

A COMPARATIVE STUDY OF MODELS FOR CORRELATED BINARY DATA WITH APPLICATIONS TO HEALTH SERVICES RESEARCH

Gemechis Dilba^{1,*} and Marc Aerts²

¹Department of Statistics, Faculty of Science, Addis Ababa University, PO Box 1176
Addis Ababa, Ethiopia. E-mail: dilba@bioinf.uni-hannover.de

²Centre for Statistics, Limburgs Universitair Centrum
Universitaire Campus–building D, 3590 Diepenbeek, Belgium

ABSTRACT: Various methods of modeling correlated binary data are compared as applied to data from health services research. The methods include the standard logistic regression, a simple adjustment of the standard errors of logistic regression by a single inflator, the weighted logistic regression, the generalized estimating equation, the beta-binomial model, and two proposed bootstrap methods. First, these approaches are compared for a fixed set of predictors by individual tests of significance. Next, several subsets of predictors are compared through the AIC criterion, whenever applicable.

Key words/phrases: Beta-binomial, bootstrap, correlated binary data, model selection, overdispersion

INTRODUCTION

Correlated binary data appear in many areas of research. Responses are often binary in nature like in decision-making: approve/disapprove, reject/-accept and in evaluations: appropriate/inappropriate, disease status: diseased/not diseased, *etc.* A sequence of binary responses can be correlated for a number of reasons. A common example occurs in toxicological studies where malformations (present/absent) are correlated for litters of mice from the same dam. In other applications, clustering typically occurs as a result of repeated measurement on the same subject. Analysis of this type of data has gained much attention in the past.

In a first step, one approaches the problem of modelling binary responses with the standard logistic regression. This method is based on the binomial likelihood. Therefore, unless the underlying assumptions are satisfied, the analysis can result in drawing completely incorrect inferences. There are two important ways in which the binomial assumptions can be violated: heterogeneity of the success probabilities and dependence among the binary responses. The latter is the focus of the analyses in this report.

In the statistical literature a number of methods have been proposed to model correlated binary data. The methods range from a simple rescaling of the covariance matrix by a single inflator factor to a full likelihood based inference. To mention few,

Williams (1975) first proposed the use of beta-binomial model and later proposed fitting weighted logistic regression (Williams, 1982). More recently, the generalized estimating equation was proposed by Liang and Zeger (1986) in regression analysis of repeated measurement. This method can be easily adapted for binary outcomes. A more comprehensive review of methods for analysing clustered binary data is presented by Pendergast *et al.* (1996) and Aerts *et al.* (2002).

The concern of this paper is to compare commonly used methods of modelling correlated binary data as applied in health services research. As described by Lohr (1998), an important component of health services research is the study of how populations and individuals can obtain effective, appropriate, competent, and compassionate health care services in short and high-quality health care. In this regard, there are two extreme situations that need attention: the overuse of unnecessary and inappropriate services and underuse of needed and appropriate care. In this regard, two variables of interest are the length of stay (*e.g.*, in the hospital) and the number of inactive days. Gange *et al.* (1996) applied the beta-binomial distribution in modelling the effect of policy changes on appropriateness of hospital stays. The response variable is binary (inappropriate/appropriate), and it is retrospectively recorded for each patient during his or her stay in the

* Current address: University of Hannover, Bioinformatic Unit, Herrenhauser Strasse 2, D-30419 Hannover, Germany.

hospital. Obviously, the responses are dependent (repeated daily observations on the same patient). Gange *et al.* (1996) modelled both the mean response (probability of inappropriateness) and the scale parameter. The scale parameter (also related to the intracluster correlation) is used to measure the degree of propagation of inappropriateness in the hospital, with low correlation implying efficient and high-quality service. The same data is used in this paper to compare various models, and to further investigate the final model of Gange *et al.* (1996).

The data consists of 1383 patients. The response variable is binary, with two possible values: 0 (a day of stay in the hospital is appropriate) and 1 (a day of stay in the hospital is inappropriate). Other variables of interest are Length of Stay in hospital (LOS), Number of inappropriate days (r), Gender, Age at Admission, Ward (1=Medical, 2=Surgical, 3=Other), and Year (1988, 1990). The covariate Year is important because there was a policy change during the two years. All variable transformations are adopted as they were in the analysis by Gange *et al.* (1996).

The objectives of this paper are twofold. First, the paper examines, based on a fixed set of predictors, how different modelling approaches lead to different significant effects. We investigate the final model of Gange *et al.* (1996) which is based on the beta-binomial by including the quadratic effect of LOS, and we also select one or more optimal models in the sense of Akaike's information criterion (AIC) (Akaike, 1973). The second main objective is that of comparing the performance of two bootstrap methods (resampling binary values versus resampling clusters) with other methods.

DESCRIPTION OF THE PROBLEM

The logistic regression for analysing binary responses (modelling proportion) demands independence among the binary observations. As a consequence, the variability in the data would not be satisfactorily captured by fitting a logistic regression. This is called extra-binomial variation or overdispersion. Overdispersion is common in count data, where the mean and variance have a functional relationship as they do in the case of binomial and Poisson distributions. A more detailed discussion of overdispersion can be found in Collett (1991). In the next section, some methods of accommodating for overdispersion are briefly introduced.

THE MODELS

Logistic regression

Suppose that Y_{ij} is a binary response having Bernoulli (π_i) distribution, $i=1,2,\dots,N$; $j=1,\dots,n_i$. The success probability π_i is related to a set of k potential predictors through the logit transform:

$$\text{logit}(\pi_i) = \mathbf{X}_i\boldsymbol{\beta}.$$

It is assumed that the binary responses are independent and that the probabilities π_i of subjects with the same covariate pattern $\mathbf{X}_i$ are identical. For the hospital stays data, repeated measurements are taken on the patients. The responses from the same patient form a natural cluster. Basically, since the logistic regression ignores this clustering, there are some pitfalls in fitting this model. In particular, the standard errors of the regression coefficients will be underestimated, giving rise to inflated test statistics. This model will be fitted as a first trial.

Single inflator factor (scaled deviance/scaled Pearson)

This is a simple way of correcting for overdispersion. Analogously to the normal regression, the dispersion parameter ϕ can be estimated by the scaled deviance statistics, χ_D^2/df or the scaled Pearson, χ_P^2/df (see McCullagh and Nelder, 1989). The dispersion parameter for the binomial distribution is 1, as can be seen from its general exponential family representation. Therefore, an estimate of ϕ that is by much greater than 1 implies that the variability in the data is more than the binomial variation. To correct for overdispersion, we multiply the covariance matrix of $\hat{\boldsymbol{\beta}}$ by $\hat{\phi}$. This adjustment does not affect the estimates of the regression coefficients. An option in SAS system for this adjustment is Scale=D (or SCALE=P) in the PROC LOGISTIC procedure.

Williams procedure

Williams (1982) proposed using a weighted logistic regression. For the i^{th} subject, if the binomial variance is given by $n_i\pi_i(1-\pi_i)\sigma_i^2$, an estimate of the regression coefficients are obtained iteratively by using the weights $w_i = 1/\hat{\sigma}_i^2$. The weights have the effect of inflating the standard errors of the regression coefficients. More details can also be found in Collett (1991). An option in

SAS system for this adjustment is SCALE=W in the PROC LOGISTIC procedure.

Generalized estimating equations (GEE)

GEE, also known as marginal models, are a multivariate generalization of the quasi-likelihood method. Quasi-likelihood is a function, which if existing, acts like a log-likelihood (Wedderburn, 1974). In the construction of this function, the mean and variance relationship in the real likelihood is perturbed in order to accommodate for overdispersion. Briefly, the steps in GEE as applied in regression analysis of repeated measurements with binary responses are: (i) relate the mean to linear combination of the covariates, (ii) express the variance of the response as a function of the mean, and (iii) choose the structure of the working correlation matrix. Several choices are discussed by Liang and Zeger (1986), including for instance, the independence, exchangeable and unstructured working correlations. More details can be found in Stocks *et al.* (1995). In SAS, GEE models can be fitted using the PROC GENMOD procedure.

Beta-binomial model

Beta-binomial distribution is a compound distribution of the binomial distribution with a prior beta density (with parameters α and β) for the response probability. The mean and variance of a random variable R having beta-binomial distribution with parameters n , α and β are:

$$\mu_R = n \frac{\alpha}{\alpha + \beta}$$

and $\sigma_R^2 = \frac{n\alpha\beta(\alpha + \beta + n)}{(\alpha + \beta)^2(\alpha + \beta + 1)}$, respectively.

The variance of the distribution, which can be written as a product of the binomial variance and some nonnegative quantity, is used to model the *extra-binomial variation*. Let $\pi = \alpha/(\alpha + \beta)$, then σ_R^2 can be rewritten as:

$$\sigma_R^2 = n\pi(1-\pi) \left[1 + (n-1) \frac{1}{(\alpha + \beta + 1)} \right].$$

For binary responses, assuming an exchangeable correlation structure, it can be shown that

$$\sigma_R^2 = n\pi(1-\pi)[1 + (n-1)\rho],$$

where ρ is the intracluster correlation.

Thus, in modelling overdispersion using the beta-binomial model, we equate ρ to $(\alpha + \beta + 1)^{-1}$. Mathematically, as both α and β tend to zero, the density approaches a U shape and the intracluster correlation approaches 1.

Now, for a random sample of size N from the beta-binomial distribution, the log-likelihood function in terms of π_i and ρ_i is given by:

$$l(\pi_i, \rho_i) = \sum_{i=1}^N \log \binom{n_i}{r_i} + \sum_{i=1}^N \log \left[\frac{B((\rho_i^{-1}-1)\pi_i + r_i, (\rho_i^{-1}-1)(1-\pi_i) + n_i - r_i)}{B((\rho_i^{-1}-1)\pi_i, (\rho_i^{-1}-1)(1-\pi_i))} \right]$$

where $B(.,.)$ is the beta function. Since the main interest lies in studying the relationship between a given set of potential predictors and both the success probability π and the correlation ρ , two well-known transformations, namely the logit and Fisher's Z transform of correlation are used as link functions in estimating the regression coefficients (see Aerts and Claeskens, 1997). Thus, we can simultaneously estimate the models

$$\ln \left(\frac{\pi_i}{1-\pi_i} \right) = \mathbf{X}_i \boldsymbol{\beta},$$

$$\ln \left(\frac{1+\rho_i}{1-\rho_i} \right) = \mathbf{X}_i \boldsymbol{\eta},$$

where $\mathbf{X}_i$ is a design matrix, $\boldsymbol{\beta}$ and $\boldsymbol{\eta}$ are vectors of regression coefficients. Beta-binomial regression is not available in most of the commonly used standard packages. To our knowledge, it is available only in EGRET (Epidemiological, Graphics, Estimation and Testing Program). In this paper, the analysis is done by writing a program in R (an open source package), which gives the maximum likelihood estimates. The R code can be provided on request.

Bootstrap method

This is a computer intensive method in which several bootstrap samples are generated, and in each case the parameters in the standard logistic regression model are estimated as usual. A good reference for this purpose is Davison and Hinkley (1997). Based on the sequence of bootstrap estimates (empirical sampling distribution), standard errors and p-values can be estimated. Two bootstrap samples are possible depending on whether the resampling is done with regard to clusters (subjects) or binary responses (over all subjects). This latter approach ignores the intracorrelation and is expected to yield (misleading) results close to those of the first method (ordinary logistic regression).

Let $\beta_1^*, \beta_2^*, \dots, \beta_M^*$ be a sequence of bootstrap estimates (replicates) for a particular parameter β_i in the logistic regression model, where M is the number of bootstrap samples. The bootstrap standard error (BSE) of an estimate of β_i is given by

$$BSE = \left(\frac{1}{M-1} \sum_{b=1}^M (\beta_b^* - \bar{\beta}^*)^2 \right)^{1/2},$$

where $\bar{\beta}^*$ is the mean of the bootstrap replicates. For estimating the variance, $M=250$ or $M=500$ should be sufficient. The asymptotic bootstrap p-value can be calculated based on the standard normal curve and the estimated bootstrap standard error.

Other important methods of modelling correlated binary data include the fitting of finite mixture distributions and the generalized linear mixed model, also called the random effects model. The latter is useful when the interest is in the subject-specific effects.

RESULTS

The estimates of the various approaches described in the previous section are briefly presented here. In this first phase of the analysis, the same set of predictors (as in the full model of Gange *et al.*, 1996) is used across all methods to facilitate comparison. The predictors are Gender, Age, Ward, Length of Stay (transformed as $\log(\text{LOS}/10)$), Year and the interaction of Year and $\log(\text{LOS})$. In all outputs, the reference category for gender is male and for Ward is Ward 1 (Medical). Ward 2 and Ward 3 are indicator variables for the other two categories, namely, Surgical and Others, respectively. For Year the reference is 1988.

The fit of logistic regression to the seven predictors is summarized in Table 1.

Table 1. The fit of standard logistic regression.

Parameter	Estimate	SE	Sign. Prob
Constant	-0.9548	0.0394	< 0.0001
Gender	-0.0508	0.0379	0.1804
Age	0.0061	0.0010	< 0.0001
Ward 2	-0.4618	0.0379	< 0.0001
Ward 3	-0.6200	0.1606	0.0001
LogLOS10	0.5150	0.0310	< 0.0001
Year	0.6373	0.1394	< 0.0001
Year $\times$ logLOS	-0.2046	0.0462	< 0.0001

$$\chi_D^2 / df = 5.75$$

At a single glance, one observes from the last column that all terms except Gender are statistically significant. However, as pointed out earlier, there is overdispersion due to the interdependence of the binary responses. This is evident from the estimate of the dispersion parameter. The standard errors are underestimated, giving rise to small p-values.

By the single-inflator (scaled deviance) approach, the standard errors in Table 1 are inflated by a factor of 2.399 ($\sqrt{\hat{\phi}}$). The point estimates of the regression parameters remain the same (Table 2).

Table 2. Single inflator (scaled deviance).

Parameter	Estimate	SE	Sign. Prob
Constant	-0.9548	0.0945	< 0.0001
Gender	-0.0508	0.0910	0.5765
Age	0.0061	0.0025	0.0126
Ward 2	-0.4618	0.0909	< 0.0001
Ward 3	-0.6200	0.3852	0.1075
logLOS10	0.5150	0.0744	< 0.0001
Year	0.6373	0.3343	0.0566
Year $\times$ logLOS	-0.2046	0.1109	0.0650

After adjustment, the effect of Ward 3 is no longer significant, while the effect of Year and the interaction term are significant at the 10% level.

By the Williams procedure, somewhat different results are obtained compared with the scaled deviance (Table 3). Here the effect of Ward 3 is significant, but Age, Year and the interaction term are not. By scaled deviance Age is significant at 5%, but by the Williams procedure it is insignificant even at the 10% level. On the other hand, the effect of Ward 3 is insignificant by the scaled deviance but significant by the Williams procedure. This shows that by implementing the two methods one can arrive at completely different final models.

Table 3. Williams procedure.

Parameter	Estimate	SE	Sign. Prob
Constant	-0.9705	0.0869	< 0.0001
Gender	-0.0383	0.0943	0.6847
Age	0.0039	0.0026	0.1248
Ward 2	-0.4914	0.0973	< 0.0001
Ward 3	-0.8403	0.3616	0.0201
LogLOS10	0.6111	0.0787	< 0.0001
Year	0.4488	0.2782	0.1067
Year $\times$ logLOS	-0.1164	0.1105	0.2922

In GEE, the estimates in Table 1 will be input as initial values for the iteration. The output in Table

4 corresponds to the exchangeable working correlation matrix.

Table 4. GEE estimates.

Parameter	Estimate	SE*	Sign. Prob
Constant	-0.9706	0.0932	<0.0001
Gender	-0.0382	0.0996	0.7010
Age	0.0039	0.0026	0.1317
Ward 2	-0.4916	0.1025	<0.0001
Ward 3	-0.8411	0.3114	0.0069
logLOS10	0.6114	0.0768	<0.0001
Year	0.4489	0.2618	0.0864
Year×logLOS	-0.1164	0.1038	0.2623

* Empirical standard error estimate.

Both in terms of the point estimates of the coefficients and the significance probabilities, GEE results resemble much those of the Williams procedure. Unlike the single-inflator, Age and the interaction term are insignificant in the GEE analysis.

The beta-binomial model differs from the methods discussed so far in that it models not only the success probability π but also the correlation ρ . Two models are estimated under this approach. In the first case π is modeled as a function of the set of covariates under study, and ρ is estimated on its own. In the second case both π and ρ are modelled, thus two vectors of regression parameters are estimated at the same time.

Under this approach the effects of Ward, length of stay and Year are found to be significant predictors of inappropriateness of hospital stays. The intraclass correlation is estimated to be 0.46, and it is significant ($p < 0.0001$) as anticipated. Thus,

there is a real dependence among the binary responses belonging to the same cluster (subject). On the basis of the significance probabilities in Table 5, one retains the same set of predictors as in the GEE analysis.

Table 5. Parameter estimates for beta-binomial model.

	Parameter	Estimate	SE	Sign. Prob
π	Constant	-1.158	0.0910	0.0000
	Gender	-0.0721	0.0948	0.4473
	Age	0.0041	0.0026	0.1075
	Ward 2	-0.2728	0.0978	0.0053
	Ward 3	-0.5753	0.3198	0.0720
	logLOS10	0.5473	0.0794	0.0000
	Year	0.4823	0.2717	0.0758
	Year×logLOS	-0.0498	0.1090	0.6479
ρ	Constant	0.4573	0.0147	< 0.0001

Finally, the bootstrap analysis is based on 10^3 bootstrap samples. The samples are generated by resampling both clusters and binary 0-1 responses. In the first case, the number of clusters is the same as the number of patients in the study ($N=1383$). And in the second case, the total number of binary responses is equal to the total of length of stays ($\sum_1^N n_i = 14833$).

The bootstrap standard errors and p-values in Table 6 are comparable with the ones obtained by the other methods. The result based on resampling binary 0-1 responses is almost identical to that of the standard logistic regression. This is true because the dependence between the binary responses is lost while duplicating the sample in the bootstrap.

Table 6. Bootstrap results.

Parameter	Resampling clusters			Resampling 0-1 responses		
	$\hat{\beta}_i$	BSE	Sign. Prob	$\hat{\beta}_i$	BSE	Sign. Prob
Constant	-0.9548	0.1131	0.0000	-0.9548	0.0388	0.0000
Gender	-0.0508	0.1293	0.6943	-0.0508	0.0371	0.1707
Age	0.0061	0.0033	0.0597	0.0061	0.0010	0.0000
Ward 2	-0.4618	0.1302	0.0004	-0.4618	0.0378	0.0000
Ward 3	-0.6199	0.3632	0.0879	-0.6200	0.1661	0.0002
logLOS10	0.5149	0.1092	0.0000	0.5150	0.0301	0.0000
Year	0.6371	0.3747	0.0891	0.6373	0.1313	0.0000
Year×logLOS	-0.2045	0.1461	0.1615	-0.2046	0.0434	0.0000

By resampling clusters, one keeps the intracorrelation. The standard errors under this approach are larger and more reliable than those of the standard logistic regression. These results are in line with those of the beta-binomial and the GEE approach, except that age seems to be significant at 10%.

The main conclusion of this comparative analysis is the importance of using an appropriate model. It should take the real nature of the data into account. As a general guideline, we recommend the use of GEE for clustered binary data. The single inflator factor (scaled deviance/Pearson) method is too rudimentary, whereas the GEE methodology offers a general estimation paradigm. It has shown its applicability in a variety of settings and is nowadays well-established and widely available in software. In case the correlation itself is of interest (rather than being nuisance), a full likelihood approach based on the beta-binomial model is preferable. Typically, the results of GEE and the beta-binomial are close or comparable, GEE (only based on the first two moments) being more robust but consequently slightly less efficient, the beta-binomial on the other hand being somewhat more efficient and allowing direct modelling of the correlation parameters but being more sensitive to misspecification. As a final conclusion, it is good practice to fit both models, compare results and finally choose the method according to specific needs in the analysis.

MODEL SELECTION

As mentioned at the beginning, one of the aims of this paper is to select one or more optimal models. For the cases where the AIC is available, it will be used as a model selection criterion. AIC is calculated as:

$$AIC = -2\log(\text{likelihood}) + 2(\text{Number of parameters}).$$

The best model has the smallest AIC. For model selection, over forty models consisting of different combinations of terms are fitted for the logistic regression, single-inflator, Williams procedure and the beta-binomial. Only the AIC values of the logistic regression and the beta-binomial are computed from real likelihoods.

The AIC values of the standard logistic regression are observed to be much larger than those of the other modelling approaches. The AIC for the scaled deviance is calculated by adjusting the log-likelihood function of the binomial (*i.e.*, dividing the log-likelihood in the standard logistic

regression by an estimate of the dispersion parameter).

For the standard logistic regression, the smallest AIC (17470.41) corresponds to the model consisting of Gender, Age, Ward, Year, Length of Stay (LOS), quadratic terms in Age and LOS, and three interaction terms. In contrast, the smallest AIC (2931.90), after adjusting by the scaled deviance is associated with the model consisting of only Year as a predictor. By the Williams procedure, the model consisting of Gender, Ward, LOS, Year, quadratic term in LOS, and two interaction terms is identified as the best model (AIC=2675.02).

For the beta-binomial the model consisting of the covariates Ward, LOS, Year and quadratic term in LOS is found to have the smallest AIC (4482.07) when both the mean response and the correlation are modeled with the same set of predictors. In this model, since two of the coefficients in the model for correlation are not significant, the corresponding predictors are removed, giving rise to the best final model according to AIC (4478.78).

The above discussion indicates that the use of AIC for non-likelihood models has to be taken with care. Indeed, Akaike's information criterion originates from information theory and maximum likelihood estimation. A simple adjustment of the log-likelihood function, as in the case of the scaled deviance, might not result in a model selection criterion with the same (optimal) performance as the original AIC. This is confirmed here. Although the other models clearly show the significance of several variables, the scaled deviance AIC version selects a too simple model.

The final beta-binomial model identified by Gange *et al.* (1996) consists of Ward, Length of Stay and Year in modelling π and Year in modelling ρ . As shown in Table 7, it is evident that the quadratic effect of LOS and the effect of Ward are also important covariates in modelling π and ρ , respectively.

Table 7. Parameter estimates for the best beta-binomial model.

	Parameter	Estimate	SE	Sign. Prob
π	Constant	-0.9222	0.0880	0.0000
	Ward 2	-0.4133	0.0989	0.0000
	Ward 3	-0.7753	0.3115	0.0128
	logLOS	0.5147	0.0596	0.0000
	Year	0.2743	0.0972	0.0048
	(logLOS) ²	-0.2451	0.0526	0.0000
ρ	Constant	1.2912	0.0652	0.0000
	Ward 2	-0.4342	0.0738	0.0000
	Ward 3	-0.6193	0.1910	0.0012
	Year	-0.2202	0.0725	0.0024
		<i>AIC</i> = 4478.78		

The estimates indicate that the probability for a day to be inappropriate was high in 1990 in the medical ward. There was a decline in the correlation (degree of propagation) from 1988 to 1990 in the three Wards, and there was higher degree of propagation in the medical ward compared with all others.

SUMMARY AND DISCUSSION

First, various modelling approaches are compared for a given set of predictors merely through tests of significance of the individual coefficients. For the standard logistic regression, the fit is found to be poor ($D/df = 5.75$) because of the dependence between the binary observations. Consequently, the standard errors of the estimates are seriously underestimated, and all predictors except gender appeared to be significant ($p < 0.001$). As a first attempt to correct for this apparent overdispersion, the scaled deviance is used. After this adjustment the effects of four predictors are found to be insignificant.

The performance of the beta-binomial model can also be compared with all the methods to some extent. To have comparable results, the beta-binomial model consisting of the correlation parameter and the regression parameters in relating the mean response to the covariates is taken. It is observed that the significance probabilities associated with GEE and the beta-binomial are close to each other. This is partly true because of the exchangeable correlation structure assumed in both models. In the beta-binomial model, the correlation is estimated to be 0.46. This correlation is significant ($p < 0.0001$) and it confirms one's expectation of the natural clustering in the binary responses.

The application of the bootstrap method showed that the appropriate data generating mechanism is to resample the clusters rather than the binary 0–1 responses in order to preserve the dependence between responses belonging to the same cluster. One can obtain satisfactory estimates of the standard errors (which are under estimated by the standard logistic regression) by the bootstrap technique when other methods of modelling overdispersion are not readily available.

As to model selection, the AIC values associated with the standard logistic regression are found to be much larger than the AICs of other methods used to model overdispersion. Too many

covariates are identified to be important by the standard logistic regression. This is obviously due to underestimated standard errors. After correcting for overdispersion using scaled deviance, the model with only one predictor, namely, Year is selected based on the AIC values. The performance of the scaled deviance, which is often used as a quick means of handling overdispersion, appeared to be poor. Many significant covariates are not included in the best model. This correction for standard errors is based on the assumption that the subpopulation sizes are approximately equal. This is certainly not the case in the data analyzed. Therefore, though this adjustment of the standard errors is simple, one has to be cautious in using it.

For the beta-binomial model, the best model is found to be the one with the mean response (probability of inappropriateness) related to Ward, LOS, Year, and LOS^2 and the correlation (degree of propagation) related to Ward and Year. This is an improvement over the final model in the analysis of Gange *et al.* (1996), where the final model consists of Ward, LOS and Year for modelling mean response, and Year to model the correlation.

It should be remarked that the comparison in this paper is based on a single dataset. To get a more complete picture one may need to simulate correlated binary data under a wide range of experimental situations (*e.g.*, varying the intraclass correlation) and compare the models. But based on the analysis presented in this paper, we recommend the use of GEE and the beta-binomial model. As a full likelihood approach, the beta-binomial model allows the direct use of AIC as a model selection criterion as well as simultaneous modelling of probability and correlation parameter. If modelling of the correlation parameter is no issue, the GEE model, being a robust semiparametric method, offers an interesting alternative and as mentioned before, it can confirm the validity of the beta-binomial model. The same holds for the bootstrap method (resampling clusters), but its high computational cost makes it less tractable.

ACKNOWLEDGEMENTS

The authors thank Dr. Jordi Alonso of Municipal Institute of Medical Research (IMIM), Barcelona, Spain, for transferring the data. We are also thankful to the two referees for their constructive comments. The first author of this paper is supported by the Flemish Inter-university Council (VLIR).

REFERENCES

1. Aerts, M. and Claeskens, G. (1997). Local polynomial estimation in multiparameter likelihood models. *J. Amer. Statist. Assoc.* **92**:1536–1545.
2. Aerts, M., Geys, H., Molenberghs, G. and Ryan, L. (2002). *Topics in Modelling of Clustered Data*. Chapman and Hall, London.
3. Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. **In:** *2nd International Symposium on Information Theory*, pp. 267–281, (Petrov, B.N., ed.) Academia Kiado, Budapest.
4. Collett, D. (1991). *Modelling Binary Data*. Chapman & Hall, London, pp. 188–222.
5. Davison, A.C. and Hinkley, D.V. (1997). *Bootstrap Methods and their Application*. Cambridge University Press.
6. Gange, S.J., Muñoz, A., Sàez, M. and Alonso, J. (1996). Use of the beta-binomial distribution to model the effect of policy changes on appropriateness of hospital stays. *Appl. Statist.* **45**:371–382.
7. Liang, K.Y. and Zeger, S. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73**:13–22.
8. Lohr, K.N. (1998). Health services research, overview. **In:** *Encyclopedia of Biostatistics*. John Wiley & Sons Ltd, pp. 1885–1890.
9. McCullagh, P. and Nelder, J.A. (1989). *Generalized Linear Models*. 2nd ed. Chapman & Hall, London, pp. 124–128.
10. Pendergast, J.F., Gange, S.J., Newton, M.A., Linstrom, M.J., Palta, M. and Fishers, M.R. (1996). A survey of methods for analyzing clustered binary response Data. *Internat. Statist. Rev.* **64**:89–118.
11. Stocks, M.E., Davis, C.S. and Koch, G.G. (1995). *Categorical Data Analysis Using the SAS System*. SAS Institute Inc., Cary, NC.
12. Wedderburn, R.W.M. (1974). Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. *Biometrika* **61**:439–447.
13. Williams, D.A. (1975). The analysis of binary responses from toxicological experiments involving reproduction and teratology. *Biometrics* **31**:949–952.
14. Williams, D.A. (1982). Extra-binomial variation in logistic linear models. *Appl. Statist.* **31**:144–148.