International Journal of Applied Mathematics
& Statistical Sciences (IJAMSS)
ISSN(P): 2319-3972; ISSN(E): 2319-3980
Vol. 6, Issue 2, Feb-Mar 2017; 43-54

International Academy of Science, Engineering and Technology
Connecting Researchers; Nurturing Innovations

MULTILEVEL ANALYSIS OF SURVEY DATA

A. SRIVIDYA[1] & M. KUMARAN[2]

[1]Vector Control Research Centre, Pondicherry, India

[2]Department of Statistical Sciences, Kannur University, Kannur, Kerala, India

ABSTRACT

Most public health related surveys are based on complex survey methods that result in data with hierarchical structure leading to dependence between observations. Conventional statistical methods are used to analyse such data thereby result in imprecise model estimates and inferences. Hierarchical model modelling represents statistical method used to analyse nested data. We have used one such survey data on one of the vector borne disease namely lymphatic filariasis, for which multistage cluster sampling is used. Both single level conventional logistic regression model and models accounting for the hierarchical data structure were fitted to the data. Comparisons were made in terms of estimated coefficients, their standard errors and goodness of fit measures. Random effects models showed that 25% of the variation in micro filarial status was accounted due to the differences between villages. Two level models performed better than the single level model. The choice of using a multilevel model for small area data and its limitations are discussed.

KEYWORDS: Logistic Regression, Multilevel Regression, Hierarchical Data, Filariasis

INTRODUCTION

Most of the epidemiological and health related studies collect data using multi stage sampling designs[1-6]. These data give raise to a hierarchical data structure: patients within hospitals, individuals within villages, and therefore need to be given special attention, while planning analysis and interpretation[7-11]. In this hierarchical structure, the clustering of patients within hospitals, individuals within villages lead to observations that are no longer independent as the lower level units within the higher level units are correlated to each other. Hence it is only logical to apply multilevel analytical methods to analyse such data to account for the effect of clustering[12] and to study the variability at different levels. While analysing so, one can simultaneously examine the effects of different levels on the outcome variable measured at the lowest level prior to making any inferences[11]. Particularly, though it is known that disease causation and its distribution are influenced by social context (namely environment, neighbourhood), often it has been assumed that the determinants of health are characterized only by the characteristics of the individuals themselves[13]. More often, when group of individuals are being used to collect data, it is often assumed that there is absolutely no interaction between the attributes of individuals and the attributes of the group in which the individual lives which is not true. And particularly, when we examine differences between groups, we need to take into consideration of the composition of the units in the group[10, 14].Multilevel statistical analysis is an approach that can deal with data having a natural hierarchical structure. With this approach, the higher level factors will be accounted for while addressing the causation factors at lower level units[11].

Vector borne diseases are of public health importance and lymphatic filariasis is one of the most debilitating vector borne disease causing severe morbidity to the individuals and economic loss to the nation[15, 16].Nearly 1.4 billion people in 73 countries worldwide are endemic for lymphatic filariasis, a parasitic infection that leads to a disease

commonly known as elephantiasis[17].In India, mass drug administration (MDA) programme with albendazole (400 mg) together with diethyl car bamazine citrate (DEC) (6 mg/kg)was initiated in 2000 in certain endemic districts, with an aim to elimination of the disease and eventually they were up scaled to 255 districts[18]. The entire population in the district is administered the drugs annually. The impact of this control strategy is measured through microfilaria surveys conducted annually prior to each round of MDA, right from baseline, prior to the start of MDA[18]. These surveys use multistage sampling design where villages/wards are randomly selected within a district and within each of these selected villages/wards, a random sample of individuals are surveyed for the presence of microfilaria, the parasite causing the disease.

In this study we have used the data from one such a microfilaria survey carried out prior to MDA (from the Department of Public Health Tamil Nadu) in one district of Tamilnadu to compare the results of analyses done with the traditional models with those obtained using a two level multilevel model. Comparisons are made in terms of coefficients (odds ratios) obtained from the fitted models and their standard errors, significance of the predictors both at village and individual level in relation to occurrence of microfilaria in different villages/wards and discussed the advantage of the latter over the former.

DATA AND METHODS

The data is from a cross sectional microfilaria survey carried out annually in the district of Thiruvannamalai carried out in the year 2000, prior to the implementation of MDA with DEC/ALB. Villages/wards were randomly selected and surveys were carried out in a random sample of individuals of these selected villages/wards. Individual level data like age, sex, microfilaria count were available. At the village level, the variable that was available was population density of the selected village/ward (obtained from Census India Website).This data has two levels: level 1 – individuals and level 2 – villages and hence was used demonstrate the application of hierarchical models and compare the results with those obtained through the conventional methods like logistic regression analysis. The outcome variable was mfstatus (Y_{ij}) (categorical variable: 0 = no filariasis, 1=filariasis) of the ith individual in the jthvillage and predictor variables were age(denoted by $x1_{ij}$), gender(denoted by $x2_{ij}$), type of residence (urban or rural)(denoted by $x3_{ij}$)for the ith individual in the jth village and population density(z_j)(village/ward level)

Our objective here is to see the possible association of the above mentioned variables on the occurrence of microfilaria. As the outcome variable is binary, we have attempted to fit two forms of logistic models: the traditional models that does not account for multilevel data structure and other ones that accounts for the multilevel data structure. The coefficients estimated of these models with their standard errors and 95% limits are estimated and the model fits are compared to see the performance of these models. Forth coming paragraphs briefly describe the basics of the models.

Logistic Regression Model

The most commonly used generalized linear model of a dichotomous response variable is a logistic regression and is specified by the probability distribution is binomial (μ) with mean μ, a linear predictor is the multiple regression equation η ie $\eta = \beta_0 + \beta_1 x_1 + \beta_2 x_2$, and the link function is the logit function given by η=logit(μ). The outcome in logistic regression analysis is often coded as 0 or 1, where 1 indicates that the outcome mf status is positive, and 0 indicates that the mf status is negative. If we defineπ_i as the probability of being mf positive then

$$logit(\pi_i) = logit\left(\frac{p_i}{(1-p_i)}\right) = \beta_0 + \beta_1 x1_i + \beta_2 x2_i + \beta_3 x3_i \quad (1)$$

Where p_i is the probability of being mf positive and 1-p_i is the probability of being mf negative. In other words, the log odds of the outcome, is modelled as a linear combination of the predictor variables in the data. The regression coefficients are estimated by the method of maximum likelihood. We have fitted logistic regression model with robust errors (unclustered assumption) and cluster robust errors (accounting for clustering).

Models Accounting for Multilevel Structure

A generalized linear model for a dichotomous response variable accounting for the multilevel structure is specified by probability distribution for π_{ij} is binomial(μ, n_{ij}) with mean μ; the linear predictor is the multiple regression equation η eg $\eta = \gamma_{00} + \gamma_{10} x_{ij} + \gamma_{01} z_j + \gamma_{11} z_j x_{ij} + u_{1j} x_{ij} + u_{0j}$ and the link function is the logit function given by η=logit(μ)

Accordingly, the two level model can be written as

$$logit(\pi_{ij}) = \gamma_{00} + \gamma_{10} x1_{ij} + \gamma_{20} x2_{ij} + \gamma_{30} x3_{ij} + \gamma_{01} z_j + u_{0j} \quad (2)$$

u_{0j} ,are the errors associated with random effects, namely the communities, with mean 0 and variance denoted by $\sigma^2_{u0.}$ While the fixed part is specified by $\gamma_{00}, \gamma_{10}, \gamma_{20}, \gamma_{01}$, the random part is specified by $u_{0j.}$ Estimation of variance of the random effect terms helps in understanding the variation in the response variable that is occurring due to the communities ie the level-2 units. As an outcome of this multilevel analysis, a quantity termed as intra cluster correlation (ICC) ρ is also estimated. ICC is defined as the proportion of variance explained by the grouping structure in the population and is given by

$$\rho = \frac{\sigma^2_{u_0}}{\left(\sigma^2_{u_0} + \sigma^2_e\right)} \quad (3)$$

Where $\sigma^2_{u_0}$ is the variance of the level-2 errors, $u_{0j.}$The intra class correlation ρ can also be interpreted as the correlation between two randomly selected level-1 units in a randomly selected level-2 unit.. The possible approaches to model the random effects ie level-2 effects using various types of models are briefly described below.

A Random Intercept Log it Model Approach

This is basically a generalized linear log it model where the intercept is modelled in such a way that it is allowed to vary across the level-2 units , while controlling for level-1 predictors. In our case, from the equation (2),

$$logit(\pi_{ij}) = \boldsymbol{\gamma_{00}} + \gamma_{10} x1_{ij} + \gamma_{20} x2_{ij} + \gamma_{30} x3_{ij} + \gamma_{01} z_j + \mathbf{u_{0j}} \quad (2)$$

The term $\gamma_{00} + u_{oj}$is considered the intercept where u_{0j} varies with village. The parameters estimated in these models are cluster specific, here community specific parameters.

A Latent Variable Approach

The above mentioned random intercept log it model can also be modelled by introducing a latent variable (unobserved variable) that is continuous and measures the effect of level-2 units on the response variable. This model is specified as follows.

Suppose that y_{ij}^* is an unobserved variable (e.g., \propensity" to contract diseases), and that we observe y_{ij} as

$$y_{ij} = \begin{matrix} 1 \; if \; y_{ij}^* > 0 \\ 0 \; if \; y_{ij}^* \leq 0 \end{matrix} \quad (4)$$

Then the threshold model is given by

$$y_{ij}^* = \beta_{0j} + \beta_1 x1_{ij} + \beta_2 x2_{ij} + u_j + e_{ij}^* \quad (5)$$

With

$$var\left(y_{ij}^* \middle| x_{ij}, u_j\right) = var\left(u_j\right) + var(e_i^*) = \sigma_u^2 + 3.29$$

where the residual variance is fixed if e_{ij}follows a standard logistic distribution with mean 0 and variance is ($\pi^2/3 = 3.29$).Further the quantity variation partitioning coefficient (VPC) is given by

$$VPC = \frac{\sigma_{u_0}^2}{\left(\sigma_{u_0}^2 + \left({}^{\pi}/_{3}\right)\right)} \quad (6)$$

and is termed as the marginal intra class correlation between the 'latent responses'.

Marginal or Population Averaged Model Approach

Another alternative to a random effects model allowing for clustering is a marginal model. This alternative approach, accounts for clustering and adjusts standard errors.

$$logit\left(\pi_{ij}\right) = \gamma_{00} + \gamma_{10}\text{x1}_{\text{ij}} + \gamma_{20}\text{x2}_{\text{ij}} + \gamma_{30}\text{x3}_{\text{ij}} + \gamma_{01} z_j \quad (7)$$

Here the clustering effect is seen as a nuisance and hence there is no parameter representing the variation between level-2 unit and no estimation of group effects. Marginal models are usually estimated using a method called Generalised Estimating Equations (GEE), and the models themselves are sometimes called GEE models. Also, here the parameters that are estimated are population averaged parameters.

We carried out sensitivity analyses by comparing three single level model (logistic regression models with option of robust errors and errors adjusted for clustering) with the two level random logit model, latent variable model and marginal model.

Statistical Software

Three logistic models (with default errors, robust errors and cluster adjusted errors), two level random effects models, a latent variable model using gllamm (Generalized latent linear and mixed models)[10, 19]and population averaged model were fitted using the statistical software STATA SE (version 9.0), Stat Corp, U.S.A. We fitted these models to the filariasis survey data and compared the performance of these models to see, how the effect of hierarchical structure of the data on the regression coefficients and variance components. In these models we have assumed no interactions between predictors at levels 1 and 2.

RESULTS

Basic Characteristics

Data from the microfilaria surveys were conducted in 14 villages/wards of Thiruvannamalai district in 2000 prior to the implementation of MDA were used for this analysis.

The details that were available for each individual were age, gender, type of residency. As for the village/ward, the only variable available was population density(Table 1).The details of villages/wards and their microfilaria rates with village\ward wise sample with microfilaria positives are shown in Table 2. The response variable as the microfilaria (Mf) status in an individual. A total of 2882 individuals were examined in the survey from 14 villages/wards of Thiruvannamalai health unit district in 2000. Though the overall microfilaria was 3.3% it ranged between 0.0 to 11.4% across these villages/wards.

MODEL FITS

Logistic Regression Model

As mentioned we started out with fitting the logistic regression model with predictors like age, sex, type of residence and population density of the village/ward.The results of the fit of single level logit model with robust errors and errors adjusted for group (clustering) variable are summarized in Table 3. It may be seen that while only gender was not significantly associated with microfilaria prevalence, age, type of residency and population density were significantly associated with occurrence of Mf for the model assuming robust errors and that living in rural areas increased the risk of being Mf positive by 8.3 times.For the model in which the errors are adjusted for clustering, only gender and type of residency were significant.It also showed that the risk of Mf increased by 1.5 times in males when compared to females and 8.3 times higher among those who lived in rural areas than those living in urban areas.

All the single level models were significant, as observed by the significant chi square value of 56.51 (with p value<0.0001).

RANDOM EFFECTS MODELS ACCOUNTING FOR TWO LEVEL STRUCTURE

Random Intercept Log It Model

This two level model accounts for the level 1 individuals nested within level 2 village/ward. To start with a null two level model with constant alone was fitted with default 12 integration points. It was observed that the log-odds of being mF positive in an "average" village is -3.50andthe variance of the random errors u_{0j} is estimated as$\sigma_u{}^2$ = 1.05 with a rho value of 0.24 which was statistically significant with a chi square = 44.2and p value <0.001 (not shown in Table 3). Figure 1 shows the caterpillar plot that depicts the village (random) effects with rank of predicted village level random effect in the X axis and the predicted the village random intercept with 95% confidence intervals in the Y axis. It may be seen of the 14 villages/ward, some villages are below 0, some overlap over 0 and some are significantly above the 0, clearly indicating difference sat village level.

With the inclusion of other predictors like age (centered), sex and the type of residency in the subsequent model the between village level variance significantly reduced to 0.25 (0.154).Of all the predictors, only type of residency was a significant predictor, the risk increased from 8.3 to 10.7 for those living in rural areas. There was also a substantial increase in their standard errors too (Table 3). Though the residual intra class correlation reduced from 0.24 to 0.072, it was still significant, showing that 7.2% of the differences in the microfilaria rates were attributable to differences between

villages/centres. Further, both -2log likelihood and AIC, measure of goodness of fit of models, indicate that a two-level model is a better option than a single level model(Table 3).

Random Intercept Model as a Latent Variable Model

The effect of the village on the response variable is an unobserved variable and the results of this latent variable model were very similar to those obtained to two level random intercept model as seen in Table 3. The odds ratios and the model fit results are almost the same as that of the two-level random intercept log it model. Here the variation partitioning coefficient (VPC), also known as the marginal ICC also had almost the same values as those obtained through the random effects log it model.

Marginal or Population Averaged Models

The odds ratios estimated under this model is shown in Table 2. It may be seen that the values estimated under this model is less that those estimated by the multilevel models, however the only predictor significant was the type of residency. It may be seen that there are separate estimates of variances due to level-2 random effects, however, the standard errors are adjusted for the clustering variable, namely villages/wards.

DISCUSSIONS

This study compared two analytical methods to evaluate the differences in the microfilaria rates, namely the traditional single level models vs two-level hierarchical models. The results have shown that the effects of villages/wards were sensitive to the type of method used for analysis. The two level hierarchical model confirmed the findings of the single level model for most predictors and how misleading the inferences could be if one used the traditional model. The two level model showed that population density was no longer a significant predictor and also resulted in higher odds ratio values for the type of residency compared to the single level model. The analysis above has shown clearly that when a data with the hierarchical structure is analysed ignoring the nested nature of the data, there is a possibility losing information on the variable we are interested in. The conventional single level model, assumes independence of observations and the errors and this leads to spurious significance to predictors providing smaller standard errors (Table 3). Studies elsewhere have shown that age is a significant factor associated with microfilaria[20-22] but our study did not support that finding. Though the single level model showed age as a significant predictor, while adjusting for clustering that significance is lost suggesting the influence of the village level aspects. It has been reported elsewhere multilevel modelling techniques are the most appropriate statistical method for dealing with outcomes collected from individuals clustered within groups and in particular when there is a great heterogeneity in sample sizes[23-26] as has been observed in our case (sample size range: 100-500).

Application of the multilevel model to this data has brought out facts that could have been missed while using a conventional single level model. The finding that the 7.1% the variation in microfilaria rates were attributable to difference between the villages/wards indicate that there are some village/ward level factors that result in heterogeneous pattern of microfilaria rates within a district. First and the foremost is that it allows us to take into account of the clustering effect i.e the interdependence between the observations in a village and has provided an estimate measuring the correlation among them in terms of significant intra correlation coefficient (ICC). Microfilaria surveys conducted by the national programme to assess the impact of MDA are of hierarchical in nature.

In this analyses, four models namely the cluster adjusted logistic model, random intercept log it model, latent variable model and the marginal model accounted for the random effects due to clustering phenomena. While the first one is a single level model adjust the standard errors of the regression coefficients for effects clustering, the next two provide an estimate for the variation due the unobserved level-2 (here village/ward) clustering effects, the last one models the clustering effect as considered a nuisance and the standard errors of the model coefficients are adjusted for overall grouping effect. As the goodness of fit measures namely AIC and the -2log likelihood are lowest for the two-level models (random effects log it and latent variable model), this analysis shows the most suitable model for this type of data are the multilevel models. The choice of models to a hierarchical data depends on what the objective of the study is. If the inferences are to be made at population level which does not require the group specific effects, then it would be advisable to use population averaged models. If on the other hand, one is interested in the group specific effects on an outcome variable, it is only more appropriate to use multilevel models with random effects or random coefficients.

This analysis has highlighted the possibility of operation of village level factors for the observed heterogeneity in the microfilaria rates among the villages/wards using multilevel analytical method with few predictors. These results may be corroborated with heterogeneous spatial distribution of filarial infection in the villages of Tamilnadu reported earlier[27]. This finding has clearly pointed out that in addition to the individual level variation, the effect of the community level factors [28]where these individuals live in also matter for the occurrence of micro filariaas has been reported elsewhere. An analysis with this approach to the data collected prior to the introduction of the large scale MDA programme has shown the microfilaria rates are impacted by village/ward effects. Analysis of post MDA data in similar lines can be carried to see if the same trend continues or the impact of MDA has reduced the differences between villages ie bring down the microfilaria rates uniformly to a low level, which also could be indirectly considered as the success of the MDA programme.

The limitation of this study is that we had only data for 14 villages/wards and also information of only few predictors at individual level and only one predictor at village level. As this data was collected as a part of baseline evaluation by the health department, we could not get more information like other household factors like proximity to mosquito breeding, and village level factors like clogged drains, presence of stagnant water bodies that aid breeding of vector mosquitoes on the surveyed individuals. However, the results of this analyses have provided as scope for application of multilevel analysis on filariasis data and sed to explore the role of macro level factors on the occurrence of microfilaria in a village\ward.

ACKNOWLEDGEMENTS

The authors acknowledge the Director, Department of Public Health and Family Welfare, Tamilnadu for providing the data on the microfilaria surveys conducted in the district of Thiruvannamalai Health Unit district for this exercise. Profound thanks are due to Dr. Soumya Swaminathan, Director, National Institute for Research in Tuberculosis, without whose help accessing of this data would have been possible.

Table 1: Characteristics and Codes of the Data Used

Variables	Description of Variables
Dependent variable	
Mf status-Y_{ij}	mf status: Mf positive=1 ; mf negative=0
Predictors	
At Individual level	
Age - $X1_{ij}$	Age of the ith individual sampled in jth village/ward
Gender -$X2_{ij}$	gender status : Female=0 and male=1 of the ith individual in jth village/ward
Type of residence-$X3_{ij}$	urban=0 and rural=1; residency status of the ith individual in the jth village/ward
At village level	
Population density (per km^2)	Population density of the j^{th}village/ward

Table 2: Microfilaria Rates among the Surveyed Villages/Wards in 2000

No	Village / Wards	No. Surveyed	No mf pos	Mf Rate (%)
1	DEVANOOR	100	8	8.0
2	JADATHARIKUPPAM	100	11	11.0
3	NEIVANATHAM	100	8	8.0
4	POONDI	114	5	4.4
5	SU.VALUVETTI	200	5	2.5
6	TANDARAI	200	11	5.5
7	THANIPADI	500	8	1.6
8	THIRUMALAI	177	10	5.6
9	ULAGAMPET	100	10	10.0
10	VILVARANI	103	9	8.7
11	WARD 1	444	6	1.4
12	WARD 14	324	0	0.0
13	WARD 24	125	0	0.0
14	WARD 7	295	5	1.7
Grand Total		**2882**	**96**	3.3

Table 3

	Single Level Model				Two Level Models				Population Averaged Model	
Variables	**Logistic Model (Robust Errors)**		**Logistic Model (Cluster Adjusted)**		**Random Effects Model**		**Latent Variable Model**			
	Odds Ratio	**p value**	**Odds Ratio**	**p value**	**Odds Ratio**	**p value**	**Odds Ratio**	**p value**	**Odds Ratio**	**p value**
Fixed effects										
Age (Centered around mean)	1.012 (0.005)	0.028	1.012 (0.006)	0.055	1.012 (0.006)	0.049	1.012 (0.006)	0.049	1.011 (0.006)	0.053
Gender (Male coded 1)	1.479 (0.315)	0.066	**1.479 (0.138)**	**0.000***	1.466 (0.313)	0.074	1.466 (0.313)	0.073	1.427 (0.283)	0.072
Residency (Rural coded 1)	**8.333 (2.787)**	**0.000***	**8.333 (3.973)**	**0.000***	10.744 (6.187)	**0.000***	10.716 (6.121)	**0.000***	9.005 (6.589)	**0.003***
Population density	0.999 (0.0004)	0.003*	0.999 (0.001)	0.198	0.999 (0.001)	0.203	0.999 (0.001)	0.200	0.999 (0.001)	0.506

(Centered around mean)										
Random effects										
Village - σ_{u0}^2					0.2540 (0.165)		0.2558 (0.169)			
ICC (rho)					0.0718 (0.044)					
Variance partitioning coefficient (VPC)							0.0721			
-2log likelihood value	785.4192		785.4192		**775.5600**		**775.5600**			
LR chisqr	49.99		50.01							
p value	<0.00001		<0.00001							
AIC	795.42		795.42		**787.56**		**787.56**			
LR test for rho, chi sqr					9.9000					
p value					<0.0001					

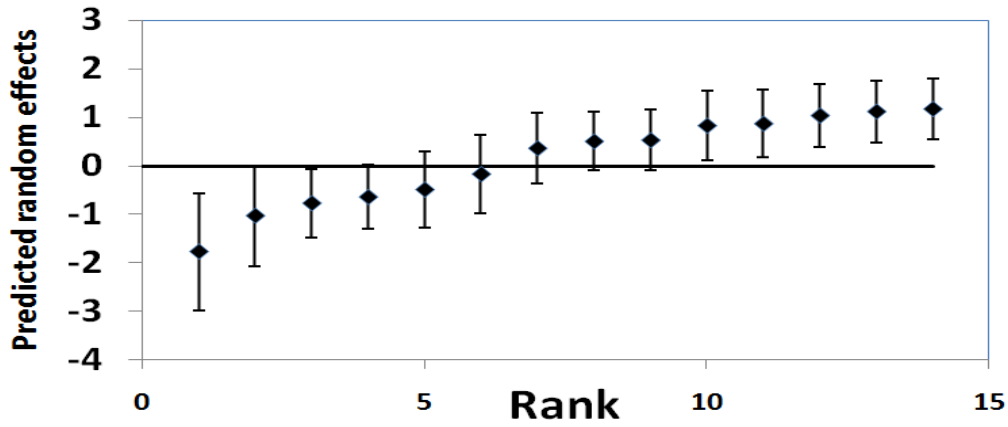

Figure 1: Predicted Random Effects of a Null Model

REFERENCES

1. Bingenheimer, J.B., *Multilevel models and scientific progress in social epidemiology.* J Epidemiol Community Health, 2005. **59**(6): p. 438-9.
2. Ha, I.D. and Y. Lee, *Multilevel mixed linear models for survival data.* Lifetime Data Anal, 2005. **11**(1): p. 131-42.
3. Lumme, S., A.H. Leyland, and I. Keskimaki, *Multilevel modeling of regional variation in equity in health care.* Med Care, 2008. **46**(9): p. 976-83.
4. Rice, N. and A. Jones, *Multilevel models and health economics.* Health Econ, 1997. **6**(6): p. 561-75.
5. Rice, N. and A. Leyland, *Multilevel models: applications to health data.* J Health Serv Res Policy, 1996. **1**(3): p. 154-64.

6. Xing, J.N., S.S. Qian, and L. Wang, *[Application and progress of multilevel models in epidemiological research]*. Zhonghua Liu Xing Bing Xue Za Zhi, 2013. **34**(1): p. 98-100.
7. Behm, J.E., et al., *Multilevel statistical models and the analysis of experimental data*. Ecology, 2013. **94**(7): p. 1479-86.
8. Chung, H. and S.N. Beretvas, *The impact of ignoring multiple membership data structures in multilevel models.* Br J Math Stat Psychol, 2012. **65**(2): p. 185-200.
9. Clarke, P., *When can group level clustering be ignored? Multilevel models versus single-level models with sparse data*. J Epidemiol Community Health, 2008. **62**(8): p. 752-8.
10. Rabe-Hesketh, S.a.S., A, *Multilevel modelling of complex survey data*. J. R. Statist. Soc. A, 2006. **169**(Part 4): p. 805–827.
11. Snijders, T.a.B., RJ, *Multilevel analysus. An introduction to basic and advanced multilevel modeling*. 2011.
12. Blundell, R. and F. Windmeijer, *Cluster effects and simultaneity in multilevel models.* Health Econ, 1997. **6**(4): p. 439-43.
13. Fine, P.E., *Herd immunity: history, theory, practice.* Epidemiol Rev, 1993. **15**(2): p. 265-302.
14. Carle, A.C., *Fitting multilevel models in complex survey data with design weights: Recommendations.* BMC Med Res Methodol, 2009. **9**: p. 49.
15. Ramaiah, K.D., et al., *Treatment costs and loss of work time to individuals with chronic lymphatic filariasis in rural communities in south India*. Trop Med Int Health, 1999. **4**(1): p. 19-25.
16. Ramaiah, K.D., et al., *The impact of lymphatic filariasis on labour inputs in southern India: results of a multi-site study.* Ann Trop Med Parasitol, 2000. **94**(4): p. 353-64.
17. Ichimori, K. and A. Crump, *Pacific collaboration to eliminate lymphatic filariasis.* Trends Parasitol, 2005. **21**(10): p. 441-4.
18. NVBDCP, *National vector borne disease control programme: annual report 2014*. 2014, Department of Health and Family Welfare: New Delhi.
19. Skrondal, A.a.R.-H., S, *Some applications of generalized linear latent and mixed models in epidemiology: Repeated measures, measurement error and multilevel modeling.* Norsk Epidemiologi, 2003. **13**(2): p. 265-278.
20. Brabin, L., *Sex differentials in susceptibility to lymphatic filariasis and implications for maternal child immunity.* Epidemiol Infect, 1990. **105**(2): p. 335-53.
21. Das, P.K., et al., *Frequency distribution of Wuchereria bancrofti microfilariae in human populations and its relationships with age and sex.* Parasitology, 1990. **101 Pt 3**: p. 429-34.
22. Vanamail, P., et al., *Estimation of age-specific rates of acquisition and loss of Wuchereria bancrofti infection.* Trans R Soc Trop Med Hyg, 1989. **83**(5): p. 689-93.

23. Bingenheimer, J., Raudenbush SW., *Statistical and substantive inferences in public health issues in the application of multilevel models.* Ann Rev Public Health, 2004. **25**: p. 53-77.

24. Diez-Roux, A., *Multilevel analysis in publich health research.* Annu Rev Public Health, 2000. **21**: p. 171-192.

25. Greenland, S., *Principles of multilevel modeling.* Int J Epidemiol, 2000. **29**: p. 159-67.

26. Leyland, A.G., H, *Multilevel modelling of health statistics.* 2001: Wiley, New York.

27. Srividya, A., et al., *A geostatistical analysis of the geographic distribution of lymphatic filariasis prevalence in southern India.* Am J Trop Med Hyg, 2002. **67**(5): p. 480-9.

28. Bonfim, C., et al., *A socioenvironmental composite index as a tool for identifying urban areas at risk of lymphatic filariasis.* Trop Med Int Health, 2009. **14**(8): p. 877-84.