# Estimating the national and regional prevalence of drinking or eating more than usual during childhood diarrhea in Malawi using the bivariate sample selection copula regression

**Alfred Ngwira** Corresp., [1] , **Francisco Chamera** [2] , **Matrina Soko** [2]

[1] Basic Sciences, Lilongwe University of Agriculture and Natural Resources, Lilongwe, Malawi, Malawi

[2] Lilongwe University of Agriculture and Natural Resources, Lilongwe, Malawi, Malawi

Corresponding Author: Alfred Ngwira
Email address: alfngwira@yahoo.com

**Background.** Estimation of prevalence of feeding practices during diarrhea using the conventional methods may be biased as the sample is non-random due to selection procedure. The study aimed at re-estimating the prevalence of feeding practices using sample selection model to correct for non randomness of sample. **Methods.** The study used 2015/16 Malawi demographic health survey (MDHS) data which had 16246 children records who had diarrhea or not. A bivariate Joe copula regression model with $90^{0}$ rotation was fitted to either drinking or eating more with diarrhea as a sample selection outcome in the bivariate models. The prevalence of either drinking or eating more was estimated using both survey weights and the predictor of second outcome in the bivariate models. The prevalences were then compared with the prevalences estimated using the conventional method. The regional prevalences were compared by comparing the maps of the prevalences. **Results.** There was a substantial increase in the re-estimated national prevalence of drinking more (40.0 % (31.7, 50.5)) or eating more food (20.46 % (9.87, 38.55)) using the bivariate model, as compared to the one estimated by the conventional method, that is, 28.9 % (27.0, 30.7) and 13.1 % (12.0, 15.0) respectively. The maps of the regional prevalences showed similar results where the prevalences estimated by the sample selection model were relatively higher than those estimated by the conventional method. The presence of diarrhea was somehow weakly correlated with either drinking or eating more food. **Conclusion.** The estimation of prevalence of drinking or eating food during diarrhea should take into account sample selection procedure to minimize bias. In addition, strategies to improve feeding practices during diarrhea episode should target improving prevalence of eating more food to avert malnutrition.

# Estimating the national and regional prevalence of drinking or eating more than usual during childhood diarrhea in Malawi using the bivariate sample selection copula regression.

Alfred Ngwira[1], Francisco Chamera[1], Matrina Mpeketula Soko[1]

[1]Basic Sciences Department, Lilongwe University of Agriculture and Natural Resources, Lilongwe, Malawi

Corresponding Author:

Alfred Ngwira[1]

Area 21, Lilongwe, Malawi.

Email: alfngwira@yahoo.com

## Abstract

**Background.** Estimation of prevalence of feeding practices during diarrhea using the conventional methods may be biased as the sample is non-random due to selection procedure. The study aimed at re-estimating the prevalence of feeding practices using sample selection model to correct for non randomness of sample.

**Methods.** The study used 2015/16 Malawi demographic health survey (MDHS) data which had 16246 children records who had diarrhea or not. A bivariate Joe copula regression model with $90^0$ rotation was fitted to either drinking or eating more with diarrhea as a sample selection outcome in the bivariate models. The prevalence of either drinking or eating more was estimated using both survey weights and the predictor of second outcome in the bivariate models. The prevalences were then compared with the prevalences estimated using the conventional method. The regional prevalences were compared by comparing the maps of the prevalences.

**Results.** There was a substantial increase in the re-estimated national prevalence of drinking more (40.0 % (31.7, 50.5)) or eating more food (20.46 % (9.87, 38.55)) using the bivariate model, as compared to the one estimated by the conventional method, that is, 28.9 % (27.0, 30.7) and 13.1 % (12.0, 15.0) respectively. The maps of the regional prevalences showed similar results where the prevalences estimated by the sample selection model were relatively higher than those estimated by the conventional method. The presence of diarrhea was somehow weakly correlated with either drinking or eating more food.

**Conclusion.** The estimation of prevalence of drinking or eating food during diarrhea should take into account sample selection procedure to minimize bias. In addition, strategies to improve feeding practices during diarrhea episode should target improving prevalence of eating more food to avert malnutrition**.**


**Key words:** geographical, feeding practices, child health, place, malnutrition

## Introduction

Diarrhea in children is defined as the occurrence of three or more loose or liquid stools per day (WHO, 2013). Diarrhea is considered as the second killer of under five children after pneumonia causing an estimated 1.5 million under five deaths every year (WHO, 2009). The prevalence of diarrhea in Malawi as of 2016 was 22% (NSO, 2017). Diarrhea has a detrimental effect on child's nutrition status due to decrease in food intake caused by anorexia (Huffman et al, 1991). To reduce deaths, dehydration and minimise the effects of diarrhea on nutritional status, mothers are encouraged to improve on feeding practices especially to give children more food and fluids than usual. The prevalence of feeding practices during child diarrhea from the Malawi demographic health survey (MDHS) report as of 2015/16 was 31% for drinking more fluids and 13% for eating more food than usual (NSO, 2017).

The estimation of prevalence of the feeding practices during childhood diarrhea using DHS data is based on the sample of children with diarrhea living at the time of survey, that is, selection into the sample is based on whether the child has diarrhea or not. This means that those selected (with diarrhea) are likely to have poor house hold characteristics, for example, poor parental education and large house hold size (Asfaha et al, 2018; Samwel et al, 2014). Such poor house hold characteristics in turn are associated with poor feeding practices during diarrhea episode, for example, children of less educated mothers and of households with large number of children are less likely to drink or eat more food (Fikadu and Girma, 2018). It is hypothesized therefore that there is negative association between the presence of diarrhea and feeding practices (drinking or eating more) during childhood diarrhea. The fore going presentation means that if selection of children into the sample to estimate prevalence of drinking or eating more food during diarrhea episode is not taken into account, the prevalences are prone to be underestimated.

The main aim of this study was to re-estimate the prevalence of drinking or eating more than usual during diarrhea episode using sample selection copula model. The advantage of copula regression over the standard bivariate normal regression is that it offers flexibility in the marginal distribution of the outcome variables (Parsa and Klugman, 2011). Specifically, the study sought to find out if diarrhea and drinking or eating more than usual were independent and also it aimed at re-estimating the prevalence of eating or drinking more than usual using the sample selection model. The significance of the study was that it would reveal if the estimates of prevalence of eating or drinking more during diarrhea reported in the DHS report (NSO, 2017) were biased or not due to sample selection. This would ensure correct policy making using the correct prevalences.

The following is the article layout. First, methods regarding the data and statistical analysis used are presented. This is followed by the presentation of results and then discussion of the results. Finally conclusions on the key research objectives are made.

## Materials & Methods

### Data

The study used secondary data, the 2015/16 Malawi DHS child record data. Permission to use the data was granted after applying to use the data through DHS web site (www.dhsprogram.com/data set_admin). The 2015 MDHS study was ethically approved by Malawi Health Research Committee, Institutional Review Board of ICF Macro, Center for Disease and Control (CDC) in Atlanta, GA, USA. No any other permission was needed to publish the results of the study. The MDHS according to NSO (2017) was a two stage cluster sampling with stratification where clusters were stratified by residence (urban/rural) and then in each cluster, households were randomly selected. In the first stage, 850 clusters, comprising of 173 clusters of urban areas and 677 clusters of rural areas were selected by probability proportional to size (PPS) cluster sampling method. In the second stage, 30 households from each urban cluster and 33 households from each rural cluster were selected by systematic sampling. The data from households was then collected using the four questionnaires that is, the woman, man, household and then biomarker questionnaire. This study used the child record data mainly collected by the woman questionnaire. The response variables in the final data set were the presence of child diarrhea (yes/no) in the last two weeks and the feeding practice variables, eating food more than usual (yes/no) and drinking fluids more than usual (yes/no). The independent variables were child age (in months), household size, maternal education (no education, primary, secondary, higher), water source (safe/unsafe), toilet (yes/no), house floor (cement/sand), district code of the child and interviewer identity code. The total number of records in the final data set was 16246.

**Statistical analysis**

A frequency distribution table of feeding practice in terms food and fluids during childhood diarrhea was made. A multiple variable bivariate sample selection model as defined by Marra and Radice (2017) was then fitted, that is, let $y_{i1}$ be the sample selection outcome, that is, diarrhea (yes/no) and $y_{i2}$ the second outcome observed only if child is selected, that is, has diarrhea, in this study say drinking more fluids or eating more food than usual. Then the flexible additive predictor of the bivariate model of diarrhea (yes/no) and drinking more or diarrhea and eating more is defined as $\eta_{vi} = \beta_{v0} + \sum s_{vk_v}(z_{vk_v i}),\ i = 1,2,3,\ldots,n$ where $v = 1,2,c$ The third additive predictor $\eta_{ci}$ is for the copula parameter, $\theta_i$, the dependency parameter between the two outcomes involved. The bivariate distribution used in this study was the Joe copula with 90 degrees rotation which was opted for so as to model the negative dependency between dirrhoea and the feeding practice variables as per the hypothesis and after the standard bivariate normal copula revealed a negative dependency between the two outcome variables. The range of the copula parameter with this bivariate distribution was $(-\infty,-1)$ where $\theta = -1$ means independence and dependence otherwise. The $s_{vk_v}(z_{vk_v i})$ represents the generic effect of the independent variable and is specified according to type of covariate considered. The main objective in this study was to estimate the prevalence of the observed outcome, $y_{2i}$, given the selected sample, say the prevalence of drinking more fluids or eating more food during diarrhoea episode defined as $P(Y_2 = 1)$. This was computed by the formula:

$$P(Y_2 = 1) = \frac{\sum_{i=1}^{n} w_i[1 - F_2(\hat{\eta}_{2i})]}{\sum_{i=1}^{n} w_i}$$

where $w_i$ were the survey weights, the woman individual sample weights. Model estimation was by the penalized maximum likelihood estimation (PMLE) considering that the usual maximum likelihood estimation (MLE) could lead to over fitting due to the presence of smooth functions (Filippou et al, 2017). Model fitting, prevalence estimation and mapping of regional prevalence was implemented by the GJRM package in R (Marra and Radice, 2017).

## Results

Table 1 presents the percentage of children under five in terms of feeding practices during diarrhoea as reported from MDHS report (NSO, 2017). Thirty one percent of children with diarrhoea were given more than usual fluids and thirty four percent were given less fluids which is a concern. Five percent of the sick children were not given any fluids. Sixty one percent of the sick children were given recommended liquid as compared to fourty five percent children that were given recommended food. These percentages were used as a bench mark as the new percentages using the sample selection model were estimated. The focus however was on the percentage of children feeding or drinking more than usual during diarrhoea.

After fitting the bivariate sample selection model with the Joe copula, the estimated national prevalence of drinking more fluids during diarrhea using the conventional method, that is, imputation method was 28.9 % with confidence interval (27.0, 30.7). Adjusting for sample selection, using the sample selection bivariate model, the estimated prevalence of drinking more fluids was 40.0 % with confidence interval (31.7, 50.5). Regarding eating more food, the estimated national prevalence of eating more during diarrhea using the conventional imputation method was 13.1 with confidence interval (12.0, 15.0), and the estimated prevalence using the bivariate sample selection model was 20.46 % with confidence interval (9.87, 38.55). The estimated copula parameter for the bivariate model of diarrhoea and drinking more fluids was -1.23 with confidence interval (-1.53,-1.09) and that of diarrhea and eating more food was -1.29 (-2.57, -1.04) both showing negative dependency. The association of drinking and eating more with diarrhea presence may be considered as weak since the copula parameter values are close to -1.

Figure 1 presents the map of prevalence of drinking more fluids during childhood diarrhea by region. The prevalence estimates by the bivariate sample selection model (Figure 1, b) are relatively higher than the imputation based estimates (Figure 1, a). The copula parameter shows relatively strong negative dependence between diarrhea and drinking more fluids in many districts (Figure 1, c). Figure 2 shows the distribution of prevalence of eating more food. The prevalence estimates by the bivariate sample selection model (Figure 2, b) are also relatively higher than the imputation based estimates (Figure 2 a). The copula parameter distribution shows many areas having weak dependence between diarrhea and eating more food (Figure 2, c).

## Discussion

The study has looked at the re-estimation of the prevalence of feeding practices during childhood diarrhea by focusing on drinking and eating more than usual. The standard estimation procedure is prone to be biased as the final selected sample is not random as individuals are selected depending on whether they have diarrhea or not. The study therefore tried to investigate the degree of biasness attached to the estimates if the standard method is used, by re-estimating the prevalences using the novel approach which takes into account sample selection process. Sample selection process was taken into account in estimation by using the bivariate sample selection model where one of the two models, modeled sample selection.

The study found that the prevalence estimates of both drinking more liquids and eating more food using the bivariate model were substantially greater than the those estimated by the conventional, imputation method. The increased prevalence was likely, since one international study (Bani et al, 2002) found much higher prevalence of mothers increasing the volume of fluids given during child diarrhea episode (75.5 %). The observed differences in the prevalence estimated by the bivariate sample selection model and the usual imputation method might be due to the non randomness of the sample selected (Marra and Radice, 2017). Non randomness of the sample would be as a result of selection process, as samples were being selected if they had diarrhoea, otherwise they were not selected. The estimates by bivariate sample selection model would be with minor bias as non randomness was corrected. The increase in prevalence estimate from standard estimation to sample selection estimation may be due to the fact that those not selected (without diarrhoea) were likely to have good education and small number of children since low education and large house hold size are positively correlated with diarrhea (Asfaha et al, 2018) and this would mean increasing drinking and eating more in case sample selection process is corrected as increased education and small house hold size increase the intake of fluids and food (Fikadu and Girma, 2018). An indication of weak association between drinking and eating more food and diarrhea is consistent with Huffman et al (1991) where having diarrhoea did not significantly affect eating habits though there was a decline in eating food as children appetite reduced.

The distribution of estimated prevalence of drinking or eating more by region (Figure 1 & 2, b) shows that there is less variation and most regions especially around the cities show increased intake of fluids and food. Generally, there is reduced intake of food compared to fluids.

The reduction in food intake may be due to decrease in appetite in the diarrhoea patients as explained by Huffman et al (1991) and Paintal and Aguayo (2016). Similar distribution of fluid and food intake across all the regions may due to the fact that all regions might have similar spatial determinants of feeding practices during diarrhoea, for example, house hold size is seen to be similar across the three main regions of Malawi with the following household sizes, 3.7 (north), 3.6 (central), and 3.7 (south) (NSO, 2017), and household size has been found to be associated with feeding practices (Fikadu and Girma, 2018).

## Conclusion

The study finds that the both estimated prevalence of drinking and eating more food by the standard method were likely to be biased considering that they deviated greatly from the estimate based on  the sample selection model found in this study. There is weak negative correlation between the presence of diarrhoea and the feeding practices during diarrhea. The implication of the results is that the prevalence of drinking  or eating  more food during childhood diarrhea should be estimated  by taking into account sample selection process (diarrhea presence) so as to correct for  the biasness that may arise due to non-randomness of the sample.

## Acknowledgements

We thank the demographic health survey (DHS) for providing the data that was used.

## References

1. WHO. Diarrheal diseases: World Health Organization, 2013. Available at: https://www.who.int/news-room/fact-sheets/detail/diarrhoeal-disease
2. UNICEF/WHO. Diarrhoea: Why children are still dying and what can be done, 2009. Available at: https://www.who.int/maternal_child_adolescent/documents/9789241598415/en/
3. National Statistical Office (NSO) [Malawi] and ICF. 2017.  Malawi Demographic and Health Survey 2015-16
4. Huffman SL,  Romana GL, Brown KH, Bentley M, Black RE. 1991. Do Child Feeding Practices Change in the Central Peruvian Highlands. *J Diarrhoeal Dis Res* 9(4): 295-300.
5. Asfaha KF, Tesfamichael FA, Fisseha GK, Misgina KH, Weldu MG, Welehaweria NB, Gebregiorgis YS. 2018. Determinants of childhood diarrhoea in Medebay Zana District,

Northwest Tigray,Ethiopia: a community based unmatched case–control study. *BMC Pediatrics* 18:120.

6. Samwel M, Eddison M, Faith N, Richard S, Elizabeth KM, Douglas N. 2014. Determinants of diarrhoea among young children under the age of five in Kenya, evidence from KDHS 2008-09. *African Population Studies* 28(2).
7. Fikadu T, Girma S. 2018. Feeding Practice during Diarrheal Episode among Children Aged between 6 to 23 Months in Mirab Abaya District, Gamo Gofa Zone, Southern Ethiopia. *International Journal of Pediatrics*.
8. Parsa, R.A.; Klugman, S.A. Copula Regression. 2011. *Variance* 5(1): 45–54.
9. Marra G, Radice R. 2017. A joint regression modeling framework for analyzing bivariate binary data in R. *De Gruyter Open* 5: 268-294
10. Filippou P, Marra G, Radice R. 2017. Penalized likelihood estimation of a trivariate additive probit model. *Biostatistics* 18(3): 569-585
11. Bani IA, Saeed AAW, Mohammed AA, Othman A. 2002. Diarrhoea and child feeding practices in Saudi Arabia. *Public Health Nutrition* 5(6), 727–731
12. Paintal K, Aguayo VM. 2016. Feeding practices for infants and young children during and after common illness. Evidence from South Asia. *Maternal & Child Nutrition* 12 (1): 39–71
13. Alebel A, Tesema C, Temesgen B, Gebrie A, Petrucka P, Kibret GD. 2018. Prevalence and determinants of diarrhea among under-five children in Ethiopia: A systematic review and meta-analysis. *PLoS ONE* 13(6): e0199684.

## Table 1 (on next page)

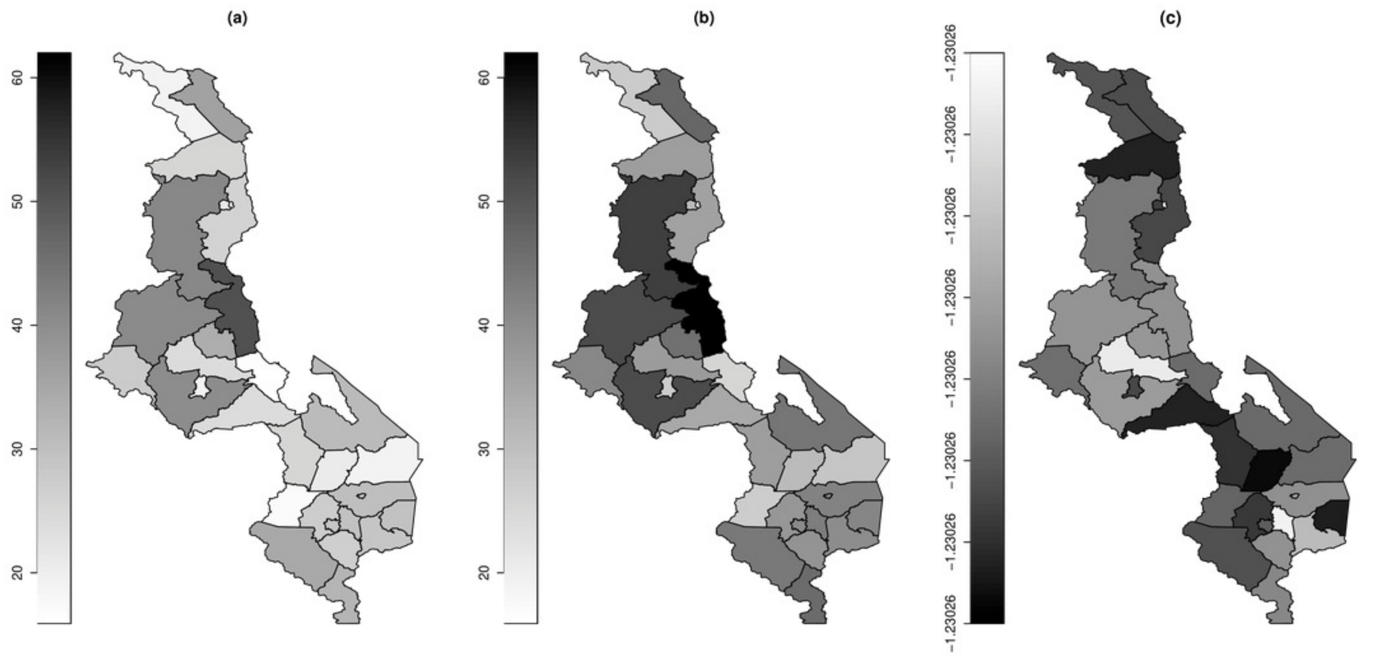
Percentage of feeding practices during diarrhoea

**Table 1:** Percentage of feeding practices during diarrhoea

| | Feeding practice | | | | |
|---|---|---|---|---|---|
| | more | usual/same | less | none | never gave |
| Liquids | 31 | 30 | 34 | 5 | 0 |
| Food | 13 | 32 | 43 | 6 | 5 |

# Figure 1

Prevalence map of drinking more fluids by district. Imputation model (a), sample selection model (b) and the copula parameter (c). Dark (high) and white (low).



# Figure 2

Prevalence map of eating more food. Imputation model (a), sample selection model (b) and the copula parameter (c). Dark (high) and white (low).