Dey & Raheem, *Cogent Mathematics* (2016), 3: 1159798
http://dx.doi.org/10.1080/23311835.2016.1159798

cogent mathematics

Received: 25 November 2015
Accepted: 24 February 2016
First Published: 01 March 2016

*Corresponding author: Enayetur Raheem, Department of Statistics, St. Anthony's College, Shillong, Meghalaya 793001, India
E-mail: enayetur.raheem@unco.edu

Reviewing editor:
Zudi Lu, University of Southampton, UK

Additional information is available at the end of the article

STATISTICS | RESEARCH ARTICLE

Multilevel multinomial logistic regression model for identifying factors associated with anemia in children 6–59 months in northeastern states of India

Sanku Dey[1§] and Enayetur Raheem[2*§]

Abstract: Objective: To examine the factors influencing the occurrence of childhood anemia in northeast India. Method: A nationally representative systematic multi-stage stratified cross-sectional sample of singleton children aged 6–59 months from all states of India. Data consist of 10,136 children in the age group 6–59 months in eight northeastern states. The level of anemia was the outcome variable with four ordinal categories (severe, moderate, mild, and non-anemic). A two-level random intercept multivariate logistic regression model was considered with state of residence as the level-2 variable. Results: About 53% of the children are anemic in the northeastern states of India. Tripura has the highest prevalence of anemia cases (74%), whereas the lowest percentage of anemia cases was in Manipur (42%). Multivariate analysis suggests that age at marriage (OR = 1.13, 95% CI: 1.05, 1.21) and the number of children ever born (OR = 1.09, 95% CI: 1.03, 1.15) have significant effects on being at or below a hemoglobin level (severely anemic). Furthermore, age

ABOUT THE AUTHORS

Sanku Dey's, PhD, research focuses on the development and application of statistical approaches to scientific and medical issues. Another area of his research encompasses both frequentist and Bayesian inference of reliability theory and general distribution theory. This area of research has real-life applications in engineering and medical sciences. Another important research in this area is to work on distribution theory using records. Record values and record times have been of interest to humans throughout history. Meteorologists frequently deal with upper and lower record temperatures and precipitation levels. Record values often appear in sporting events. I have developed several mathematical models based on distributions like Rayleigh, generalized Rayleigh, generalized exponential, and generalized inverted exponential distributions.

Enayetur Raheem's, PhD, research is focused on building better predictive models through improved estimation of regression parameters. He is interested in data/text mining/feature extraction, and machine learning. Other interests include application of statistical methods in medical/health, biological and environmental sciences.

PUBLIC INTEREST STATEMENT

A multilevel multinomial logistic regression model was considered to identify potential risk factors of anemia in children aged 6–59 months in northeastern India. Level of anemia was considered as the outcome variable with four ordinal categories (severe, moderate, mild, and non-anemic) based on hemoglobin concentration in blood. A two-level random intercept model was used with state of residence as the level-2 variable. Considering the hierarchical nature of the data, we did not find place of residence (urban vs. rural), sex of child, or household living standard (low, medium, and high SES) to be significant. Rather, mother's age at marriage and the number of children ever born were significantly associated with anemia. Our findings suggest that as the child becomes 48 months or older, likelihood of severe anemia decreases. This study throws light on the fact that anemia intervention needs to focus more at state level followed by individual level.

cogent • mathematics

of the child (OR = 0.92, 95% CI: 0.86 − 1.00) was a significant predictor, indicating that odds of severe anemia decrease if the child is 48 months or older. Conclusions: The high prevalence of mild and moderate anemia demands due emphasis in the programs and policies of the government so that the overall prevalence of anemia among children aged 6–59 months can be reduced. Comprehensive strategies need to focus more at state level followed by individual level for combating anemia.

Subjects: Mathematics & Statistics; Multivariate Statistics; Science; Statistical Theory & Methods; Statistics; Statistics & Probability

Keywords: multilevel multinomial logistic regression; child anemia; odds ratio

1. Introduction

Anemia is considered the most prevalent nutritional deficiency globally (McLean, Cogswell, Egli, Wojdyla, & de Benoist, 2009). About one-third of the global population (over two billion) is anemic (de Benoist, McLean, Egll, & Cogswell, 2008) while the prevalence of anemia among preschool children is 47.4% (IIPS, 2000). According to WHO estimates, India has the highest prevalence of anemia among the South Asian countries in all age groups. The Third National Family Health Survey (IIPS, 2007) of India revealed that 70% of the children are anemic in the age group of 6–59 months, including 3% severely anemic, 40% moderately anemic, and 26% mildly anemic (IIPS, 2000). As per five major surveys (National Family Health Survey (NFHS Assunção, Santos, de Barros, Gigante, & Victora, 2007; de Benoist et al., 2008), District Level Household Survey 2 (DLHS), Indian Council of Medical Research (ICMR) Micronutrient Survey (Togeja & Singh, 2004) and Micronutrient Survey (NNMB, 2004), over 70% of preschool children are anemic in the country.

Iron deficiency (ID) is listed as one of the “top ten risk factors contributing to death” (Dubey, 1994). Iron deficiency anemia (IDA) is prevalent in South Asia, predominantly India, Bangladesh, and Pakistan. However, the prevalence of IDA has declined substantially in neighboring countries such as Bangladesh and Pakistan (Gillespie, 1997), whereas it has plummeted from 20 to 8% within a decade in China (Lokeshwar, Manglani, Rao, Patel, & Kulkarni, 1990). Studies carried out in Egypt (Seshadri, Gopaldas, Walter, & Heywood, 1989), India (Soemantri, Gopaldas, Seshadri, & Pollitt, 1989), Thailand (Pollitt, Hathirat, Kotchabhakdi, Missell, & Valyasevi, 1989), and the USA (Pollitt, 1991) revealed that iron and folate deficiency anemia reduce the learning capacity, attentiveness, and intelligence of children aged below five years. Several studies have been carried out on anemia in India since 1980. However, very few studies focused on child anemia at the national and regional levels (Bharati, Pal, Chakrabarty, & Bharati, 2015; Dey, Goswami, & Dey, 2013; Singh & Patra, 2014). The risk factors of anemia most often cited in the literature are: low family income and low maternal level of education, lack of access to health care services, inadequate sanitary conditions, and a diet with poor quantities of iron (Oliveira, Osório, & Raposo, 2007; Osório, Lira, & Ashworth, 2004) .

Nutritional problem is very common among children below five years in all the states of India, particularly, in those states whose performances are very poor in respect of demographic and socioeconomic indicators. There are several factors responsible for IDA, for example: low dietary intake, inadequate iron (less than 20 mg/day) and folic acid intake (less than 70 mg/day), and chronic blood loss due to infection such as malaria and hookworm infestations, among others (NNMB, 2004; Togeja & Singh, 2004). A recent study on the determination of socioeconomic and demographic determinants of anemia among Indian children aged 6–59 months found an increasing trend of prevalence of anemia up to two years of age and then decreased thereafter (Goswmai & Das, 2015). Therefore, it is imperative to find out the factors that contributed to anemia and to examine the contribution of existing programs in combating child anemia, especially in the less developed areas such as northeastern states of India. In view of the above, our objective is to determine the prevalence of anemia among children (6–59 months) from the northeastern states of India. To comprehend the prevalence of anemia and to identify the significant predictors, socioeconomic differentials have been taken into consideration in our study.

2. Methods

The present paper uses the data-set from the 2005–2006 National Family Health Survey (released in 2008) for the northeastern states of India. Respondents were selected through a systematic multi-stage stratified sample survey conducted in all 29 states of India (IIPS, 2007). In each state, populations were stratified by urban and rural area of residence, and the sample size at the state level was proportional to the size of the state's urban and rural population. A uniform sample design was adopted in all states. In each state, rural sample was selected in two stages: at the first stage, primary sampling units (PSUs), which are villages, were selected with probability proportional to population size (PPS). At the second stage, households within each PSU were selected randomly. In urban areas, a three-stage sampling procedure was adopted. At first, wards were selected with PPS sampling. In the second stage, one census enumeration block (CEB) was randomly selected from each sample ward. In the final stage, households were randomly selected within each sample CEB. The study population constitutes a nationally representative cross-sectional sample of singleton children aged 0–59 months and born after January 2000 or January 2001 (n = 50,750) to mothers aged 15–49 years from all 29 states of India. Information on children was obtained by a face-to-face interview with mothers, with a response rate of 94.5%. The survey also provided district-level information on the prevalence of under-nutrition [weight-for-age, using the standard deviation (SD) classification] among children in the age group of 0–59 month(s); prevalence of anemia (Hb estimation by HemoCue Hb201 + analyzer, Angelholm, Sweden) for the children aged 6–59 months, adolescent girls aged 10–19 years, and pregnant women; household availability of iodized salt; and the coverage of vitamin A program, with appropriate dosage. To meet the objectives of the study, we have produced a data-set that pertains to the northeastern states of India. Our study comprises 10,136 children within 6–59 months of age.

3. Variables

The variables of the study are briefly described in the following.

3.1. Outcome/response variable

We create the response variable using the variables that measures hemoglobin level in blood. We split the variable into four categories as per the specification of WHO to define level of anemia. In particular, we define a new variable called hglevel to indicate level of hemoglobin in blood (g/dL). We set hglevel = 1 if hemoglobin level is below 7 g/dL and call it severely low level (indicating severe anemia), hglevel = 2 if hemoglobin level is 7–9.9 g/dL (indicating moderate anemia), hglevel = 3 if hemoglobin level is 10–10.9 g/dL (indicating mild anemia), and hglevel = 4 if hemoglobin level is above 11 g/dL indicating no anemia.

3.2. Independent variables/predictors

The study includes a set of independent variables to understand the extent and differentials in the level of anemia among children aged 6–59 months.

4. Objective of the study

Our objective is to study the relationship between some household-level predictors and some maternal variables on the probability of being at or below a hemoglobin level. Based on our classification of anemic children, we calculate odds ratio and the predictive probability of being at or below a hemoglobin level based on a polytomous logistic regression model.

In particular, we like to answer the following research questions:

(1) What is the likelihood of being at or below each level of hemoglobin in the blood (g/dL) for children at a typical household?

(2) Does the likelihood of being at or below each level of hemoglobin (i.e. anemia level) vary across state of residence?

(3) What are the relationships between the household and maternal variables, and the likelihood of a child being at or below a given hemoglobin level?

5. Statistical analysis

In the present study, we employ bivariate as well as multivariate techniques to identify the factors that are associated with anemia in children of 6–59 months of age. In the bivariate analysis, cross-tabulation was made between the potential risk factors and the presence of anemia. Pearson's chi-square test and *p*-values were used to test for the significance of each of the potential risk factor in bivariate analysis. Further analyses were carried out to study the relationship of the potential risk factors with the hemoglobin level in a multivariate setup. We use multinomial multilevel logistic regression model to predict the level of anemia as a function of mother's age at marriage, number of children ever born to mother, religion, literacy of mother, household living standard, place of residence, sex of the child, and age of the child (in months). The wealth index was constructed using household asset data. Each household asset was assigned a score generated through principal components analysis and the resulting scores were standardized and summed up for each household (IIPS, 2007).

Due to the stratified nature of data in NFHS, the children are naturally nested into mothers, mothers are nested into households, households into PSUs, and PSUs into the states. Hence, keeping in view of the hierarchically clustered nature of the survey data, we use multilevel regression model to avoid possible under-estimation of parameters from a single-level model (Griffiths, Matthews, & Hinde, 2002). The advantage of multilevel model is that they properly account for the correlation structure of the data that frequently occurs in social sciences and in multistage survey sampling.

To properly account for the hierarchical nature of the data, we consider state of residence as the level-2 variable under which the respondents are nested. Thus, respondent-level variables are the level-1 variables used in this study to predict the level of anemia in children. Our outcome variable is ordinal polytomous with four levels for anemia. The model for such data uses a multinomial distribution and cumulative logit link function to compute the cumulative odds for each category of the response (O'Connnell, Goldstein, Rogers, & Peng, 2008).

For the polytomous response variable, let us denote η_{kij} as the log odds of a child in the *i*th household in the *j*th state being at or below *k*th level of hemoglobin in blood. The model for level-1 may be written as:

$$\eta_{kij} = \log\left(\frac{P(R_{ij} \le k)}{1 - P\left(R_{ij} \le k\right)}\right) = \beta_{0j} + \beta_{1j} X_{ij} + \delta_k \tag{1}$$

Here, β_{0j} is the intercept, assumed to be random, represents the average log odds of being at or below severely low hemoglobin level; $P\left(R_{ij} \le k\right)$ represents the probability of responding at or below *k*th level of the outcome variable; X_{ij} represents level-1 predictorβ_{1j}, the slope coefficient corresponding to X_{ij} which measures the change in the probability of being at a given hemoglobin level per unit change in the level-1 predictor; and δ_k epresents the difference between the *k*th category and the preceding one. Essentially, for *k* = 1, we have $\delta_1 = 0$.

We consider the intercept to be random having the following model:

$$\beta_{0j} = \gamma_{00} + u_{0j},$$

where γ_{00} is the log odds of severe anemia relative to no anemia when the predictor variables in the model are evaluated at zero, and u_{0j} is the random error term with $u_{0j} \sim N\left(0, \tau_{00}\right)$ where $u\tau_{00}$ is the level-2 error variance.

Now, substituting the expression for β_{0j} in equation (McLean et al., 2009), we get

$$\eta_{kij} = \log\left(\frac{P(R_{ij} \leq k)}{1 - P\left(R_{ij} \leq k\right)}\right) = \gamma_{00} + \beta_{1j}X_{ij} + \delta_k + u_{0j} \tag{2}$$

This model represents a scenario with no level-2 predictor (as in our case), and one level-1 predictor. However, in our study, we have more than one level-1 predictor, and so the model is meant to be augmented by those additional covariates as they are considered in the analysis.

Model (2) gives log odds when fitted to data. Alternatively, predicted probability of the event of interest (e.g. being at or below a hemoglobin level) can be calculated using the following formula as discussed in (Ene, Leighton, Blue, & Bell, 2014):

$$\text{Predicted probability (PP) } \theta_{kij} = \frac{e^{\eta_{kij}}}{1 + e^{\eta_{kij}}} \tag{3}$$

This is simply a conversion of log odds of the event of interest to probability of the event of interest in order to provide a meaningful interpretation of the numbers. In this expression, θ_{kij} is the probability of the event (e.g. being at or below a hemoglobin level), $1 - \theta_{kij}$ is the corresponding probability of being above a given hemoglobin level, η_{kij} represents the log odds of the event of interest, and *e* is approximately 2.72.

We used PROC GLIMMIX in SAS 9.4 for multinomial distribution, and CLOGIT link function. To the best of our knowledge, there is no model diagnostic procedure available under PROC GLIMMIX for multinomial model. We considered several candidate models, and the best model was selected based on −2 log likelihood as a criterion. Details are discussed later in the paper.

5.1. Model-building strategies for multilevel logistic regression

A multilevel multinomial logistic regression model was considered to predict the probability of being at or below a hemoglobin level using the available predictors. Since the outcome variable is ordinal, we consider cumulative logit link function. The steps of the model-building process are outlined in Table 3. We consider random intercept model where the model is fitted with only the intercept (Model 1). Then, in Model 2, we add maternal characteristics such as number of children ever born (2 or less, 3–4, and 5 or more) and age at marriage (below 18, 18–26, and above 26 years of age). In Model 3, we add child's age in months (<48 months, 48, or more months). To get Model 4, we add religion and literacy of mother to Model 3. Finally, we augment Model 4 by adding the sex of the child, household living standard (low, medium, high), and place of residence (urban vs. rural). We call it Model 5.

6. Results

6.1. Bivariate analysis

Of the 10,136 children included in this study, 51.6% were male and 48.4% were female. Fifty-two percent of the children were anemic (53% male and 52% female). Table 1 presents cross-tabulation of prevalence of anemia (frequency and percentages) by level of anemia for northeastern states of India. The last column shows total number of children in the study sample from each state along with the percentages of children with anemia. The last column of Table 1 shows crude prevalence rate of each type of anemia per hundred children in the sample. Overall, 12% of the children have severe anemia, 34% have moderate anemia, 6% have mild anemia, and 48% have no anemia. We observe that Tripura has the highest percentage (74%) of anemia cases among the children in the sample. Of the 361 children in Tripura, about 19% are severely anemic, 50% moderate, and 6% have mild anemia. Sikkim falls next to Tripura with about 71% anemia cases (22% severe, 43% moderate,

Table 1. Prevalence of anemia by state of residence

State	Anemia level				
	Severely anemic	Moderately anemic	Mildly anemic	Non-anemic	Total (% anemic)
Sikkim	159	301	42	206	708
	22.46%	42.51%	5.93%	29.10%	71%
Arunachal Pradesh	317	724	261	1,611	2,913
	10.88%	24.85%	8.96%	55.30%	45%
Nagaland	33	180	35	277	525
	6.29%	34.29%	6.67%	52.76%	47%
Manipur	241	526	118	1,198	2,083
	11.57%	25.25%	5.66%	57.51%	42%
Mizoram	51	258	57	296	662
	7.70%	38.97%	8.61%	44.71%	55%
Tripura	67	181	20	93	361
	18.56%	50.14%	5.54%	25.76%	74%
Meghalaya	22	55	17	105	199
	11.06%	27.64%	8.54%	52.76%	47%
Assam	334	1,189	136	1,026	2,685
	12.44%	44.28%	5.07%	38.21%	62%
Total (crude %)	1,224	3,414	686	4,812	10,136
	(12%)	(34%)	(6%)	(48%)	

Note: Overall Pearson's chi-square statistics value is 622.84 with a p-value of <0.0001. We found statistically significant association between state of residence and anemia level.

and 6% mild). Nagaland and Meghalaya each have 47% anemia cases, while Arunachal Pradesh has 45% followed by Manipur which has the lowest percentage of anemia cases (42%). Sikkim has the highest percentage of severely anemic children (about 23%) followed by Tripura (about 19%). Mizoram has the lowest rate of severely anemic children with only 8% prevalence rate.

Pearson's chi-square test for association shows that the place of residence is associated with the level of anemia ($\chi^2 = 622.84$, $p < 0.0001$). In other words, prevalence of anemia among children varies significantly among the states.

Anemia is slightly more prevalent among males (53%) than females (52%). Among the severely anemic children, 85% are in rural areas compared to 15% in urban areas. Similarly, 79% of moderately anemic children and 75% of mildly anemic children live in rural areas. Muslim children have the highest prevalence rate of any form of anemia (62%) followed by 56% among Hindus. Christians have the lowest prevalence of anemia (48%). However, among the severely anemic child, 44% are Hindu, 7% Muslim, 29% are Christian (Table not shown).

Of the severely anemic cases, 62% are from households with low living standard, 27% are from medium, and 10% are from high living standard households. Interestingly, prevalence of any form of anemia is low among the mothers who cannot read or write. However, the association is not statistically significant ($\chi^2 = 6.43$, $p = .09$). Among the mothers of severely anemic children, 46% have two or less children, 36% have three to four children, and 18% have more than five children. Majority of the mothers of severely anemic children (63%) were married between the ages 18 and 26 years, 32% were married below 18 years, while 5% were married when they were above 26 years of age.

Results of bivariate analysis are presented in Table 2. We observe that there is a statistically significant association between the level of anemia (severe, moderate, mild, and none) and religion of the respondents ($p < 0.0001$), and place of residence ($p < 0.0001$), and between the level of anemia and age of mother at marriage ($p = 0.0425$). There is no statistically significant association found between anemia level and household living standard, sex of child, literacy of mother, total children ever born to mother, and age of the child (months). Owing to this, we investigate the relationship between anemia level and the variables cited above using multilevel polytomous logistic regression.

6.2. Multivariate analysis

In multivariate analysis, the fitted models along with the estimated effects and their standard errors are presented in

Table 4. We used Laplace estimation method so that we can compare the models based on negative 2 log likelihood (−2LL) or the deviance test. The last row of Table 4 shows the p-values for the deviance test based on Chi-square statistic.

We consider five difference models and they are presented in Table 3. Notice that Model 1 is nested under Model 2, which in essence is nested under Model 3, and this in turn is nested under Model 4, and again Model 4 nested under Model 5. This allows us to compare these models based on −2LL. We found Model 3 to be better than Model 1 as the drop in −2LL is statistically significant ($p < 0.0001$), and Model 3 to be better than Model 2 ($p = 0.0452$) as well. Models 4 and 5 include more variables and thus reduce the −2LL values further. However, the gain in reduction of −2LL is not statistically significant enough to consider them for our analysis. We, therefore, resort to Model 3 and use this model to answer our research questions.

In regard to our first research question about the likelihood of being at or below a given hemoglobin level (indicative of anemia level) at a given household, we consider the fixed effects estimates provided under Model 1 of Table 4. The estimated effects in Table 4 (Model 1) are the log-odds of being at or below a hemoglobin level with typical background characteristics (no covariates in the model). We use these estimates to calculate the predictive probability (PP) of a child being at or below a hemoglobin level. For example, the probability of being at or below severely low hemoglobin level (i.e. severely anemic) for a given child with average background characteristics (with no covariate in the model) is

$$\text{P (being at or below severe anemic level)} = \frac{e^{\eta_{ij}}}{1+e^{\eta_{ij}}} = \frac{e^{-1.95}}{1+e^{-1.95}} = \frac{0.1423}{1+0.1423} = 0.1247$$

Similarly, the probability of being at or below moderately anemic level is 0.4825 and at or below mildly anemic is 0.5523. These are cumulative probabilities of being at or below a given level of anemia. In order to obtain the exact probability of being at a given level, we have to subtract the adjacent probabilities. For example, the predicted probability of being at or below severely anemic level is 0.1247, at the moderately anemic level is (0.4825 − 0.1247=) 0.3578, and at the mild anemic level is (0.5523 − 0.4825=) 0.0698. Also, probability of being non-anemic is 1 − 0.5523 = 0.4477.

In the second research question, we wanted to know if the likelihood of being at or below each level of hemoglobin (i.e. anemia level) varies across state of residence. To answer this question, we estimated the covariance parameter for the variable "state." The estimated intercept for "state" is 0.2015 ($z = 1.94$, $p = 0.0263$). Thus, there exists a statistically significant variation across states for the likelihood of being at or below a hemoglobin level. Additionally, the intra-class correlation coefficient (ICC) was computed using the covariance parameter estimate for the state (0.2015). The ICC indicates how much of the total variation in the probability of being at or below a hemoglobin level is due to the variation among the states. Following the procedure of (Snijders & Bosker, 1999), we calculate ICC as:

Table 2. Frequency distribution of anemia level by the predictors of anemia. The first column shows chi-square test for independence; the cells represent the frequency (first row), row percentage (second row). The last column shows crude odds ratio for overall anemia

Variables (Chi-square and p-value)		Severely anemic	Moderately anemic	Mildly anemic	Non-anemic	Crude odds ratio
Place of residence Chi-square = 27.16 $p < 0.0001$	Rural	1,035	2,705	516	3,818	1.04
		12.82%	33.50%	6.39%	47.29%	
	Urban (ref)	189	709	170	994	
		9.17%	34.38%	8.24%	48.21%	
Religion Chi-square = 166.35 $p < 0.0001$	Hindu	537	1,330	238	1,667	1.23*
		14.24%	35.26%	6.31%	44.19%	
	Muslim	83	442	48	350	1.59*
		8.99%	47.89%	5.20%	37.92%	
	Christian	352	1,050	245	1,824	0.88
		10.14%	30.25%	7.06%	52.55%	
	Others (ref)	252	592	155	971	
		12.79%	30.05%	7.87%	49.29%	
Household living standard Chi-square = 8.33 $p = 0.2146$	Low (ref)	764	2,031	400	2,956	
		12.42%	33.02%	6.50%	48.06%	
	Medium	336	999	199	1,312	1.01
		11.81%	35.10%	6.99%	46.10%	
	High	124	384	87	544	1.08
		10.89%	33.71%	7.64%	47.76%	
Sex of child Chi-square = 2.08 $p = 0.5548$	Male	636	1,772	368	2,453	1.08
		12.16%	33.89%	7.04%	46.91%	
	Female (ref)	588	1,642	318	2,359	
		11.98%	33.46%	6.48%	48.07%	
Literacy of Mother Chi-square = 6.43 $p = 0.923$	Can read & write (ref)	692	2,024	385	2,771	
		11.78%	34.47%	6.56%	47.19%	
	Can't read & write	497	1,257	276	1,867	0.97
		12.75%	32.26%	7.08%	47.91%	
Total children ever born to mother Chi-square = 9.72 $p = 0.1366$	Up to two children)	548	1,444	276	2,014	0.91
		12.80%	33.72%	6.45%	47.03%	
	Three or four children	423	1,214	271	1,805	0.97
		11.39%	32.70%	7.30%	48.61%	
	Five or above children (ref)	218	623	114	819	
		12.29%	35.12%	6.43%	46.17%	
Age at marriage Chi-square = 13.03 $p = 0.0425$	Below 18 Years	384	1,016	183	1,556	1.03
		12.23%	32.37%	5.83%	49.57%	
	18 to 26 years	748	2,090	445	2,846	0.91
		12.20%	34.10%	7.26%	46.43%	
	Above 26 Years (ref)	57	175	33	236	
		11.38%	34.93%	6.59%	47.11%	
Age of child in months Chi-square = 5.07 $p = 0.1669$	>48 months	499	1,477	313	2,034	1.03
		11.54%	34.17%	7.24%	47.05%	
	<48 months	725	1,937	373	2,778	
		12.47%	33.32%	6.42%	47.79%	

*Significant results based on 95% confidence interval for the odds ratio.

Table 3. Model-building strategies

Model 1	Model 2	Model 3	Model 4	Model 5
No predictors but only random effects for the “state”	Model 1 + level-1 fixed effects that are related to maternal characteristics such as number of children ever born (2 or less, 3–4, and 5 or more) and age at marriage (below 18, 18–26, and above 26 years).	Model 2 + age of the child (months)	Model 3 + level-1 fixed effects religion and literacy of mother	Model 4 + sex of child (male/female), household living standard (low, medium, and high), and place of residence (urban/rural)
Results used to determine the percentage variation in anemia level explained by the level-2 units (state of residence)	Results indicate the relationship between level-1 (household level) predictors and the anemia level	Results indicate if addition of child’s age improves the fit.	Results indicate whether addition of demographic variables improves the model fit	Results indicate whether addition of SES variables improves the model fit

Table 4. Estimates for two-level multinomial logistic regression models for predicting anemia level (*N* = 10,136)

	Model 1	Model 2	Model 3	Model 4	Model 5
Fixed effects					
Intercept 1 (severely anemic)	−1.95* (0.16)	−2.29* (0.19)	−2.35* (0.19)	−2.42* (0.20)	−2.23* (0.23)
Intercept 2 (moderately anemic)	−0.07 (0.16)	−0.42 (0.19)	−0.48* (0.19)	−0.55 (0.20)	−0.37 (0.23)
Intercept 3 (mildly anemic)	0.21 (0.16)	−0.14 (0.19)	−0.19 (0.19)	−0.27 (0.20)	−0.08 (0.23)
Age at marriage		0.12* (0.04)	0.12* (0.04)	0.12* (0.04)	0.12* (0.04)
Total children ever born		0.08* (0.03)	0.09* (0.03)	0.08* (0.03)	0.08* (0.03)
Age of child (<48 months)			−0.08* (0.04)	0.08* (0.04)	0.08* (0.04)
Religion				0.03 (0.02)	0.02 (0.02)
Literacy of mother				0.02 (0.04)	0.003 (0.04)
Household living standard					−0.02 (0.03)
Place of residence					−0.06 (0.05)
Sex of child					−0.03 (0.04)
Error variance					
Intercept	0.2015* (0.10)				
Model fit					
−2LL	23,063.22	22,228.13** ($p < 0.0001$)	22,224.12** ($p = 0.0452$)	22,222.03 ($p = 0.3517$)	22,218.83 ($p = 0.3618$)

Note: Entries in the table are estimated effects while the standard errors are reported in the parenthesis.

*Level of significance at $p < 0.05$.

**Significant LR test; ICC = 0.0577. PROC GLIMMIX in SAS 9.4 with Laplace estimation method was used.

$$\text{ICC} = \frac{\tau_{00}}{\tau_{00} + 3.29} = \frac{0.2015}{0.2015 + 3.29} = 0.0577$$

The ICC = 0.0577 indicates that approximately 6% of the total variation in the response is accounted for by the states. Thus, the remaining 94% variability is due to the variation within the respondents/households and other unknown factors.

Finally, to answer our last research question, we use the best fit model (Model 3) as discussed earlier. Model 3 contains three significant covariates (age at marriage, number of children ever born, and age of the child in months). We find significant effect of age at marriage ($\hat{\beta} = 0.12$, $p = 0.04$, number of children ever born ($\hat{\beta} = 0.09$, $p = 0.03$), and age of the child ($\hat{\beta} = 0.08$, $p = 0.04$. The estimates for mother's age at marriage and the number children ever born have positive effects on the log odds, whereas age of child has a negative effect. This implies if the age at marriage increases by one unit, the corresponding change in the log odds is 0.12. Similarly, a unit increase in the number of children ever born to mother increases the log odds by 0.09, and getting 48 months or older reduces the log odds by 0.08 units. In other words, age at marriage and number of children born to women increase the probability of being at the severely anemic level. On the other hand, increase in age of the child decreases the probability of being at a given anemia level. Since we have four levels for the outcome variables, we have three intercepts as shown in Table 4. We performed significance tests for these fixed intercepts and they were all found to be significant (result not shown). We calculated odds ratio for these predictors for easier interpretation. For age at marriage (OR = 1.13, 95% CI: 1.05, 1.21), one unit increase in age at marriage results in 1.13 times increase of having lower hemoglobin level (i.e. becoming more anemic). For number of children even born (OR = 1.09, 95% CI: 1.03, 1.15), one unit increase in the number of children born to women inflates the odds of having low hemoglobin level (indicating severely anemic) by 1.09. On the other hand, age of the child (OR = 0.92, 95% CI: 0.86 − 1.00) indicates odds of being at a lower hemoglobin level and decreases if the child is 48 months or older.

Using Model 3, the predicted probability of being at a given anemia level can be calculated following the formula given in Equation (3).

7. Discussion

Detection of the risk factors is detrimental to planning and implementation of programs to eradicate child anemia, especially in those groups where prevalence is very high. Our analysis suggests that childhood anemia is highly prevalent in the northeastern states of India. The overall prevalence of anemia in children in the northeastern states is 53%. There are important and significant relationships between anemia and some of the selected potential risk factors. Bivariate analysis suggests that there is no significant association between anemia level and household living standard, sex of child, literacy of mother, total children ever born to mother, and age of the child (months), whereas results from multivariate analysis show that age of the child, mother's age at marriage, and number of children ever born have significant effects on child anemia. We also found statistically significant variation across states for the chance of being at or below a certain hemoglobin level. The intra-class correlation coefficient, ICC = 0.0577, indicates approximately 6% of the total variation in the response variable explained by the state of residence. This is a sizeable contribution in explaining the total variability in the response variable.

In contrary to our findings, several other studies (le Cessie et al., 2002; Unsal, Bor, Tozun, Dinleyici, & Erenturk, 2007) indicated that male children were at greater risk of anemia than female children. Also, children living in rural areas were at greater risk of anemia as compared to their urban counterparts. Results also show that children belonging to households with low and medium standards of living index are more susceptible to anemia compared to their counterparts which corroborates the findings in Brazil and other countries (Assunção et al., 2007; Mamiro et al., 2005). We also notice that there is a declining tendency of the prevalence of anemia after two years of age. This may be due to the iron intake which improves with age as a result of a more varied diet, including the introduction of meat and other hemoglobin-containing diets.

Interestingly, the prevalence of any form of anemia is low among illiterate mothers. Results also indicate that the highest prevalence rate of any form of anemia was among the Muslim children (62%) followed by Hindus (56%) and Christians (48%). The reason might be that the children from Muslim and Hindu religions are from lowest quintiles and illiterate mothers. The fertility of the mother impacted anemia in children i.e. the higher number of children of the mother increases the requirement for childcare, demand for food, and inadequate supply of nutritional diet to all the children which ultimately make the children more vulnerable to the risk of anemia (Singh & Patra, 2014)

Mother's age at marriage had also a considerable effect on anemia in their children. Children of women who got married between 18 and 26 years were at greater risk of anemia (Agho, Dibley, D'Este, & Gibberd, 2008; Unsal et al., 2007).

The results from this study indicate that ignoring the hierarchical nature of the data could result in over statement of the significance of some of the variables included in the model. Most importantly, the standard errors would be biased downward. Anemia is a widespread health problem in India, especially in less developed regions like northeastern states of India. The strength of the paper is that it provides important information for policy-makers, program planners, and implementers that seek to reduce childhood anemia in northeastern states of India. The study findings have some important and relevant policy messages. High prevalence of mild and moderate anemia demands multiple interventions and strategies to tackle the burden of anemia. Special policies must be formulated and executed for prevention of anemia.

The present study demonstrates the endemicity of childhood anemia in northeastern states of India which may be due biological, social, and cultural factors. However, in this study, mother's age at marriage and the number of children ever born are the factors significantly associated with anemia in children. Results also suggest that age of the child is also a significant predictor, as the child ages to 48 months or more, the likelihood of severe anemia decreases. This study throws light on the fact that anemia intervention needs to focus more at state level followed by individual level. The limitations of the study are that indicators such as mother's anemia level, serum ferritin, dietary intake, worm infestations, malaria, and infectious diseases (Sinha, Deshmukh, & Garg, 2008) have not been considered.

Acknowledgments
All suggestions from the associated editors and reviewers will be acknowledged as appropriate.

Funding
The authors received no direct funding for this research.

Author details
Sanku Dey[1]
E-mail: sanku_dey2k2003@yahoo.co.in
Enayetur Raheem[2]
E-mail: enayetur.raheem@unco.edu
ORCID ID: http://orcid.org/0000-0001-8880-3649
[1] Department of Statistics, St. Anthony's College, Shillong, Meghalaya 793001, India.
[2] Applied Statistics and Research Methods, University of Northern Colorado, Greeley, CO 80639, USA.
[§] Sanku Dey identified the research problem. Both authors have developed the plan of analysis. Enayetur Raheem carried out the analysis. Both authors have contributed in writing the revising the manuscript.

Citation information
Cite this article as: Multilevel multinomial logistic regression model for identifying factors associated with anemia in children 6–59 months in northeastern states of India, Sanku Dey & Enayetur Raheem, *Cogent Mathematics* (2016), 3: 1159798.

References

Agho, K. E., Dibley, M. J., D'Este, C., & Gibberd, R. (2008). Factors associated with haemoglobin concentration among Timor-Leste children aged 6–59 months. *Journal of Health, Population and Nutrition, 26*, 200–209.

Assunção M. C. F., Santos, I., de Barros A. J. D, Gigante, D. P., & Victora, C. G. (2007). Anemia in children under six: Population-based study in Pelotas, Southern Brazil. *Revista de Saúde Pública, 41*, 328–335. http://dx.doi.org/10.1590/S0034-89102007000300002

Bharati, S., Pal, M., Chakrabarty, S., & Bharati, P. (2015). Socioeconomic determinants of iron-deficiency anemia among children aged 6 to 59 months in India. *Asia-Pacific Journal of Public Health, 27*, NP1432–NP1443. doi:10.1177/1010539513491417

de Benoist, B., McLean, E., Egll, I., & Cogswell, M. (2008). Worldwide prevalence of anaemia 1993–2005: WHO global database on anaemia, vi–41. Retrieved from http://apps.who.int/iris/bitstream/10665/43894/1/9789241596657_eng.pdf

Dey, S., Goswami, S., & Dey, T. (2013). Identifying predictors of childhood anaemia in North-East India. *Journal of Health, Population and Nutrition, 31*, 462–470.

Dubey A. (1994). Iron deficiency anaemia: Epidemiology, diagnosis and clinical profile. In H. Sachdev & P. Choudhury, (Eds.), *Nutrition in children: Developing country concerns* (pp. 217–235). New Delhi: B.I. Publications.

Ene, M., Leighton, E. A., Blue, G. L., & Bell, B. A. (2014). *Multilevel models for categorical data using SAS® GLIMMIX: The* *basics. SAS Global Forum*. Retrieved from http://analytics.ncsu.edu/sesug/2014/SD-13.pdf

Gillespie S. (Ed.) (1997). *Malnutrition in South Asia: A regional profile* (189 p). Kathmandu: United Nations Children's Fund.

Goswmai, S., & Das, K. K. (2015). Socio-economic and demographic determinants of childhood anemia. *Jornal de Pediatria, 91*, 471–477. http://dx.doi.org/10.1016/j.jped.2014.09.009

Griffiths, P., Matthews, Z., & Hinde, A. (2002). Gender, family, and the nutritional status of children in three culturally contrasting states of India. *Social Science & Medicine, 55*, 775–790. http://dx.doi.org/10.1016/S0277-9536(01)00202-7

IIPS. (2000). *National Family Health Survey (NFHS-2), 1998–99*. Retrieved from http://dhsprogram.com/pubs/pdf/FRIND2/FRIND2.pdf

IIPS. (2007). *National Family Health Survey (NFHS-3), 2005–06*, International Institute for Population Sciences (IIPS) and Macro International. Retrieved from http://pdf.usaid.gov/pdf_docs/PNADK385.pdf

le Cessie, S., Verhoeff, F. H., Mengistie, G., Kazembe, P., Broadhead, R., & Brabin, B. J. (2002). Changes in haemoglobin levels in infants in Malawi: Effect of low birth weight and fetal anaemia. *Archives of Disease in Childhood - Fetal and Neonatal Edition, 86*, F182–F187. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11978749

Lokeshwar, M., Manglani, M., Rao, S., Patel, S., & Kulkarni, M. (1990). Iron deficiency—Clinical manifestation and management. In M. Mehta & M. Kulkarni, (Eds.) *Proceedings of national symposium cum work shop on child nutrition: The Indian scene* (pp. 269–275). Bombay: Division of Gastroenterology and Nutrition, LTM Medical College and Hospital.

Mamiro, P. S., Kolsteren, P., Roberfroid, D., Tatala, S., Opsomer, A. S., & Van Camp, J. H. (2005). Feeding practices and factors contributing to wasting, stunting, and iron-deficiency anaemia among 3–23-month old children in Kilosa district, rural Tanzania. *Journal of Health, Population and Nutrition, 23*, 222–230.

McLean, E., Cogswell, M., Egli, I., Wojdyla, D., & de Benoist, B. (2009). Worldwide prevalence of anaemia, WHO Vitamin and Mineral Nutrition Information System, 1993–2005. *Public Health Nutrition, 12*, 444–454. http://dx.doi.org/10.1017/S1368980008002401

NNMB. (2004). *Micronutrient survey*. Hyderabad. Retrieved from http://nnmbindia.org/NNMBMNDREPORT2004-Web.pdf

O'Connell, A. A., Goldstein, J., Rogers, H. J., & Peng, C. Y. (2008). Multilevel logistic models for dichotomous and ordinal data. In A. A. O'Connell & D. B. McCoach (Eds.) *Multilevel modeling of educational data* (pp. 199–242). Charlotte, NC: Information Age Publishing,

Oliveira, M. A., Osório, M. M., & Raposo, M. C. (2007). Socioeconomic and dietary risk factors for anemia in children aged 6 to 59 months. *Jornal de Pediatria, 83*, 39–46. http://dx.doi.org/10.2223/JPED.1579

Osório, M. M., Lira, P. I., & Ashworth, A. (2004) Factors associated with Hb concentration in children aged 6–59 months in the State of Pernambuco, Brazil. *British Journal of Nutrition, 91*, 307–314. http://dx.doi.org/10.1079/BJN20031042

Pollitt E. (1991). Effects of a diet deficient in iron on the growth and development of preschool and school aged children. *Food and Nutrition Bulletin, 13*, 110–118.

Pollitt, E., Hathirat, P., Kotchabhakdi, N. J., Missell, L., & Valyasevi, A. (1989). Iron deficiency and educational achievement in Thailand. *American Journal of Clinical Nutrition, 50*, 687–697.

Seshadri, S., Gopaldas, T., Walter, T., & Heywood, A. (1989). Impact of iron supplementation on cognitive functions in preschool and school-aged children: The Indian experience. *American Journal of Clinical Nutrition, 50*, 675–686.

Singh, R. K., & Patra, S. (2014). Extent of anaemia among preschool children in EAG states, India: A challenge to policy makers. *Anemia, 2014*, 868752, 9 p. doi:10.1155/2014/868752

Sinha, N., Deshmukh, P. R., & Garg, B. S. (2008). Epidemiological correlates of nutritional anemia among children (6–35 months) in rural Wardha, Central India, *Indian Journal of Medical Sciences, 62*, 45–54.

Snijders, T. A. B, & Bosker, R. J. (1999). Multilevel analysis: An introduction to basic and advanced multilevel modeling. *Comparative and General Pharmacology*, viii, 266. Retrieved from http://www.amazon.com/Multilevel-Analysis-Introduction-Advanced-Modeling/dp/0761958908

Soemantri, A. G., Gopaldas, T., Seshadri, S., & Pollitt, E. (1989). Preliminary findings on iron supplementation and learning achievement of rural Indonesian children. *American Journal of Clinical Nutrition, 50*, 698–702.

Togeja, G., & Singh, P. (2004). *Micronutrient profile of Indian population*. New Delhi: Indian Council of Medical Research.

Unsal, A., Bor, O., Tozun, M., Dinleyici, E. C., & Erenturk, G. (2007). Prevalence of anemia and related risk factors among 2–11 months age infants in Eskisehir. *Turkish Journal of Medical Sciences, 7*, 1335–1339.