

RESEARCH

Open Access



Application of longitudinal multilevel zero inflated Poisson regression in modeling of infectious diseases among infants in Ethiopia

Bezalem Eshetu Yirdaw^{1*}, Legesse Kassa Debusho¹ and Aregash Samuel²

Abstract

Background In sub-Saharan African countries, preventable and manageable diseases such as diarrhea and acute respiratory infections still claim the lives of children. Hence, this study aims to estimate the rate of change in the log expected number of days a child suffers from Diarrhea (NOD) and flu/common cold (NOF) among children aged 6 to 11 months at the baseline of the study.

Methodology This study used secondary data which exhibit a longitudinal and multilevel structure. Based on the results of exploratory analysis, a multilevel zero-inflated Poisson regression model with a rate of change in the log expected NOD and NOF described by a quadratic trend was proposed to efficiently analyze both outcomes accounting for correlation between observations and individuals through random effects. Furthermore, residual plots were used to assess the goodness of fit of the model.

Results Considering subject and cluster-specific random effects, the results revealed a quadratic trend in the rate of change of the log expected NOD. Initially, low dose iron Micronutrient Powder (MNP) users exhibited a higher rate of change compared to non-users, but this trend reversed over time. Similarly, the log expected NOF decreased for children who used MNP and exclusively breastfed for six months, in comparison to their counterparts. In addition, the odds of not having flu decreased with each two-week increment for MNP users, as compared to non-MNP users. Furthermore, an increase in NOD resulted in an increase in the log expected NOF. Region and exclusive breastfeeding also have a significant relationships with both NOD and NOF.

Conclusion The findings of this study underscore the importance of commencing analysis of data generated from a study with exploratory analysis. The study highlights the critical role of promoting EBF for the first six months and supporting children with additional food after six months to reduce the burden of infectious diseases.

Keywords Infectious disease, Longitudinal analysis, Multilevel zero inflated Poisson regression, Quadratic trend, Infants

Introduction

Promoting well-being and ensuring healthy lives for children lies at the heart of the United Nations Sustainable Development Goals (SDGs). While considerable progress has been made in several nations towards meeting these SDGs by the year 2020, it is important to note that Ethiopia continues to rank among the top five countries responsible for nearly half of the world's under-five

*Correspondence:

Bezalem Eshetu Yirdaw
12962805@mylife.unisa.ac.za

¹ Department of Statistics, University of South Africa, c/o Christiaan de Wet Road & Pioneer Avenue, Johannesburg 1709, Florida, South Africa

² Ethiopian Public Health Institute, Gulele Sub City, Addis Ababa, Ethiopia



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

mortality rates [1]. Recent report from the World Health Organization (WHO) illustrate infectious diseases such as diarrhea and acute respiratory infections (ARIs) are among the leading causes of under-five mortality [1]. Approximately 9% of all deaths among children under the age of five are attributed to diarrhea. Notably, the prevalence of under-five years deaths due to diarrhea is particularly elevated in sub-Saharan African countries [2]. Furthermore, as reported by the WHO in 2019, ARI was one of the predominant childhood diseases, ranking among the top seven common causes of death in children and bearing a notably elevated morbidity rate [3]. The incidences of ARI and diarrheal diseases are high during the first two years of a child's life, casting a shadow over their physical growth. This early-life challenge can potentially lead to adverse health outcomes in adulthood [4, 5]. A recent study in 2021 shows that the prevalence of diarrhea in Ethiopia was 17%, indicating that this issue still requires attention and intervention [6].

Researchers worldwide have conducted extensive studies on the risk factors associated with diarrhea and ARI. A study from Ethiopia showed that child age, drinking water source, family size and exclusive breastfeeding (EBF) status of a child directly associated with childhood diarrhea [6]. In addition, respiratory infections in children under the age of five years are associated with a range of factors, including demographics (education and employment status of the mother), socioeconomic conditions, nutrition status, health-related aspects, and environmental influences [7]. Studies conducted across different regions have shown that childhood diarrhea is associated with various factors, including the socio-economic status of a household [8, 9], EBF [10, 11], education status of a caregiver [9, 10, 12] sanitation [13], age of a mother [12], sex of a child, age of a child [9], water source [14], family size [13] and nutrition status of a child [10]. On the other hand, a study from Pakistan reveals that diarrhea, EBF and gender are significantly associated with ARI [15]. A previous study shows the simultaneous occurrence of childhood diarrhea and ARI [16]. Furthermore, a study from Nepal revealed a causal relationship between diarrhea and ARI [17].

Although the Ethiopian Ministry of Health and regional health offices have prioritized the enhancement of a child health, a considerable number of children continue to lose their lives in Ethiopia due to preventable and manageable cases of diseases such as diarrhea and ARI. Hence, identifying the underlying factors behind diarrhea and ARI has a paramount importance in achieving the SDGs and advancing a nation's development, because healthy children can subsequently play dynamic roles in their communities, contributing in various ways. Hence, the primary objective of this study is to estimate the rate

of change in the expected log number of days a child suffered from diarrhea (NOD) and number of days a child suffered from flu/common cold (NOF) and identify the associated risk factors among young children in Ethiopia.

A previous study investigated the longitudinal prevalence of diarrhea and flu/common cold and the impact of the intervention, low dose iron Micronutrient Powder (MNP), on the prevalence rates [18]. The study used a generalized Poisson linear mixed effects model, and reached to the conclusion that MNP usage increased the longitudinal prevalence of both diarrhea and flu [18]. On the other hand, a previous study shows that MNP usage did not elevate the risk of infectious diseases like diarrhea and ARI [19]. Another study also suggested that supplementing children with micronutrient, including vitamin A and zinc, could reduce the severity of infectious diseases [20]. These conflicting results have prompted our decision to start the analysis from exploratory analysis. Therefore, our study aims to investigate the impact of using exploratory analysis for model specification.

Various studies have highlighted the positive impacts of EBF in reducing the risk of infectious diseases [10, 11, 15, 21, 22]. Another study provided evidence of a reduced risk of infectious diseases among infants who breastfed exclusively for six months [23]. Furthermore, a study indicated that the use of MNP, which contains vitamin A and zinc, could mitigate the severity of infectious diseases [20]. Therefore, according to the findings of [20, 23] studying the interaction of EBF month \times MNP usage on NOD and NOF can help policy makers in formulating guidelines and recommendations for infant feeding practices. Therefore, this study explore the effect of the interaction of EBF month \times MNP usage on NOD and NOF.

Methodology

Study data

In this study, we used secondary data originally collected to evaluate the effectiveness of MNP in improving child morbidity and growth among young children in Ethiopia [18]. The data were collected from the Oromia and Southern Nations, Nationalities, and Peoples' Region (SNNPR) of Ethiopia, spanning from March 2015 to May 2016. The dataset displays a hierarchical structure, with observations taken from each individual at two-week intervals throughout the data collection period. Data were collected from 2356 children from 35 villages (clusters) at baseline, and after excluding children with only one measurement, we analyzed data from 2283 children. The dataset comprises three main components: child morbidity, anthropometric measurements, and the iron status of the children. For this study, our focus is mainly on child morbidity.

The outcome variables, diarrhea and flu, were assessed every two weeks and a total of 18 measurements for each individual were used for this study. Thus, the number of days a child suffered from diarrhea (NOD) and the number of days a child suffered from flu/commen cold (NOF) within the two weeks period was counted for each of the 18 observation times. Furthermore, the covariates included were MNP usage, EBF months, gender, baseline age of a child, region, wealth index, age of a mother (AOM) and educational status of a mother (ESM). According to WHO, a child should breastfed for the first six months of life and should continue breastfeeding along with nutritionally adequate and safe complementary (solid/liquid) foods after six months [24]. Thus, EBF months were classified as "0" if a child's EBF months were less than or greater than six months and "1" if a child's EBF months were equal to six months. Further, wealth index were calculated using the principal component analysis using stata version 14 by considering family assets, number of domestic animals, family size, type of toilet and water source. Further insights into the sampling procedure can be found in Samuel et al's work [18]. In addition, it is important to note that the study specifically considered children aged 6 to 11 months at the baseline assessment.

Methods

A count outcome can be modeled using the generalised Poisson (GP), negative binomial (NB) or zero inflated Poisson (ZIP) regression models. The GP model mainly used under the assumption of mean equals to variance or when there is no overdispersion and underdispersion. The NB and the ZIP regression models can be used when the data exhibit overdispersion and excessive zeros that go beyond what is expected from the Poisson distribution, respectively. However, it is always advisable to start from the simple GP model. The probability mass function for the Poisson random variable Y_t with parameter λ_t is given by

$$p(Y_t = y_t) = \frac{\lambda_t^{y_t} e^{-\lambda_t}}{y_t!}, \quad t = 0, 1, 2, \dots \tag{1}$$

Thus, the GP model with p covariates is given by

$$\log(\lambda_t) = \eta_t = \beta_0 + \beta_1 x_{t1} + \dots + \beta_p x_{tp}, \tag{2}$$

where $\beta_0, \beta_1, \dots, \beta_p$ are regression coefficients. However, the data in the current study exhibit a hierarchical structure, with individuals nested within cluster measured

multiple times. This structure may violate the assumption of independence in a GP model, as measurements within the same individual or cluster are likely to be correlated. The multilevel generalised Poisson (MGP) regression model is a model which account this correlation. Thus, a subject-level random effect was introduced first to capture the subject specific correlation. Subsequently, a cluster-level random effect was added to account cluster specific correlation. Let the count outcome $Y_{tli} (i = 1, 2, \dots, m; l = 1, 2, \dots, n_i; t_i = 1, 2, \dots, n_{ij})$ represents NOD or NOF of a child which represent the t^{th} observation of l^{th} individual in the i^{th} cluster, the probability mass function for the random variable Y_{tli} with parameter λ_{tli} is given by

$$p(Y_{tli} = y_{tli} | u_{li}, v_i) = \frac{\lambda_{tli}^{y_{tli}} \exp^{-\lambda_{tli}}}{Y_{tli}!}, \quad y_{tli} = 0, 1, 2, \dots, \tag{3}$$

where u_{li} and v_i represents the subject- and cluster-specific random effects. Thus, the MGP model for the outcome Y_{tli} with parameter λ_{tli} for p covariates is given by

$$\log(\lambda_{tli}) = \eta_{tli} = \beta_0 + \beta_1 time_{tli} + x'_{tli} \beta + z'_{tli} u_{li} + z'_{tli} v_i, \tag{4}$$

where $time_{tli}$ is the time points in which individuals were measured and $z'_{tli} = [1, time_{tli}]$. In addition, $u_{li} \sim \mathcal{N}(0, \Sigma(\theta_u))$ and $v_i \sim \mathcal{N}(0, \Sigma(\theta_v))$, where $\Sigma(\theta_u)$ and $\Sigma(\theta_v)$ represents a symmetric and positive semi-definite variance-covariance matrix of random effects u_{li} and v_i parameterised by vector of variance-covariance components θ_u and θ_v , respectively. Furthermore, the outcomes NOD and NOF in the current study exhibit excess zeros (see Fig. 1). The Negative Binomial model is often a good starting point for modeling overdispersed count data, including datasets with more zeros than would be expected under a Poisson model. The probability mass function for the random variable Y_t from NB distribution is given by

$$p(Y_t = y_t) = \frac{\Gamma(y_t + \phi_t)}{\Gamma(\phi_t)\Gamma(y_t + 1)} \left(\frac{\phi_t}{\phi_t + \lambda_t}\right)^{\phi_t} \left(\frac{\lambda_t}{\phi_t + \lambda_t}\right)^{y_t}, \tag{5}$$

where ϕ_t and λ_t are dispersion parameter and average count of the outcome Y_t , respectively. However, similar to the GP model in Expression 2, NB model may also fail to account the dependency within subject at the subject-level and the dependency within cluster at the cluster-level. Let $Y_{tli} | u_{li}, v_i \sim NB(\lambda_{tli}, \phi_{tli})$, the probability mass function of the random variable Y_{tli} is given by

$$p(Y_{tli} = y_{tli} | u_{li}, v_i) = \frac{\Gamma(y_{tli} + \phi_{tli})}{\Gamma(\phi_{tli})\Gamma(y_{tli} + 1)} \left(\frac{\phi_{tli}}{\phi_{tli} + \lambda_{tli}}\right)^{\phi_{tli}} \left(\frac{\lambda_{tli}}{\phi_{tli} + \lambda_{tli}}\right)^{y_{tli}} \tag{6}$$

where ϕ_{tli} and λ_{tli} are dispersion parameter and average count of the outcome. Hence, the multilevel negative binomial model for the outcome Y_{tli} is given by

$$\text{and } \begin{aligned} \log\left(\frac{\phi_{tli}}{1-\phi_{tli}}\right) &= \xi_{tli} = \alpha_0 + \alpha_1 \text{time}_{tli} + a'_{tli} \alpha + r_{li} + s_i \\ \log(\lambda_{tli}) &= \eta_{tli} = \beta_0 + \beta_1 \text{time}_{tli} + x'_{tli} \beta + z'_{tli} u_{li} + z'_{tli} v_i, \end{aligned} \tag{9}$$

$$\log(\lambda_{tli}) = \eta_{tli} = \beta_0 + \beta_1 \text{time}_{tli} + x'_{tli} \beta + z'_{tli} u_{li} + z'_{tli} v_i, \tag{7}$$

where the distributional assumption for the random effects u_{li} and v_i is similar to the assumption of MGP model in Expression 4. The NB regression model offer a better fit when there is overdispersion, however, it may fail to adequately address the substantial number of zero counts observed (see Fig. 1). The ZIP model is specifically designed to account for the overabundance of zeros possessed by a count outcome. The Zero-Inflated Poisson (ZIP) regression model was built under the assumption that the data originates from two distinct processes. One process generates the zero counts, while the other process generates both the zero and non-zero counts. Consequently, the ZIP model combines logistic regression and Poisson regression to model the zero part based on the presence or absence of the outcome, and its count part, respectively. However, the ZIP model may still fail to account the correlation between measurements within the same individual or cluster. The multilevel zero inflated Poisson (MZIP) regression model is an extension of ZIP model which can be used in the case of correlated observations. According to [25] the probability mass function of an outcome Y_{tli} from a ZIP distribution with the average count λ_{tli} and probability of extra zeros ϕ_{tli} is given by

$$p(Y_{tli} = y_{tli} | r_{li}, s_i, u_{li}, v_i) = \begin{cases} \phi_{tli} + (1 - \phi_{tli})e^{-\lambda_{tli}}, & \text{if } y_{tli} = 0 \\ (1 - \phi_{tli}) \frac{\lambda_{tli}^{y_{tli}} e^{-\lambda_{tli}}}{Y_{tli}!}, & y_{tli} = 1, 2, \dots \end{cases} \tag{8}$$

In the MZIP models, the fixed effect covariates for the zero and count part of the model are not necessarily the same [26]. For this study, we have assumed a correlation between random intercept and random slope at the Poisson component of the model. However, based on the parsimonious principle, we assumed no correlation between the random effects in the Poisson and logistic components of the model. Let the count outcome Y_{tli} represents NOD or NOF of a child, where Y_{tli} represent the t^{th} observation of i^{th} individual in the i^{th} cluster, and m, n_i and n_{ij} represent total number of clusters, total number of individuals in the i^{th} cluster and total number of observations for the i^{th} individual in the i^{th} cluster, respectively. Hence, the linear predictors for the zero and non-zero part of the model which linked to the logit and log link functions are given by

respectively, where α_0 and β_0 are the respective baseline log odds and log expected of the outcome at the logistic and Poisson components of the model, after keeping the other covariates in the model constant. In addition, a_{tli} and x_{tli} are a vector of fixed effect covariates present in the logistic and Poisson components of the model, accompanied by their respective regression coefficients α and β . Furthermore, r_{li} and s_i are the subject and cluster specific random effects at the logistic component of the model. Further, $r_{li} \sim \mathcal{N}(0, \sigma_{r_0})$, $s_i \sim \mathcal{N}(0, \sigma_{s_0})$, $u_{li} \sim \mathcal{N}(0, \Sigma(\theta_u))$, and $v_i \sim \mathcal{N}(0, \Sigma(\theta_v))$, where σ_{r_0} and σ_{s_0} are variances of subject specific and cluster specific random intercepts in the logistic part of the model. While $\Sigma(\theta_u)$ and $\Sigma(\theta_v)$ are symmetric and positive semi-definite variance-covariance matrix of random effects in the Poisson part of the model parameterised by vector of variance-covariance components θ_u and θ_v , respectively. For this study, the collection of parameters $\Theta = (\alpha^t, \beta^t, \sigma_{r_0}, \sigma_{s_0}, \theta_u, \theta_v)$ were estimated using the maximum likelihood estimation method, where α^t and β^t are vector of regression coefficients including the regression coefficient of time_{tli} at the logistic and Poisson components of the model, respectively.

Model selection

Let Y represents a vector of discrete count outcome. When modeling count data, one of the suitable methods to be employed is a Poisson regression. However, count data sometimes exhibit overdispersion and excessive zeros, that go beyond what is expected from the Poisson distribution. For instance, the response variables NOD and NOF in this study had many zeros (Fig. 1). In cases like this, a standard Poisson model may not be the most appropriate choice for further analysis. Therefore, it is advisable to begin with a generalized Poisson (GP) regression model and assess whether this model can effectively handle the observed excess zeros. Following the approach outlined by [27, 28], we utilized quantile residuals rather than ordinary residuals to select the correct family distribution. This was accomplished via simulation, similar to the principles of Bayesian p -values and parametric bootstrapping,

which converts residuals to scaled residuals using the |DHARMA| package in |R| [28]. Thus, if the fitted model is accurate, the scaled/quantile residuals simulated from the fitted model should conform to a uniform [0,1] distribution. In addition, it is advisable to perform a dispersion test to test overdispersion or underdispersion in the model [28]. Hence, we initially employed the GP model, transitioned to the negative binomial model, and ultimately arrived at a zero-inflated Poisson regression model (see Exploratory analysis section for further detail). Furthermore, Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) were considered to further validate the model selection. These provided additional evidence that supported the superiority of the selected model by balancing the model fit with complexity.

Having the zero-inflated Poisson regression model in mind, it is essential to recognize that the data in this study exhibits a longitudinal and hierarchical structure. Furthermore, the individual and cluster profile plots in Fig. 4 show variations in both intercepts and slopes. We employed a likelihood ratio test to examine zero variance components. This test is based on the hypothesis $H_0 : \sigma^2 = 0$ versus $H_1 : \sigma^2 > 0$ and follows an asymptotic mixture distribution of $0.5\chi_0^2 + 0.5\chi_1^2$, where σ^2 is a variance of subject specific random intercept or slope, a covariance between random intercept and random slope at individual level or variance of cluster specific random intercept or slope and covariance between random intercept and random slope at cluster level [29].

Result

Exploratory analysis

This study considered two outcomes, namely NOD and NOF. According to Walker et al. (2013), the joint modelling of diarrhea and ARI would sound appropriate. However, upon reviewing the data, we found that the

correlation between the two outcomes is relatively low (< 0.2). In addition, the number of times that a child was suffering from flu and diarrhea at the same time were limited. This sparse nature of the data makes fitting the joint modelling of the two outcomes difficult. Given the hierarchical nature of the data, the estimation of random effects at the individual and cluster levels for the joint modeling of the two outcomes becomes problematic due to the insufficient number of observations in these cases. Furthermore, given the low correlation between the two outcomes, a joint model may not capture meaningful associations and could lead to unstable estimates. On the other hand, Newman et al., (2020), revealed a causal relationship between diarrhea and ARI among infants. Hence, including this relationship in a joint model could complicate the analysis without providing additional insight. Therefore, this study considered a separate analysis of NOD and NOF.

The outcomes in this study exhibit a count nature, and as demonstrated by the histogram in Fig. 1, both outcomes namely NOD and NOF, exhibit numerous zero. Various studies have indicated that a generalized Poisson linear mixed effects model may not be suitable for such data types [26, 30]. However, the presence of many zeros does not necessarily indicate zero-inflation, it could be explained by the inclusion of explanatory variables. Therefore, it is advisable to begin with a generalized Poisson (GP) regression model that incorporates all fixed-effect covariates and investigate whether the model can account for the observed excess zeros in the data. Furthermore, the plots in Fig. 2 show that the rate of change for both outcomes follows a half-inverted U shape emphasizing the importance of considering quadratic time effect. Furthermore, the plot for NOD shows that the two lines intersected at the end of the study (Fig. 2a). Hence, it is crucial to account for interactions $time \times MNP$ usage and $time^2 \times MNP$ usage to assess

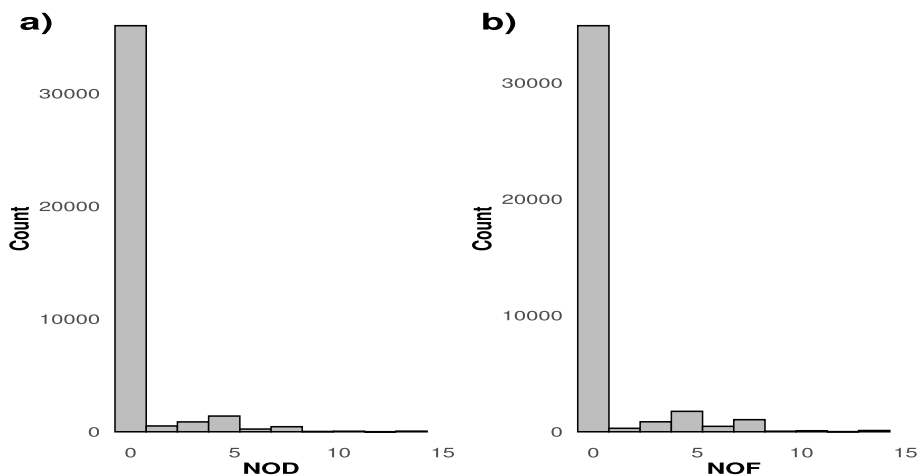


Fig. 1 A histogram showing the observed frequency distribution for the number of days a child suffered from diarrhea and flu/common cold

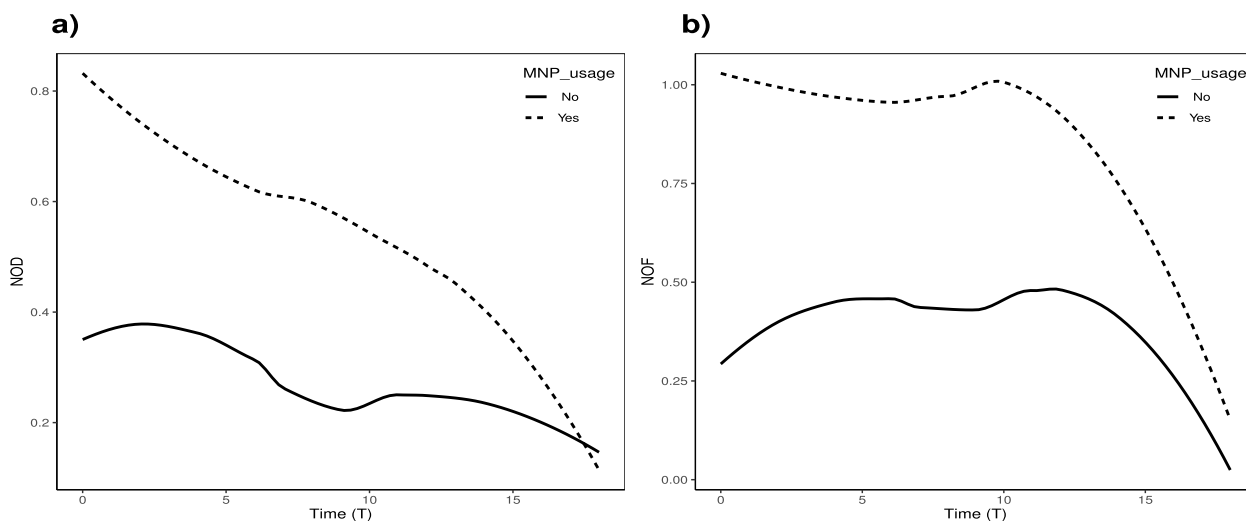


Fig. 2 An interaction plot showing the relationship between MNP usage and outcomes over time: (a) NOD, (b) NOF

the rate of change in the log expected NOD and NOF for children who used MNP compared to those who did not.

Thus, the GP model for NOD (M_{D1}) is given by

$$\begin{aligned} \eta_d = & \beta_0 + \beta_1 \text{time} + \beta_2 \text{time}^2 + \beta_3 \text{MNP usage} \\ & + \beta_4 \text{MNP usage} \times \text{time} + \beta_5 \text{MNP usage} \times \text{time}^2 \\ & + \beta_6 \text{EBF months} + \beta_7 \text{EBF months} \times \text{MNP usage} \\ & + \beta_8 \text{Age} + \beta_9 \text{gender} + \beta_{10} \text{region} + \beta_{11} \text{AOM} \\ & + \beta_{12} \text{ESM} + \beta_{13} \text{wealth index} \\ = & \mathbf{X}\boldsymbol{\beta}, \end{aligned} \tag{10}$$

where η_d represents the linear predictor, *time* is the time points in which the observations were recorded, *MNP usage* represents whether or not a child used MNP, *EBF months* is exclusive breastfeeding months of a child, *Age* is age of a child, *AOM* is age of a mother and *ESM* is educational status of a mother. However, the model validation test for the GP model for NOD does not indicate a good fit (see Table 1 and Fig. 3a). Specifically, the

zero inflation test appeared significant suggesting that the model failed to account for the observed excess zeros in the data ($p - \text{value} < 0.001$). Furthermore, the overdispersion parameter is far greater than one and highly significant. Moreover, the QQ-plot in Fig. 3a reveals substantial deviations from the expected distribution.

On the other hand, the Negative Binomial (NB) regression model is a well-suited model for count data that exhibit overdispersion compared to what would be expected from the GP model. Therefore, we considered a NB regression model (M_{D2}) with the same mean structure as in Expression 10. While this model effectively handles the excessive zeros in the observed data, a dispersion test indicates a highly significant underdispersion problem (see Table 1). In addition, the QQ plot in Fig. 3b confirms deviations from the expected distribution. Another viable option that can effectively capture the characteristics of the data is a Zero-Inflated Poisson (ZIP) Regression model (M_{D3}). Even though the dispersion test remains significant, the dispersion parameter is very close to one. Furthermore, the QQ plot in Fig. 3c shows no deviations from the expected distribution, indicating a good fit compared to M_{D1} and M_{D2} . Moreover, the values of AIC and BIC for M_{D3} are significantly smaller than those for M_{D1} and M_{D2} .

Similarly, for the outcome NOF we started from the GP model (M_{F1}) (see Expression 11) and followed the same procedure as we did for NOD. However, the zero-inflation and overdispersion tests were significant for the GP model (Table 1). In addition, the QQ plot in Fig. 3d shows deviation from the expected distribution. Furthermore, the NB regression model M_{F2} did not fit the data very well (see Table 1 & Fig. 3e). Conversely, the ZIP model (M_{F3}) demonstrates a good fit as compared to M_{F1} and M_{F2} (Table 1 & Fig. 3f).

Table 1 Model validation tests for the outcomes NOD and NOF

Models	Zero-inflation test	Dispersion test	AIC	BIC
M_{D1}	1.34**	5.00**	85017.1	85179.6
M_{D2}	1.00	0.47**	41136.3	41307.3
M_{D3}	1.00	1.05*	38127.1	38298.0
M_{F1}	1.52**	5.31**	109738.5	109900.9
M_{F2}	1.01*	0.43**	51778.6	51738.6
M_{F3}	1.00	1.04*	47520.1	47691.1

** $p - \text{value} < 0.001$; * $p - \text{value} < 0.05$

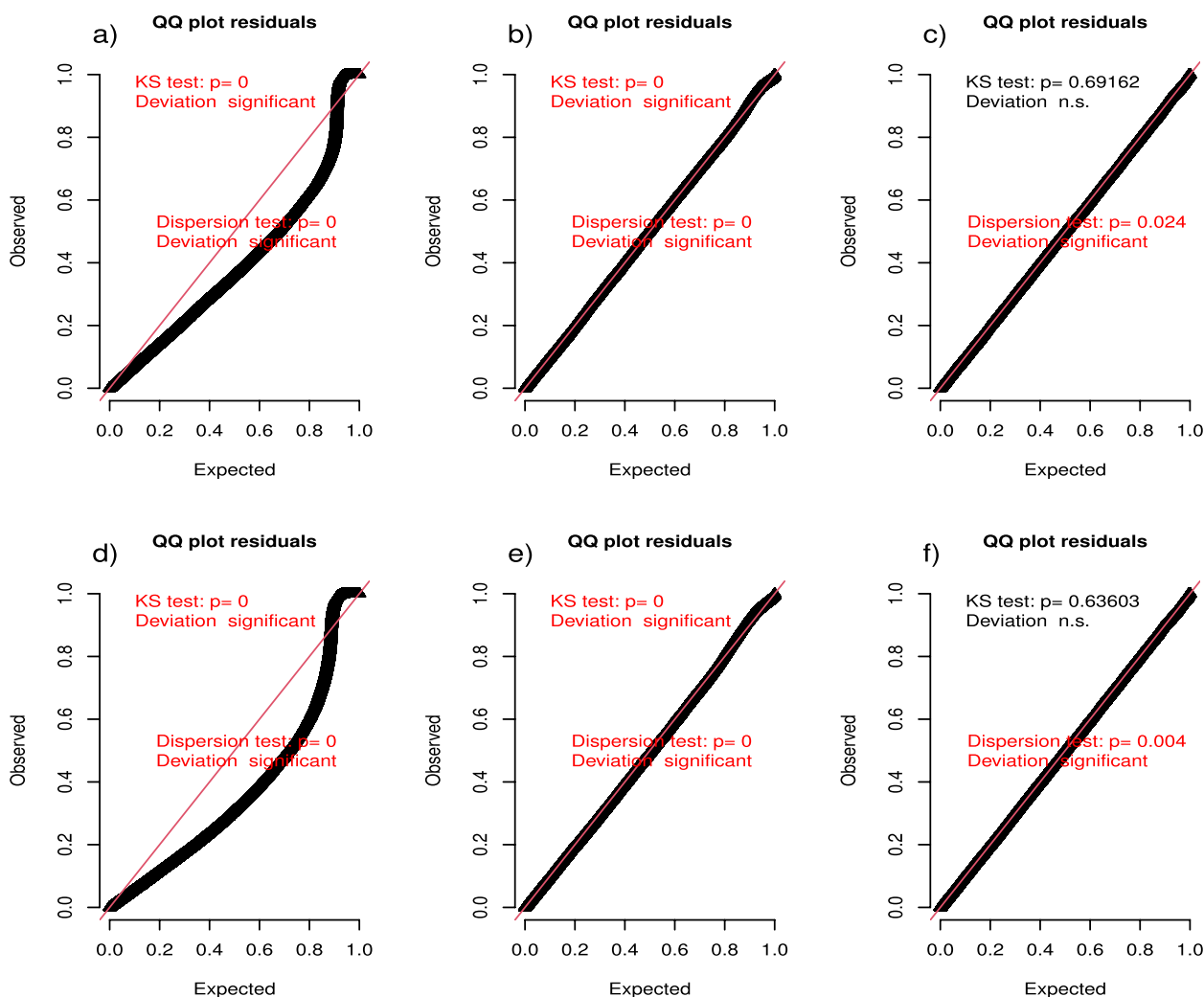


Fig. 3 Residual plots of: (a) GP model for NOD, (b) NB model for NOD, (c) ZIP model for NOD, (d) GP model for NOF, (e), NB model for NOF, (f) ZIP model for NOF

$$\begin{aligned}
 \eta_f &= \beta_0 + \beta_1 \text{time} + \beta_2 \text{time}^2 + \beta_3 \text{MNP usage} \\
 &+ \beta_4 \text{MNP usage} \times \text{time} + \beta_5 \text{MNP usage} \times \text{time}^2 \\
 &+ \beta_6 \text{EBF months} + \beta_7 \text{EBF months} \times \text{MNP usage} \\
 &+ \beta_8 \text{Age} + \beta_9 \text{sex} + \beta_{10} \text{region} + \beta_{11} \text{MOA} \\
 &+ \beta_{12} \text{MES} + \beta_{13} \text{wealth index} + \beta_{14} \text{NOD} \\
 &= X\beta,
 \end{aligned}
 \tag{11}$$

Moreover, the significant overdispersion problem observed in both NOD and NOF may be attributed to the variation due to subject specific and cluster specific random effects. For instance, the individual and cluster profile plots for NOD in Fig. 4a and b show the existence of substantial variations in both intercept and rate of change in NOD across time among the observed individuals and clusters, respectively. Similarly, the individual and cluster profile plots for NOF in Fig. 4c and d also demonstrate

variations in both intercept and slope across time among the observed individuals and clusters, respectively. To adequately address these variations, it is important to consider random intercepts and slopes at both the individual and cluster levels.

Selection of variance-covariance structure for random effects

In this section we conducted a mixture chi-square test to select a variance-covariance structure at both subject and cluster levels (Tables 2 & 3). Starting from a ZIP regression model with no random effects, eight models were fitted to test the variances of random effects and the covariance between the random intercept and random slope at each level of the hierarchy.

For the outcome NOD, the test for the subject specific random intercept ($H_0 : \sigma_{u_0} = 0$ against $H_1 : \sigma_{u_0} > 0$) yielded a

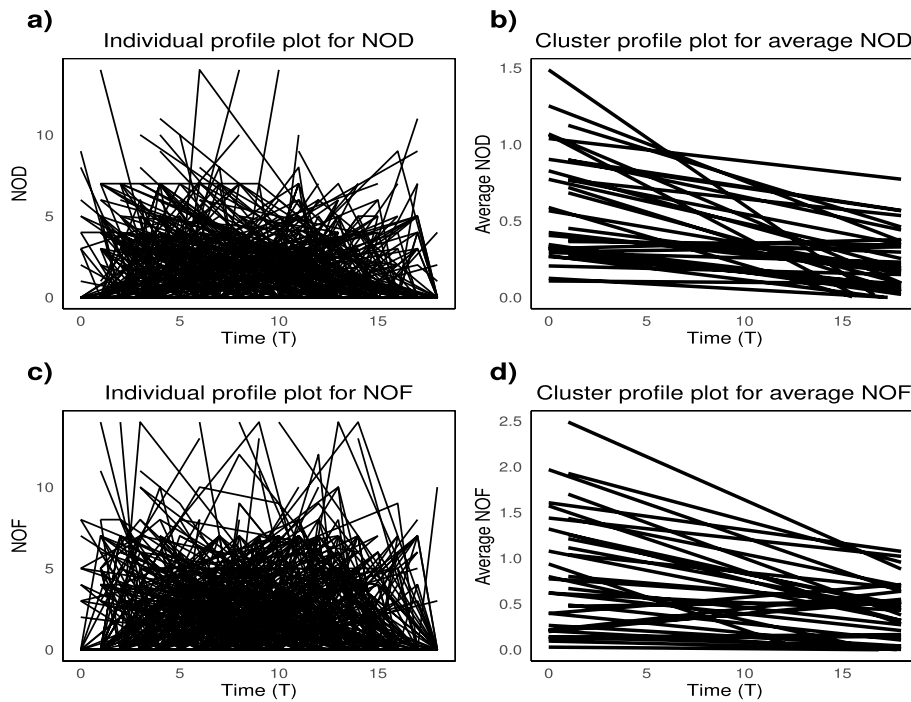


Fig. 4 Individual and cluster profile plots for NOD and NOF

Table 2 Longitudinal multilevel zero inflated Poisson regression models for selection of variance-covariance structure for NOD

Models	Poisson Part	Logistic part	-2log(Lik)	AIC	BIC	p-value
M ₁	$\eta = X\beta$	$\xi = A\alpha$		37327.7	37599.3	
M ₂	$\eta = X\beta + u_0$	$\xi = A\alpha$	44.6	37283.1	37575.2	< 0.001
M ₃	$\eta = X\beta + u_0 + v_0$	$\xi = A\alpha$	16.3	37268.8	37559.5	< 0.001
M ₄ , $\hat{\sigma}_{u_{01}} = 0$	$\eta = X\beta + u_0 + u_1 \text{ time} + v_0$	$\xi = A\alpha$	-31.3	37302.10	37601.3	1.00
M ₅ , $\hat{\sigma}_{v_{01}} = 0$	$\eta = X\beta + u_0 + v_0 + v_1 \text{ time}$	$\xi = A\alpha$	13.4	37257.4	37556.7	< 0.001
M ₆ , $\hat{\sigma}_{v_{01}} \neq 0$	$\eta = X\beta + u_0 + v_0 + v_1 \text{ time}$	$\xi = A\alpha$	2.1	37257.3	37565.1	0.250
M ₇ , $\hat{\sigma}_{v_{01}} = 0$	$\eta = X\beta + u_0 + v_0 + v_1 \text{ time}$	$\xi = A\alpha + r_0$	254.1	37005.3	37313.1	< 0.001
M ₈ , $\hat{\sigma}_{v_{01}} = 0$	$\eta = X\beta + u_0 + v_0 + v_1 \text{ time}$	$\xi = A\alpha + r_0 + s_0$	341.3	36666.0	36982.3	< 0.001

Table 3 Longitudinal multilevel zero inflated Poisson regression models for selection of variance-covariance structure for NOF

Models	Poisson Part	Logistic part	-2log(Lik)	AIC	BIC	p-value
M ₁	$\eta = X\beta$	$\xi = A\alpha$		45698.0	45954.5	
M ₂	$\eta = X\beta + u_0$	$\xi = A\alpha$	26.6	45673.4	45938.4	< 0.001
M ₃	$\eta = X\beta + u_0 + v_0$	$\xi = A\alpha$	45.8	45629.6	45903.2	< 0.001
M ₄ , $\hat{\sigma}_{u_{01}} = 0$	$\eta = X\beta + u_0 + u_0 \text{ time} + v_0$	$\xi = A\alpha$	0.00	45631.6	45913.8	1.000
M ₅ , $\hat{\sigma}_{v_{01}} = 0$	$\eta = X\beta + u_0 + v_0 + v_1 \text{ time}$	$\xi = A\alpha$	27.3	45604.3	45886.4	< 0.001
M ₆ , $\hat{\sigma}_{v_{01}} \neq 0$	$\eta = X\beta + u_0 + v_0 + v_1 \text{ time}$	$\xi = A\alpha$	30.7	45575.6	45866.3	< 0.001
M ₇ , $\hat{\sigma}_{v_{01}} \neq 0$	$\eta = X\beta + u_0 + v_0 + v_1 \text{ time}$	$\xi = A\alpha + r_0$	299.1	45278.5	45577.7	< 0.001
M ₈ , $\hat{\sigma}_{v_{01}} \neq 0$	$\eta = X\beta + u_0 + v_0 + v_1 \text{ time}$	$\xi = A\alpha + r_0 + s_0$	434.1	44846.5	45154.2	< 0.001

significant result ($-2\log(Lik) = 44.62, p - value < 0.001$), suggesting that subject specific random intercept should be included in the Poisson part of the model. Furthermore, the test also verified the significance of the variances of cluster-specific random intercept (σ_{v_0}) in the Poisson part of the model ($-2\log(Lik) = 16.27, p - value < 0.001$). However, including the subject-specific random slope was unnecessary, as the test ($H_0 : \sigma_{u_1} = 0$ against $H_1 : \sigma_{u_1} > 0$) did not appear significant (Table 2). To test cluster specific random slope, two models one with subject specific random intercept and cluster specific random intercept (M_3), and the other with subject specific random intercept, cluster specific random intercept and cluster specific random slope (M_5) were fitted and the test ($H_0 : \sigma_{v_1} = 0$ against $H_1 : \sigma_{v_1} > 0$) ensured that the inclusion of cluster specific random slope was crucial ($-2\log(Lik) = 13.37, p - value < 0.001$). However, the non-rejection of the null hypothesis $H_0 : \sigma_{v_{01}} = 0$ with $-2\log(Lik) = 2.09$ and $p - value = 0.250$ suggests that considering the covariance between the random intercept and slope at the cluster level of the model (M_6) was not necessary. Furthermore, the result from testing the hypothesis $H_0 : \sigma_{r_0} = 0$ against $H_1 : \sigma_{r_0} > 0$ and $H_0 : \sigma_{s_0} = 0$ against $H_1 : \sigma_{s_0} > 0$ in models M_7 and M_8 , respectively, confirmed that the inclusion of subject and cluster specific random intercepts at the logistic part of the model was important. Moreover, the model with the smaller AIC and BIC was M_8 , which support the LRT (Table 2).

Similar tests were conducted to evaluate the random effects for the outcome NOF. The test $H_0 : \sigma_{u_0} = 0$ against $H_1 : \sigma_{u_0} > 0$ in M_1 (no random intercept) yielded a test statistic value of 26.63 with a $p - value < 0.001$.

This provides compelling evidence against the null hypothesis H_0 , indicating significant subject-specific variability in the Poisson part of the model. Furthermore, the test $H_0 : \sigma_{v_0} = 0$ against $H_0 : \sigma_{v_0} > 0$ also confirmed a statistically significant difference among clusters ($-2\log(Lik) = 45.77, p - value < 0.001$). However, the subject-specific random slope was not statistically significant. To test the cluster-specific random slope, we considered models M_3 and M_5 with the hypothesis $H_0 : \sigma_{v_1} = 0$ against $H_1 : \sigma_{v_1} > 0$, and the test was highly significant ($-2\log(Lik) = 27.33, p - value < 0.001$), suggesting the inclusion of cluster-specific random slope. The test ($H_0 : \sigma_{v_{01}} = 0$ against $H_1 : \sigma_{v_{01}} > 0$) with $-2\log(Lik) = 30.66$ and $p - value < 0.001$ indicates a substantial improvement in the fit of M_6 compared to M_5 , favoring the model with an unstructured covariance structure at the cluster level over the model with a diagonal covariance structure. Similarly, the LRT, the AIC and BIC values supported the inclusion of both subject-specific and cluster-specific random intercepts in the logistic part of the model (Table 3). This leads to the conclusion that the next section will be based on model M_8 .

Multilevel zero inflated Poisson regression analysis

In this section, the results from the multivariable MZIP regression model are presented for both NOD and NOF (Tables 4 & 5). For the outcome NOD, considering subject and cluster specific variations, the covariates significant at a 5% level, including region, EBF month, time, MNP usage, time², time × MNP usage and time² × MNP usage in the Poisson part of the model, and gender, region, MNP usage, time and time × MNP usage

Table 4 Parameter estimates (standard errors, s.e) of multilevel zero inflated Poisson regression model for NOD

	Poisson part		Logistic part	
	Estimate (s.e)	p-value	Estimate (s.e)	p-value
Fixed Effects				
Intercept	1.530 (0.052)	< 0.001	2.456(0.172)	< 0.001
Gender, Female	-	-	0.084(0.042)	0.045
Region, SNNPR	0.060 (0.031)	0.054	-0.512(0.177)	0.004
EBF month, = 6	-0.138(0.029)	< 0.001	0.106(0.069)	0.124
MNP usage, Yes	0.026(0.053)	0.619	-0.765(0.185)	< 0.001
Time	-0.025(0.011)	0.027	0.059 (0.006)	< 0.001
Time ²	0.002(6 × 10 ⁻⁴)	0.006	-	-
Time × MNP usage, Yes	0.035(0.014)	0.013	0.020(0.007)	0.012
Time ² × MNP usage, Yes	-0.002 (8 × 10 ⁻⁴)	0.004	-	-
Variance components				
σ_0 (subject)	$\sigma_{u_0} = 0.130$		$\sigma_{r_0} = 0.432$	
σ_0 (cluster)	$\sigma_{v_0} = 0.052$		$\sigma_{s_0} = 0.502$	
σ_1 (cluster)	$\sigma_{v_1} = 0.007$		-	

Table 5 Parameter estimates (standard errors, s.e) of multilevel zero inflated Poisson regression model for NOF

	Poisson part		Logistic part	
	Estimate (s.e)	p-value	Estimate (s.e)	p-value
Fixed Effects				
Intercept	1.421 (0.048)	< 0.001	2.903(0.197)	< 0.001
Region, SNNPR	0.116 (0.026)	< 0.001	-1.193(0.215)	< 0.001
NOD	0.017 (0.003)	< 0.001	-	-
EBF month, = 6	0.034 (0.034)	0.315	0.141(0.062)	0.009
MNP usage, Yes	0.163 (0.048)	< 0.001	-1.068(0.220)	< 0.001
Time	0.003 (0.004)	0.429	0.043 (0.003)	< 0.001
EBF month, = 6 × MNP usage, Yes	-0.106(0.045)	0.019	-	-
Time × MNP usage, Yes	-	-	0.023(0.007)	< 0.001
Variance components				
$\sigma_0(\text{subject})$	$\sigma_{u_0} = 0.087$		$\sigma_{r_0} = 0.409$	
$\sigma_0(\text{cluster})$	$\sigma_{v_0} = 0.132$		$\sigma_{s_0} = 0.615$	
$\sigma_1(\text{cluster})$	$\sigma_{v_1} = 0.021$		-	
$\sigma_{01}(\text{cluster})$	$\rho_{v_{01}} = -0.940$		-	

in the logistic part of the model, were retained for further analysis. Keeping the effects of other covariates constant, the odds of not having diarrhea for girls were higher than that of boys ($\alpha = 0.084$, *s.e.* = 0.042, *p* – value = 0.004). In addition, the odds of not having diarrhea for children in the SNNPR region was lower than those from the Oromia region ($\alpha = -0.512$, *s.e.* = 0.177, *p* – value = 0.004). Furthermore, the log of expected NOD was 0.138 unit lower for children whose EBF months were six months compared to those with less than or greater than six months ($\beta = -0.138$, *s.e.* = 0.029, *p* – value < 0.001).

The interaction terms, Time × MNP usage and Time² × MNP usage, reveal non-monotonic changes in log expected NOD among children who used MNP as time progresses. The positive coefficient for Time × MNP usage ($\beta = 0.035$, *s.e.* = 0.014, *p* – value = 0.013) and the negative coefficient for Time² × MNP usage suggest an inverted U-shaped relationship between NOD and time, confirming the pattern seen in Fig. 2a in exploratory analysis. This indicates that the log of expected NOD increases by 0.035 for every two-week increment in time among MNP users compared to non-users which could be an initial adverse reaction to the MNP at the beginning of the study. However, as time progresses, the log of expected NOD decelerates for each two-week increment among MNP users compared to non-users. The interaction term Time × MNP usage in the logistic part of the model revealed that the odds of not having diarrhea for children who used MNP increased for two weeks increment in time ($\alpha = 0.020$, *s.e.* = 0.007, *p* – value = 0.012) (Table 4). These findings suggest that, even if the difference

is minimum due to various nuisance factors, the use of MNP contributed to the well-being of the children.

The results from MZIP regression model for the outcome NOF is presented in Table 5. The covariates significant at 5% level of significance were region, NOD, EBF month, MNP usage, time and EBF months × MNP usage in the Poisson part of the model, and region, EBF month, MNP usage, time and time × MNP usage at the logistic part of the model. After adjusting the effects of other covariates, subject specific and cluster specific random effects, we observed that the log of expected NOF was 0.116 unit higher for children living in the SNNPR region compared to children in the Oromia region ($\beta = 0.116$, *s.e.* = 0.026, *p* – value < 0.001). Furthermore, in the logistic part of the model, it was revealed that the odds of not having the flu were lower for children living in the SNNPR region as compared to children from Oromia region ($\alpha = -1.193$, *s.e.* = 0.215, *p* – value < 0.001).

The log expected NOF was expected to increase by 0.017 for a unit increase in NOD ($\beta = 0.017$, *s.e.* = 0.003, *p* – value < 0.001). Notably, the interaction term EBF month × MNP usage indicated that the log expected NOF was expected to decrease by 0.106 units for children who exclusively breastfed for six months and used MNP ($\beta = -0.106$, *s.e.* = 0.045, *p* – value = 0.019) as compared to children who did not used MNP and whose EBF months less or greater than six months. In addition, the odds of not having flu/common cold was higher for children who exclusively breastfed for six months as compared to the counterparts ($\alpha = 0.141$, *s.e.* = 0.062, *p* – value = 0.009). Moreover, the logistic part of the model in Table 5 reveals that the odds of not having the

flu/common cold increased as time increased by two weeks for children who used MNP as compared to those who did not ($\alpha = 0.023$, $s.e. = 0.007$, $p - value < 0.001$). These interpretations hold after adjusting subject and cluster specific random effects.

Model diagnostic

To assess the goodness of fit of the model for both NOD and NOF, we employed scaled/quantile residuals generated from the fitted models. As depicted in Fig. 5a and b, the QQ plots for NOD and NOF, respectively, show that the residuals align the straight line. This indicates that the residuals for both models follow a uniform distribution over the range [0, 1]. In addition, the dispersion and KS test appeared non significant revealing a good fit.

Discussion

This study primarily focused on the longitudinal analysis of NOD and NOF among young children drawn from 35 clusters within the SNNPR and Oromia regions, Ethiopia. The study employed exploratory analysis, i.e., data driven approach to identify the suitable family distribution for modeling the two outcomes and emphasized the importance of such analysis in achieving appropriate model specification.

The results obtained from the longitudinal MZIP model reveal that, after adjusting for subject and cluster-specific

random effects, several covariates are associated with both the log of the expected value of NOD and the log odds of not having diarrhea. The covariates region, EBF months, and gender were found to have significant associations with the log of the expected value of NOD. In addition, significant associations were observed for the interactions between linear time and MNP usage, as well as between the quadratic time and MNP usage in relation to the log of the expected value of NOD. Furthermore, gender, region, and the interaction of time by MNP usage showed significant relationship with the log odds of not having diarrhea. The study shows that region has a significant association with log of the expected value of NOD and odds of not having diarrhea. This observation is consistent with the findings of a systematic review research that highlighted regional disparities in the prevalence of diarrhea among under five children in Ethiopia [31]. Furthermore, our study reveals that EBF months are significantly associated with the expected log of NOD, aligning with the results of several previous studies [10, 11, 21, 22, 32, 33]. It was observed that the log odds of not having diarrhea were higher for female children compared to their male counterparts. This finding is consistent with prior research conducted in Ethiopia that reported a significant association between a child's sex and the odds of experiencing diarrhea [34].

One of the interesting findings of this study underscores the importance of starting with exploratory analysis and the impact of selecting an appropriate

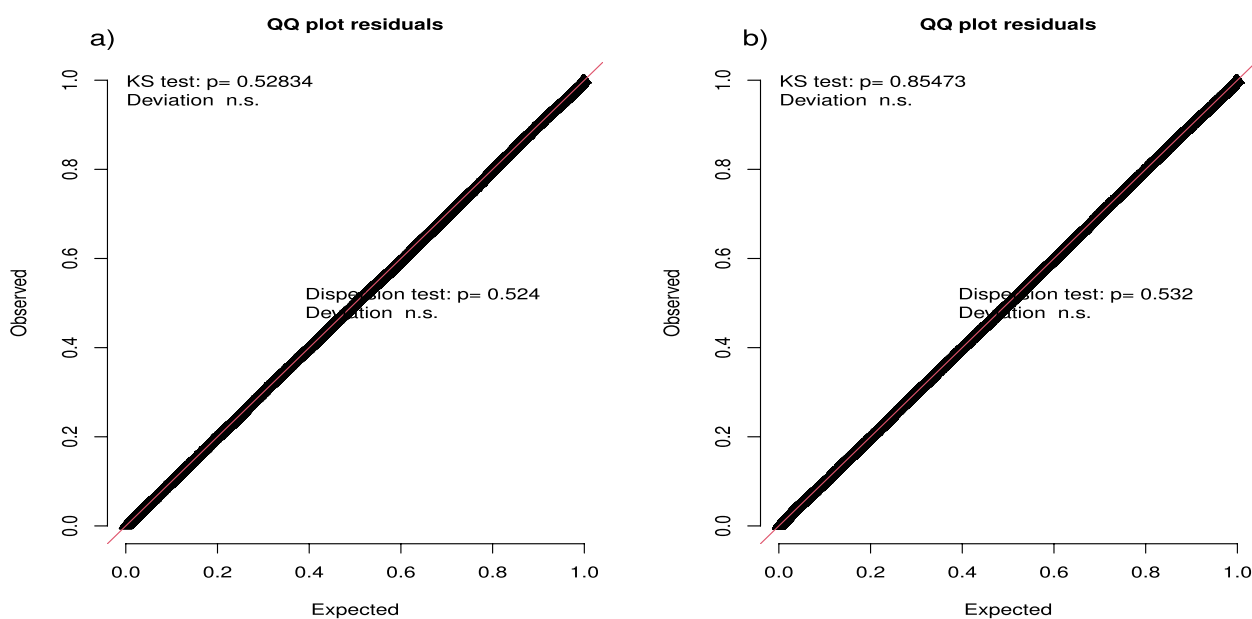


Fig. 5 Residual plots of MZIP models for: (a) NOD, (b) NOF

model specification on study outcomes i.e., a prior study employing the same dataset as the current study concluded that longitudinal diarrhea prevalence was higher among children who used MNP [18]. However, the results of the current study contradict these findings by revealing an inverted U-shaped trend in the log-expected NOD as time increases for children who used MNP. This indicates that the log-expected NOD for children who used MNP increases in comparison to those who did not as time increases until a turning point is reached. After this turning point, the negative coefficient in the quadratic term of the interaction implies that the log-expected NOD decreases as time increases for children who used MNP, compared to those who did not. In addition, the odds of not having diarrhea was initially very low for children who used MNP at the start of the study. The exploratory plot also reveals a high prevalence of diarrhea among children who used MNP at the beginning of the study, which may explain the inverted U-shaped rate of change in the expected log of NOD.

The variances of the individual and cluster level random effects for the outcome NOD indicate that there is more variability between individuals than between clusters. This may be due to individual-level factors such as poor hygiene, poor sanitation, or poor nutrition practices [14]. Therefore, since these variables were not included in the model, the variation due to these factors can be controlled by the subject specific random effects which may increase the variance of subject specific random effects. In addition, the individual reaction to the low-dose iron micronutrient powder may be different at the beginning of the study. A previous study shows that micronutrient with iron supplies can increase the incidence of diarrhea [35]. Therefore, some children may experience a sensitive reaction to the micronutrient powder, which may contribute to high variability between individuals within the same cluster.

Moreover, the study demonstrates that the odds of not having diarrhea for children who used MNP increase as time progresses compared to those who did not use MNP. Thus, we can assert that the use of MNP do not increase the longitudinal prevalence of diarrhea. Our findings align with another study that demonstrated providing children with MNP did not increase the risk of childhood infectious diseases such as diarrhea and lower respiratory infections [19]. Furthermore, a different study indicated that supplementing children with micronutrient, including vitamin A and zinc, can reduce the severity of diarrhea [20]. Based on this finding, we can say that selecting a data driven model plays a pivotal role in drawing reliable conclusion.

The finding of this study for the outcome NOF reveal that after adjusting for the subject and cluster

specific random effects, region, NOD and the interaction MNP usage \times EBF months had a significant association with the expected log of NOF. In addition, region, EBF months and the interaction MNP usage \times time had a significant association with the odds of having flu/common cold. A significant relationship between NOF and region was identified in both part of the model. This finding is consistent with the results reported in a prior investigation [7]. Furthermore, the current study supported the causal relationship between diarrhea and ARI which aligned with earlier findings [17]. It was observed that the log expected NOF increases with NOD. This result was supported by previous findings which shows diarrhea as risk factor of ARI [7, 15, 15]. Various studies also show the simultaneous occurrence of ARI and diarrhea [16, 17].

Moreover, this study indicates a significant association between the interaction of EBF month and MNP usage with the log expected NOF. While there are no studies specifically addressing the effects of the interaction between EBF month and MNP usage, various studies have highlighted the benefits of EBF for the first six months of a child's life [15, 21, 23]. In addition, a study has demonstrated that providing children with MNP support can reduce the risk of infectious diseases [20]. Furthermore, this study indicates that the odds of not having flu/common cold increased with each two-week increment in time for children who used MNP, compared to children who did not. This finding aligns with earlier studies [19, 20]. However, it contradicts a previous study that demonstrated a higher longitudinal prevalence of flu/common cold among children who used MNP compared to those who did not [18]. The variances of the individual and cluster level random effects for the outcome NOF indicate that most of the variability in NOF is occurring at the cluster level rather than the individual level, which indicates that the occurrence of flu/common cold is mostly influenced by cluster-level factors such as environmental conditions of the villages as it is an easily transmittable disease. In addition, flu/common cold is associated with seasonal changes that can affect the entire cluster [36]. These factors may increase the variance of the cluster specific random effects. The current study could not control seasonal effects because of the limitation of the data. Thus, We recommend that future studies control for seasonal effects, as this may provide significant additional insights.

Conclusion

In this paper, we conducted a multilevel zero-inflated Poisson regression analysis on longitudinal data focusing on infectious diseases such as diarrhea and flu among children aged 6 to 11 months. Our study underscores the vital

role of commencing with exploratory analysis to select the appropriate statistical model for the data. After adjusting for random effects, we observed that children who used MNP exhibited an initially higher rate of change in the expected log of NOD compared to those who did not use MNP. However, over time, this rate declined. In addition, region, EBF month, and gender demonstrated significant associations with NOD. Similarly, children who used MNP and were EBF for six months shows a decrease in the log-expected NOF compared to their counterparts. Furthermore, the odds of not having flu/common cold were higher for children who used MNP for two weeks increment in time. Region and NOD were also found to have significant associations with NOF. In light of these findings, we emphasize the importance of starting from exploratory analysis as a fundamental step in statistical modeling. In addition, we recommend raising awareness about the critical importance of EBF for the first six months to mitigate the impact of infectious diseases. Policymakers and health practitioners should encourage MNP usage, with regular monitoring and adaptation programs over time for the better outcome. Further, developing comprehensive strategies considering the joint influence of MNP usage and exclusive breastfeeding along with interventions for childhood diarrhea may assist in reducing morbidity and mortality associated with comorbidity of diarrhea and flu. It is important to consider studying the dependency between infectious disease with malnutrition indicators over time in future studies as malnourished children are more vulnerable to these diseases.

Abbreviations

AIC	Akaike information criterion
AOM	Age of a mother
ARIs	Acute respiratory infections
BIC	Bayesian information criteria
EBF	Exclusive breastfeeding
EPHI	Ethiopian Public Health Institute
ESM	Educational status of a mother
GP	Generalized Poisson
LRT	Likelihood ratio test
MNP	Micronutrient Powder
MZIP	Multilevel Zero-Inflated Poisson
NB	Negative Binomial
NOD	Number of days a child suffered from diarrhea
NOF	Number of days a child suffered from flu/common cold
SDGs	Sustainable Development Goals
SNNPR	Southern Nations, Nationalities, and Peoples' Region
WHO	World Health Organization
ZIP	Zero-Inflated Poisson

Acknowledgements

We thank the Ethiopian Public Health Institute for giving access to the data. Mrs. Bezalem Eshetu Yirdaw would like to thank the University of South Africa for providing a good working environment office. She would like to thank the National Research Foundation (NRF) of South Africa for the partial scholarship, the Schlumberger foundation faculty for the future program for the financial support provided during her study. The L'Oreal Unesco for women in science foundation for the endowment and award towards her PhD project in 2022.

Authors' contributions

BEY reviewed literature, performed conceptualization, statistical methods, statistical analyses, interpretations and compiled the manuscript. LKD conceptualize the research problems, suggested the statistical methods applied in the paper, supervised and reviewed the findings of data analyses and compilation of the manuscript. AS reviewed the manuscript.

Funding

The authors declare that there was no funding to conduct this research.

Availability of data and materials

The data that support the findings of this study are available from the authors upon reasonable request and with permission of Ethiopian public health institute.

Declarations

Ethics approval and consent to participate

Ethical approval for this study was obtained from the University of South Africa School of Science Ethics Review Committee with reference number 2023/CSET/SOS/014. The data for this study is a secondary data taken from Ethiopian public health institute and the Ethical approval for the original study was obtained from the Ethiopian National Research Ethics Review Committee (NRERC). Signed consent was obtained from caregivers of the study children before participation in the study. The study was registered at <http://www.clinicaltrials.gov/> with clinical trials identifier of NCT02479815. All methods were carried out in accordance with relevant guidelines and regulations/Declaration of Helsinki.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 1 April 2024 Accepted: 27 August 2024

Published online: 07 September 2024

References

- (WHO) WHO. Child mortality (under 5 years). 2020. <https://www.who.int/news-room/fact-sheets/detail/levels-and-trends-in-child-under-5-mortality-in-2020>. Accessed 23 Sep 2023.
- (UNICEF) UNICEF. Diarrhea. 2022. <https://data.unicef.org/topic/child-health/diarrhoeal-disease/>. Accessed 23 Sep 2023.
- (WHO) WHO. The top 10 causes of death. 2020. <https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death>. Accessed 23 Sep 2023.
- Dewey KG, Mayers DR. Early child growth: how do nutrition and infection interact? *Matern Child Nutr*. 2011;7:129–42.
- Kundu S, Kundu S, Al Banna MH, Ahinkorah BO, Seidu AA, Okyere J. Prevalence of and factors associated with childhood diarrhoeal disease and acute respiratory infection in Bangladesh: an analysis of a nationwide cross-sectional survey. *BMJ Open*. 2022;12(4):e051744.
- Feleke Y, Legesse A, Abebe M. Prevalence of diarrhea, feeding practice, and associated factors among children under five years in Bereh District, Oromia, Ethiopia. *Infect Dis Obstet Gynecol*. 2022. p. 1–13. <https://doi.org/10.1155/2022/4139648>.
- Merera AM. Determinants of acute respiratory infection among under-five children in rural Ethiopia. *BMC Infect Dis*. 2021;21(1):1–12.
- Tareke AA, Enyew EB, Takele BA. Pooled prevalence and associated factors of diarrhea among under-five years children in East Africa: A multilevel logistic regression analysis. *PLoS ONE*. 2022;17(4):e0264559.
- Apanga PA, Kumbeni MT. Factors associated with diarrhoea and acute respiratory infection in children under-5 years old in Ghana: an analysis of a national cross-sectional survey. *BMC Pediatr*. 2021;21(1):1–8.

10. Bbaale E. Determinants of diarrhoea and acute respiratory infection among under-fives in Uganda. *Australas Med J*. 2011;4(7):400.
11. Saeed OB, Haile ZT, Chertok IA. Association between exclusive breastfeeding and infant health outcomes in Pakistan. *J Pediatr Nurs*. 2020;50:e62–8.
12. Demissie GD, Yeshaw Y, Alemine W, Akalu Y. Diarrhea and associated factors among under five children in sub-Saharan Africa: Evidence from demographic and health surveys of 34 sub-Saharan countries. *PLoS ONE*. 2021;16(9):e0257522.
13. Tesfaye TS, Magarsa AU, Zeleke TM. Moderate to severe diarrhea and associated factors among under-five children in Wonago District, South Ethiopia: a cross-sectional study. *Pediatr Health Med Ther*. 2020;11:437–43. <https://doi.org/10.2147/PHMT.S266828>.
14. Hailu B, Ji-Guo W, Hailu T. Water, sanitation, and hygiene risk factors on the prevalence of diarrhea among under-five children in the rural community of Dangila district, northwest Ethiopia. *J Trop Med*. 2021;2021.
15. Mir F, Ariff S, Bhura M, Chanar S, Nathwani AA, Jawwad M, et al. Risk factors for acute respiratory infections in children between 0 and 23 months of age in a peri-urban district in Pakistan: A matched case-control study. *Front Pediatr*. 2022;9:704545.
16. Walker CLF, Perin J, Katz J, Tielsch JM, Black RE. Diarrhea as a risk factor for acute lower respiratory tract infections among young children in low income settings. *J Glob Health*. 2013;3(1).
17. Newman KL, Gustafson K, Englund JA, Khatry SK, LeClerq SC, Tielsch JM, et al. Risk of respiratory infection following diarrhea among adult women and infants in Nepal. *Am J Trop Med Hyg*. 2020;102(1):28.
18. Samuel A, Brouwer ID, Feskens EJ, Adish A, Kebede A, De-Regil LM, et al. Effectiveness of a program intervention with reduced-iron multiple micronutrient powders on iron status, morbidity and growth in young children in Ethiopia. *Nutrients*. 2018;10(10):1508.
19. Lemaire M, Islam QS, Shen H, Khan MA, Parveen M, Abedin F, et al. Iron-containing micronutrient powder provided to children with moderate-to-severe malnutrition increases hemoglobin concentrations but not the risk of infectious morbidity: a randomized, double-blind, placebo-controlled, noninferiority safety trial. *Am J Clin Nutr*. 2011;94(2):585–93.
20. Fischer Walker CL, Black RE. Micronutrients and diarrheal disease. *Clin Infect Dis*. 2007;45(Supplement_1):S73–S77.
21. Mulatu T, Yimer NB, Alemnew B, Linger M, Liben ML. Exclusive breastfeeding lowers the odds of childhood diarrhea and other medical conditions: evidence from the 2016 Ethiopian demographic and health survey. *Ital J Pediatr*. 2021;47(1):1–6.
22. Hossain S, Mhrshahi S. Exclusive breastfeeding and childhood morbidity: A narrative review. *Int J Environ Res Public Health*. 2022;19(22):14804.
23. Duijts L, Jaddoe VW, Hofman A, Moll HA. Prolonged and exclusive breastfeeding reduces the risk of infectious diseases in infancy. *Pediatrics*. 2010;126(1):e18–25.
24. (WHO), W.H.O.: Infant and young child feeding (2021). <https://www.who.int/data/nutrition/nlis/info/infant-and-youngchild-feeding>. Accessed 28 Sept 2023.
25. Lambert D. Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics*. 1992;34(1):1–14.
26. Lee AH, Wang K, Scott JA, Yau KK, McLachlan GJ. Multi-level zero-inflated Poisson regression modelling of correlated count data with excess zeros. *Stat Methods Med Res*. 2006;15(1):47–61.
27. Dunn PK, Smyth GK. Randomized quantile residuals. *J Comput Graph Stat*. 1996;5(3):236–44.
28. Hartig F. DHARMA: Residual Diagnostics for Hierarchical (Multi Level / Mixed) Regression Models. (2022). R package version 0.4.6. <http://florihartig.github.io/DHARMA/>. Accessed 26 Sept 2023.
29. Stram DO, Lee JW. Variance components testing in the longitudinal mixed effects model. *Biometrics*. 1994;1171–7.
30. Moghimbeigi A, Eshraghian MR, Mohammad K, Mcardle B. Multilevel zero-inflated negative binomial regression modeling for over-dispersed count data with extra zeros. *J Appl Stat*. 2008;35(10):1193–202.
31. Alebel A, Tesema C, Temesgen B, Gebrie A, Petrucka P, Kibret GD. Prevalence and determinants of diarrhea among under-five children in Ethiopia: a systematic review and meta-analysis. *PLoS ONE*. 2018;13(6):e0199684.
32. Mhrshahi S, Ichikawa N, Shuaib M, Oddy W, Ampon R, Dibley MJ, et al. Prevalence of exclusive breastfeeding in Bangladesh and its association with diarrhoea and acute respiratory infection: results of the multiple indicator cluster survey 2003. *J Health Popul Nutr*. 2007;25(2):195.
33. Hamer DH, Solomon H, Das G, Knabe T, Beard J, Simon J, et al. Importance of breastfeeding and complementary feeding for management and prevention of childhood diarrhoea in low- and middle-income countries. *J Glob Health*. 2022;12.
34. Anteneh ZA, Andargie K, Tarekegn M. Prevalence and determinants of acute diarrhea among children younger than five years old in Jabithennan District, Northwest Ethiopia, 2014. *BMC Public Health*. 2017;17(1):1–8.
35. Soofi S, Cousens S, Iqbal SP, Akhund T, Khan J, Ahmed I, et al. Effect of provision of daily zinc and iron with several micronutrients on growth and morbidity among young children in Pakistan: a cluster-randomised trial. *Lancet*. 2013;382(9886):29–40.
36. Vidal K, Sultana S, Patron AP, Salvi I, Shevlyakova M, Foata F, et al. Changing epidemiology of acute respiratory infections in under-two children in Dhaka. *Bangladesh Front Pediatr*. 2022;9:728382.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.