Modelling Distribution of Under-Five Child Diarrhoea across Malawi

Tsirizani M. Kaombe and Jimmy J. Namangale

Mathematical Sciences Department, Chancellor College - University of Malawi, P.O. Box 280, Zomba, Malawi

Abstract: Analysis of diarrhoea data in Malawi has been commonly done using classical methods. However, different approaches, such as Bayesian methods, have been introduced in literature. This study aimed at trying out semi-parametric methods in comparison with classical ones, as well as how each isolates risk factors for child diarrhoea. This was done by fitting Logit, Poisson, and Bayesian models to 2006 Malawi Multiple Indicator Cluster Survey data. The comparison between Logit and Poisson models was done via chi-square's goodness-of-fit test. Confidence and Credible Intervals were used to compare Logit/Poisson and Bayesian model estimates. Modelling and inference in Bayesian method was done through MCMC techniques. The results showed agreement in significance and direction of estimates from Bayesian and Poisson/Logit models, but Poisson provided better fit than Logit model. Further, all the models identified child's age, breastfeeding status, region of stay and toilet-sharing status as significant factors for determining the child's risk. The models ruled out effects of mother's education, area of residence, and source of drinking water on the risk. Bayesian model separately proved significant closeness to lake/river factor. The findings imply that classical and semi-parametric models are equally helpful when estimating the child's risk to diarrhoea.

Keywords: Child Diarrhoea, Bayesian methods, Logit model, Poisson model.

1. Introduction

Diarrhoea is an increase in volume of stool or frequency of defecation. It is one of the most common clinical signs of gastrointestinal diseases, but can also reflect primary disorders outside digestive system [1]. The disease can take different levels of clinical intensity; acute to chronic or severe stage. Acute diarrhoea appears rapidly and may last from 5 to 10 days, while chronic diarrhoea lasts much longer than this [2]. Children are more susceptible to complications of diarrhoea because a smaller amount of fluid loss leads to dehydration, compared to adults [3]. An estimated 3.5 million deaths each year are due to diarrhoea worldwide, 80% of which occur in under-5 children [4]. In Malawi, the disease accounts for 11% of deaths in under-5 annually [5].

Improving a child's health is one of the eight Millennium Development Goals (MDGs) that were adopted by governments at the United Nations Millennium Summit in 2000. In MDG 4, countries had committed to reducing child mortality rates by three quarters from their national baseline rates between 1990 and 2015 [4]. To that effect, Malawi government recently adopted UNICEF/ WHO's seven-point plan for diarrhoea control, which include: fluid replacement to prevent dehydration; zinc treatment; rotavirus and measles vaccinations; promotion of early and exclusive breastfeeding and vitamin A supplementation; promotion of hand washing with soap; improved water supply quantity and quality, including treatment and safe storage of household water; and community-wide sanitation promotion [6]. The 2004 Malawi Demographic and Health Survey (DHS) indicated that 22% of children had diarrhoea at some time in two weeks preceding the survey and that the disease problem was not

---

Corresponding author: Tsirizani Mwalimu Kaombe, MSc., Lecturer in Statistics, Chancellor College, University of Malawi; Research field: Biostatistics; Post address: PO Box 280, Zomba, Malawi. E-mail: tsirizanikaombe@yahoo.com.

uniform across age groups (highest in age 6-11 months, 41%), across districts (most prevalent in Salima, Kasungu, and Thyolo, $\geq 27\%$), across areas of residence (most prevalent in rural areas, 23%), among other background characteristics of the child. Hence, there was need to model the distribution of the disease across the country in order to capture the proper pattern for better control programmes.

Further, efforts to control childhood diarrhoea in sub-Saharan Africa, including Malawi, have been primarily hampered by knowledge gap on prevention strategies by citizens of the region. For instance, it has been reported that morbidity and mortality from childhood diarrhoea were compounded by inappropriate household case management and frequent misuse of antibiotics by citizens of countries such as Gambia, Guinea-Bissau, Kenya, Malawi, Nigeria, Tanzania, DRC and Sudan [7]. More recently, a study by [1] in Temeke Municipality, Dar es Salaam-Tanzania has revealed poor Mothers' knowledge on predisposing factors of childhood diarrhoea, which was directly correlated with education level. It is reported that only about one-third of respondents were aware of risk factors of the disease, which cited poor sanitation and water as main factors. Besides, diarrhoea episodes were perceived wrongly as normal growth stage and that they were caused by several other ''illnesses''. These results agree with findings from a qualitative study by [8] done in Rumphi, Malawi, where it is reported that mothers and caregivers wrongly associated child diarrhoea to child's teeth development and breastfeeding by a pregnant mother. Thus, a quantitative follow up study to that of [8] was necessary at national level to establish the risk factors.

Therefore, this study aimed at applying statistical models to investigate variations in risk of diarrhoea in under-5 children in Malawi. Specifically, the study: 1) compared estimates found using classical and semi-parametric models; 2) identified appropriate classical model to use when explaining the child's risk;

and 3) evaluated influence of socio-economic and bio-demographic factors on the risk.

2. Materials and Methods

2.1 Study Design and Statistical Methods

The study applied quantitative Bayesian semi-parametric additive, Logistic, and Poisson regression models on 2006 Malawi Multiple Indicator Cluster Survey (MICS) data. The national data was used in anticipation of accurate statistical estimates due to large sample size and cross-sectional child information from all districts that was available.

Furthermore, logistic model was used due to the fact that the outcome variable, two-week total number of under-5 diarrhoea cases, was believed to follow binomial distribution. The model, introduced in the 1940s, is an example of a Generalized Linear Model where the random component is a Bernoulli variable, whose distribution is specified by probabilities; $P(Y = 1) = \pi$ of success, and $P(Y = 0) = 1 - \pi$ of failure, Y denoting diarrhoea outcome on a child, with 1 representing presence and 0 absence of diarrhoea. The total number of cases $\sum(Y = 1)$ in one child observation, which was either 1 or 0, was a Bernoulli random variable with prevalence parameter π, $\sum(Y = 1) \sim Ber(\pi)$. Hence, for some fixed number n of children observed, the total number of cases, $\sum(Y = 1)$ had a Binomial distribution with parameters n and π, $\sum(Y = 1) \sim Ber(n, \pi)$. The distribution's specification is:

$$f(y; n, \pi) = \exp\left[\log\binom{n}{y} + n\log(1 - \pi) + y\log\frac{\pi}{1 - \pi}\right]$$

which is an exponential form with natural parameter, $\log\left[\frac{\pi}{1 - \pi}\right]$.

The actual value of diarrhoea prevalence π in the population of children was believed to be a quantity that depended on unique contributions by some

separate characteristics of the children. The relationship between the characteristics, x and the parameter, π was assumed to be nonlinear, resembling a logistic function, given by:

$$\pi(x) = \frac{\exp(\beta_0 + \beta_1 x_1 + ... + \beta_p x_p)}{1 + \exp(\beta_0 + \beta_1 x_1 + ... + \beta_p x_p)} \quad ,$$

where $\beta_0, . . . ,\beta_p$ are regression parameters going with characteristics and were to be estimated from the data y.

The logistic function given above was linearized by the logit transform below;

$$\hat{\pi}(x) = \log\left(\frac{\pi(x)}{1 - \pi(x)}\right) = \beta_0 + \beta_1 x_1 + ... + \beta_p x_p$$

which is the binomial distribution's natural parameter stated before.

The estimation of regression parameters was done by Maximum Likelihood Estimation (MLE) technique. The MLE estimates of parameters β, and consequently of probabilities $\pi(xi) = g(xi^T\beta)$, are obtained by maximizing the log-likelihood function;

$$l(\pi; y) = \sum_{i=1}^{n}\left[\log\binom{n}{y} + n\log(1 - \pi) + y\log\frac{\pi}{1 - \pi}\right]$$

using iterative weighted least squares procedure [9].

Since the total number of diarrhoea cases $\sum(Y = 1)$, at any time was a positive integer, then Poisson distribution was also assumed for the sum of cases throughout the 4-months period of survey. If $\sum(Y = 1)$ is a Poisson random variable with mean $\mu = E\left[\sum(Y = 1)\right] = Var\left[\sum(Y = 1)\right]$, then:

$$f(y; \mu) = \exp[\mu - \log(y!) + y\log\mu], \ y=0,1,2,...$$

Here, the natural parameter was $\log\mu$, and thus linearized regression relation was;

$$\log\mu = \alpha + \Sigma\beta x.$$

The Poisson model parameters were estimated using the same MLE technique.

To allow for correlation within households, robust standard errors were calculated using residuals at cluster level.

For the Bayesian semi-parametric model, continuous covariates were treated non-parametrically with help of smoothing functions whereas categorical variables were related parametrically to the response variable [10, 11]. In general, a Bayesian model is considered to be a regression model in which regression parameters are in turn given a probability model [12]. The use of semi-parametric model was therefore thought of in order to capture both linear effects of discrete covariates and nonlinear effects of continuous covariates on the child's risk to diarrhoea. The usual logistic model was assumed first, then parameters β were assumed to be distributed as gamma random variables with fixed location and scale parameters, $u = v = 0.001$, except for child age whose parameters were assumed to have normal prior distributions with 0 means and inverse gamma distributed variances.

Briefly, it was assumed that $y = \sum(Y = 1)$ belongs to an exponential family, with mean $\mu = E(y \mid u, \gamma)$ linked to a linear predictor η by $\mu = h(\eta)$, $\eta = u'\gamma$, where h denoted known response function, and γ unknown regression parameters. The following structured additive predictor model was adapted, see [13]:

$$\eta_r = f_1(x_{r1}) + ... + f_p(x_{rp}) + u'_r\gamma ,$$

where r is a generic observation index; x_{rj} covariates of different type and dimension; and $f_j (j = 1, 2, ..., p)$ are functions of the covariates. The functions f_j comprised of the nonlinear effects of continuous covariate child age.

These estimators, which can be regarded from a Bayesian perspective as posterior means or from a Frequentist perspective as "Best Linear Unbiased Predictors", often have better properties than simple sample-based estimators using only data from the unit in question. This makes them useful in the problem of "small-area estimation", that is, making estimates for

units or domains for which there is very limited amount of information [14].

2.2 Geographic Location and Population Distribution

As earlier alluded to, MICS was conducted in all districts in Malawi, a country that is located in south-east Africa, landlocked between Mozambique to its eastern and southern sides, Zambia to its western side, and Tanzania to its northern side. The country covers an area of 118, 484 km^2, of which 94, 276 km^2 (79.6%) is land and 24, 208 km^2 (20.4%) water. By 2008, the country had 13, 077,160 people. About 90% of the population lives in rural areas where, among other things, access to health services and poverty are major hardships. The country's population had 7, 157, 985 (45.1%) people aged between 0 and 14 years as of October, 2011 [5, 15, 16].

2.3 Study Population and Sampling Techniques

The MICS sampled 23,238 under-five children, of which questionnaires for 22,994 (98.9%) were completed. Further, 27,073 women (aged 15 - 49 years) were identified from sampled households, of which 26,259 (97%) were interviewed. With an aim to obtain estimates, at district level, on key indicators related to wellbeing of children and women, MICS study targeted a sample of size 1,200 households (HHs) per district. By then, there were 28 districts in Malawi, two of which (Likoma and Neno) were too small to draw 1,200 HHs out of total available HHs. As a result, Likoma was merged with Nkhata Bay and Neno with Mwanza [17].

A two-stage cluster sampling design was used to select the households, where within each district 40 census enumeration areas (identified as clusters) were selected, and within each cluster a systematic sample of 30 households was drawn. A total of 31,200 HHs (26 districts multiplied by 1,200 HHs) were selected in 1,040 clusters (26 districts multiplied by 40 clusters) in that process. All the selected clusters were visited during fieldwork [17].

The target population in this study was children aged at most 5 years, with outcome variable of interest being cases/ non-cases of diarrhoea as in 2006 MICS. The explanatory variables included child's age, breastfeeding status (weaned or still breastfeeding), area of residence (rural or urban), region of stay (northern, central or southern), toilet facility (shared between families or not), mother's education, source of drinking water, and closeness to lake/ river.

2.4 Instrumentation and Data Collection

The MICS study, conducted from July to November, 2006, used the under-five children questionnaire, administered to mothers or caretakers to collect information on Vitamin A, breastfeeding status, care of illness, diarrhoea, malaria, immunization, and anthropometry, among other general information on a child.

2.5 Data Analysis Procedures

The data were examined in Stata version 10 package. The children with incomplete data were dropped from analysis, with randomness assumption. Further, the analysis involved estimating baseline characteristics of children such as totals and their percentages. In addition, variable-specific point estimates of two-weeks diarrhoea incidences were calculated. Besides, crude odds ratios (ORs) estimating a child's risk to diarrhoea given two levels of data were calculated, with the aim of foreshadowing model estimates. Finally, logistic (with Logit link), Poisson (with log link), and Bayesian semi-parametric models were fitted to data, using Stata package for the first two models, and BayesX package for the Bayesian model. The results from logistic model are reported as odds ratios (ORs) of effects of levels of factors together with their corresponding 95% Confidence Intervals (CIs), those of Poisson as relative risks with 95% CIs, while for Bayesian model they are reported as posterior mean effects of factors, with their 95% Credible Intervals (CrIs).

Furthermore, consistency of estimates between Bayesian and Logit/ Poisson models was checked through estimates of CrIs and CIs. While, logistic and Poisson models were compared based on chi-square's goodness-of-fit test results.

For the Bayesian model, adequacy was checked via Deviance Information Criterion (DIC) and posterior predictive checking was done via posterior credible intervals. The DIC is a generalization of the Akaike information criterion (AIC) and Bayesian information criterion (BIC). It is most applicable in Bayesian model selection problems where posterior distributions have been obtained through (MCMC) simulation. Deviance is defined as $D(\theta) = -2\log(p(y|\theta)) + C$, where y is the data, θ unknown model parameters, and $p(y|\theta)$ is the likelihood function, while C is a constant that cancels out in all calculations that compare different models and, which therefore, does not need to be known. The expectation $\overline{D} = E[D(\theta)]$ is a measure of how well the model fits data; the larger this is, the worse the fit. The effective number of model parameters is computed as $p_D = \overline{D} - D(\overline{\theta})$, where $\overline{\theta}$ is expectation of θ. The larger this is, the better the model fit. The DIC is calculated as $DIC = p_D + \overline{D}$. The idea is that models with smaller DIC should be preferred to those with larger DIC. Since $\overline{D}$ will decrease as the number of model parameters increases, the p_D term compensates for this effect by favoring models with smaller number of parameters. Hence, DIC is a compromise between model fit and complexity [18].

3. Results

3.1 Child Baseline Characteristics and Diarrhoea Cross-Classification Results

There were a total of 15, 018 (65.3%) children who had complete information on studied variables in the 2006 MICS data and hence formed part of these results. The results in Table 1 show that the study involved almost equal numbers of male and female children. Further, most of the children were in age group 12-23 months, with mean age of 28 and standard deviation of 16 months. In addition, a large proportion of the children was weaned, with more residing in rural areas.

On diarrhoea incidence, the results indicate that 21.85% of the children suffered from diarrhoea at some time in two weeks preceding the survey, with proportional rates in males and females. The rate was highest in age 12-23 months, and higher in breastfed children.

The crude odds ratios show that female children were as likely as male children to catch diarrhoea, although sex was not a significant factor. Further, it is shown that weaned children had 59.8% reduced odds of catching diarrhoea than the breastfed.

Children aged 12-23 months had 84.1% higher odds of catching diarrhoea than those aged 0-11 months. While children aged 24-35; 36-47; and 48-59 months had respectively 18.8%; 52.8%; 65.4% reduced odds of catching diarrhoea compared to those aged 0-11 months. Furthermore, children from rural areas had 20.7% higher odds of catching diarrhoea than those from urban areas.

3.2 Logistic and Poisson Model Results

The results in Table 2 show that chi-square's goodness-of-fit test leads to rejection of the binomial distribution assumption, at 5% level (GoF = 1019, p = 0.0015). For Poisson model, output for unconditional mean and variance of diarrhoea cases provided mean of 0.2185 and variance of 0.1708. The values indicated slight under-dispersion though not substantial and thus predictor variables could be of help. Further, using Microsoft Excel 'rand' function, random samples of 100, 1000, 5000, 10000 and 15000 generated from diarrhoea variable produced prevalence rates of 0.27, 0.251, 0.242, 0.241, and 0.239 respectively, indicating that increasing sample size resulted in reduction of

Table 1 Baseline child characteristics, diarrhoea incidence, and crude odds ratios.

Characteristic		Total children (%)	Diarrhoea Incidence (%)	Crude OR (95% CI, p-value)
Overall		15, 018(100)	3,282 (21.85)	
Gender	Male (ref)	7,450 (49.61)	1,648(10.97)	
	Female	7,568(50.39)	1,634 (10.88)	0.969 (0.897-1.047,0.432)
Age	0-11 (ref)	2,826 (18.82)	682 (4.54)	
	12-23	3,458 (23.03)	1,277 (8.5)	1.841 (1.646-2.058,<0.001)
	24-35	3,400 (22.64)	698 (4.65)	0.812 (0.72-0.916, <0.001)
	36-47	3,054 (20.34)	399 (2.66)	0.472 (0.412-0.542,<0.001)
	48-59	2,280 (15.18)	226 (1.5)	0.346 (0.293-0.408,<0.001)
Breastfeeding	Breastfed (ref)	6,585 (43.85)	2,013 (13.40)	
	Weaned	8,433 (56.15)	1,269 (8.45)	0.402 (0.371-0.436,<0.001)
Area of residence	(ref)Urban	1,936(12.89)	371 (2.47)	
	Rural	13,082(87.11)	2,911 (19.38)	1.207 (1.07-1.362, 0.002)
Altitudinal locale	Highland(ref)	7,037 (46.86)	1,651 (10.99)	
	Lake/river shore	7,981 (53.14)	1, 631(10.86)	0.838 (0.775-0.905,<0.001)
Region	Northern (ref)	3,650 (24.30)	604 (4.02)	
	Central	5,429(36.15)	1,405 (9.36)	1.761(1.582-1.96, <0.001)
	Southern	5,939(39.55)	1,273 (8.48)	1.376 (1.236-1.532,<0.001)
Mother education	Primary(ref)	12,779(85.09)	2,834 (18.87)	
	Secondary	2,165 (14.42)	437 (2.91)	0.887 (0.793-0.994, 0.038)
	Higher	74 (0.49)	11 (0.07)	0.613 (0.322-1.164, 0.13)
Family toilet	Not shared (ref)	9,348 (62.25)	1,869 (12.45)	
	Shared	5,670 (37.75)	1,413 (9.41)	1.328 (1.228-1.437,<0.001)
Drinking water source	Piped(ref)	10,766(71.69)	2,294 (15.28)	
	Protected well	818 (5.45)	188 (1.25)	1.102 (0.93-1.305, 0.26)
	Unprotected well	2,455 (16.35)	584 (3.89)	1.153 (1.039-1.279, 0.007)
	Surface water	979 (6.52)	216 (1.44)	1.045 (0.893-1.224, 0.58)

Table 2 Logit and Poisson model's adjusted OR and RR results.

Variable		Logit, OR (95%CI, p)	Poisson, RR (95%CI, p.)
Breastfeeding	ref(breastfed)	0.695 (0.6-0.8, p<0.001)	0.768 (0.693 - 0.85, p<0.001)
Age in months	ref(0-11)		
	12-23	1.928 (1.17-2.16, p < 0.001)	1.57 (1.452-1.699, p<0.001)
	24-35	1.09 (0.93-1.29, p = 0.29)	1.055(0.936-1.19, p=0.378)
	36-47	0.669 (0.55-0.81, p < 0.001)	0.7(0.604-0.813, p<0.001)
	48-59	0.496 (0.4-0.61, p < 0.001)	0.538(0.453-0.637, p<0.001)
Area of residence	ref(urban)	1.122 (0.985, 1.28, p=0.08)	1.09(0.989-1.202, p=0.083)
Altitudinal locale	ref(highland)	0.918 (0.84, 1.002, p=0.055)	0.939(0.881-1.001, p=0.052)
Region	ref(Northern)		
	Central	1.678(1.495-1.88, p<0.001)	1.472 (1.348-1.608, p<0.001)
	Southern	1.365(1.22-1.528, p<0.001)	1.272 (1.167-1.387, p<0.001)
Mother education	ref(primary)		
	Secondary	0.922 (0.818, 1.04, p=0.185)	0.941(0.861-1.029, p=0.182)
	Higher	0.783 (0.404, 1.52, p=0.47)	0.822 (0.485-1.395, p=0.468)
Family toilet	ref(not shared)	1.273 (1.17-1.38, p<0.001)	1.192 (1.123-1.266, p<0.001)
Drinking water source	ref(piped)		
	Protected well	0.997(0.84, 1.19, p=0.97)	0.995 (0.878-1.129, p=0.942)
	Unprotected well	1.036(0.93, 1.16, p=-0.53)	1.025 (0.947-1.109, p=0.539)
	Surface water	1.084(0.92, 1.28, p=0.343)	1.063 (0.943-1.2, p=0.321)
Overall model fit		GoF=1019, p=0.002; LRT=985, p<0.001	GoF=9225, p=1.00; W=973, p<0.001

prevalence rate, making mean relatively constant. So, it was reasonable to approximate binomial model with logit link by Poisson model with log link. The goodness-of-fit test accepted Poisson assumption, at 5% level (GoF= 9225, p = 1.00), implying that the data give no evidence that diarrhoea variable did not follow Poisson distribution.

The estimates from Logit and Poisson models show that, adjusting for other factors, a weaned child had respectively 30.5% and 23.2% reduced odds and risk of catching diarrhoea compared to a breastfed child. In addition, children who were living in central region had respectively 67.5% and 47.2% higher odds and risk of catching diarrhoea than those from northern region. Likewise, children from southern region had respectively 36.5% and 27.2% adjusted higher odds and risk of catching diarrhoea compared to those from the north. Furthermore, it is indicated that adjusted odds and risk of catching diarrhoea were higher by 27.3% and 19.2% respectively in children whose families shared toilets compared to those whose families did not. The results also show that adjusted odds and risk of catching diarrhoea in children aged 12-23 months were respectively higher by 92.8% and 57% than in children aged 0-11 months, while there was no difference in odds or risk between age groups 24-35 and 0-11 months. However, the adjusted odds and risk were respectively lower by 33.1% and 30% in children aged 36-47 and lower by 50.4% and 46.2% in age 48-59 compared to those aged 0-11 months.

Finally, the two models showed no evidence of difference in adjusted odds and risk of catching diarrhoea between children living in rural and urban areas, lakeshore/ riverside and highland areas, primary and higher than primary educated mothers, and in children drinking from piped water and other sources.

These results agreed with crude ORs reported in section 3.1.

3.3 Bayesian Semi-Parametric Model Results

The results presented in this section were run in

BayesX package Version 2.0.1, using the following code:

b.regress diarrhoea = child-age(psplinerw2) + breastfeeding + mother-edu. + toilet-facility + water-source + lakeshore/highland + residential-area + region, family=binomial iterations=12000 burnin=2000 step=10 predict using d.

The MCMC simulations were run on the set of full conditional posterior distributions in order to derive full posterior estimates for parameters, see [19]. The options *iterations*, *burnin* and *step* define the total number of iterations, the burn in period, and the thinning parameter of the MCMC simulation run [13]. Specifying *step=10* forced BayesX to store only every 10th sampled parameter which leads to a random sample of length 1000 for every parameter in this case. Therefore, a sample of 10000 random numbers was obtained with the above specifications. The model presented in Table 3 has DIC based on un-standardized deviance results, $\bar{D}$ = 14847.347, pD = 11.887, DIC = 14871.121, and based on saturated deviance results, $\bar{D}$ = 14847.347, pD = 11.887, DIC = 14871.121.

The results show that posterior mean amount of diarrhoea cases was expected to be low in weaned children; in children whose mother's education was higher than primary; and in children who lived along the shores of Lake Malawi or Shire River. This means that there were low chances of suffering from diarrhoea in weaned children, children whose mothers attained post primary education, and those who were staying along lake/river shores, compared to their respective counterparts. On the contrary, the results show that posterior mean number of cases was expected to be high in children whose families were sharing toilets, drinking from non-piped water source, living in rural areas, and living in other regions than northern. This implies that there was high risk to diarrhoea in children whose families shared toilets, were drinking from non-piped water sources, living in rural areas, and living in other regions than the north.

Table 3 Bayesian model results

Variable		Posterior mean	95%CrI	80%CrI
Constant		-1.137	-1	-1
Breastfeeding	ref(breastfed)	-0.376	-1	-1
Mother's education	ref (primary)	-0.111	0	-1
Family toilet	ref(not shared)	0.258	1	1
Drinking water source	ref(piped)	0.029	0	1
Altitudinal locale	ref(highland)	-0.202	-1	-1
Area	r ef(urban)	0.110	0	1
Region	ref(northern)	0.123	1	1
	0-11	-0.016	0	0
	12-23	0.624	1	1
Age group in months:	24-35	0.078	0	1
	36-47	-0.411	-1	-1
	48-59	-0.714	-1	-1

Note: 95% or 80% CrI; ± 1 ~ significant posterior mean; 0~insignificant

However, the 95% CrIs show that mother's education, source of drinking water and area of residence were not significant factors for determining a child' risk to diarrhoea, but the rest variables were significant.

For age, expected posterior mean number of cases was low in age groups 0-11 months, 36-47 months, and 48-59 months, but high in age groups 12-23 and 24-35 months. This once again shows that age groups 0-11, 36-47, and 48-59 had reduced risk to diarrhoea, while ages 12-23 and 24-35 were at high risk. However, the 95% CrIs show that age groups 0-11 and 24-35 months had no significant effects.

These results agreed with those from Logit and Poisson models.

4. Discussion

4.1 Consistency of Bayesian and Logit/ Poisson Models Estimates

The results have shown general agreement in significance and direction of estimates from Bayesian and Poisson/ Logit models. The three models have coincidentally provided evidence of usefulness of child's age, breastfeeding status, toilet sharing status, and region of stay in determining a child's risk to diarrhoea, but together ruled out effects of mother's education, area of residence (rural or urban), and source of drinking water on the risk.

4.2 Risk Factors for Child Diarrhoea

4.2.1 Breastfeeding status of child

The lower risk in weaned compared to breastfed children could reflect low possibilities of diarrhoea germs' transmission between a weaned child and his/her mother, since the child chooses what to put to the mouth independent of the mother. It may also reflect low hygiene considerations in breastfeeding mothers when giving food items to babies.

4.2.2 Child age

In addition, the variations across age groups could reflect breastfeeding stages of a child. The weaning time which is reported to pose more threats of diarrhoea attacks to a child [20], is around 17.6 months in Malawi [21] which is spanned in the observed high risk age 12-23 months. Further, low risk in age 0-11 months could reflect presence of exclusively and predominantly breastfed children in the data, who are reported to be at lower risk of diarrhoea than general breastfed children [2, 3, 22, 23, 24, 25]. The 2006 MICS reported that about 56% of children aged below 6 months were exclusively breastfed. Thus, age 0-11 months overlaps age 0-6 months which may explain the observed low risk in age 0-11 months.

Furthermore, the results have shown a shift of most risky age group to diarrhoea upward from 6-11

months reported in 2004 Malawi DHS or age 6-8 months for 1999 and 2003 Nigerian DHS, see [26], to age 12-23 months reported in this study. The shift seems to follow weaning time and move away from exclusive breastfeeding bracket. Reports indicate increased diarrhoea during and following weaning time among exclusively breastfed infants reportedly weaned at 6 months, see [27]. As per tradition, weaning entails introduction of complementary infant foods which may in turn spread diarrhoea to the child if not hygienically prepared by the mother. This is why researchers have recommended emphasis on hygienic preparation of weaning foods and water purification in resource-limited settings [28].

4.2.3 Toilet sharing

The findings suggest high risk in children from families that shared toilets than those whose families did not. This may reflect high possibilities of diarrhoea germs transmission from other people who use the same toilet as the child or her mother.

4.2.3 Region of stay

The causes for observed higher risk to diarrhoea in children from central and southern regions compared to those from northern region can be far from speculation, however the findings agree with those reported by [28] and those from 2004 Malawi DHS.

4.2.4 Staying along lake/river shore

The results suggest low risk in children living along Lakeshore or river banks compared to highlands. Although the opposite was expected, the results may reflect high utilization of water sanitation interventions rolled out by government and other stakeholders in recent years, such as free water guard in drinking water [29] and improved drinking water sources, such as piped water and boreholes, targeted to lakeshore/riverine dwellers who were previously believed to be at high risk of diarrhoea than highlanders. Thus, high use of safe and clean drinking water by residents of lakeshore or shire valley has reversed the old trend of child diarrhoea cases between highlands and lakeshore areas.

4.2.5 Mother education

The insignificance of mother's education is not a surprise; it may reflect the way the MICS study was designed, which just sought differences in academic qualification and not health education of mothers. Although other studies in sub-Saharan Africa have supported influence of mother's academic qualification on a child's risk of diarrhoea [26], there cannot be any immediate reason as to why one can think that mere differences in levels of academic or formal education achievements (other than health/hygiene education) can result in differences in children's risk to diarrhoea. What is felt to have an effect on a child's health is the mother's knowledge in hygiene, which is richly provided in the primary education curriculum in Malawi. But also, basic health education literacy of the mother that can be attained through attendance of antenatal or postnatal care services and not necessarily formal or professional education.

5. Conclusions

The findings suggest that estimating a child's risk to diarrhoea using Bayesian semi-parametric model is as good as using logistic or Poisson model. The two groups of models have isolated same significant as well as insignificant factors for determining the child's risk to diarrhoea. In addition, Poisson model is an appropriate assumption for the 2006 MICS diarrhoea dataset. It can further be concluded that region from which a child comes, child's age, whether or not a child is still breastfeeding, whether or not a child comes from a family that shares toilet with another family, and closeness to lake/river are statistically significant factors for determining the child's risk to diarrhoea. However, under-five child diarrhoea has little (if any) to do with area of residence, source of drinking water, and mother's education.

Recommendations

We recommend fitting Bayesian semi-parametric

model in parallel with classical models as a checking tool for classical model assumptions.

More interventions in child diarrhoea are needed in central region of the country by government and other stakeholders in health in order to contain the problem in the region.

Malawi's Ministry of Health and other stakeholders should continue mobilizing for high hygiene practices in breastfeeding mothers in the country, especially around weaning period.

Another study is needed to find causes of high risk to diarrhoea in children from central region of Malawi.

Acknowledgements

References

[1] K.D. Mwambete, R. Joseph, Knowledge and perception of mothers and caregivers on childhood diarrhoea and its management in Temeke Municipality, Tanzania, Tanzania Journal of Health Research 12 (1) (2010).

[2] C.G. Victora, S.R. Huttly, S.C. Fuchs, et al, Deaths due to dysentery, acute and persistent diarrhoea among Brazilian infants, Acta Paediatr suppl 381 (1992) 7-11.

[3] WHO, Collaborative study team on the role of breastfeeding on the prevention of infant mortality, effect of breastfeeding on infant and child mortality due to infectious diseases in less developed countries: A pooled analysis, Lancet 355 (2000) 451-5.

[4] WHO, UNICEF, UNFPA & World Bank, Maternal and infant mortality estimates, 2005, [Online] (2007), http://www.who.int/making-pregnancy-safer/topics/maternal-mortality/en/ (Accessed Jul. 26, 2011).

[5] Malawi Demographic and Health Survey, 2004, National Statistical Office of Malawi, Calverton, MD, ORC Macro, 2005.

[6] PATH, Charting the course for integrated diarrhea control in Malawi: A way forward for policy change, Washington, DC, 20001, USA, 2011.

[7] H.D, Hamer, J. Simon, D. Thea, G.T. Keusch, Childhood diarrhoea in sub-Saharan Africa, Child Health Research Project Special Report, April 1998.

[8] A.C. Munthali, Change and continuity in the management of diarrhoeal diseases in under-five children in rural Malawi, Malawi Med Journal 16 (2) (2005) 43-46.

[9] A.J. Dobson, An Introduction to Generalized Linear Models, (2nd ed.), Chapman & Hall/CRC, New York, 2002.

[10] A. Jerak, S. Wagner, Estimating probabilities of EPO patent oppositions in a Bayesian semiparametric regression framework. Department of Statistics, Munich: University of Munich, 2003.

[11] N.B. Kandala, et al, Semiparametric analysis of the socio-demographic determinants of under nutrition in two African countries, Research in Official Statistics, EUROSTAT 4 (1) (2001) 81-100.

[12] A. Gelman, Multilevel (hierarchical) modelling: what it can and can't do, Columbia University, New York, 2005.

[13] A. Brezger, T. Kneib, S. Lang, BayesX: Analyzing Bayesian structured additive regression models, Journal of Statistical Software, 14(11) (2005).

[14] C.J. Skinner, D. Holt, T.M.F. Smith, Analysis of complex surveys, Wiley, Chichester, 1989.

[15] Population and Housing Census, 2008 National Statistical Office of Malawi, Zomba, 2008.

[16] Index Mundi, Malawi age structure-demographics [Online], 2011, http://www.indexmundi.com, (Accessed Feb. 10, 2012).

[17] Monitoring the Situation of Children and Women: Malawi Multiple Indicator Cluster Survey 2006, FINAL REPORT, National Statistical Office of Malawi & UNICEF-Malawi, Lilongwe & Zomba, Malawi, 2008.

[18] T. Mesele, Bayesian approach to identify predictors of children nutritional status in Ethiopia, Master's Thesis, Unpublished, Addis Ababa University, Addis Ababa, 2009.

[19] A. Ferreira da Silva, cudaBayesreg: Bayesian computation in CUDA, The R Journal 2(2) (2010).

[20] R. B. Kerr, P. R. Berti, & M. Chirwa, Breastfeeding and mixed feeding practices in Malawi: timing, reasons, decision makers, and child health consequences, Food and Nutrition Bulletin 28(1) (2007).

[21] L. N. Kazembe, Spatial Modelling of Initiation and Duration of Breastfeeding: Analysis of Breastfeeding Behaviour in Malawi – I, World Health Population. 10(3) (2008) 14-31.

[22] S. Arifeen, R. E. Black, G. Antelman, A. Baqui, L. Caulfield, S. Becker, Exclusive breastfeeding reduces acute respiratory infection and diarrhoea deaths among infants in Dhaka slums, Pediatr, 108 (2001) 67.

[23] A. P. Betran, M. Onis, J. A. Lauer, J. Villar, Ecological study of effects of breastfeeding on infant mortality in Latin America, Br Med J 323 (2001) 1-5.

[24] L. A. Hanson, R. Ashraf, S. Zaman, et al, Breast feeding is a natural contraceptive and prevents disease and death in infants, linking infant mortality and birth rates, Acta Paediatr *83* (1994) 3-6.

[25] P. W. Yoon, R. E. Black, L. H. Moulton, S. Becker, Effects of not breastfeeding on the risk of diarrhoea and respiratory mortality in children under 2 years of age in Metro Cebu, The Philippines, American Journal of Epidemiology 143 (1996) 1142-1148.

[26] N. B. Kandala, C. Ji, N. Stallard, S. Stranges, F. P. Cappuccio, Morbidity from diarrhoea, cough and fever among young children in Nigeria, Annals of Tropical Medicine & Parasitology, 102(5) (2008) 427-445.

[27] P. Clayden, Diarrhoea in uninfected infants of HIV-positive mothers who stop breastfeeding at 6 Months, HIV i-Base (2007).

[28] L. N. Kazembe, L. S. Muula, C. Simoonga, Joint spatial modelling of common morbidities of childhood fever and diarrhoea in Malawi, Health & Place 15(1) (2009) 165-172.

[29] S. Kumwenda, Assessment of water guard use at household level in Chikwawa district, Masters' Thesis, Unpublished, College of Medicine, University of Malawi, 2009.