Original Article

Multilevel modeling of geographically distributed vitamin A deficiency

Girdhar G. Agarwal, Sumi Khare*

Department of Statistics, University of Lucknow, Lucknow, India.

Summary **Vitamin A deficiency is a common health problem in developing countries like India. Present study involves data on children aged between 6-36 months from northern part of India collected geographically, to find prevalence and important factors for risk of night-blindness. Both traditional logistic models and multilevel logistic models were applied to achieve our aim. All individual level variables vitamin A diet intake, age, vitamin A capsule intake and awareness about vitamin A were found significant for risk of night-blindness ($p < 0.05$) in individual level analysis. The effect of risk factors for night-blindness was smaller in multilevel modeling as compared to individual level model. The reason is that the previous model takes into account the within-block as well as among-block variations. Multilevel analysis, did not find, individual level variables vitamin A diet intake, awareness of vitamin A and vitamin A capsule intake significant for the outcome variable ($p > 0.10$). There was about 139% change in odd-ratio for vitamin A capsule taken once. Block level variable, average age of subjects in blocks comes out as significant factor ($p = 0.01$) for night-blindness. Thus, this paper demonstrates the usefulness of multilevel modeling in the analysis for epidemiology of disease risk, which is structured in a hierarchiary, with particular reference to geographical analyses of small area data.**

Keywords: Vitamin A deficiency, Multilevel models, Variance components

Introduction

Vitamin A plays an important role in body's defenses against infection. Children are vulnerable to vitamin A deficiency from the time they are born right up to three years of age. During this time, vitamin A deficiency is more lethal as it can cause permanent blindness, even death (*1*). The risks become less in older children, but vitamin A deficiency reduces overall immunity and makes all children susceptible to diseases like measles and diarrhoea. UNICEF estimates that vitamin A deficiency is a public health concern in 72 countries in Asia and Africa including India. The first repeat survey of the National Nutrition Monitoring Bureau (NNMB) in India, conducted during 1988-90 in the same villages that were surveyed earlier during 1975-79 showed that the prevalence of bitot's spot has declined from 1.8 percent to 0.7 percent (*2*). However, the second repeat survey conducted in 1996-97 showed no further improvement (*3*) and the prevalence is still above 0.5 percent, which is the WHO cut off level for a public health problem (*4*). The national averages do not give a full picture because the prevalence rates vary widely, not only between the states but also within a state (*5*). In many areas of public health research including night-blindness prevalence, the data structures are often hierarchical in nature. Generally, two statistical procedures are used to deal with these types of data. The first is to disaggregate all higher order variables to the individual level and carry out the analysis at individual level. Thus, the assumption of the independence of observations that is basis to the classical statistical technique becomes invalid. The other is to aggregate individual's level variables to higher level and does the analysis at higher level. Thus, all the within group variation, which may account for as much as 80% or 90% of the total variation is discarded before the analysis is carried out. Consequently, relations between aggregated variables are often much stronger giving

**Correspondence to:* Dr. Sumi Khare, Department of Statistics, University of Lucknow, Lucknow-226007, Uttar Pradesh, India;
e-mail: ksumi_20@rediffmail.com

distorted interpretation at the individual level (*6-15*). A number of papers in the epidemiological literature contend that ignoring the above issues (unobserved group effects) in data analysis produces downward biases in the standard errors of the estimated parameters which lead to erroneous estimates of the impact of the individual variable. This leads in some instances to faulty conclusions (*8-10*).

In this paper, our aim is two fold: (i) find the factors responsible for night-blindness and (ii) use hierarchical model to take into account group effects. The data was collected in a multistage scheme consisting of villages within blocks. The clusters of individuals within villages and villages within blocks naturally tend to have similar outcomes within clusters than between clusters. When the design of the study includes clusters, the inherent correlational structure can effect significance levels if it is not appropriately incorporated in the analysis (*16*). Therefore, two types of analyses were applied for achieving our aim, an individual level analyses as well as multilevel model approach. The factors that were found significant in individual level analysis, does not remain significant when multilevel models was applied. The present work illustrates how the use of multilevel modeling provides a greater insight than is possible from a single level approach that considers hierarchical structure and makes it possible to incorporate variables from all levels which leads to correct analysis and proper interpretation of data. Unobserved block and village effects were taken into account in multilevel models. It also takes into account individual level variable and block level variable of the subjects under study. Therefore, we define our objectives as: (i) to find block wise and overall prevalence of night-blindness using two types of model, and (ii) to illustrate why multilevel analyses are more appropriate and considered before doing any such analyses.

Methods and Materials

Data

The present study involved data from Hardoi district, a rural part of North India. The district, with an area of 5,986 sq km has a population of 3,397,414 (*17*). This population is distributed in 19 administrative block and 1883 villages. In first stage, the selection of eight blocks was done using the method of simple random sampling. In the second stage, 25 villages were randomly selected from each block. Finally, 16 households were chosen using systematic and purposeful sampling from each village. Only those households that had at least one child in the eligible age range were selected. Within the selected household, one child was randomly chosen from all the eligible children. Thus, 3,200 children aged 0.5-3 years were selected for the study (*18*). Four well trained teams-each comprising a medical officer and a nonmedical research assistant-conducted the survey. The medical officer examined the child for night-blindness and interviewed the parents for night blindness, immunization and actual doses of vitamin taken by the child during the past year. A questionnaire is used for obtaining data from the parents concerning their awareness of night blindness and dietary habits, focusing on the quantitative estimate of intake of vitamin A rich food. The medical officer examined the child for Xeropthalmia which includes conjunctival Xerosis, Bitot's spot, Active corneal Xerosis, corneal ulceration, cornear Scars. The presence of night-blindness in subject was taken as outcome variable, for which prevalence was found. The possible confounders (lowest level variables) for night-blindness considered were (i) weekly vitamin A diet intake by the child, (ii) age of child (iii) vitamin A capsule intake of child (iv) parents' awareness about vitamin A. The block (highest) level variables considered were (i) number of family health centre, and (ii) the average age in blocks. The later inclusion in the model was done to evaluate the group effect on the coefficients of individual level variables and the variance components.

Model description

A hierarchical structure is presented here. Since only one child was selected from each household, the lowest level was child or household. The households are nested within villages and villages are nested within blocks. This data structure implies that a multilevel analysis may be appropriate, and individual, village and blocks will be taken as first, second and third level, respectively of a multilevel model hierarchy. The multilevel logistic model (*12*) used to estimate the individual and block level variable's effect on outcome variable is

$$\log it\left[p_{ijk}\right] = \gamma_o + \sum_{h=1}^{4} \gamma_h x_{hijk} + \sum_{l=1}^{2} \mu_l z_{lk} + u_{jk} + v_k$$

where p_{ijk} is the probability that i^{th} individual in j^{th} village of k^{th} block has night-blindness, x_h (h = 1, 2, 3, 4) represents individual level variables and z_{lk} (l = 1, 2) represents block level variables in the model. Thus a unit difference between the x_h values of two individuals in the same block is associated with a difference of γ_h in their log odds, or equivalently, a ratio of exp (γ_h) in their odds. u_{jk} is the error term for the deviation from average proportion for j^{th} village in k^{th} block which is assumed to have mean zero and variance τ^2. Similarly v_k is the error term for k^{th} block which is assumed to have mean zero and variance φ^2. It doesn't include a level one residual because it is an equation for the probability p_{ijk} of the outcome rather than for the outcome. The individual-level model does not contain the last three terms of above equation.

The purpose of using this model is to control for

the correlation between subjects in a particular block. In this study, the model is estimated using computer software MLwin (version 1.1) for multilevel analysis (*19*). However, simple logistic regression analysis (*20*) was carried out using SPSS (version 11) software.

Results

The prevalence (in %) for night-blindness among different blocks were given in Table 1. The proportion was as low as 0.3% for block Kachauna whereas it was as high as 7.0% for block Tadiyawan. The overall prevalence rate for night blindness was 2.8%. The overall bitot's spot was found in 0.8% subjects, which is about 1.6 times WHO cut-off point, indicating a public health problem (*4*).

Table 2 presents statistical result when three separate analyses were applied on risk of night blindness. Model A is a simple logistic regression model with four individual level explanatory variables as total vitamin A intake, vitamin A capsule intake, age and awareness about vitamin A. Model B in table II is random intercept model with same variables as in Model A, and model C presents the results when block level variables, number of family health centers and average age of subjects in blocks was added to model B. Each model presents, the odds ratios and 95% confidence intervals for the fixed effects. However, models B and C, also shows the village-level and the block-level variance components with their standard errors. The categories in parentheses in Table 2 were reference groups and the odd-ratios for other categories were considered relative to them. All variables included in model A were found to be significant for the outcome variable (with a maximum *p*-value of 0.04). The night-blindness was two times more likely to be present in the individual whose mother was not aware of vitamin A than the one whose mother has knowledge about vitamin A. There was a protective effect of age, *i.e.* as the age increases the child was more likely to suffer from night blindness. All odd-ratios for individual level factor changed with addition of random intercept in model B *i.e.* when the intercept was allowed to vary between blocks as the blocks differ in proportion of outcome variable. The confidence interval of odd-ratio for all variables gets wider with maximum of 240% for vitamin A capsule taken once. There was 24% reduction in the risk of night-blindness for those who were not aware about vitamin A as compared to model A. Although there was only 1% change in the odd-ratio for the weekly vitamin A diet intake, its confidence limits gets wider by 50% making it non-significant ($p > 0.10$) for the outcome variable. The most dramatic change was in the protective effect of those who had taken vitamin A capsule once, as it was reduced by 139%. Except for the last two categories of age-groups all independent level variables vitamin A diet intake, awareness of vitamin A and the vitamin A capsule intake were found non-significant for the night-blindness in model B with a maximum *p* value change of 0.86 for vitamin A capsule taken once. This change

Table 1. Distribution of night blindness with blocks

Block	Prevalence (%)[a]
Ahirori	2.75
Behendar	3.25
Kachauna	0.25
Madhoganj	0.75
Mallawan	5.78
Sandi	2
Sandila	0.5
Tadiyawan	7
Total	2.78

[a] Total number of subjects in each block is 400.

Table 2. Odd ratios and 95% confidence intervals for fixed effects and coefficients and standard errors for random effects

	Model A	Model B	Model C
Fixed part	*ORs* (*95%CI*)	*ORs* (*95%CI*)	*ORs* (*95%CI*)
Weekly vitamin A diet intake	0.953 (0.913-0.994)[a]	0.97 (0.664-1.495)	0.973 (0.932-1.016)
(Aware about vitamin A)	1.00	1.00	1.00
Not Aware about vitamin A	1.962 (1.012-3.806)[a]	1.50 (0.660-3.400)	1.595 (0.715-3.556)
(Never taken Vitamin A capsule)	1.00	1.00	1.00
Once in last six months	0.383 (0.175-0.835)[a]	0.420 (0.146-1.190)	0.382 (0.139-1.050)
Once in life	0.384 (0.178-0.828)[a]	0.918 (0.297-2.840)	0.784 (0.269-2.282)
Age 6-12 months	0.510 (0.271-0.963)[a]	0.530 (0.271-1.010)	0.509 (0.265-0.979)[a]
13-18 months	0.490 (0.253-0.948)[a]	0.600 (0.311-1.160)	0.579 (0.298-1.125)
19-24 months	0.235 (0.097-0.573)[b]	0.260 (0.190-0.640)[b]	0.250 (0.101-0.618)[a]
25-30 months	0.427 (0.213-0.858)[a]	0.470 (0.240-0.940)[a]	0.455 (0.228-0.907)[a]
(31-36 months)	1.00	1.00	1.00
Block level variable			
Number of Family health centre			0.886 (0.704-1.118)
Average age of block			0.367 (0.165-0.819)[a]
Random part		*Variance component (S.E)*	*Variance component (S.E)*
Φ^2 (at block level)		0.656 (0.469)	0.443 (0.363)
τ^2 (at village level)		4.509 (0.701)	3.468 (0.630)

[a] $p < 0.05$, [b] $p < 0.01$.

would attribute to average proportion difference of night-blindness between the blocks and villages. There was seven times less variation between blocks than between villages within blocks. Here the usefulness of multilevel modeling comes out as if the variation at village and block level were not considered, the factors that were not important taken to be significant, thus affecting the outcome variable. All individual level interaction effects were found to be non-significant for the models considered.

Adding block-level variables in model C, the variability at block-level reduces by 33% and at village-level by 23%. It indicates that a definite amount of variability at block-level was explained by the block-level variable number of family health centers and the average age of subjects in blocks, which was found significant for the outcome variable ($p = 0.01$). There was almost no change in odd-ratio of weekly vitamin A intake, but its confidence limits reduced by 32% with addition of the block-level variables. The risk of night blindness increased by 6% for those who were not aware about vitamin A, but the relationship was still not significant ($p = 0.33$). Interactions between individual and block-level were explored but none was detected.

Discussion

There have been numerous individual/district/state level analyses dealing with prevalence percentage of night-blindness (*21,22*), but to our knowledge, in India, no study so far tried to analyze data using the more appropriate procedure of the present study. It is well known that studies involving a large number of important variables categorized suitably combined with the appropriate analytical procedure, will provide more valid and stable results. This study shows the significant factors for outcome variable along with the consequences of not choosing appropriate analysis. Hence, we say that age was significantly affecting the occurrence of night-blindness along with weekly vitamin A diet intake. There is a slight difference in the interpretation of an odds ratio estimated from a multilevel logistic regression model compared with the one obtained from a standard logistic regression model owing to the addition of the random variation at village and block-level. As for former the coefficients are generalized for a wider population of blocks, not restricted to eight blocks only. The effect of risk factors for night-blindness was smaller in multilevel modeling as it takes into account the within-block as well as among-block variation. The effect of the individual level variable, vitamin A capsule intake on the risk of night-blindness was increased with the addition of block effects. However, this variable was non-significant for night-blindness variable in multilevel analysis ($p > 0.10$). The independent variables were considered to have fixed effect. As, no further improvement in model was observed, when random slope was added to the model *i.e.* when the relationship between night-blindness and independent variables varied across blocks In general, while the direction of effects were similar between two methods considered, multilevel modeling led to wider the confidence intervals of fixed effects, specially of weekly vitamin A intake.

Austin *et al*. paper demonstrated that the confidence intervals of group-level variables gets wider when hierarchical structure of data was incorporated in analysis, whereas it was shown in our case that confidence interval of individual-level variables would also get wider when the variability among villages and among blocks was added to the residual variability among individuals. The implication of this will be that, a variable, which were significant in individual-level analysis does not remain significant when multilevel models, are used.

In summary, this study has demonstrated the potential usefulness of multilevel modeling in epidemiology analysis of disease risk measured across a heterogeneous population. Estimating a relationship between risk of night-blindness and individual level variables without taking into account the hierarchical nature of the data was shown to be mistaken. So, before reporting results for analyses where data has hierarchical structure, multilevel models should always be considered.

Acknowledgements

The USAID micronutrient program, India, supported this research.

References

1. Sommer A, Hussaini G, Tarwotjo I, Susanto D. Increased mortality in children with mild vitamin A deficiency. Lancet 1983; 2:585-588.
2. Ghosh S. Combating vitamin A deficiency - A rational public health approach. NFI Bulletin. January 2001; 22:1.
3. Report of Second Repeat Survey-Rural National Nutrition Monitoring Bureau, National Institute of Nutrition, Hyderabad. 1996-1997.
4. Toteja GS, Singh P. Micronutrient profiles in Indian population. ICMR press, New Delhi, India, 2002; pp. 5-6.
5. Vir SC. Linkage of vitamin A massive dose supplementation with pulse polio immunization– an overview. In: A report on benefits and safety of administration of vitamin A to preschool children and pregnant and lactating women (Srivastava CU, ed.). New Delhi, India, 2000.
6. Bryk AS, Raudenbush SW. Hierarchical linear models. Applications and Data Analysis Methods. Sage Publications, London, UK, 1992.
7. Goldstein H. Multilevel statistical models. Edward Arnold, London, UK, 1995.
8. Drevett RF, Goldstein H. Modeling lactation using an

inverse polynomial in a multilevel statistical model. Stat Med 1994; 13:1643-1655.
9. Steele F, Diamond I. Immunization uptake in rural Bangladesh: A multilevel analysis. J Roy Statist Soc A 1996; 159:289-299.
10. Austin PC, Jack VT, David AA. Comparing hierarchical modeling with traditional logistic regression analysis among patients hospitalized with acute myocardial infarction: should we be analyzing cardiovascular outcomes data differently? Am Heart J 2003; 145:27-45.
11. Lisa A. Student disengagement and the socialization styles of high schools. Social Forces 2005; 84:2.
12. Snijder TAB, Bosker J. Multilevel analysis: An introduction to basic and advanced multilevel modeling. Sage Publications, London, UK, 1999.
13. Goldstein H, Healy MJR. The graphical presentation of a collection of means. J Roy Statist Soc A 1995; 158:175-177.
14. Kreft IG, Leeuw JD. Introducing multilevel modeling. Sage Publications, London, UK, 1998.
15. Dennis.D, Brownie C. Comparing variance and other measures of dispersion. Biometrika 2004; 19:571-578.
16. DeLong E. Hierarchical modeling: Its time has come. Am Heart J 2003; 145:16-18.
17. Agnihori S. Uttranchal and Uttar Pradesh at a glance. Jagran Research Centre Press, Kanpur, India, 2003; pp. 10-11.
18. Agarwal GG, Awasthi S, Walter SD. Intraclass correlation estimates for assessment of vitamin A intake in children. J Health Popul Nutr 2005; 23:66-73.
19. Rasbash J. Multilevel models project 2001. Institute of Education, University of London.
20. Hosmer DW, Lemeshow S. Applied logistic regression. John Wiley & Sons, New York, USA, 1989.
21. Toheja GS, Singh P, Dhillon BS, Saxena BN. Vitamin A deficiency in 16 districts in India. Indian J Pediatr 2002; 69:603-605.
22. Swami H, Jhakur JS, Bhatia, Ahuja R. Rapid assessment and delivery of vitamin A to slum children by using national immunization day in Chandigarh. Indian J Pediatr 2001; 68:719-723.

(*Received December 15, 2007; Revised December 28, 2007; Accepted January 9, 2008*)