International Journal of Sciences: Basic and Applied Research (IJSBAR)

ISSN 2307-4531
(Print & Online)

http://gssrr.org/index.php?journal=JournalOfBasicAndApplied

Joint Regression and Association Modelling of Child Comorbidities in Uganda

Leonard K. Atuhaire*

School of Statistics and Planning, Makerere University, Kampala, Uganda
Email: leonard.atuhaire@mak.ac.ug

Abstract

Though morbidity in children living in low-income countries is commonly characterized by more than one health condition, studies to determine the determinants of child morbidity usually study one of the illnesses or a combination of the illnesses, but independently, thus ignoring possible dependencies. In many such cases a multivariate regression approach would be appropriate. Thus, in this paper, we specifically aimed at comorbidity among under-five children using joint response model that accommodates the interdependence between three illnesses (Diarrhoea, Acute Respiratory Infection (ARI) and fever) in assessing their risk factors. We considered child illness as a trivariate binary outcome (Y_1, Y_2, Y_3) and carried out a trivariate copula regression model to jointly model Diarrhoea, ARI and fever. Older ages(3&4 years) mother having primary or higher education and not being in the poorest quintile are associated with reduced prevalence of ARI; while cooking with wood/straw and other fuels, rather than charcoal, is associated with increased prevalence of ARI. Being aged 1,2 or 4 years, higher level of education for the mother, urban residence, and not being in the poorest quintile are associated with reduced prevalence of fever; while, surprisingly having an improved source of drinking water is associated with increased prevalence of fever. Being 2 years or older , being in the three upper quintiles of household wealth and using charcoal for cooking are associated with reduced prevalence of diarrhea. Male children have higher prevalence of diarrhea. The three illnesses are strongly correlated with each other even after accounting for covariates at marginal level.This study has illustrated an easily applicable technique to joint modelling of common childhood illnesses which allows exploration of the correlations between them. We strongly recommend the use of this method for correlated outcomes, whatever the type of variable.

Keywords: Comorbidity; copula regression; joint modelling.

----

* Corresponding author.

1. Introduction

Diarrhoea, cough and fever are the leading causes of childhood morbidity and mortality in sub-Saharan Africa [1], Uganda included. Though morbidity in children living in low-income countries is commonly characterized by more than one health condition [2], studies to determine the determinants of child morbidity usually study one of the illnesses or a combination of the illnesses, but independently. Examples of such studies in sub-Saharan Africa include Kandala and his colleagues [3] who considered diarrhoea, cough, and fever in Nigeria ; Senyonga and his colleagues [4] who considered diarrhea in Uganda; Niringiye and Douglason [5] who considered Malaria in Uganda; Roberts and Mathews [6] who considered malaria in Uganda; Gaston and Ramroop [7] who considered malaria in Malawi; Bbaale [8] who considered Acute Respiratory Infection and Diarhoea in Uganda; Gebertsadik , Worku, and Berhane [9] who considered ARI in Ethiopia; Kabinga, Mbewe, and Banda [10] who considered diarrhoea in Zambia; and Zgambo and his colleagues [11] who considered malaria in Malawi. Such methods make the implicit assumption that the three indicators of morbidity are independent variables, which is empirically quite implausible. The problem of comorbidity due to overlapping risk factors such as nutrition, sanitation, and overcrowding is highly prevalent among children under five years. This comorbidity may actually be the result of disease dependencies. While comorbidity may be obvious, dependencies can also be ascertained fairly easily. When the researcher is interested in modelling more than one response, univariate regression will not yield valid inferences if there is residual dependence between the outcomes conditional on the covariates [12]. Where comorbidity and disease dependencies exist, a multivariate regression approach would be appropriate. To be fair, a number studies have fitted joint models for comorbidities in sub-saharan Africa. For example, Fenn and his colleagues [2] used a joint model for diarrhea, dysentery, pneumonia, malaria, measles in Ghana; Kazembe and his colleagues [13] used a joint model for fever and diarrhea in malawi; Gawayan and his colleagues [14] used a joint model for malaria and non-malarial febrile illness in Nigeria; Adebayo and his colleagues [15] used a joint model for malaria and anemia in Nigeria; Habyarimana and his colleagues [16] used a joint model for Poverty and malnutrition in Rwanda; Kinyoki and his colleagues [17] used a joint model for ARI, diarrhea, and stunting in Somalia; Roberts and Zewotir [18] used a joint model Malaria and anaemia in Kenya, Malawi, Tanzania and Uganda. In this paper our aim was to illustrate the applicability of joint regression models for correlated binary data to the study of child comorbidities. Specifically, we made use of a joint model approach to explore the correlation between diarrhoea, fever and Acute Respiratory Infection (ARI) in young children in Uganda, while accounting for the effects of mother and child factors. To our knowledge, no studies have jointly modelled these three illnesses for Uganda. The objective was to examine the relationship between of mother and child characteristics on child health outcomes using household data. Specifically the study examined the effect of mother and child characteristics on case prevalence of diarrhea, acute respiratory infection, and fever (quite often, malaria) using a joint binary response model that accommodates the interdependence between the three diseases. Thus, we specifically aimed at comorbidity among under-five children using a joint response model that accommodates the interdependence between the three infections in assessing their risk factors

2. Materials and methods

2.1 Data

The data used for the present study were from the 2016 Uganda Demographic and Health Surveys which collected information on a nationally representative sample of women in child-bearing age (15-49) [19]. The survey collected a large number of indicators for the respondent, her partner, the household she resides in, and her children who were born within the five years preceding the survey. This study is based on 14,710 children with valid responses to questions on prevalence of (i) acute respiratory infection, (ii) fever, and, (iii) diarrhea. These children belonged to 9,886 women. The study has three outcome variables, namely (i) having had symptoms of acute respiratory infection (ARI) in the 2 weeks preceding the survey, (ii) having had a fever in the 2 weeks preceding the survey, and (iii) having had diarrhoea in the 2 weeks preceding the survey. Fever is a symptom of malaria but is also associated with other childhood illnesses. The independent variables considered are (i) child's age in years, (ii) child's sex, (iii) mother's smoking status, (iii) Cooking fuel, categorized as charcoal, wood/straw, and other; (iv) place of residence(rural/Urban);(v) mother's level of education ; (vi) household wealth quintile; (vii) Source of drinking water (improved/unimproved); (viii) toilet facility(improved/unimproved). The choice of the independent variables was entirely determined by the variables against which the prevalences of the three illnesses are tabulated by the UDHS but excluding the geographical variables (Region etc.).

2.2 Approaches to multivariate regression for correlated binary responses

Multivariate models consist of two parts: (i) a regression model for each marginal response,

$$g\{E(Yij)\} = \underline{x}_{ij}^{t}\,\underline{\beta} \qquad (1)$$

where g() is suitable link function, and (ii) an association structure to account for correlation between the multiple response variables. In this paper we illustrate the application of such joint regression and association modeling for the three illnesses using DHS data. We therefore consider child illness as a trivariate binary outcome (Y_1, Y_2, Y_3) and carry out a multivariate regression for correlated binary responses. Because each of the three variables is binary, logistic regression is usually employed for this purpose. we also considered models based on the logit link. A number of approaches to implementing multivariate or joit regression models for correlated binary data have been mentioned in the literature. We mention only a few.

GEE models

One approach is to use the generalized estimation equation (GEE) approach initially introduced for the analysis of longitudinal data [20]. This approach allows one to carry out simultaneous inference on marginal response rates and pair-wise correlations [21]. Under this approach, using the present example, ARI, fever, and diarrhoea would be taken as 'repeated' measures of illness. However, the GEE approach cannot lead to estimated joint probabilities [22] and the effect of covariates on dependence cannot be measured. The advantage with this approach is that the models can easily models can be fitted with any GEE routine, for example the **xtgee** command in *stata* version 15 [23].

Dependency ratio models

Another approach is to use the dependency ratio, a novel and interesting association structure introduced by Ekholm and his colleagues [24,25,26]. The dependence ratio is defined as the joint success probability divided by the joint success probability assuming independence. For example, the second order dependence ratio is

$$\tau_{12} = \Pr(Y_1=1, Y_2=1)/\{\mathrm{pr}(Y_1=1).\Pr(Y_2=1)\} \qquad (2)$$

In words τ_{12} measures how many times more probable it is to observe both (Y_1=1) and (Y_2=1) than would be expected if the two events were independent. This generalizes to dependence ratios of higher orders. These dependency ratios provide 'simple and sensible association models for third- and higher- order moments' [27]. A package 'drm' [28] in R [29] is available for fitting such models. The major advantage, and one could say, a unique characteristic of this approach is that it can estimate three and higher order dependencies.However, this approach does not appear to have attracted much use.

Copula models

By definition, an *m*-variate copula $C(u_1, . . . ,u_m)$ is a cumulative distribution function (cdf) with uniform marginals on the unit interval [0,1,30]. If $F_j\ (y_j\ )$ is the cdf of a univariate random variable Y_j, then $C(F_1(y_1), . . . , F_m(y_m))$ is an *m*-variate distribution for $\mathbf{Y}=(Y_1, . . . ,Y_m)$ with marginal distributions F_j , j =1, . . . ,m. Conversely, if H is an *m*-variate cdf with univariate marginal cdfs $F_1, . . . , F_m$, then there exists an *m*-variate copula C such for all $\mathbf{y}=(y_1, . . . , y_m)$,

$$H(y_1, . . . , y_m)=C(F_1(y_1), . . . , F_m(y_m)) \qquad (3)$$

Copulae provide a powerful framework to build multivariate distributions, which can then be used for formulating multivariate regression models. A copula based multivariate binary regression model is obtained by letting $F_j(y_j) = F_j(\underline{x}^t\ \underline{\beta}\ )$ for suitable choice of $F_j(.)$,

2.3 The proposed approach

We proposed the use of a trivariate copula regression model to jointly model Diarrhoea, ARI and fever. Following Nikoloulopoulos and Karlis [31] and Filippou and his colleagues [12], we proposed a trivariate Gaussian copula model with logistic margins. The model is based on a trio of responses and a copula specification for the dependence structure between the three responses.

$$C(F_1(\eta_{1i}), F_2(\eta_{2i}), F_3(\eta_{3i})) = \Phi_3\ (\Phi^{-1}\ \{F_1(\eta_{1i})\}\ , \Phi^{-1}\ \{F_2(\eta_{2i})\}\ , \Phi^{-1}\ \{F_3(\eta_{3i})\}\ ; 0, \Sigma_i), \qquad (4)$$

where Φ^{-1} is the quantile function of a standard normal, $F_m(\eta_{mi})$ is derived in this case from the logistic univariate cdf which is defined as

$$F_m(\eta_{mi}) = \exp(\eta_{mi})/\{1 + \exp(\eta_{mi})\} \qquad (5)$$

The elements of Σ_i are $\sigma_{rs,i}$, where $\sigma_{rs,i}$, is the correlation between the r^{th} and the s^{th} responses for individual i

Each coefficient in Σ_i is allowed to be expressed as a function of an additive predictor.

Function gjrm() in the R package GJRM [32] which is a function for fitting various generalised joint regression models with several types of covariate effects and distributions is used to implement this joint model.

We therefore specified the following six models for input into the gjrm() function:

$$\eta_j = \underline{x}^t \underline{\beta}_j \qquad j=1,2,3 \tag{6}$$

$$\sigma_{rs} = \underline{x}^t \underline{\beta}_{rs} \qquad r=1,2;\ s=2,3;\ r<s \tag{7}$$

3. Results

3.1 Descriptive statistics

Table 1 gives the number of children, and prevalence of ARI, fever and diarrhea by some of the variables known to be related to the three illnesses.

Briefly, referring to Table 1, for all the three outcomes children aged one year appear to be worse off than the infants, but thereafter, prevalence reduces with increasing age of the children. And for all three outcomes male children have higher prevalence than the female children. As far as ARI is concerned, the results for smoking are a bit surprising, showing increased prevalence for children of mothers who smoke. However, smoking is a rarity, probably confined to women in the richer class. The results for cooking fuel show reduced prevalence for children in households where charcoal is used. This is probably confounded with urban residence. Urban residence, in general shows reduced prevalence of all three outcomes. ARI and fever show dramatic reduction in prevalence with increasing level of education. The results for diarrhea are not so clear cut. For all three outcomes, we note that there is reduced incidence with increasing wealth, particularly for ARI and fever. For diarrhea, the results for source of drinking water are rather surprising, while the results for toilet facility show the expected reduction in prevalence for children using improved facilities.

Table 1: Number of children, and prevalence of ARI, fever and diarrhea by selected mother/child characteristics

Mother/Child Characteristic	**Number of children**	**% with** ARI	% with fever	**% with diarrhoea**
Overall	14,710	9.67	34.25	19.87
Child's age				
0	3,091	10.74	32.09	29.96
1	2,922	12.08	40.52	30.77
2	2,916	10.22	36.80	18.86
3	2,861	7.76	33.10	11.29
4	2,920	7.47	28.84	7.71
Child's sex				
Male	7,384	9.98	34.74	21.10
Female	7,326	9.36	33.76	18.63
Mother smokes				
Yes	113	7.96	36.28	20.35
No	14,597	9.69	34.23	19.87
Cooking fuel				
Charcoal	3,057	5.99	24.24	16.69
Wood/straw	11,122	10.65	37.03	19.92
Other	531	10.36	33.71	19.59
Place of residence				
Urban	2,673	6.96	22.33	17.28
Rural	12,037	10.28	36.89	20.45
Region	Details omitted			
Mother's education level				
No education	1,937	13.42	38.26	18.74
Primary	9,187	9.68	36.36	20.27
Secondary	2,801	8.32	28.74	19.92
Higher	785	5.22	19.36	17.83
Wealth Quintile				
Poorest	3,915	13.64	44.37	21.84
Poorer	3,210	10.16	36.85	20.81
Middle	2,824	8.99	32.54	19.55
Richer	2,461	7.96	31.13	19.06
Richest	2,300	4.91	18.83	16.48
Source of drinking water				
Improved	10,950	9.62	35.46	19.88
Unimproved	3,760	9.84	30.72	19.84
Toilet facility				
Improved	2,034	7.52	29.25	17.06
Unimproved	12,216	10.01	35.06	20.31

3.2 Comorbidity and dependency

As earlier stated earlier, comorbidity, if it exists, can easily by shown by a crosstabulation. Table 2 shows the frequencies in the crosstabulation of the three illnesses, and it shows considerable comorbidity, including some cases with all three conditions.

Table 2: Frquency of Comorbidity

Fever	ARI	Diarrhea	Frequency	% of Total
N	N	N	8,071	54.87
N	N	Y	1,122	7.63
N	Y	N	359	2.44
N	Y	Y	120	0.82
Y	N	N	2,831	19.25
Y	N	Y	1,263	8.59
Y	Y	N	526	3.58
Y	Y	Y	418	2.84
Total			14,710	100.00

To determine whether the three illnesses are independent or not we fitted a saturated log-linear model for the frequencies, whose generic form for three factors A,B, and C is:

$$\log(E_{ijk}) = a + A_i + B_j + C_k + AB_{ij} + AC_{ik} + BC_{jk} + ABC_{ijk}$$

Here, A_i , B_j , and C_k are the 'main effects' of the of the ^{i}th level of factor A, the ^{j}th level of factor B, and the ^{k}th level of factor C, respectively. AB_{ij} is the effect of interaction between the ith level of factor A and the jth level of factor B. AC_{ik} , BC_{jk} , and ABC_{ijk} have similar interpretations. Here we are using a variation of the notation popularized by Bishop, Fienberg, and Holland [33]. Dependency is established by significance of the interaction terms. In the present case the factor A, B, and C would stand for fever, diarrhoea, and Acute Respiratory Infection. The results of the partial Chi-square test for the significance of the interaction terms are shown Table 3

Table 3:.Significance of Interaction Terms

Effect	df	Partial Chi-Square	Sig.
dia*fev	1	672.926	0
dia*ari	1	110.59	0
fev*ari	1	513.083	0
Dia*fev*ari	1	5.019	0.025

The results showed significant association (both two- and three-way) and justified the use of a joint regression model.

3.3 Results of independent logistic regression models

Table 4 gives the results of the independent logistic regression for each of the three illnesses. The main purpose is to compare the results of the independent models with the results of the joint model.

Table 4: Results of Independent Logistic Regressions.

Mother/Child Characteristic	**Independent logistic regression for Acute Respiratory Infection**				**Independent logistic regression fever**				**Independent logistic regression for diarrhoea**			
	Coef	SE	Z	p-value	Coef	SE	Z	p-value	Coef	SE	Z	p-value
Child's age 0 1 2 3 4	0.156 -0.052 -0.356 -0.391	0.083 0.086 0.093 0.093	1.874 -0.606 -3.826 -4.190	0.061 0.545 0.000 0.000	0.388 0.217 0.052 -0.155	0.056 0.056 0.057 0.058	6.917 3.845 0.905 -2.672	0.000 0.000 0.366 0.008	0.048 -0.591 -1.207 -1.592	0.057 0.063 0.072 0.080	0.835 -9.433 -16.668 -19.797	0.404 0.000 0.000 0.000
Child's sex Female Male	0.067	0.057	1.169	0.242	0.035	0.036	0.975	0.330	0.155	0.043	3.578	0.000
Mother smokes No Yes	-0.285	0.352	-0.810	0.418	-0.080	0.208	-0.382	0.702	0.002	0.252	0.009	0.993
Cooking fuel Charcoal Wood/straw Other	0.213 0.775	0.105 0.387	2.040 2.002	0.041 0.045	0.033 0.200	0.061 0.283	0.541 0.705	0.588 0.481	-0.233 -0.494	0.071 0.367	-3.305 -1.345	0.001 0.179
Place of residence Rural Urban	-0.001	0.098	-0.015	0.988	-0.341	0.061	-5.554	0.000	-0.128	0.072	-1.783	0.075
Mother's education level No education Primary Secondary Higher	-0.254 -0.152 -0.400	0.079 0.107 0.193	-3.210 -1.423 -2.071	0.001 0.155 0.038	0.032 -0.016 -0.272	0.054 0.070 0.115	0.586 -0.226 -2.368	0.558 0.821 0.018	0.093 0.132 0.073	0.068 0.085 0.127	1.363 1.552 0.571	0.173 0.121 0.568
Wealth Quintile Poorest Poorer Middle Richer Richest	-0.290 -0.393 -0.488 -0.898	0.078 0.085 0.100 0.143	-3.715 -4.612 -4.853 -5.868	0.000 0.000 0.000 0.000	-0.295 -0.471 -0.505 -1.002	0.051 0.055 0.062 0.090	-5.804 -8.622 -8.126 -11.10	0.000 0.000 0.000 0.000	-0.089 -0.172 -0.224 -0.476	0.062 0.067 0.076 0.105	-1.435 -2.591 -2.958 -4.533	0.151 0.010 0.003 0.000
Source of drinking water Unimproved Improved	0.023	0.068	0.329	0.742	0.339	0.044	7.668	0.000	0.015	0.052	0.278	0.781
Toilet facility Unimproved Improved	0.001	0.096	0.014	0.989	0.081	0.058	1.404	0.160	-0.093	0.070		

3.4 Results of marginal logistic regression models

Table 5 shows the results of the marginal regressions for the three illnesses from the joint copula-based regression model.

Table 5: Results of Marginal Regressions from the Joint Model

Mother/Child Characteristic	Marginal logistic regression for Acute Respiratory Infection				Marginal logistic regression fever				Marginal logistic regression for diarrhoea			
	Coef	SE	Z	p-value	Coef	SE	Z	p-value	Coef	SE	Z	p-value
Child's age												
0												
1	0.162	0.083	1.938	0.053	0.389	0.056	6.952	0.000	0.038	0.057	0.660	0.509
2	-0.041	0.086	-0.479	0.632	0.221	0.056	3.929	0.000	-0.598	0.063	-9.542	0.000
3	-0.345	0.093	-3.711	0.000	0.052	0.056	0.982	0.326	-1.214	0.072	-16.779	0.000
4	-0.384	0.093	-4.118	0.000	-0.154	0.058	-2.648	0.008	-1.605	0.081	-19.779	0.000
Child's sex												
Female												
Male	0.071	0.057	1.236	0.216	0.043	0.036	1.180	0.238	0.159	0.043	3.690	0.000
Mother smokes												
No												
Yes	-0.201	0.349	-0.580	0.562	-0.064	0.206	-0.312	0.755	0.022	0.246	0.088	0.930
Cooking fuel												
Charcoal												
Wood/straw	0.216	0.105	2.058	0.040	0.022	0.061	0.362	0.718	-0.238	0.071	-3.373	0.001
Other	0.788	0.383	2.058	0.040	0.191	0.283	0.677	0.498	-0.432	0.360	-1.200	0.230
Place of residence												
Rural												
Urban	0.003	0.098	0.026	0.979	-0.340	0.062	-5.535	0.000	-0.129	0.071	-1.799	0.072
Mother's education level												
No education												
Primary	-0.252	0.079	-3.180	0.001	0.034	0.054	0.621	0.535	0.091	0.068	1.343	0.179
Secondary	-0.152	0.106	-1.426	0.154	-0.012	0.070	-0.173	0.863	0.128	0.085	1.496	0.135
Higher	-0.392	0.192	-2.041	0.041	-0.266	0.115	-2.321	0.020	0.066	0.127	0.518	0.604
Wealth Quintile												
Poorest												
Poorer	-0.279	0.078	-3.575	0.000	-0.298	0.051	-5.867	0.000	-0.077	0.062	-1.247	0.212
Middle	-0.385	0.085	-4.531	0.000	-0.475	0.055	-8.715	0.000	-0.167	0.066	-2.521	0.012
Richer	-0.482	0.100	-4.795	0.000	-0.504	0.062	-8.125	0.000	-0.230	0.076	-3.049	0.002
Richest	-0.896	0.152	-5.883	0.000	-1.024	0.090	-11.34	0.000	-0.460	0.105	-4.390	0.000

Source of drinking water Unimproved Improved	0.035	0.068	0.514	0.607	0.341	0.044	7.740	0.000	0.012	0.052	0.221	0.825
Toilet facility Unimproved Improved	0.005	0.096	0.050	0.960	0.084	0.058	1.452	0.146	-0.091	0.070	-1.308	0.191

In the event the results from the independent models are quite similar to the results from the marginal models from the joint regression. The following observations are based on the joint regression model

*ARI:*Older ages(3&4 years) mother having primary or higher education and not being in the poorest quintile are associated with reduced prevalence of ARI; while cooking with wood/straw and other fuels, rather than charcoal, is associated with increased prevalence of ARI.

*Fever:*Being aged 1,2 or 4 years, higher level of education for the mother, urban residence, and not being in the poorest quintile are associated with reduced prevalence of fever; while, surprisingly having an improved source of drinking water is associated with increased prevalence of fever.

*Diarrhoea:*Being 2 years or older , being in the three upper quintiles of household wealth and using charcoal for cooking are associated with reduced prevalence of diarrhea. Male children have higher prevalence of diarrhea.

3.5 Fitted correlations

The following residual correlations were estimated by the model:

$\sigma_{\text{diarrhea, fever}}$ = 0.419(0.303,0.515)

$\sigma_{\text{diarrhea, ARI}}$ = 0.293(0.129,0.429)

$\sigma_{\text{fever, ARI}}$ = 0.425(0.273,0.543)

Generally, the three binary outcomes are strongly correlated with each other even after accounting for covariates at marginal level. The implication of such positive correlations is that the interventions to reduce the prevalence of one illness will also reduce the prevalence of the correlated illness.

3.6 Results of linear regression models for the correlations

Since the copula model allows for each of the correlations to be expressed as a function of an additive predictor, we fitted such models to the correlations. Table 6 shows the results of linear regressions of the correlations on the same covariates that were considered for the marginal models.

Table 6: Results of Linear Regressions for the Correlations

Mother/Child Characteristic	**Linear regression for the correlation between diarrhea and fever**				**Linear regression for the correlation between diarrhea and ARI**				**Linear regression for the correlation between fever and ARI**			
	Coef	SE	Z	p-value	Coef	SE	Z	p-value	Coef	SE	Z	p-value
Child's age												
0												
1	-0.154	0.052	-2.935	0.003	0.024	0.064	0.381	0.703	-0.078	0.068	-1.149	0.251
2	-0.099	0.056	-1.778	0.075	0.066	0.067	0.996	0.320	-0.097	0.069	-1.402	0.161
3	-0.014	0.063	-0.226	0.821	0.189	0.076	2.473	0.013	0.031	0.077	0.402	0.688
4	-0.051	0.055	-0.777	0.437	0.113	0.080	1.416	0.157	0.078	0.078	0.998	0.318
Child's sex												
Female												
Male	-0.009	0.037	-0.235	0.814	0.035	0.046	0.770	0.441	0.049	0.046	1.053	0.292
Mother smokes												
No												
Yes	0.684	0.420	1.631	0.103	0.346	0.311	1.114	0.265	-0.073	0.328	-0.222	0.825
Cooking fuel												
Charcoal												
Wood/straw	-0.011	0.060	-0.191	0.848	-0.107	0.082	-1.301	0.193	-0.003	0.082	-0.037	0.970
Other	0.163	0.337	0.485	0.627	0.748	0.765	0.978	0.328	-0.878	0.597	-1.471	0.141
Place of residence												
Rural												
Urban	0.027	0.061	0.443	0.658	0.132	0.080	1.652	0.099	0.042	0.077	0.539	0.590
Mother's education level												
No education												
Primary	-0.020	0.060	-0.336	0.737	0.061	0.069	0.887	0.375	0.011	0.070	0.152	0.879
Secondary	-0.093	0.073	-1.271	0.204	-0.067	0.089	-0.750	0.453	-0.075	0.089	-0.847	0.397
Higher	-0.091	0.113	-0.811	0.417	0.064	0.147	0.433	0.665	-0.130	0.141	-0.919	0.358
Wealth Quintile												
Poorest												
Poorer	-0.034	0.055	-0.615	0.539	0.032	0.065	0.497	0.619	-0.035	0.067	-0.523	0.601
Middle	-0.096	0.057	-1.679	0.093	0.079	0.071	1.121	0.262	-0.072	0.070	-1.028	0.304
Richer	-0.155	0.064	-2.423	0.015	-0.079	0.081	-0.970	0.332	-0.030	0.081	-0.374	0.709
Richest	-0.144	0.090	-1.599	0.100	-0.290	0.120	-2.417	0.016	-0.010	0.118	-0.083	0.934
Source of drinking water												
Unimproved												
Improved	-	0.046	-	0.912	0.062	0.056	1.119	0.263	-0.040	0.056	-0.703	0.482

	0.005		0.110									
Toilet facility Unimproved Improved	0.108	0.061	1.775	0.076	0.044	0.074	0.592	0.554	-0.102	0.074	-1.375	0.169

The correlation between diarrhea and fever appears to reduce among children aged one year and among children from the 'Richer' wealth quintile. The correlation between diarrhea and ARI appears to increase among children aged three years but to reduce among children from the 'Richest' wealth quintile. The correlation between fever and ARI does not seem to vary with any of the covariates considered here. Overall there does not appear to be much variation of the correlations and they could safely have been assumed constant.

4. Discussion

This study aimed to explore the relationship between Diarrhoea, cough and fever in young children in Uganda by making use of a joint trivariate copula regression model. This has been achieved and we have used the copula regression model to identify the significant factors associated with each of the three illnesses. The results are in broad agreement with other studies that have considered these three illnesses in sub-Saharan African countries. The effects of the child's age, household wealth and mother's education on the prevalence of ARI have also been observed, at least in part, by Kandala and his colleagues [3], Bbaale [8] and Gebertsadik and his colleagues [9] . Kandala and his colleagues [3] studying diarrhoea, cough and fever among young children in Nigeria, like us found parental education, residence (urban/rural), household economic status, and the child's age to be significantly associated with the childhood morbidity investigated. However in their case they also found significant variables which we did not have in our study, namely, place of birth (hospital v. other), type of feeding (breastfed only v. other), maternal visits to antenatal clinics, and marital status of mother. Bbaale [8] considering diarrhoea and acute respiratory infection among under-fives in Uganda found that, similar to this study, significant factors associated with the occurrence of both diseases to include mother's education, wealth status and child age. However in his case he also found significant variables which we did not have in our study, namely, type of dwelling, mother's occupation, breastfeeding and child nutritional status. These are, however, arguably, related to household wealth. Also, unlike in this study, he found place of residence significant. Gebertsadik and his colleagues [9] considered ARI in Ethiopia and found that ARI was significantly associated with socioeconomic category and nutritional status. One can argue that this is in agreement with our finding on household wealth. Likewise, the effects of the child's age, mother's education, household wealth, place of residence and source of drinking water on prevalence of fever/malaria have also been observed, at least in part, by Kandala and his colleagues [3] (*vide supra*) , Roberts and Mathews [6], Zgambo and his colleagues [11,7]. Reference [6] considering malaria in Uganda found that closely related to the risk of malaria were main floor material, main wall material and availability of electricity in the household. These are arguably proxies for household wealth. Also significant were the child's age and their caregiver's education level. This is in agreement with the present study. However in their case they also found significant variables which we did not have in our study, namely the event of indoor residual spraying (IRS) and the altitude of the location of the household. Zgambo and his colleagues [11] in Malawi found that the child's age and household wealth.were associated with having malaria. This is in complete agreement with the present study. Reference [7] also in Malawi found that mother's education level, wealth index, child's age, the, place of residence were significantly associated with malaria. Also significant were toilet facility and electricity, which are a reflection of household wealth. However in their case they also found significant variables which we did not have in our study, namely anaemia altitude of the place, region Our results, however, disagree with Niringiye and Douglason [5] who, curiously using Ordinary Least Squares, found no relationship between malaria prevalence and environmental and socio-economic variables in Uganda. And finally, the effects of the child's age and household wealth on diarrhea prevalence was observed, at least in part by Kandala and his colleagues [3] (*vide supra*), Ssenyonga and his colleagues [4], Bbaale [8] (*vide supra*), Kabinga and his colleagues [10].Ssenyonga and his colleagues [4] found child's age, mother's education significantly associated with diarrhoea in Uganda. They also found significant factors not considered in our study ,namely region of residence, and duration of breastfeeding . In their case, unlike in the present study, source of drinking water was found to be significant. Interestingly significant comorbidity with fever was found. Kabinga and his colleagues [10] in Zambia found that, similar to our study, diarrhoea was associated with the child's age and household wealth. But again, unlike in this study source of drinking water was also significant. The approach we have used allows the correlation between the three responses to be estimated while controlling for the effects of independent variables. We indeed found that the three illnesses are strongly correlated with each other even after accounting for covariates at marginal level. Copula based multivariate regression methods have not been much used in this region (sub-Saharan Africa) and therefore few studies have looked at correlations among illnesses when there is comorbidity. We can only cite Kinyoki and his colleagues [17], who also found a high significant correlation between ARI and diarrhea in among children in Somalia. In the present study we used the same set of covariates for the three marginal models and for the models for the correlations. This need not be the case, and in fact each of the six models could have had its own set of covariates. We used a Logit link for the three marginal models. Other links like the probit and complementary log-log are possible, and in fact each marginal model can have its own link function. Though the present study considered the case of three binary variables, the copula method, and the GJRM package are available for other types outcome (continuous variable, count (discrete) variable, survival time) in any combination. The major limitation of the study is that illness was as reported by the caregiver (usually the mother) and therefore not medically validated.

5. Conclusion

This study has illustrated an easily applicable technique to joint modelling of common childhood illnesses which allows exploration of the correlations between them. We strongly recommend the use of this method for correlated outcomes, whatever the type of variable.

Acknowledgements

The author would like to thank the MEASURE DHS project for their support and for free access to the UDHS data. The author would also like to thank the College of Business and Management Sciences, Makerere University for a paper writing grant.

References

[1]. Adeokun, S. T., Yata, S. (2020) Childhood morbidity and its determinants: evidence from 31 countries in sub-Saharan Africa. *BMJ Global Health, 5(10)*. http://dx.doi.org/10.1136/bmjgh-2020-003109.

[2]. Fenn, B., Morris, S.S. & Black, R.E. (2005). Comorbidity in childhood in northern Ghana: magnitude, associated factors, and impact on mortality. *Int. J. Epidemiol*, **34**(2), 368–375, DOI 10.1093/ije/dyh335.

[3]. Kandala, N-B., Ji, C., Stallard, N., Stranges, S., Cappuccio, F. (2008) Morbidity from diarrhoea, cough

and fever among young children in Nigeria. *Ann Trop Med Parasitol 102(5):*427–445

[4]. Ssenyonga, R., Muwonge, R., Twebaze, F. B. N., Mutyabule, R. (2009) Determinants of Acute Diarrhoea in Children Aged 0 – 5 Years in Uganda. *The East African Medical Journal, 86(11)*, 513-519.

[5]. Niringiye, A. and O.G. Douglason, O. H. (2010) Environmental and Socio-economic Determinants of Malaria Prevalence in Uganda *Res. J. Environ. Earth Sci., 2(4): 194-198, 2010*

[6]. Roberts, Danielle and Mathews, Glenda (2016) Risk factors of malaria in children under the age of five years old in Uganda *Malaria Journal* **15**:246. https://doi.org/10.1186/s12936-016-1290-x

[7]. Gaston, R.T., Shaun Ramroop, S.(2020) Prevalence of and factors associated with malaria in children under five years of age in Malawi, using malaria indicator survey data. *Heliyon, 6(5).*e03946,https://doi.org/10.1016/j.heliyon.2020.e03946.

[8]. Bbaale, E.(2011) Determinants of diarrhoea and acute respiratory infection among under-fives in Uganda. *Australas Med J. 2011;4(7)*:400-9. doi: 10.4066/AMJ.2011.723. Epub 2011 Jul 31.

[9]. Gebertsadik, A., Worku, A., Berhane, Y. (2015) Factors associated with acute respiratory infection in children under the age of 5 years: evidence from the 2011 Ethiopia Demographic and Health Survey. *Pediatric Health, Medicine and Therapeutics Volume 6 9-13.* **DOI:** https://doi.org/10.2147/PHMT.S77915

[10]. Kabinga, F.M., Mbewe, A., Banda, J. (2016)Environmental health factors associated with diarrhoea in the under five children in Mtendere, Lusaka, Zambia. *International Journal of Infectious Diseases,Volume 53, Supplement*, Page 69. DOI: https://doi.org/10.1016/j.ijid.2016.11.173

[11]. Zgambo, M., Mbakaya, B.C., Kalembo, F.W. (2017) Prevalence and factors associated with malaria arasitaemia in children under the age of five years in Malawi: A comparison study of the 2012 and 2014 Malaria Indicator Surveys (MISs). *PLoS ONE 12(4)*: e0175537. https://doi.org/10.1371/journal.pone.0175537

[12]. Filippou, P., Kneib, T., Marra, G. and Radice, R. (2018). A trivariate additive regression model with arbitrary link functions and varying correlation matrix. *Journal of Statistical Planning and Inference, 199*, pp. 236-248. doi: 10.1016/j.jspi.2018.07.002

[13]. Kazembe, L. N., Muula, A., Simoonga, C. (2008) Joint spatial modelling of common morbidities of childhood fever and diarrhea in Malawi. *Health & Place 15(1):*165-72. DOI: 10.1016/j.healthplace.2008.03.009

[14]. Gayawan, E., Arogundade, E. D., Adebayo, S. B. (2014) A Bayesian multinomial modeling of spatial pattern of co-morbidity of malaria and non-malarial febrile illness among young children in Nigeria. *Trans R Soc Trop Med Hyg 108(7):*415–424

[15]. Adebayo, S. B., Gayawan, E., Heumann, C., Seiler, C. (2016) Joint modeling of anaemia and malaria in children under five in Nigeria. *Spat Spatio-Temp Epidemiol 17*:105–115

[16]. Habyarimana, F., Zewotir, T., Ramroop, S. (2016) Joint Modeling of Poverty of Households and Malnutrition of Children Under Five Years from Demographic and Health Survey Data: Case of Rwanda. *Journal of Economics and Behavioral Studies (ISSN: 2220-6140) Vol. 8, No. 2, pp.* 108-114.

[17]. Kinyoki, D. K., Manda, S. O., Moloney, G. M., Odundo, E. O., Berkley, J. A.,Noor, A. M., Kandala, N-B. (2017) Modelling the Ecological Comorbidity of Acute Respiratory Infection, Diarrhoea and

Stunting among Children Under the Age of 5Years in Somalia. *International Statistical Review (2017)*, 85, 1, 164–176 doi:10.1111/insr.12206

[18]. Roberts, D.J., Zewotir, T.(2020) Copula geoadditive modelling of anaemia and malaria in young children in Kenya, Malawi, Tanzania and Uganda. *J Health Popul Nutr. 2020 Nov 6;39(1):*8. doi: 10.1186/s41043-020-00217-8.PMID: 33158460

[19]. Uganda Bureau of Statistics (UBOS) and ICF (2018) *Uganda Demographic and Health Survey 2016: Key Indicators Report*. Kampala, Uganda: UBOS, and Rockville, Maryland, USA: UBOS and ICF.

[20]. Liang, K. Y. & Zeger, S. L. (1986) Longitudinal data analysis using generalized linear models. *Biometrika, 73,* 13-22

[21]. Prentice, R.L. (1988), Correlated Binary Regression with Covariate Specific to Each Binary Observation. *Biometrics, 44*, 1033-1048.

[22]. Lindsey J, Lambert P. (1998)On the appropriateness of marginal models for repeated measurements in clinical trials. *Statistics in Medicine ; 17*:447–469.

[23]. StataCorp (2015) *Stata StR.atistical Software*: Release 15. College Station, TX: Stata Corporation.

[24]. Ekholm, A., Smith, P.W.F., McDonald, J.W. (1995). Marginal regression analysis of a multivariate binary response. *Biometrika, 82,* 847–854.

[25]. Ekholm, A., McDonald, J.W., Smith, P.W.F. (2000). Association models for a multivariatebinary response. *Biometrics 56,* 712–718.

[26]. Ekholm, A., Jokinen, J., McDonald, J. W., amd Smith, P. W. F. (2003) Joint Regression and Association Modelling of Longitudinal Ordinal Data. *Biometrics, 59,* 795-803

[27]. Jokinen, J. (2006) Fast estination algorithm for likelihood-based analysis of repeated categorical responses. *Computational Statistics and Data Analysis, 51*, 1509-1522.

[28]. Jokinen, J. (2009) *drm:regression and association models for repeated categorical data*. http://www.helsinki.fi/~jtjokine/drm

[29]. R Development Core Team (2006) *R:A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org.

[30]. Nelsen, R. B. (2006) *An Introduction to Copulas*. Springer: New York.

[31]. Nikoloulopoulos, A. K., Karlis, D.(2008) Multivariate logit copula model with an application to dental data. *Stat Med. ;27*:6393–406.

[32]. Marra G, Radice R. GJRM: Generalised Joint Regression Modelling. R package version 0.1-2. 2017. Available on CRAN. https://rdrr.io/cran/

[33]. Bishop, Y. M. M., Fienberg, S. E., Hlland, P.(1975) *Discrete Multivariate Analysis: Theory and Practice.* Cambridge, Mass.: MIT Press