

RESEARCH ARTICLE

Adjusting for a cluster effect in the logistic regression model: an illustration of theory and its application

R.V. Jayatillake¹, M.R. Sooriyarachchi^{1*} and D.L.P. Senarathna²

¹ *Department of Statistics, Faculty of Science, University of Colombo, Colombo 03.*

² *South Asian Clinical Toxicology Research Collaboration, Faculty of Medicine, University of Peradeniya, Peradeniya.*

Revised: 20 December 2010; Accepted: 18 March 2011

Abstract: When a response having two outcomes is modelled using a logistic model the responses on each observation are often considered to be independent of each other. This assumption may not always be valid and the responses may be correlated with each other as in the case of clustered data, which can occur especially in the case of survey data. When responses are correlated as explained, then the ordinary logistic regression model is unsuitable as the standard errors will be biased, and therefore this model should be adjusted for the cluster effect. In this paper one of the many methods of adjustment suggested in the literature, which is based on robust standard error estimation for cluster sampling data is examined. The objective of this paper is to illustrate the theory and mode of application of this theory by way of using a survey on Paraquet poisoned patients in Sri Lanka. Here, the patients are clustered within hospitals and therefore adjustment of the logistic model for the clustering effect is discussed. Adjusting for the hospital (cluster) effect significantly reduced the standard error of four out of thirty three odds ratios given by the model. That is four odds ratios that were not significant before adjustment became significant after adjustment. There was no change in significance in the other twenty nine odds ratios. On average there is a reduction of 2.29% in the standard error of the odds ratios after adjustment for the cluster effect. This indicates that it is effective and important to adjust for the cluster effect.

Keywords: Cluster effect, Huber formula, logistic regression, mixed model, random effect.

INTRODUCTION

A logistic regression model (Agresti, 1990; Collett, 1994; Hosmer & Lemeshow, 2000) is usually fitted when the response variable has two levels. When observations are selected based on strata a cluster effect can be anticipated within strata. Usually in data modelling, stratification

or clustering is ignored. If however, stratification is indeed effective, the ignoring of stratification will usually increase the resulting variance of the estimators (Parson, 1992) as the independence between patients is violated due to the correlation within strata, so that the usual maximum likelihood method for estimating standard errors is not valid. In order to overcome this problem a robust method of estimation of standard errors of the parameter estimates has to be used. Therefore an adjustment is required for the cluster (strata) effect in estimating the standard errors of the parameter estimates. In the past various adjustments have been suggested by several authors, the more popular ones being Kacker and Harville(1984), Kenward and Roger(1997), and Huber(1967). Huber's adjustment is now available in the statistical package STATA(Liu, 1998) and is explained here. This method was developed by Huber (1967) and is based on bootstrap jackknife method.

The objective of this paper is to illustrate the theory of adjustment for cluster effects in a logistic model and its application by way of an example to determine the factors contributing to the transfer of poisoned patients from peripheral hospitals to secondary hospitals where the hospital is considered as a cluster. The study (Senerathna, 2006) used is a cross sectional study. The patients admitted to peripheral hospitals in Anuradhapura and Polonnaruwa Districts during the period of July 2005 to December 2005 have been considered for the study. The study population consists of 1024 patients from 36 hospitals. A total of 40 variables were measured on each patient. Table 1 gives the broad areas about which information was collected and the variables measured in the study.

*Corresponding author (roshini@mail.cmb.ac.lk)

Table 1: Information collected and variables measured

Patient details	Hospital details
Demographics	Location
Age	Distance to the primary hospital
Gender	Annual number of admissions
Clinical details	Drugs available
Type of poison ingested	Fuller's earth
Pulse rate	Activated charcoal
Respiratory rate	Pralidoxime
Blood pressure	
Consciousness	Equipments available
Appearance of the pupils	Nasogastric (NG) tubes
	Airways
Treatments given	ET tubes
Gastric decontamination	ECG recorder
Fuller's earth	
Diazepam	Staff
Intravenous fluids	No. of medical officers
	No. of nurses
	No. of attendants

METHODS AND MATERIALS

The logistic regression model: Consider a single categorical explanatory variable, A, with I levels. Let p_i be the probability of belonging to the response category of interest for an observation having the i^{th} level of A. The logistic regression model (Collett, 1994), which gives the relationship between the response (p_i) and the i^{th} level of a factor A, is as follows.

$$\text{logit} \left(\frac{p_i}{1-p_i} \right) = \alpha + \beta_i^A \quad ; \quad i=1,2,\dots,I \quad \dots(1)$$

Here α is the intercept and β_i^A is the effect of the i^{th} level of the variable A on the response.

Adjusting for cluster effect: One approach to adjust for the cluster effect is to introduce it as a covariate to the model. Suppose β_j^B represents the effect of the j^{th} cluster, then the model can be given as

$$\text{logit} \left(\frac{p_{ij}}{1-p_{ij}} \right) = \alpha + \beta_i^A + \beta_j^B \quad \dots(2)$$

Although it is possible to estimate all cluster effects, it becomes tedious and not feasible when the numbers of clusters are very large (Paterson & Goldstein, 1991).

In addition, it would be useful to be able to generalize inferences to all the clusters in the population than to restrict it to those in the sample drawn.

Therefore, the other approach (Paterson & Goldstein, 1991) would be to introduce the cluster effect (u_j) to the model as a normally distributed random variable with zero mean and a constant variance. This is given by the model:

$$\text{logit} \left(\frac{p_{ij}}{1-p_{ij}} \right) = \alpha + \beta_i^A + u_j \quad \dots(3)$$

where, $u_j \sim N(0, \sigma_u^2)$

This is a generalized linear mixed model where A is a fixed effect and u is a random effect. More details about this type of model, model fitting and interpretation are given in Brown & Prescott (1999).

Fixed and random effect standard errors are calculated using a formula that is based on a known variance matrix, V. However, because V is in fact estimated it is known that in most situations there will be some bias in the standard errors given by the maximum likelihood method (Brown & Prescott, 1999). Various adjustments for bias have been suggested and Huber's adjustment is now available in the statistical package STATA, which is used for the analysis of the example data set used in this paper (Annex).

Huber (1967) method of adjustment of estimates of standard errors: Huber's (1967) method of adjustment for a cluster effect in the estimates of standard errors in a logistic model has been described by Liu (1998) and is briefly explained here.

Let us denote the logistic model,

$$\ln \left(\frac{p_i}{1-p_i} \right) = x_i \beta \quad \dots(4)$$

Where p_i is the probability of an event for the i^{th} unit, x_i is the design matrix for the i^{th} unit, β is the vector of regression coefficients and $i = 1, \dots n$.

Thus p_i can be denoted by,

$$p_i = \frac{e^{x_i \beta}}{1 + e^{x_i \beta}} \quad \dots(5)$$

Then the log likelihood function L for a probability density function is

$$L = \log \left(\prod f(y_i) \right) = \sum \log(f(y_i)) = \sum l(y_i) \quad \dots(6)$$

Define the dummy variables y_i such that $y_i = 1$ if the i^{th} unit has the event of interest and $y_i = 0$ otherwise.

Then

$$\begin{aligned} L &= \sum \log [p_i^{y_i} (1-p_i)^{1-y_i}] \\ &= \sum [y_i \log p_i + (1-y_i) \log(1-p_i)] \quad \dots(7) \end{aligned}$$

[where $f(y_i) = p_i^{y_i} (1-p_i)^{1-y_i}$ is the density function of the Bernoulli distribution].

For the i^{th} observation the Score function can be denoted as

$$S_i = \frac{\partial(L)}{\partial(x_i, \beta)} \quad \dots(8)$$

and the Hessian can be denoted as

$$H_i = \frac{\partial^2 L}{\partial(x_i, \beta)^2} \quad \dots(9)$$

This robust estimation is based on asymptotic bootstrap or jackknife (Efron & Tibshirani, 1993) method. The bootstrap method is based on estimating from re-sampling with replacement from the original sample. The jackknife method, similarly estimates by systematical re-computing the estimate by dropping one observation at a time from the original sample. Liu (1998) shows that the amount of the shift in the estimate of the standard error when i^{th} observation is dropped is $-D^{-1} S_i x_i^T$ where the matrix

$$D = \sum_i H_i (x_i^T x_i) \quad \dots(10)$$

It is assumed that no single observation has very large effect in the fitting, and then dropping two observations is roughly the sum of dropping each observation individually. The same is extended to a cluster, where dropping all members of the cluster is roughly equivalent to dropping each member individually.

By Huber (1967) formula the standard variance estimate

$$\text{Var}(\hat{\beta}) = D^{-1} \left(\sum S_i x_i^T x_i S_i \right) D^{-1} \quad \dots(11)$$

Using equation (7) for logistic model the score

$$S_i = \frac{\partial(L)}{\partial(x_i, \beta)} = y_i - p_i \quad \dots(12)$$

where y_i is the observed response of the i^{th} unit and $y_i \sim \text{bin}(n_i, p_i)$ and Hessian

$$H_i = \frac{\partial^2(L)}{\partial(x_i, \beta)^2} = p_i(1-p_i)$$

Substituting these in equation (11)

$$\begin{aligned} \text{Var}(\hat{\beta}) &= \left\{ \sum_i [p_i(1-p_i) (x_i^T x_i)] \right\}^{-1} \left[\sum_i (y_i - p_i) x_i^T x_i (y_i - p_i) \right] \\ &\quad \left\{ \sum_i [p_i(1-p_i) (x_i^T x_i)] \right\}^{-1} \quad \dots(13) \end{aligned}$$

Suppose there are C clusters and g_j observations in each cluster ($j=1,2,\dots,C$). Then the robust standard error is

$$\text{Var}(\hat{\beta}) = \tilde{D}^{-1} \left(\sum_j U_j^T U_j \right) \tilde{D}^{-1} \quad \dots(14)$$

where $\tilde{D} = \sum_{j=1}^C \sum_{k=1}^{g_j} (H_{jk} x_{jk}^T x_{jk})$ and $U_j = \sum_{k=1}^{g_j} x_{jk} S_{jk}$ is the contribution to score from each cluster. Liu (1998) suggests two finite sample corrections to adjust the estimate more closely to the population values. This requires a multiplier q_c , which is considered to be $\frac{N-1}{N-p-1} * \frac{C}{C-1}$ or asymptotically $\frac{C}{C-1}$ where N is the number of observations, p is the number of parameters to be estimated and C is the number of clusters.

For large samples, both adjustments are the same (As for large N , $p \ll N$ and $(N-p-1) \approx (N-1)$).

Selecting variables to the model: Initially a study could consist of a large number of variables and therefore, variables that are significantly associated with the outcome have to be identified using an appropriate univariate test for association i.e. Chi square test (Agresti, 2002). As an univariate test is unadjusted for other factors, a liberal significance level of about 0.2 should be used in selecting variables for modelling (Collett, 1994). The variables selected at the univariate stage should then be examined after adjustment for other variables using a logistic model with all other chosen variables as fixed effects and the cluster variable as a random effect. Forward selection method (Collett, 1994) with a significance level of 5% (0.05) is usually used to select a suitable model. When selecting important variables to the model, the Wald statistic associated with the variable was used instead of the likelihood ratio test (LRT) statistic, which is the standard statistic used in the basic logistic regression model. This is because when a cluster effect is present,

the standard errors of the parameter estimates are adjusted and thus the LRT is invalid (Rana *et al.*, 2010).

The example: The study data consists of poisoned patients who were admitted to 36 peripheral hospitals in the North Central Province of Sri Lanka between July 2005 and December 2005 (Senarathna, 2006). The data collected was classified into two groups, namely, patient details and hospital details. Patient data such as age, gender, clinical symptoms and treatments received were obtained from the 'bed head ticket' and the hospital details were collected from relevant authorities of each hospital.

The main response of interest in this study is the outcome for the patient. During the study period only a few deaths were recorded due to poisoning. Due to lack of data, death was excluded as an outcome. Thus only two outcomes, discharged or transferred to a secondary hospital, were considered as outcomes for the patient. It was of interest to determine factors affecting the transfer of poisoned patients from peripheral hospitals (Eddleston, 2006).

This is a binary outcome and the suitable model to fit for this type of data is the Binary Logistic regression model. However, as patients were collected from hospitals, patients from the same hospital tend to be correlated (Guittet *et al.*, 2006). In addition, corresponding information on the hospital where the patient is admitted to is the same for patients from the same hospital. Therefore in estimating the standard error of the parameter estimates in a logistic model, we cannot use the usual maximum likelihood method, which assumes the observations to be independent. Therefore an adjustment is required for this hospital effect in estimating the standard errors of the parameter estimates.

Thus the methods discussed are illustrated in this example in order to explain its advantages.

RESULTS AND DISCUSSION

Logistic model for the patient outcome

From univariate analysis, association of the variables

with the outcome for the patient is analysed one variable at a time. However it is important to analyse the effect of these variables when present together. In addition it is also important to identify the most significant factors that affect the outcome. Therefore, to achieve these objectives a logistic mixed model is fitted to the data, which adjusts for the cluster (hospital) effect.

The variable selection indicated that the following fixed effects were included in the logistic regression model. The variable name is given within brackets.

- Poison type ingested (cpoison)
- Appearance of the patient's pupil (cpupils)
- Consciousness of the patient (conscious)
- Intravenous (IV) access is given to the patient (ivaccess)
- Atropine is given to the patient (atropine)
- IV fluids are given to the patient (ivfluids)
- Patient is being monitored (monitoring)
- Availability of ambuventilation facilities in the hospital (ambuventilation)
- Availability of fuller's earth at the hospital (fullersearth)
- Distance from peripheral hospital to the secondary hospital (distance)
- Availability of diazepam IV in the hospital (diazepamiv)
- Availability of nasogastric (Ng) tubes at the hospital (ngtubes)
- Availability of adrenaline in the hospital (hadrenaline)
- Availability of 2-pyridine aldoxime methiodide (PAM) in the hospital (hpam)
- Number of sanitary labourers in the hospital (slabourer)
- Availability of dextrose in the hospital (dextrose)

Of these variables, distance and slabourer are continuous variables while all the others are categorical variables. The cluster effect was fitted using the 'cluster' option in the STATA package. This implies that the cluster (hospital) effect is considered to be a random effect. The standard errors are adjusted based on the Huber formula. The model selected by the forward selection method is

$$\begin{aligned} \text{Logit}(p_{ijklmnopqrstuv}) = & \beta_0 + \beta_i^{\text{cpoison}} + \beta_j^{\text{ambuventilation}} + \beta_k^{\text{cpupils}} + \beta_l^{\text{ivaccess}} + \beta_m^{\text{atropine}} \\ & + \beta_n^{\text{ivfluids}} + \beta_o^{\text{monitoring}} + \beta_p^{\text{fullersearth}} + \beta_1 \times \text{distance} \\ & + \beta_q^{\text{diazepam}} + \beta_r^{\text{ngtubes}} + \beta_s^{\text{hadrenaline}} + \beta_t^{\text{hpam}} \\ & + \beta_2 \times \text{slabourer} + \beta_u^{\text{dextrose}} + \beta_v^{\text{conscious}} + \beta_w^{\text{hospital}} \end{aligned} \quad \dots(15)$$

Where $\beta_w^{\text{hospital}}$ follows a $N(0, \sigma_H^2)$ distribution where σ_H^2 is the between hospital variance. The model fitting was limited to the main effects due to large number of variables considered and to keep the complexity of the model to the minimum. The model with these main effects (given by equation 15) was then checked for adequacy using diagnostic methods.

Table 2 shows the odds ratios, robust standard errors of the odds ratios and 95% confidence intervals for the odds ratios for the logistic model with mixed effects (equation 15). Additionally, it also includes the

percentage decrease in the standard error of the odds ratio when compared to the model unadjusted for cluster effect.

The odds ratios of cpoison_2, ivfluids_2, hadrenalin_2 and conscious_2 are significant after adjusting for the cluster effect while these were not significant before adjustment. The significance of all other odds ratios at 5% level are the same before and after adjustment. After adjustment for the cluster effect, the average decrease in standard error of the odds ratios overall is 2.3%. The change in significance of 4 odds ratios indicates that the

Table 2: Odds ratios estimated for the parameters of the logistic model with hospital (random) effect

Parameter	Odds ratio	Robust standard error (S.E.)	% reduction in S.E. after adjustment	95% confidence interval	
Cpoison_2*	2.7143	1.2506	14.13	1.10021	6.697
Cpoison_3	1.2815	0.5142	15.15	0.58374	2.813
Cpoison_4	1.0414	0.4879	-14.91	0.41574	2.608
Cpoison_5	1.8772	1.1244	-8.75	0.58031	6.072
Cpoison_6	1.0549	0.4306	-9.96	0.47397	2.348
Cpoison_7	5.109	1.5479	21.76	2.82122	9.252
Cpoison_8	10.2926	7.8431	12.37	2.31151	45.83
Cpoison_9	0.2111	0.1241	-3.85	0.06671	0.668
Cpoison_10	1.0981	0.3614	0.61	0.57616	2.093
Cpoison_11	0.7957	0.2786	-2.24	0.4006	1.58
ambuventilation	0.0548	0.0231	9.41	0.02396	0.125
cpupils_2	6.4665	3.6843	2.28	2.11687	19.754
cpupils_3	1.198	0.3315	-11.06	0.69648	2.061
Ivaccess_2	19.9231	17.6008	28.45	3.52672	112.549
Ivaccess_3	0.8285	1.0992	13.60	0.0615	11.16
Atropine_2	3.0759	1.0057	-14.27	1.62055	5.838
Atropine_3	1.0911	0.3538	12.69	0.57792	2.06
ivfluids_2*	0.1467	0.1176	34.48	0.03046	0.706
ivfluids_3	0.7325	0.7829	25.74	0.09017	5.95
monitoring_2	0.4239	0.249	65.60	0.13404	1.341
monitoring_3	2.4886	0.5514	-19.81	1.61204	3.842
fullersearth_2	0.1890	0.0559	-59.26	0.10566	0.337
Distance	0.9737	0.0091	-56.89	0.95598	0.992
diazepamiv_2	0.0122	0.0106	11.67	0.00221	0.067
Ng tubes_2	3.7745	1.4693	-38.78	1.75997	8.095
hadrenalin_2*	2.7467	0.6232	60.34	1.76074	4.285
hpam_2	3.6974	1.0998	-38.11	2.06406	6.623
Slabourer	0.8621	0.031	-5.08	0.80338	0.925
Dextrose_2	0.4959	0.0893	17.31	0.34843	0.706
Dextrose_3	1.1261	0.6729	-10.00	0.34909	3.633
conscious_1	0.8719	0.2534	7.85	0.49327	1.541
conscious_2*	6.5458	6.0322	20.07	1.07534	39.845
conscious_3	4.4969	4.8627	-7.33	0.54009	37.441

The baseline level for the calculating odds ratios has been considered as the first level for all categorical variables. The * indicates the variables corresponding to the odds ratios that became significant after adjustment for cluster effect.

adjustment for cluster effect is important. Interpretation of the odds ratios given in Table 2 is included in another section.

Diagnostics of the model

One method of determining the fit of the model is to observe how the model can classify the outcome for the patient for each observation, which is a function of the observed responses and estimated values of the covariates. Table 3 shows the classification given by the model against actual outcome for the patient. The model fitted, correctly identifies the outcome of 78.42% of the patients. This is nearly 80% and shows that the model classifies the outcome well (Rana *et al.*, 2010) and thus indicates its goodness of fit.

Comparing the outcome without adjusting for cluster (hospital) effect

One matter of interest is to compare the change in the standard errors of the odds ratios with and without adjusting for the cluster/hospital effect. Table 4 gives the odds ratios, robust standard errors of the odds ratios and the 95% confidence intervals for the odds ratios for logistic model without adjusting for the hospital (random) effect.

The standard errors that have higher values compared to corresponding robust standard errors when adjusted for cluster effect is indicated by bold letters. It can be observed that out of the 33 parameters estimated in the model, 18 parameters have a higher standard error when unadjusted for cluster effect. When significance of the odds ratios are compared before and after adjustment, this shows that four odds ratios that were not significant in the unadjusted analysis become significant after the adjustment, indicating that adjustment for cluster effect is important in drawing correct conclusions.

Table 3: Classification of the model against actual observations

Status	Observed		Total
	Discharged	Transferred	
Classified			
Discharged	600	131	731
Transferred	90	203	293
Total	690	334	1024
Correctly classified			78.42%

Interpreting the odds ratios

The type of poison ingested by the patient is a significant factor that contributes to the transfer of patients. The odds of transfer for patients with Oleander, Paraquat and spray poisons are approximately 5, 10 and 0.2 times the odds for transfer with organophosphate poisoned patients. This is the case even if there is no adjustment for cluster effect. The adjustment for cluster effect makes the carbamate poisoned patients also have significantly

Table 4: Odds ratios estimated for the parameters of the logistic model without hospital (random) effect

Parameter	Odds ratio	Robust standard error	95% confidence interval	
cpoison_2	2.7143	1.4565	0.94825	7.77
cpoison_3	1.2815	0.606	0.50724	3.238
cpoison_4	1.0414	0.4246	0.46833	2.316
cpoison_5	1.8772	1.0339	0.63783	5.525
cpoison_6	1.0549	0.3916	0.50958	2.184
cpoison_7	5.109	1.9783	2.39186	10.913
cpoison_8	10.2926	8.9506	1.87201	56.59
cpoison_9	0.2111	0.1195	0.06961	0.64
cpoison_10	1.0981	0.3636	0.57385	2.101
cpoison_11	0.7957	0.2725	0.40668	1.557
ambuventilation	0.0548	0.0255	0.02201	0.136
cpupils_2	6.4665	3.7701	2.06253	20.274
cpupils_3	1.198	0.2985	0.73519	1.952
ivaccess_2	19.9231	24.5993	1.7716	224.051
ivaccess_3	0.8285	1.2723	0.04084	16.805
atropine_2	3.0759	0.8801	1.75558	5.389
atropine_3	1.0911	0.4052	0.52699	2.259
ivfluids_2	0.1467	0.1794	0.01334	1.613
ivfluids_3	0.7325	1.0543	0.04362	12.301
monitoring_2	0.4239	0.7238	0.01493	12.037
monitoring_3	2.4886	0.4603	1.73191	3.576
fullersearth_2	0.1888	0.0351	0.13115	0.272
distance	0.9737	0.0058	0.96236	0.985
diazepamiv_2	0.0122	0.012	0.00178	0.084
ngtubes_2	3.7745	1.0579	2.17917	6.538
hadrenaline_2	2.7467	1.5712	0.89517	8.428
hpam_2	3.6974	0.7963	2.42432	5.639
slabourer	0.8621	0.0295	0.80607	0.922
dextrose_2	0.4959	0.108	0.32354	0.76
dextrose_3	1.1261	0.6117	0.3883	3.266
conscious_1	0.8719	0.275	0.46984	1.618
conscious_2	6.5458	7.5467	0.68328	62.707
conscious_3	4.4969	4.5305	0.62423	32.395

The baseline level for the calculating odds ratios has been considered as the first level for all categorical variables.

higher odds for transfer (approximately 3 times) compared to organophosphate patients.

When cluster effect is adjusted for, it makes those patients that were given IV fluids have lower (0.15) odds of being transferred compared to patients that were not given IV fluids.

When cluster effect is adjusted for, it makes the odds for transfer for a patient admitted to hospitals without adrenalin approximately 3 times higher compared to a patient admitted to hospital with adrenalin.

After adjustment for the cluster effect, a patient that has been unconscious has approximately 6 times higher odds of being transferred compared to a patient that has been conscious.

CONCLUSION

Results indicate that while levels 7, 8 and 9 of poison ingested are significantly different from level 1 of poison ingested in the absence of the adjustment for cluster effect. When cluster effect is adjusted for, additionally level 2 of poison ingested also becomes significant. Adjustment for cluster effect also makes the IV access, adrenalin and consciousness factors significant. Therefore 4 (out of 33) parameters that were not significant before adjustment for cluster effect became significant with the adjustment, indicating the importance of its implementation.

REFERENCES

1. Agresti A. (1990). *Categorical Data Analysis*. John Wiley and Sons Inc., New York, USA.
2. Brown H.K. & Prescott R. (1999). *Applied Mixed Models in Medicine*. John Wiley and Sons Inc., Chichester, USA.
3. Collet D. (1994). *Modeling Binary Data*. Chapman and Hall Publishers, New York, USA.
4. Eddleston M., Sudarshan K., Senthilkumaran M., Reginald K., Karalliedde L., Senarathna L., de Silva D., Sheriff M.H.R., Buckley N.A. & Gunnell D. (2006). Patterns of hospital transfer for self-poisoned patients in rural Sri Lanka: implications for estimating the incidence of self-poisoning in the developing world. *WHO Bulletin* **84**(4): 257–336.
5. Efron B. & Tibshirani R.J. (1993). *An Introduction to the Bootstrap*. Chapman and Hall Publishers, New York, USA.
6. Guittet L., Ravaud P. & Giraudeau B. (2006). Planning and a cluster randomized trials with unequal cluster sizes: practical issues involving continuous outcomes. *Medical Research Methodology* **6**(17): 1–15.
7. Hosmer D.W. & Lemeshow S. (2000). *Applied Logistic Regression*. John Wiley and Sons Inc., New York, USA.

8. Huber P.J. (1967). The behavior of maximum likelihood estimates under non-standard conditions. *Proceedings of the Fifth Berkley Symposium on Mathematical Statistics and Probability* **1**: 221–233.
9. Kenward M.G. & Roger J.H. (1997). Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics* **53**(3): 983–997.
10. Kracker R.N. & Harville D.A. (1984). Approximations for standard errors of estimators for fixed and random effects in mixed linear models. *The Journal of the American Statistical Association* **79** (388): 853–862.
11. Liu H. (1998). Robust standard error estimates for cluster sampling data: a SAS/IML macro procedure for logistic regression with huberization. *Proceedings of the 23rd Annual SAS Users Group International*, Nashville, Tennessee, USA, March 22–25, pp. 205–210.
12. Parson V.L. (1992). Using the sampling design in logistic regression analysis of NCHS survey data- some applications. *Proceedings of the American Statistical Association, Section on Survey Research Methods*, American Statistical Association, Alexandria, USA, pp.327–332.
13. Paterson L. & Goldstein H. (1991). New statistical methods for analyzing social structures: an introduction to multilevel models. *British Educational Research Journal* **17** (4): 387–393.
14. Rana S., Midi H. & Sarka S.K. (2010). Validation and performance analysis of the binary logistic regression model. *Proceedings of the WSEAS International Conference on Environmental, Medicine and Health Sciences*, Penang, Malaysia, March 23–25, The World Scientific and Engineering Academy and Society, Wisconsin, USA, pp. 51–55.
15. Senarathna D.L.P. (2006). How the level of resources and hospital staff attitude affect poisoning patient outcome from primary care hospitals in rural Sri Lanka. *M.Sc. thesis*, University of New Castle, Australia.

ANNEX

Stata commands used in the analysis:

/* Stata commands used in fitting the fitted model with main effects adjusting for cluster effect. This programme gives the results on the parameter estimates and robust standard errors of the estimates and the 95% confidence interval of the estimates for the adjusted analysis. */

```
xi:logit outcome i.ncpoison i.ambuventilation ///
i.ncpupils i.nivaccess i.natropine i.nivfluids ///
i.nmonitoring i.fullersearch distance i.diazepamiv ///
i.ngtubes i.hadrenaline i.hpam slabourer ///
i.ndextrose i.conscious, cluster(hospitals)
```

/* Stata commands used in fitting the fitted model with main effects adjusting for cluster effect. This programme

gives the results of the odds ratios, robust standard errors of the odds ratios and the 95% confidence interval of the odds ratios for the adjusted analysis. */

```
xi:logistic outcome i.ncpoison i.ambuventilation ///
i.ncpupils i.nivacess i.natropine i.nivfluids ///
i.nmonitoring i.fullersearth distance i.diazepamiv ///
i.ngtubes i.hadrenaline i.hpam slabourer ///
i.ndextrose i.conscious, cluster(hospitals)
```

/*Stata commands used in fitting the fitted model with main effects unadjusted for cluster effect. This programme gives the results on the parameter estimates and robust standard errors of the estimates and the 95% confidence interval of the estimates for the unadjusted analysis. */

```
xi:logistic outcome i.ncpoison i.ambuventilation ///
i.ncpupils i.nivacess i.natropine i.nivfluids ///
i.nmonitoring i.fullersearth distance i.diazepamiv ///
i.ngtubes i.hadrenaline i.hpam slabourer ///
i.ndextrose i.conscious
```

/*Stata commands used in fitting the fitted model with main effects unadjusted for cluster effect. This programme gives the results of the odds ratios, robust standard errors of the odds ratios and the 95% confidence interval of the odds ratios for the unadjusted analysis. */

```
xi:logistic outcome i.ncpoison i.ambuventilation ///
i.ncpupils i.nivacess i.natropine i.nivfluids ///
i.nmonitoring i.fullersearth distance i.diazepamiv ///
i.ngtubes i.hadrenaline i.hpam slabourer ///
i.ndextrose i.conscious
```