JP Journal of Biostatistics
Volume 9, Number 2, 2013, Pages 67-81
Published Online: July 2013
Available online at http://pphmj.com/journals/jpb.htm
Published by Pushpa Publishing House, Allahabad, INDIA

COMPARING LOGISTIC RANDOM EFFECTS MODELING FOR CROSSOVER DESIGNS IN INFERTILITY

Boikanyo Makubate

College of Science
Botswana International University of Science and Technology
P/Bag B041, Gaborone, Botswana
e-mail: makubateb@biust.ac.bw

Abstract

Background. Within the field of infertility, hierarchically structured data are not hard to find. For example, in the assisted reproductive technology whereby a crossover design is conducted, observations will be clustered within couples. The random-effects logistic regression model is a very popular choice for the analysis of multilevel data. The purpose of this article is to compare different statistical software implementations of random-effects logistic regression models using the multilevel dataset from the crossover trial in infertility.

Methods. Sixty-two couples with primary or secondary infertility due to male factor entered the study. The sixty-two couples were randomly equally divided into two groups. Each group began one of the two treatment modalities (controlled ovarian hyperstimulation in conjunction with timed intercourse or intrauterine insemination) for three consecutive cycles and then switched to the alternative treatment after one rest cycle, if pregnancy was not achieved. Random effects logistic models were fit to a dataset of couples undergoing assisted

Keywords and phrases: crossover designs, logistic regression, random effects, multilevel, conditional estimates.

Received April 17, 2013

reproductive technology. The estimates obtained were compared among the four statistical packages.

Results. The parameter estimates (protocol) obtained from all the four statistical packages were not very much dissimilar. The software differs considerably in computing time. SAS© and R© took few seconds to compute the estimates.

Conclusions. In comparing the four statistical packages, it was found that the estimates were not very much dissimilar. Thus, there seems to be no explicit preference for either a frequentist or Bayesian approach. The choice for a particular implementation may largely depend on the computing time.

Background

What is multilevel modeling? Multilevel modeling is designed to explore and analyze data that come from populations which have a complex structure. Within the field of social, medical and biological sciences, multilevel or hierarchical structures are pervasive. Children may be nested within classes and classes nested within schools, patients may be nested within hospital centers and centers nested within towns. Multilevel data structures also arise in longitudinal studies where measurements are clustered within an individual [1] as it is the case in the crossover trials. Crossover designs are trials in which patients are allocated to sequences of treatment with the purpose of studying differences between individual treatments [2-6]. The commonest of crossover designs is the $AB : BA$ crossover design in which approximately half of the patients are first given treatment A and on a subsequent occasion treatment B whereas the rest of the patients are first given treatment B and on a subsequent occasion treatment A.

In longitudinal studies, repeated measurements of a response variable and a set of covariates are made on subjects across occasions. Because the within-subject measurements are likely to be positively correlated, the correlation must be accounted for by analysis appropriate to the longitudinal data. The standard logistic regression model described in Neter et al. [7] fails in its assumptions to accurately characterize the dependence in the data.

Basically, the standard logistic regression model assumes that the observations are independent, which they clearly are not when they are clustered within individuals. One solution to this problem is to generalize the model to the case of a combination of fixed (e.g., treatment) and random effects. The random effects allow the correlation between the repeated measurements to be incorporated into the estimates of parameters.

Regression estimates obtained from this technique are subject-specific, that is, they describe the individual's response (conditional estimates, conditional on the random effect). Conditional estimates represent the effect of a regressor on the outcome controlling for or holding constant the value of the random subject effect. On the other hand, the estimates from the standard logistic regression are "marginal" or "population-averaged" estimates. Marginal estimates represent the effect of a regressor averaging over the population of subjects.

In this article, a hierarchical dataset from the crossover trial in infertility with a binary outcome (pregnant or not pregnant) is used. Statistical methods that explicitly take into account hierarchically structured data have gained popularity in recent years, and there now exist several special purpose statistical packages designed specifically for estimating multilevel models. The purpose of this article is to compare different statistical software implementations, with regard to estimation results and computing time. The implementations include both frequentist and Bayesian approaches. Statistical software for hierarchical models has been compared already by Zhou et al. [8] and Guo and Zhou [9] about 15 years ago, and recently by Li et al. [1]. This article is different from previous reviews in that it is using the data set from designs that have a built-in tendency to produce missing data. It is the purpose of this article to show that even in data set from crossover trial in infertility the estimates obtained using different statistical softwares are not very much dissimilar.

Methods and Materials

The dataset used here is from Gregoriou et al. [10] based on a crossover

study to compare the pregnancy rates achieved by intrauterine insemination (IUI) and timed intercourse (TI) in gonadotrophin (hCG) stimulated cycles. In this dataset, sixty-two couples were randomly equally divided into two groups: group *A* or group *B*. Couples randomized to group *A* will begin with protocol 1 before switching to protocol 2. Couples randomized to group *B* receive protocols in the reverse order. For all couples, controlled ovarian hyper-stimulation (COH) was performed with the help of gonadotrophin (hCG), and either timed intercourse (TI) or intrauterine insemination (IUI) was employed. In protocol 1, timed intercourse (TI) was employed, while in protocol 2, intrauterine insemination (IUI) was employed. Couples stayed in the same protocol for at most 3 cycles before they can switch to the alternative protocol, with each couple receiving in total at most 6 cycles of protocols. For more details on this study, we refer to Gregoriou et al. [10]. The permission to use the patient data used in this study was obtained from the principle investigators of the original studies.

Logistic regression

To formulate the logistic model, let p_i represent the probability of a positive outcome $(\text{i.e., } Y_i = 1)$ for the *i*th individual. The probability of a negative outcome $(\text{i.e., } Y_i = 0)$ is then $1 - p_i$. Denote the set of covariates as $\mathbf{x}_i = (1, x_{i1}, ..., x_{ip})$, where $\beta = (\beta_0, \beta_1, ..., \beta_p)'$ is a $(p+1) \times 1$ vector of corresponding regression coefficients. Then the logistic regression model is written as

$$p_i = \frac{\exp(\mathbf{x_i'}\beta)}{1 + \exp(\mathbf{x_i'}\beta)} = \frac{1}{1 + \exp(-\mathbf{x_i'}\beta)} = \Psi(-\mathbf{x_i'}\beta), \tag{1}$$

where $\Psi(\cdot)$ is the logistic cumulative distribution function, namely,

$$\Psi(z) = \frac{1}{1 + \exp(-z)}. \tag{2}$$

This model can also be represented in terms of the log odds or logit of the probabilities, namely,

$$\log \frac{p_i}{1-p_i} = \mathbf{x_i'}\beta. \tag{3}$$

For estimation, with Y_i as a binary outcome variable from Bernoulli distribution, we have

$$Pr(Y_i) = \Psi^{Y_i}[1-\Psi]^{1-Y_i}, \tag{4}$$

where $\Psi = \Psi(-\mathbf{x_i'}\beta)$ as in equation (2). Following Hogg and Craig [11], the likelihood for a sample of *m* independent observations can be written as the product of equation (4) over the *m* individuals, i.e.,

$$L = \prod_{i=1}^{m} \Psi^{Y_i}[1-\Psi]^{1-Y_i}. \tag{5}$$

Thus, the log-likelihood function becomes

$$\log L = \sum_{i=1}^{m} [Y_i \log \Psi_i + (1-Y_i)\log(1-\Psi_i)]. \tag{6}$$

The log-likelihood above can now be maximized to obtain the maximum likelihood estimates (MLEs).

Random effects model

Let *i* denote the individuals and *j* denote the treatment. Let Y_{ij} be the value of the dichotomous outcome variable, coded 0 or 1, associated with treatment *j* nested within individual *i*. The logistic regression model is written in terms of the log odds (i.e., the logit) of the probability of a response, denoted p_{ij}. Considering a random-intercept model, augmenting the standard logistic regression model 3 with a single random effect yields:

$$\log \frac{p_{ij}}{1-p_{ij}} = \mathbf{x_{ij}'}\beta + v_i, \tag{7}$$

where $\mathbf{x_{ij}}$ is the $(p+1)\times 1$ covariate vector (includes a 1 for the intercept),

β is the $(p+1)\times 1$ vector of unknown regression parameters, and v_i is the random subject effect. These are assumed to be distributed in the population as $N(0, \sigma^2)$.

For estimation, we assume that there are $i = 1, ..., m$ subjects as before, each with $j = 1, 2, ..., n_i$ repeated observations. In the case of p predictor variables, the data are given by

$$(\mathbf{x}_1, \mathbf{Y}_{1j}), (\mathbf{x}_2, \mathbf{Y}_{2,j}), ..., (\mathbf{x}_m, \mathbf{Y}_{mj}),$$

where $\mathbf{Y}_{ij} = (Y_{i1}, Y_{i2}, ..., Y_{in_i})$ represents the value of the dichotomous response variables, and $x_i = (x_{i1}, ..., x_{ip})$ is the value of the predictor variables for the *i*th subject. The single response Y_{ij} is the *j*th response of the *i*th subject. Furthermore, $Y_{ij} = 1$ for success and $Y_{ij} = 0$ for failure. We consider estimation of a random-intercepts model, that is,

$$\log \frac{p_{ij}}{1 - p_{ij}} = \mathbf{x}'_{\mathbf{ij}}\beta + v_i. \tag{8}$$

The next step is to assume that the within-subject measurements are conditionally independent given the random subject effect (i.e., the random effects account completely for the correlation of the data within subjects). This assumption is critical and is known in Agresti [12], as the conditional independence assumption. Because the within-subject measurements are assumed to be conditionally independent, the conditional likelihood of n_i measurements within the *i*th subject is given by:

$$\ell(\mathbf{Y_i} \mid v) = \prod_{j=1}^{n_i} [\Psi(z_{ij})]^{Y_{ij}} [1 - \Psi(z_{ij})]^{1-Y_{ij}}. \tag{9}$$

To get the likelihood of the n_i response patterns for all the *m* subjects, we need to have an expression for the likelihood of $\mathbf{Y_i}$ that does not depend on the random effects. We can arrive at such an expression by integrating over the distribution of the random effects. This yields the marginal

probability for $\mathbf{Y_i}$ in the population of subjects as:

$$h(\mathbf{Y_i}) = \int_v \ell(\mathbf{Y_i} \mid v) g(v) dv, \tag{10}$$

where $g(v)$ represents the population distribution of the random effects *v*, namely, $N(0, \sigma^2)$.

We can now form the marginal likelihood of the response patterns $\mathbf{Y_i}$ from all subjects, and thus the total sample, by multiplying each of the subject's marginal likelihoods together.

Namely,

$$L = \prod_{i=1}^{m} h(\mathbf{Y_i}) \tag{11}$$

or

$$\log L = \sum_{i=1}^{m} \log h(\mathbf{Y_i}). \tag{12}$$

It is easier to manipulate the log-likelihood in equation (12). If we choose the values of the parameters that maximize the log-likelihood in equation (12), then those same values will also maximize the likelihood in equation (11).

Integration over the random effect distribution

In order to solve the above likelihood solutions, integration over the random effects distribution must be performed. Various approximations for evaluating the integral over the random effects distribution have been proposed in the literature. Perhaps the most frequently used methods are based on the first or second order Taylor expansions [13]. Numerical integration can also be used to perform the integration over the random-effects distribution. Specifically, if the assumed distribution is normal, then Gauss Hermite quadrature can approximate the above integral to any practical degree of accuracy [14]. The integration is approximated by a

summation on a specified number of quadrature points Q for each dimension of the integration.

Consider the Gaussian integration formula for the Hermite polynomial in [14]. The Hermite polynomials are defined over $[-\infty, \infty]$ and the weighting function of Hermite polynomials is

$$w(x) = e^{-x^2}. \tag{13}$$

Therefore, Gauss Hermite quadrature naturally gives the integration for

$$\int_{-\infty}^{\infty} f(x)e^{-x^2}dx. \tag{14}$$

Thus, Gauss Hermite quadrature can be naturally associated with normal distribution as follows: suppose we were to evaluate

$$\int_{-\infty}^{\infty} f(y)e^{-\frac{(y-u)^2}{\sqrt{2}\sigma}}dy, \tag{15}$$

where $f(y)$ is a function of y. Substituting

$$x = \frac{y-u}{\sqrt{2}\sigma} \tag{16}$$

in equation (16) yields

$$\int_{-\infty}^{\infty} f(y)e^{-\frac{(y-u)^2}{\sqrt{2}\sigma}}dy = \int_{-\infty}^{\infty} f(\sqrt{2}\sigma x + \mu)e^{-x^2}\sqrt{2}\sigma dx$$

$$= \sqrt{2}\sigma \sum_{i=1}^{N} w_i f(\sqrt{2}\sigma x + \mu), \tag{17}$$

where following [14], $\{x_i\}_{i=1}^{N}$ are the roots of order N Hermite polynomial $P_N(x)$ and $\{w_i\}_{i=1}^{N}$ are:

$$w_i = \int_{-\infty}^{\infty} \prod_{\substack{j=1 \\ j\neq i}}^{N} \frac{x - x_j}{x_i - x_j} e^{-x^2}dx. \tag{18}$$

Therefore, if y is normally distributed with mean μ and variance σ^2, the expected value of $f(y)$, $(E[f(y)])$ is given by:

$$\frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty} f(y)e^{-\frac{(y-u)^2}{\sqrt{2}\sigma}}dy = \frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty} f(\sqrt{2}\sigma x+\mu)e^{-x^2}\sqrt{2}\sigma dx$$

$$= \frac{1}{\sqrt{\pi}}\sum_{i=1}^{N} w_i f(\sqrt{2}\sigma x+\mu), \qquad (19)$$

where $\{x_i\}_{i=1}^{N}$ and $\{w_i\}_{i=1}^{N}$ are the same as above.

With the numerical Gaussian quadrature integration, the approximation to the marginal likelihood gets better as the number of quadrature points increases. However, as the dimension of the random effects increases, the number of quadrature points increases exponentially; the total number of quadrature points required for all the random effects Q^r, where r is the number of random effects. The numerical quadrature becomes computationally burdensome when there are more than 5 random effects [15].

Statistical software

The **NLMIXED** in SAS© procedure fits nonlinear mixed models, that is, models in which both fixed and random effects are permitted to have a nonlinear relationship to the response variable. These models can take various forms, but the most common ones involve a conditional distribution for the response variable given the random effects. **PROC NLMIXED** enables us to specify such a distribution by using either a keyword for a standard form (normal, binomial, Poisson) or SAS© programming statements to specify a general distribution.

PROC NLMIXED fits the specified nonlinear mixed model by maximizing an approximation to the likelihood integrated over the random effects. Different approximations to the integral (10) are available, and the two principal ones are the one we used, Gaussian quadrature and a first order

Taylor series approximation. There are a variety of alternative optimization techniques; the default is the Newton Raphson described in the previous sections. Standard errors are obtained by the Delta method. For the theory and computational techniques of **PROC NLMIXED,** Pinheiro and Bates [16] are strongly recommended. Other softwares which are capable of handling the random effects include GenStat©, Stata©, MathCad©, R©, S-Plus© and WinBugs©. In this article, attention is focused on the use of SAS©, R©, GenStat© and WinBugs© to perform random effects model.

Analysis

The data can be analyzed as the $AB : BA$ crossover design, where the first 3 cycles constitute the first period and the second 3 cycles constitute the second period. That is to say that if one couple conceives under cycle 1 and the other couple conceives under cycle 2 or 3, then those two couples are regarded as having conceived under the first period. Similarly, if one couple conceives under cycle 4 and the other couple conceives under cycle 5 or 6, then those two couples are regarded as having conceived under the second period. It is possible to have treated each cycle as a period. Here, we cannot do that as we are restricted by the way the data is presented. The way the data is presented does not allow us to treat each cycle as a period. The data is analyzed using the random effects model described above. We consider four types of models: (1) empty or null model, (2) the model involving protocol only, (3) the model involving period effects only and (4) the model involving both protocol and period effects.

Results

Descriptive statistics

Of the 62 couples enrolled, 20 couples conceived, of which 12 conceived during the first 3 cycles and 8 conceived in the second 3 cycles. Couples left the study if and only if they conceived. Table 1 is taken from Gregoriou et al. [10] shows the achieved pregnancy rates in each group of attempts by the two protocols. Table 2 gives the summary results of fitting various models

for the dataset in Gregoriou et al. [10] using different fitting methods. The models are, in order of fitting: a null model, a model involving protocol (treatment) only, a model involving period only and a model fitting protocol and period. Four statistical packages: SAS©, R©, WinBugs© and GenStat© are illustrated. SAS© and R© give similar results in all the four models since the general fitting criterion (maximum likelihood) is the same and only details of numerical implementation are different. In fact, the deviances are in good agreement between SAS©, R© and GenStat©. All the four methods indicate that IUI is more effective than TI, and that couples undergoing IUI have nearly four times higher odds of conception than couples undergoing TI.

Table 1. Pregnancy rates achieved in each group of attempts by the two protocols

	Method of treatment									
	COH + TI					COH + IUI				
Attempts	Patients	Cancelled cycles	Completed cycles	Conceptions	PR per cycle (%)	Patients	Cancelled cycles	Completed cycles	Conceptions	PR per cycle (%)
	Group A					*Group B*				
1-3	31	15	73	4	5.5	31	14	70	8	11.4
	Group B					*Group A*				
4-6	23	14	55	1	1.4	27	13	60	7	11.7
Total	54	29	128	5	3.9	58	27	130	15	11.5

Table 2. Comparative analysis of the Gregoriou data

Model	Method	Intercept	Protocol	Period	Deviance
Null	SAS©	–1.7140 (0.5055)			104.5
	R©	–1.7111 (0.2904)			104.5
	WinBugs©	–1.544 (0.01676)			-
	GenStat©	–1.525 (0.2457)			104.811
Protocol	SAS©	–1.9607 (0.6267)	–0.6725 (0.3269)		99
	R©	–1.9777 (0.3518)	–0.6778 (0.3216)		99.04
	WinBugs©	–1.726 (0.2889)	–0.6516 (0.2868)		-
	GenStat©	–1.166 (0.3070)	–0.890 (0.5305)		91.507
Period	SAS©	–1.6761 (0.5830)		0.04682 (0.3338)	104.5
	R©	–1.67013 (0.28528)		0.04942 (0.26754)	104.5
	WinBugs©	–1.586 (0.2574)		0.1249 (0.2563)	-
	GenStat©	–1.6595 (0.3829)		0.2300 (0.4993)	104.646
Protocol and period	SAS©	–1.8976 (0.6157)	–0.6639 (0.3196)	0.08964 (0.3157)	99
	R©	–1.9018 (0.3408)	–0.6643 (0.3133)	0.0873 (0.2868)	98.96
	WinBugs©	–1.77 (0.2973)	–0.6629 (0.2915)	0.156 (0.2645)	-
	GenStat©	–1.3232 (0.4207)	–0.9100 (0.5326)	0.2883 (0.5104)	96.267

Discussion

Although the parameter estimates were not very much dissimilar between the four software implementations, a considerable variation in computing time was observed. SAS© **NLMIXED** and R© **lme4** were the fastest taking only seconds to produce the results. GenStat© **HGANALYSE** was slower than both SAS© and R©, however, it was very much faster than WinBUGS©. WinBUGS© was invariably slower than the frequentist approaches, which is due to the computational intensive MCMC approach and that convergence is much harder to judge than in a classical frequentist sense.

Conclusions

In this article, we used a likelihood-based approach to the statistical analysis of pregnancy data from a crossover design. This approach was based on the logistic random effects model incorporating both protocol and period effects. In both scenarios (the model with protocol only and the model with protocol and period effects), the protocol estimates obtained under the four statistical packages were not very much dissimilar. The parameter estimates from logistic random effects regression models should not be influenced by the choice of the statistical package. The choice of the statistical implementation in a clinical trial should depend on other factors such as speed and desired flexibility. This study shows that if there is no prior acquaintance with a certain package frequentist approach has to be given first preference. SAS© **NLMIXED** and R© **lme4** are highly recommended because of their efficiency. This study also supports the findings by Li et al. [1].

Acknowledgement

I thank my employer (Botswana International University of Science and Technology) for providing me with time and finances to publish this article.

References

[1] B. Li, H. F. Lingsma, E. W. Steyerberg and E. Lesaffre, Impact of environment, stress, occupational, and other hazards on sexuality and sexual behavior, Environmental Health Perspectives 101 (1993), 101-107.

[2] M. J. Campbell, Medical Statistics: A Commonsense Approach, 3rd ed., Wiley, 1999.

[3] B. Makubate, The place of crossover designs in infertility trials, Ph.D. Thesis, University of Glasgow, 2009.

[4] B. Makubate and S. Senn, Planning and analysis of cross-over trials in infertility, Stat. Med. 29(30) (2010), 3203-3210.

[5] J. McDonnell, A. Goverde and J. Vermeiden, The place of the crossover design in infertility trials: a maximum likelihood approach, Hum. Reprod. 19 (2004), 2537-2544.

[6] S. Senn, Cross-over Trials in Clinical Research, 2nd ed., Wiley, New York, 2002.

[7] J. Neter, M. Kutner, C. Nachtsheim and W. Wasserman, Applied Linear Statistical Models, 4th ed., McGraw-Hill Companies, 1996.

[8] X. Zhou, A. J. Perkins and S. L. Hui, Comparisons of software packages for generalized linear multilevel models, Amer. Statist. 53 (1999), 282-290.

[9] G. Guo and H. Zhao, Multilevel modelling for binary data, Ann. Rev. Sociology 26 (2000), 441-462.

[10] O. Gregoriou, N. Vitoratos, N. Papadias, S. Konidaris, A. Gargaropoulos and D. Rizos, Pregnancy rates in gonadotrophin stimulated cycles with timed intercourse or intrauterine insemination for the treatment of male subfertility, European J. Obstetrics, Gynecology, and Reproductive Biology 64 (1996), 213-216.

[11] R. Hogg and A. Craig, Introduction to Mathematical Statistics, 5th ed., Prentice Hall, Upper Saddle River, New Jersey, 1995.

[12] A. Agresti, Categorical Data Analysis, 2nd ed., John Wiley and Sons, New Jersey, 1989.

[13] A. Skrondal and S. Rabe-Hesketh, Generalized Latent Variable Modeling: Multilevel Longitudinal, and Structural Equation Models, Chapman and Hall/ CRC, New York, 2004.

[14] A. Stroud and D. Secrest, Gaussian Quadrature Formulas, Prentice Hall, Englewood Cliffs, 1966.

[15] E. Lesaffre and B. Spiessens, On the effect of the number of quadrature points in a logistic random-effects model: an example, Applied Statistics 50 (2001), 325-335.

[16] J. Pinheiro and D. Bates, Approximations to the log-likelihood function in the nonlinear mixed effects model, J. Comput. Graph. Statist. 4 (1995), 12-35.