food better than households with no land or land they do not use.

	Variable description	Reason for Inclusion
8	Land planted previous season	Households where land was planted in the previous season might have enough food stock and have better FCS scores than those who did not plant.
9	Ownership of livestock	Livestock ownership could be a ready source of food high in micronutrients and protein thus improving the dietary diversity and food consumption frequency of households over those that do not own livestock.
10	Usual migration of households	Instability of households could cause strains on household budgets and food consumption.
11	Number of harvests in one year	Farming households harvesting bi-annually should have food available throughout the year and could have a better dietary diversity of micronutrient rich food than other households.
12	Months harvest food lasted	It is worthwhile finding out how households fared in terms of food consumption with regard to stocking farmed food.
13	Availability of vegetable plot or home garden	Availability of a plot of land or a vegetable garden increases the potential of a household's dietary intake of vitamin rich foods. Therefore the FCS is expected to be better for families with home gardens than those without.
14	Main sources of livelihood	It is out of question that the main source of livelihood is supposed to influence food consumption positively in that the stronger the weight of the main source, the better the access to a diversity of diets.
15	Number of meals per day	Having more meals a day increases the diversity of diets. Hence it could be worthwhile knowing the magnitude of its influence

	Variable description	Reason for Inclusion
		and significance.
16	Main source of sorghum and millet	Cereals represent the main staple for most countries. It is therefore important to investigate the source(s) that improve or worsen the probability of it being in a better food consumption group.
17	Experience of shock or strain	It is expected that households that had experienced shock or strain might accordingly experience inadequate dietary intake.
18	Incidence of receipt of food aid in the last 3 months	It is worthwhile investigating whether receipt of food aid had a significant effect on FCS. It is normal to expect that a household receiving food aid would have better FCS than one that does not. This notion could be wrong in circumstances when a household entirely dependent on food aid has lower dietary diversity than a household that does not receive food aid.
19	Wealth Index Quintiles	Wealth Index Quintiles are proxy measures of poverty and hence, by extension, indirect measures of food insecurity.

The state was included as there are inter-state variations (Government of Southern Sudan, 2008). The variables labelled 2 to 18 are persistently investigated in food security vulnerability assessment surveys. The demographic variables, namely: household type, number of household members, gender of household head, level of education of household head, and migration or stability of households, are consistently included in food security surveys. The variables for livelihoods, subsistence and resilience to shocks, labelled 14 to 18, are included in numerous studies. Similarly, variables on ownership and use of land, ownership of livestock, availability of home gardens and incidence and availability of stock of harvests, are consistent in food security vulnerability and assessment surveys. These variables were extensively reported in the 2006 Sudan Household Health Survey report (Government of Southern Sudan, 2008). This underscores that these variables are regarded as important factors affecting food security of households and vulnerable population groups.

An important variable derived by calculation was the Wealth Index Score (*WIS*), with values classified into quintiles (i.e., 20th percentiles of the wealth index), to create Wealth Index Quintiles. The wealth index scores and quintiles were extracted from the original dataset. The calculation of the index was based on ownership of assets and weightings determined using Principal Component Analysis. The procedure to do this is explained in the 2006 Sudan Household Health Survey report as well as in the WFP Comprehensive Food Security and Vulnerability Analysis Report (WFP, 2007a).

Households were classified into five ordered categories. Seventy-one *per cent* of households fell into the 'poorest' to 'moderate' wealth index brackets. This was not surprising given the post-conflict setting of Southern Sudan and that Sudan has successively ranked among the 50 poorest nations in terms of the Human Development Index (United Nations' Development Programme, 2009). As the wealth index was an indirect measure of food in/security, it was important to examine the statistics of the relationship between the Food Consumption Score and Wealth Index Quintiles.

3.4 Data analysis methods

The SPSS Version 15 Ordinal Regression technique was used to:

- investigate the goodness-of-fit of the model
- generate parameter estimates for determining difference between categories of the response variable
- calculate fitted probability values
- give model inspection goodness-of-fit statistics and
- produce tests of hypothesis of the significance of the relationships or association between categories of the response variable and levels of the significant predictors.

Tests of hypothesis were based on the Likelihood Ratio Test, the Score Test and the Wald's Chi-square Test. Interpretation of the outputs of the Ordinal Regression procedures enabled an in-depth understanding of the results and findings.

Although it is possible to analyse the data using the Ordinary Linear Regression (OLR) model, by treating the calculated values of the FSC as those of a continuous or interval scale variable, the Ordinal Logistic Regression model, otherwise known as the Proportional Odds Model (POM), was preferred. This was mainly because interest was centred on modelling the ordered groups of the Food Consumption Scores (i.e. the variable FCG instead of FCS), rather than the continuous variable.

3.5 The Logistic Regression model for ordered categorical data

Logistic Regression is a member of a family of Generalized Linear Models – a methodology developed by McCullagh and Nelder (1989) and uses a generalisation of Linear Regression for prediction of the cumulative probabilities of the ordered categories of the response variable. The method enables fitting a set of equations for each category of the ordered dependent variable with each equation, giving predicted probabilities of being in the corresponding category or any of the categories that are higher in rank (SPSS, 2006).

3.5.1 The concept of 'proportional odds'

There are a number of methods available to model ordered categorical data. Some of the more widely used are: linear-by-linear models, continuation ratio *Logits,* and proportional odds (Collet, 2003; Hosmer & Lemeshow, 2000). In this study, there were three categories for the response variable: 'poor', 'borderline poor' and 'good' food consumption. Because of the nature of the data, only the Proportional Odds Model (POM) is discussed here.

The Proportional Odds Model can be understood as an extension of logistic regression or, as proposed by Collet (2003), a generalisation of the Logistic Regression Model. Therefore, calculation and interpretation of the model parameters and deviance statistics are the same as for Logistic Regression for binary response data. The technique allows modelling of ordered data by converting the data into a number of dichotomies. A Binary Logistic Regression models one dichotomy whereas the proportional odds model uses a number of dichotomies. The ordered data are arranged as a series of binary comparisons. For the study dataset, a three-category ordered variable (coded 1, 2 and 3) was represented as two comparisons: category 1 compared to categories 2 and 3; categories 1 and 2 compared to category 3. This method of modelling is referred to as the 'all possible Logistic Regression model' (Collet, 2003, p 325-6).

The Proportional Odds Model, rather than predicting the actual cumulative probabilities, predicts a function of their values called the 'Link Function'. The SPSS Ordinal Regression Procedure (or PLUM) provides a choice among several link functions - the most commonly used are *Logit*, *Probit* and the *Complementary Log-Log* Link Functions (SPSS, 2006). An important assumption distinguishing the Proportional Odds Model is that the predicted values of categories of the ordered variable are a set of parallel lines. This enables the testing of an alternative hypothesis of non-parallel lines. The test of non-parallel lines is suitable for the Ordinal Regression and so was used in the model applied in this study.

3.5.2 Model formulation

The Proportional Odds Model of a relationship between m independent variables each with h levels, and one response variable with k ordered categories, is derived by Collet (2003). In this text, the k ordered categories of the response variable Y, are denoted by $C_1, C_2, \ldots, C_k$, where $k \geq 2$ and where a response in category C_j can be described as "worse than" one in $C_{j'}$; if $j < j'$. Thus for the food consumption score with the responses (or outcomes) labelled "poor", "borderline poor" or "good", the categories would be C_1, C_2 and C_3 so that $C_1 < C_2 < C_3$.

If it is supposed that Y_i is a categorical response variable for the i^{th} household with k levels, then Y_i takes the value j if the response is in category C_j, $j=1,2\ldots,k$. If x_i denotes a value of an explanatory (or independent) variable X, the probability that the i^{th} household responds in category C_j, is denoted by p_{ij}, such that:

$p_{ij} = P(Y_i = j)$ = P[household i responds in category C_j].

It follows that the cumulative probability of a response in category C_j or worse, denoted as γ_{ij} is $\gamma_{ij} = p_{i1} + p_{i2} + \cdots + p_{ij}$. As a result, $\sum_{j=1}^{k} p_{ij} = \gamma_{ik} = 1$ (Collet, 2003).

More understanding of the theory on formulation of the Proportional Odds Model (other texts use 'Cumulative Odds Ratios') can be found in Agresti (2002), McCullagh (1980), Peterson and Harrel (1990) and SPSS (2006).

3.5.3 Test of hypothesis featuring the deviance

Selection of important variables to include in a *k*-category ordered model depends on the values of the deviance of a model. According to Ashby, Pocock and Shaper (1986, p 292): 'For two models with the same subjects, one with *p* independent variables, and the other with an extra *q* independent variables, twice the difference in the maximised log-likelihood is distributed asymptotically as χ^2 on q degrees of freedom, under the null hypothesis that the extra q variables do not discriminate between variables'. The raw deviance calculated from grouped binary data, is defined as:

$$D = -2 \log L = -2 \sum_{ij} w_{ij} \log \hat{p}_{ij}$$

where $\hat{p}_{ij}$= fitted probability for j^{th} cell in group i and $\hat{w}_{ij}$ = weight (count) for j^{th} cell in group *i*. The *deviance* is used in calculating the likelihood ratio statistic for testing the null hypothesis of no difference between groups.

To test $H_0 \colon \theta = 0$, where θ is the common value of the odds ratio, the likelihood ratio test statistic is considered, as given by the difference of deviances:

$$\chi^2 = D(0) - D(\theta) = -2\ell(0) - [-2\ell(\hat{\theta})].$$

3.5.4 Model checking

After selecting a model, it is important to check whether the Proportional Odds Model is appropriate for the data. The techniques that follow involve examination or diagnoses of fitness of the proportional odds (parallel regression lines for cumulative *Logits*) assumption by looking at the case of grouped data. Fitted probabilities and testing the proportional odds assumption is looked at in the following sub-sections. The text on model checking will be primarily based on Collet (2003) who methodically discussed the assessment of Logistic Regression assumptions.

3.5.5 Significance tests

The SPSS output included a result of a special test featuring the Score Test. A significant Chi-square value indicates lack of fit of the proportional odds assumption, while a non-significant test shows goodness of fit and the hypothesis that the regression lines for cumulative *Logits* were parallel, cannot be rejected. A significant Chi-square test indicated that the proportional odds assumption was not justified.

As it was established that the proportional odds model used cumulative *Logits* of ordered categorical data, there was need to inspect how closely the fitted cumulative proportions of each of the three categories were to the observed proportions (Agresti, 2004; Ashby *et al*., 1986; Collet, 2003). The fitted and observed proportions indicated that the fitted model was good for estimating relationships. Like Logistic Regression, Ordered *Logit* uses Maximum Likelihood methods, and finds the best set of regression coefficients to predict the values of the *Logit*-transformed probability that the dependent variable falls into one category rather than another. Logistic Regression assumes that if the fitted probability, p, is greater than 0.5, the dependent variable should have value 1 rather than 0. Ordered *Logit* does not have such a fixed assumption. Instead, it fits a set of cut-off points. Because there were three levels of the dependent variable (1 to 3), the Logistic Regression found that $3\text{-}1 = 2$ cut-off values k_1 to k_2, such that if the fitted value of $Logit(p)$ was below k_1, the dependent variable was predicted to take the value 0. If the fitted value of $Logit(p)$ was between k_1 and k_2, the dependent variable was predicted to take the value 1, and so on. As with Logistic Regression, an overall Chi-square for the goodness-of-fit of the entire fitted model was determined. The Chi-square test was used to assess the improvement due to adding an extra independent variable or group of independent variables.

4. DATA ANALYSIS AND RESULTS

4.1 Examination of the distribution of the response variable and choice of Link Function

The distribution of the response variable was examined before choosing the Link Function. A bar chart distribution of food consumption group values (Figure 1) showed that the ‘borderline poor’ consumption (23.8 *per cent*) group was the smallest, followed by ‘poor’ consumption (31.5 *per cent*) and ‘good’ consumption (44.6 *per cent*). Values below the mean food consumption score were regarded as ‘poor’ or ‘borderline poor’.

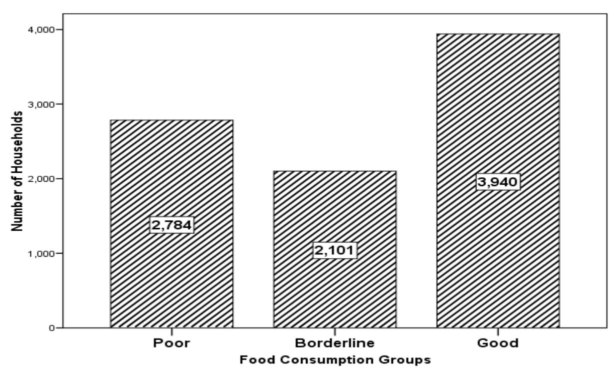

Figure 1: Distribution of households by categories of Food Consumption Scores in Southern Sudan, 2006 (N=9220).

It was also observed that the ungrouped distribution of the data has a very wide range, with extreme values of 0.5 and 105 for the Food Consumption Scores. These extreme values prompted a careful choice of the suitable Link Function. A Link Function transforms the cumulative probabilities for estimation of the model. SPSS version 15 avails five Link Functions for the ordinal regression model. The five Link Functions, their notational formulations and application were shown (Table 7). Of all the listed Link Functions, the *Probit* Link Function was found to be the most appropriate based on the 95 *per cent* confidence interval.

Table 7 **Summary of five Link Functions used in Ordinal Regression (SPSS version 15, 2006)**

Function	**Form**	**Typical application**
Logit	$\log(\xi/1-\xi)$	Evenly distributed categories
Complementary *log-log*	$\log(-\log(1-\xi))$	Higher categories more probable
Negative *log-log*	$-\log(-\log(\xi))$	Lower categories more probable
Probit	$\phi^{-1}(\xi)$	Latent variable is normally distributed
Cauchit (inverse *Cauchy*)	$\tan(\xi-0.5))$	Latent variable has many extreme values

4.2 Testing the goodness-of-fit of the model

The aim of fitting an Ordinal Logistic Regression model was to predict the ordinal outcome of the Food Consumption Scores that had three categories: 'poor', 'borderline' and 'good' consumption. The model fitted all of the 19 possible predictors, including sixteen factors and three covariates. Two predictor variables were found not to be statistically significant and so were dropped from the initial model.

The location-only ordinal regression model was fitted to the data. In specifying options for the model, the Link Function chosen was the *Probit* and 95 *per cent* confidence interval. The PLUM procedure of SPSS version 15 allows building of a model, generating predictions and evaluating the importance of various predictors where the dependent variable is categorically ordered

A model fitted with the *Probit* Link Function (Table 8) showed an improvement on the parameter estimates and their corresponding significance values over one fitted with the *Complementary Log-Log* or the *Cauchit* Link Functions. The model showed impressive goodness-of-fit statistics. Both Pearson's Chi-square and Deviance Chi-square values were not significant, which indicated that the data and the model predictions were similar. The model passed the goodness-of-fit test, despite the presence of many empty cells as a result of including the three covariates: household size, number of months during which harvest food lasted, and number of meals eaten daily. The test of parallel lines evaluates the proportional odds assumption. The test checks the assumption that model parameters (estimates of slope coefficients or $\hat{\beta}_j$ values) are the same for all categories of the response variable. This test enables comparison of the estimated model with one set of the model coefficients for all categories to a model with a separate set of coefficients for each category (SPSS, 2006; Hosmer & Lemeshow, 2000). It was assumed that with the same slope throughout the categories of food consumption groups and different intercepts, log-linear equations of the relationship of an explanatory variable with food consumption groups, if plotted on a plane with the *Logits* of *FCS* as the dependent variable and levels of a factor as the independent variable, will result in parallel lines.

The highly significant test of parallel lines indicated that the null hypothesis was to be rejected. That is, the coefficients differed across the categories of Food Consumption Scores so much so that no two lines of the same slope, for different categories, were parallel. This fit problem could be due to improper ordering of the categories of the response variable. The middle category of the Food Consumption Score, borderline consumption, would need to be widened to include values from the category 'good consumption', which could have been over-lumped; meaning the category is too wide. Therefore, there was adequate evidence that the model needed to be refitted with revised food consumption score categories.

Table 8 Proportional Odds Model test statistics and 95 per cent significance values (S. Sudan, 2006; N=9220)

Procedure [a]	Statistic	Estimate	Significance
Model fitting information	Chi-square (49 degrees of freedom)	615.522	0.000
Goodness-of-fit	Pearson's Chi-square	5607.304	0.394 [NS]
	Deviance Chi-square (49 degrees of freedom)	5409.561	0.952 [NS]
Test of parallel lines	Chi-square	118.733	0.0000

[a] Model with *Probit* Link Function.

4.3 Fitted model and determination of the magnitude and significance of influence of predictor variables

The influence of each predictor variable on the response variable was determined by examining the coefficients of each factor or covariate. The interpretation of values of coefficients differs between factors and covariates. For covariates, positive coefficients indicate positive relationships between predictors and outcomes. Negative coefficients indicate negative relationships. A covariate with an increasing positive value of a coefficient corresponds to an increasing probability of being in one of the upper level categories of the cumulative response. For factors, a factor level with a higher coefficient indicates a greater probability of being in one of the upper level categories of the cumulative response. A factor with a negative sign (indicator) indicates that its level had a negative effect on the corresponding category of the response variable. The converse is true for a factor with a positive sign. Therefore, interpretation of the model was based on the parameter estimates, as shown in Table 9.

SPSS version 15 tabulates parameter estimates of the *Logit* coefficients of the model, gives lower and upper limits of the 95 *per cent* confidence intervals, displays values of the Wald Test, degrees of freedom (df), and the significance probability. Whether an estimate is significant or not determines whether to reject or accept the null hypothesis. The null hypothesis states that each parameter estimate contributes zero to the relationship with the response variable. For an estimate to differ significantly from zero, its significance value must be less than the significance level of 0.05. Therefore, according to the model, twelve variables (including all three covariates and nine factors) had levels with significant values compared to reference levels (in case of factors). For brevity, only variables with significant *p*-values are presented and discussed here.

Table 9 Proportional Odds Model estimates and 95 per cent significance values (S. Sudan, 2006; N=9220)

Variable	Ordered Category Level	Estimate	Significance
1. State	71=Jonglei	0.661	0.000
	72=Upper Nile	0.625	0.000
	73=Unity	0.984	0.000
	81=Warrap	0.328	0.005
	82=Northern Bahr el-Ghazal	0.309	0.007
	83=Western Bahr el-Ghazal	0.508	0.000
	84=Lakes	0.365	0.000
	91=Western Equatoria	0.325	0.001
	92=Central Equatoria	1.087	0.000
2. Type of household	1=resident	0.108	0.218 [NS]

Variable	Ordered Category Level	Estimate	Significance
	2=internally displaced	0.053	0.598 NS
3. Level of education of household head	1=none (did not attend school)	0.002	0.980 NS
	2=primary	0.001	0.987 NS
	3=secondary	-0.145	0.133 NS
4. Gender of household head	1=male	-0.002	0.976 NS
5. Land ownership	1=yes	-0.172	0.012
6. Land use	1=yes	-0.301	0.001
7. Land planting	1=yes	0.147	0.006
8. Livestock ownership	1=yes	0.240	0.000
9. Migration in past 12 months	1=yes	0.068	0.206 NS
10. Number of harvests per year	1=once	-0.052	0.419 NS
11. Source of staple cereals	1=own production	0.458	0.284 NS
	2=market purchase	0.526	0.220 NS
	3=hunting/fishing/ gathering	0.660	0.163 NS
	4=exchanged for food	-0.349	0.496 NS
	5=borrowed	0.056	0.935 NS
	6=gift from relatives/friends	0.867	0.091
	7=food aid	0.402	0.357 NS

Variable	Ordered Category Level	Estimate	Significance
12. Owned vegetable garden	1=yes	0.282	0.000
13. Source of livelihoods	1=livestock rearing	0.648	0.000
	2=agricultural production	0.435	0.010
	3=fishing	0.519	0.012
	4=hunting/gathering	0.026	0.903 NS
	5=petty trade	0.296	0.184 NS
	6=collection of natural resources	0.003	0.987 NS
	7=unskilled labour	-0.005	0.982 NS
	8=handicrafts	0.004	0.988 NS
	9=skilled labour	0.449	0.147 NS
	10 = employed work	0.500	0.060
	11=food aid assistance	0.060	0.783 NS
14. Experienced food shock	1=yes	0.052	0.276 NS
15. Received food aid	1=yes	0.005	0.929 NS
16. Wealth Index Quintiles	1=poorest	0.121	0.139 NS
	2=poorer	0.176	0.020
	3=moderate	0.077	0.282 NS
	4=richer	0.038	0.579 NS

Variable	Ordered Category Level	Estimate	Significance
17. Household size		0.016	0.057
18. Months lasting of harvest food		0.050	0.000
19. Number of meals eaten daily		0.232	0.000

[NS] Not significant

4.3.1 Non-significant relationships

The model identified seven variables as having non-significant relationships (all levels) with any of the ordered categories of the response variable Food Consumption Scores, based on the five *per cent* level of significance. These variables were: gender of the household head, level of education of the household head, type of household, household migration, number of harvests per year, experience of food shock, and receiving food aid.

The finding that gender of the household head was not determined to be significantly related to Food Consumption Scores categories clearly indicates no inequality between female- and male-headed households in terms of food consumption. This was contrary to earlier expectations that female-headed households may be disadvantaged in terms of access to food. In the dataset, 12.9 *per cent* of the households were headed by women. Similarly, the level of education of the household head was not significantly related to the Food Consumption Score. This too is contrary to expectations. Nevertheless, the post-conflict setting of Southern Sudan could explain this peculiarity. Employment may generate insufficient income for households and this income may not have been spent on food.

Whether the household was resident in the location 12 months prior to the interview or was a returnee was also not significantly related to food consumption. This could be because of the prevalence of the kinship support system in Southern Sudan, where related people and those in the same neighbourhood liberally shared food to survive. Other factors such as rainfall, farmland, animal wealth and disposable cash could also have played a role. One would expect that households resident in one location for over a year would have attained better Food Consumption Scores than those that had recently migrated into the area. However, this was not the case, as the model showed no significant difference between the two categories of households. Whether a household harvested once or twice in a year did not significantly affect its food consumption. So too, experiencing a food shock prior to the survey (41 *per cent* of the sample) did not significantly affect food consumption. This might reflect the endemic nature of food insecurity in Southern Sudan. Receiving food aid affects the Food Consumption Score, possibly due to limited dietary diversity and the practice of eating fewer meals regularly. This could reinforce the fact that food received from aid could provide relief from hunger but would not solve food insecurity.

4.3.2 Significant relationships

The results of fitting the Ordinal Regression Model with *Probit* specification showed that 12 of the 19 fitted explanatory variables were significant at the the 5 *per cent* significance level. All three fitted covariates: household size, number of months in which harvest food lasted, and number of meals eaten by a household per day were significant at the 5 *per cent* significance interval. Of the 16 factors explored in the model, nine were significant at the 5 *per cent* significance level. For factors, comparison between levels of predictors could best be made in terms of 'odds ratios'.

Household size was significantly and positively related to food consumption. This was expected as more or fewer household members could affect food consumption. The number of months that harvest food lasted was positively related to food consumption. This was also expected, as the longer the food from harvest lasts, the greater the probability of protecting household food consumption. In addition, as expected, the number of meals eaten per day was positively related to the levels of food consumption. The probability of a household having 'good' food consumption increased as the number of meals eaten per day increased.

State of residence differed significantly from the comparator Eastern Equatoria State. The odds of Central Equatoria State were about three times better than those of Eastern Equatoria State in terms of Food Consumption Scores. Households differed significantly in relation to land ownership. Households that owned land were 0.8 times less likely than households that did not own land for being in a 'good' food consumption group – contrary to expectations. A similar result was observed for land use. There was a borderline significant difference (*p*-value=0.091) between a household that received cereals (sorghum and millet) from relatives compared to a household that did not receive food from relatives with regard to the probability of being in the 'good' food consumption group. The odds of a household receiving food from relatives attaining a 'better' food consumption score level, were twice ($\theta = exp$ [0.867 = 2.4]) as high as those from a household that did not receive food from relatives. This indicated that households that fared on their own were at risk of becoming food insecure, as they were not so well linked through the kinship support system.

Planting of farmland was significantly related to the food consumption groups (*p*-value = 0.006). The odds of a household that planted land being in the 'good' food consumption group were 1.1 times ($\theta = exp[0.15]$) higher than for a household that did not plant land. This indicates that planting land improved the likelihood of a household's food consumption classification. Having a home garden was highly significant (*p*-value = 0.000) and positively related to the food consumption groups. The odds of a household with a home garden or vegetable plot were 1.3 times ((θ = =*exp* [0.282]) higher for 'good' Food Consumption Scores. As expected, having a home garden increased the probability of being in the 'better' food consumption group. Dependence on home gardens meant higher dietary diversity or higher frequency of consumption of nutritious foods.

Ownership of livestock positively improved food consumption score levels. The odds of a household owning livestock being in the 'good' food consumption group were about 1.3 times higher compared to households that did not own livestock. This was according to expectation, as livestock ownership would provide the necessary dietary values, cash, or means of exchange for food improvement of the household's food security and livelihoods.

The main source of livelihood was an important determinant of food consumption and, by implication, food security. The software treated the category of the main source of livelihood "other", as a reference category. Analysis showed that four sources of livelihoods significantly affected food consumption: livestock rearing, agricultural production, fishing, and employment, and were related to the probability of being classified into 'good' food consumption groups. The odds of a household depending on any of these four sources of livelihoods being classified as 'good' consumers were, on average, about twice higher than those of households that stated they depended on "other" (unspecified) livelihood sources. This reinforces the need to include livestock keeping, land ownership, planting of farmland, and home gardening into a model for prediction and surveillance of food insecurity in Southern Sudan.

As regards Wealth Index Quintiles, there was significant difference between the second level and 'good' food consumption groups (*p*-value=0.020). This means that people in the 'borderline poor' Wealth Index Quintile attained 'better' food consumption. This result should be taken with a pinch of salt, as about half (48.9 *per cent)* of the households were categorised as being in the 'poorer' Wealth Index Quintile.

4.4 Examination of the strength of the model prediction

The next step was to examine how well the model predicted levels of food insecurity. This was done by cross matching the actual values of the response variable against their predicted values and determining the percentage of the correctly predicted values (Table 10).

The model correctly predicted the 'good' consumption group (81.1 *per cent*) but was not so accurate in correctly classifying the lower consumption categories, especially the 'borderline' food category. Most (60.0 *per cent*) 'borderline' category cases were classified as 'good' consumption. The model correctly classified 58.7 *per cent* of the 'poor' consumption cases. Overall, the model classification table showed a marked improvement of the *Probit* model over the two Ordinal Logistic Regression models: Complementary *Log-Log* and *Cauchit*. However, the models examined demonstrated the need to re-scale the ordinal categories of Food Consumption Scores.

Table 10 Classification table of predicted by observed categories for the *Probit* model (S. Sudan, 2006; N=9220)

Food consumption category	Predicted Response Category		Total
	'Poor' Consumption	'Good' Consumption	
'Poor' consumption	515 (58.7%)	363 (41.3%)	878 (100.0%)
'Borderline poor' consumption	275 (40.0%)	412 (60.0%)	687 (100.0%)
'Good' consumption	238 (18.9%)	1021 (81.1%)	1259 (100.0%)
Total	1028 (36.4%)	1796 (63.4%)	2824 (100.0%)

The ordering criteria were adopted from a guide by the WFP VAM Unit on calculation and use of Food Consumption Scores in food security analysis (WFP, 2008). There was reason to believe that the 'borderline' consumption category was too narrow for the wide range of Food Consumption Scores in the dataset. Some WFP publications use the categories zero to 21 for 'poor' food consumption; 21.5 to 35 for 'borderline' food consumption and 35.5 or more for 'good' food consumption (WFP, 2007). However, the categorisation scheme shown in Table 5 was used.

5. GENERAL DISCUSSION

Analysis of the results from the Ordinal Logistic Regression showed that at least 12 explanatory variables could accurately predict food insecurity. For the predictors with more than two levels such as state, main source of livelihood, main source of cereal (maize and sorghum) and Wealth Index Quintiles, only one or two levels were determined to be significantly associated with the

Food Consumption Scores. The seven explanatory variables shown to have non-significant relationships with food consumption could reflect a situation peculiar to Southern Sudan or to the period during which the survey was carried out (March, 2006). It is possible that the situation might change with the passage of time. All 19 variables could still be explored with a different dataset to confirm or negate the finding that the seven explanatory variables do or do not have non-significant relationships with food consumption levels. In fact when a new model was fitted with only the significant predictors, some of the predictor variables improved and some resulted in non-significant probabilities of association with food consumption p-values (Table 11).

Table 11 Partial output of a model fitted with all 19 variables and one fitted with significant predictors only

Variable		Level	*p*-value	
			Model with all 19 variables	New model with selected variables only
1	State	71 = Jonglei	0.000	0.003
		72 = Upper Nile	0.000	0.032
		73 = Unity	0.000	0.000
		81 = Warrap	0.005	0.440
		82 = NB el-Ghazal	0.007	0.903
		83 = WB el-Ghazal	0.000	0.005
		84 = Lakes	0.000	0.162
		91= W Equatoria	0.001	0.850
		92 = C Equatoria	0.000	0.000
2	Ownership of land	1 = yes	0.012	0.293
3	Use of land	1 = yes	0.001	0.000
4	Planting of land	1 = yes	0.006	0.000
5	Livestock	1 = yes	0.000	0.000

	ownership			
6	Ownership of vegetable gardens	1 = yes	0.000	0.000
7	Main source of livelihood	1 = livestock rearing	0.000	0.000
		2 = agricultural production	0.010	0.000
		3 = fishing	0.012	0.000
		4 = hunting/gathering	0.903	0.143
		5 = petty trade	0.184	0.001
		6 = collection of natural resources	0.987	0.195
		7 = unskilled labour	0.982	0.037
		8 = handicrafts	0.988	0.008
		9 = skilled labour	0.147	0.002
		10 = employed work	0.060	0.000
		11 = food aid assistance	0.783	0.111
8	Main source of cereals	1 = own production	0.284	0.493
		2 = market purchase	0.220	0.292
		3 = hunting/fishing/ forest gathering	0.163	0.205
		4 = exchanged for food	0.496	0.224
		5 = borrowed	0.935	0.578
		6 = gift from relatives or friends	0.091	0.458
		7 = food aid	0.357	0.513
9	Wealth Index	1 = poorest	0.139	0.000

	Quintiles	2 = poorer	0.020	0.000
		3 = moderate	0.282	0.004
		4 = richer	0.579	0.157
10	Size of household head		0.057	0.130
11	Months harvest food lasted		0.000	0.000
12	Number of daily meals		0.000	0.000

Table 11 shows a number of drastic changes in the association of some of the explanatory variables with food consumptions scores. Key among these is the finding that the main source of food cereals, ownership of land, and size of the household, were no longer influential predictors, when the seven variables earlier determined to be non-significant, were dropped. Conversely, it is interesting to note that when the *Probit* Ordinal Logistic model was fitted alone, it resulted in more levels (8) of the predictor 'main source of livelihood' being significantly associated with the Food Consumption Groups. In the earlier model, only four levels were determined to have significant probabilities. On the contrary, four states - Northern Bahr el-Ghazal, Western Equatoria, Warrap and Lakes - showed non-significant relationships to the Food Consumption Scores.

This indicated that some or all of the dropped variables could have modified the influence of some predictors to significantly or non-significantly associate with the response variable. Further analysis showed that most the factors with significant relationships with food consumption levels were expected. Ordinal Regression or the Proportional Odds Model was deemed to be more suitable for the food consumption data to determine the levels of nominal scale explanatory variables associated with the ordered categories of the response variable. Linear Regression models do not work very well in predicting ordinal responses (SPSS, 2006). In general, the Logistic Regression gives a more versatile analysis for explanatory and response variables on the categorical scale.

6. CONCLUSION AND RECOMMENDATION

The study found that twelve variables could statistically determine the outcome of food consumption and provide a basis for empirically predicting food insecurity in the post-conflict setting of Southern Sudan. The study showed that predicting food insecurity is possible and that it is easy to use the Ordinal Logistic Regression technique's *Probit* Model. The Ordinal Logistic Regression technique, otherwise known as the Proportional Odds Model (POM), was shown as valid for measuring the relationship between a set of independent predictors and Food Consumption Scores, thereby providing a practical and simple means for classifying households into vulnerable groups, monitoring changes in household food insecurity, and predicting (for early warning) levels of food insecurity. The study presents the first such analysis of food insecurity in Southern Sudan from national data and establishes a sound baseline for future surveillance and monitoring.

Further research is recommended to follow the food security situation in Southern Sudan as well as application of the technique to other countries embarking on the CAADP process.

ACKNOWLEDGEMENTS

This study would not have been possible without the support of the FAO Southern Sudan Sub-Office and FAO Rome who provided the dataset and a scholarship for this study. The authors thank the reviewers of this article for their valuable comments.

REFERENCES

AGRESTI, A (2002) *Categorical Data Analysis*, *2nd ed.*, New York, John Wiley and Sons.

AGRESTI, A (2004) *Analysis of Ordinal Categorical Data*, New York, Wiley.

ASHBY, D, POCOCK, SJ & SHAPER, AG (1986) Ordered polytomous regression: an example relating serum biochemistry and haematology to alcohol consumption. *Appl. Statistics*, 35, 289-301.

BENTLY, M & PELTO, G (1991) The household production of nutrition. *Social Science & Medicine*, 33, 1101-1102.

BICKEL, G, NORD, M, PRICE, C, HAMILTON, W & COOK, J (2000) *Guide to measuring household food security*, Alexandria, VA, Office of Analysis, Nutrition, and Evaluation, Food and Nutrition Service, United States Department of Agriculture.

BOUDREAU, T (1998) The food economy approach: a framework for understanding rural livelihoods. *Relief and Rehabilitation Network Paper No. 26*. London, Overseas Development Institute.

COATES, J, SWINDALE, A & BILINSKY, P (2007) *Household Food Insecurity Access Scale (HFIAS) for Measurement of Household Food Access: Indicator guide (V.3)*. Washington, DC, Food and Nutrition Technical Assistance (FANTA) Project. Available from: http://www.fantaproject.org/publications/hfss.shtml [cited 23 April 2003].

COATES, J, WEBB, P & HOUSER, R (2003) *Measuring Food Insecurity: Going beyond indicators of income and anthropometry*. Washington, DC, Food and Nutrition Technical Assistance (FANTA) Project, Academy for Educational Development. Available from: http://www.fantaproject.org/publications/hfss.shtml [cited 23 April 2009].

COLLET, D (2003) *Modelling Binary Data*, London, Chapman & Hall/CRC.

FOOD AND AGRICULTURAL ORGANISATION OF THE UNITED NATIONS/FOOD INSECURITY AND VULNERABILITY INFORMATION AND MAPPING SYSTEMS (2002) *Selecting indicators for national FIVIMS*. Rome, Food and Agricultural Organisation.

FOOD AND AGRICULTURAL ORGANISATION OF THE UNITED NATIONS/SUDAN INFORMATION FOR FOOD SECURITY IN ACTION (2008) Food Security Update (Jan-March 2008). Juba, Sudan, Sudan Information for Food Security in Action (SIFSIA), FAO.

FOOD AND NUTRITION TECHNICAL ASSISTANCE PROJECT AND FOOD AID MANAGEMENT (FAM) (2003). *Food Access Indicator Review.* Washington, D.C. Food and Nutrition Technical Assistance (FANTA). Available from: http://www.fantaproject.org/downloads/pdfs/accessIndicator_Dec03.pdf [cited 23 April 2003].

FRONGILLO, E & NANAMA, S (2006) Development and validation of an experience-based tool to directly measure household food insecurity within and across seasons in northern Burkina Faso. *The Journal of Nutrition*, 136, 1409-1419.

GOVERNMENT OF SOUTHERN SUDAN (GOSS) (2008) *Sudan Household Health Survey Report*. Juba, Ministry of Health, Government of Southern Sudan (GOSS).

HENDRIKS, SL (2005) The challenges facing empirical estimation of household food (in)security in South Africa, Development Southern Africa. *Development Southern Africa*, 22, 103–19.

HODDINOTT, J (1999) Choosing outcome indicators of household food security. *Technical Review No. 7. ed*. Washington D.C., International Food and Policy Research Institute (IFPRI).

HODDINOTT, J (1999a) Choosing outcome indicators of household food security. *Training materials: food security*. Washington D.C., International Food Policy Research Institute (IFPRI).

HODDINOTT, J (1999b) Operationalizing household food security in development projects: an introduction. *Training materials: food security*. Washington D.C., International Food Policy Research Institute (IFPRI).

HOSMER, DW & LEMESHOW, S (2000) *Applied Logistic Regression*, New York, Wiley.

LOKOSANG, LB (2009) Statistical analysis of determinants of household food insecurity in post conflict Southern Sudan, University of KwaZulu-Natal (Masters thesis).

MAXWELL, D, WATKINS, B, WHEELER, R & COLLINS, G (2003) The Coping Strategy Index: a tool for rapidly measuring food security and the impact of food aid programmes in emergencies. *International Workshop on Food Security in Complex Emergencies: building policy frameworks to address longer-term programming challenges*. Rome, Food and Agricultural Organisation (FAO)

MCCULLAGH, P & NELDER, JA (1989) *Generalized Linear Models*, London, Chapman & Hall.

MCCULLAGH, P (1980) Regression Models for Ordinal Data. *Journal of the Royal Statistical Society B*, 42, 109-142.

NANDY, S, KELLY, M, GORDON, D, SUBRAMANIAN, S & SMITH, G (2003) *Food & Health in Welfare: Current Issues, Future Perspectives*. Bristol, UK, Townsend Centre for International Poverty Research, University of Bristol.

PETERSON, B & HARREL, F (1990) Partial proportional odds models for ordinal response variables. *Applied Statistics*, 39, 205-217.

SAVE THE CHILDREN FUND UNITED KINGDOM (SCF-UK) (2000) *Household Economy Approach: A resource manual for practitioners*, London, Save the Children Fund (SCF) United Kingdom.

SETBOONSARNG, S (2005) *Child Malnutrition as a Poverty Indicator: An Evaluation in the Context of Different Development Interventions in Indonesia*. Asia Development Bank Institute (ADBI) Discussion Paper N. 21.

SPSS (2006) *SPSS Version 15.0 for Windows*. Chicago, SPSS Inc.

STANDARDIZED MONITORING AND ASSESSMENT OF RELIEF IN TRANSITIONS (SMART) (2006) *Measuring mortality, nutritional status and food security in crisis situations*: *SMART Methodology*, *Version 1*, Washington, D.C. and New York, United Nations Children Fund (UNICEF) and United States Agency for International Development.

SWINDALE, A & BILINSKY, P (2006) *Household Dietary Diversity Score (HDDS) for measurement of Household Food Access: Indicator Guide*. Washington D.C., Food and Nutrition Technical Assistance (FANTA) Project.

SWINDALE, A & OHRI-VACHASPATI, P (2005) *Measuring household food access: A Technical Guide*. Washington DC, Food and Nutrition Technical Assistance (FANTA) Project.

THE AFRICAN UNION (2009) Comprehensive Africa Agriculture Development Programme (CAADP): *Pillar III Framework for African Food Security (FAFS)*, Midrand, AU/NEPAD Secretariat.

UNITED NATIONS CHILDREN FUND (UNICEF) (2006) *Annual Report*. Geneva, United Nations Children Fund.

UNITED NATIONS CHILDREN'S FUND (UNICEF) (2008) *Multiple Indicator Cluster Surveys (MICS)*. New York, NY, Statistics and Monitoring Division of Policy and Practices, UNICEF. Available from: www.childinfo.org/mics.html [cited on 29 August 2009).

UNITED NATIONS CHILDREN'S FUND (UNICEF) (2008) *Multiple Indicator Cluster Surveys (MICS)*. New York, NY, Statistics and Monitoring Division of Policy and Practices, UNICEF. Available from: www.childinfo.org/mics.html [cited on 29 August 2009).

UNITED NATIONS DEVELOPMENT PROGRAMME (UNDP) (2009) *Human Development Report 2009: Overcoming barriers*. New York, NY, Palgrave Macmillan.

UNITED STATES CENSUS BUREAU, INTERNATIONAL PROGRAMS CENTER, POPULATION DIVISION (2007) *CSPro User's Guide: Version 3.0* Washington, D.C., United States Census Bureau.

WEBB, P & ROGERS, B (2003) *Addressing the "In" in Food Insecurity. Occasional Paper No. 1*. Washington DC, USAID Office of Food for Peace.

WEBB, PJ, COATES, E, FRONGILLO, B, ROGERS, A, SWINDALE, A & BILINSKY, P (2006) *Measuring Household Food Insecurity: Why It's So Important and Yet So Difficult to Do*. Washington DC, Food and Nutrition Technical Assistance (FANTA) Project.

WOLFE, SW & FRONGILLO, EA (2000) *Building household food security measurement the ground up: background paper*. Washington DC, Food and Nutrition Technical Assistance Project.

WOOLFE, B (1955) On estimating the relation between blood group and disease. *Human Genetics*, 19, 251-3.

WORLD FOOD PROGRAMME OF THE UNITED NATIONS (2007) *CFSVA Methodology Workshop Report*. Rome, WFP.

WORLD FOOD PROGRAMME OF THE UNITED NATIONS/VULNERABILITY ASSESSMENT AND MAPPING (2007a) *Sudan: Comprehensive food security and vulnerability analysis (CFSVA)*. Rome, World Food Programme.

WORLD FOOD PROGRAMME OF THE UNITED NATIONS/VULNERABILITY ASSESSMENT AND MAPPING (2007b) *Sudan: Southern Sudan Comprehensive Food Security and Vulnerability Analysis (CFSVA)*. Rome, Italy, World Food Programme.

WORLD FOOD PROGRAMME OF THE UNITED NATIONS/VULNERABILITY ASSESSMENT AND MAPPING (2008) *Food consumption analysis: Calculation and use of the food consumption score in food security analysis*. Rome, World Food Programme, Vulnerability Analysis and Mapping Branch (ODAV).

WORLD HEALTH ORGANISATION (1995) *Physical status: the use and interpretation of anthropometry*. Geneva, WHO Technical Report No. 854.