Journal of Biostatistics and Epidemiology

J Biostat Epidemiol. 2017; 3(1): 7-12.

Original Article

Multilevel modeling of clustered grouped survival data of dental implant failures; a Bayesian approach

Abbas Rahimiforoushani[1], Nooshin Akbari-Sharak[1*], Mohammadjavad Kharazi-Fard[1]

[1] Department of Epidemiology and Biostatistics, School of Public Health, Tehran University of Medical Sciences, Tehran, Iran

ARTICLE INFO

Received 17.05.2016
Revised 04.08.2016
Accepted 12.09.2016
Published 04.03.2017

Key words:
Grouped clustered failure time;
Intracluster correlation;
Monte Carlo Markov Chain;
Non-informative prior;
Bayesian approach;
OpenBUGS

ABSTRACT

Background & Aim: In clinical dental studies, each participant has usually several visits, and since the review and ongoing monitoring of the subjects are often expensive or even impossible, so people are examined periodically during regularly pre-scheduled visits. Therefore, discrete or grouped clustered failure time data are collected. We aimed to show the use of Monte Carlo Markov Chain (MCMC) and the non-informative prior in a Bayesian framework in multilevel modeling of clustered grouped survival data.

Methods & Materials: A two-level model with additive variance components model for the random effects was considered. Both the grouped proportional hazards model and logistic regression with logit link function model were used. Using grouped proportional hazards method, we could approximate intracluster correlation of the log failure times. The statistical package OpenBUGS was adopted to estimate the parameter of interest based on the MCMC method. A cohort study was used in which 1011 persons visited at clinic dentistry of Tehran University of Medical Sciences, Iran, between the years 2002 and 2013 for dental implant and 2368 implants were placed for them in total. Clinical status of dental implants was evaluated in three periods after placement, thus clustered grouped failure times of the dental implants were recorded.

Results: The grouped proportional hazards model showed that clustering effect among the log failure times of the different implants from the same person was fairly strong (correlation = 0.99). Complication and biomaterial variables had no effect on the implant failure, and there was no difference in the failure times related to the molar, premolar, canine, primary, and incisor since 95% credible interval (CI) included 0. The CI related to the gender and place of teeth not including 0, so these variables were significant in the model. The estimates of the baseline parameters (γ_1, γ_2, and γ_3) were increasing indicating increasing hazard rates from interval 1-3. Results of logistic regression were similar to grouped proportional hazards model with wider confidence intervals.

Conclusion: The use of MCMC approach and non-informing prior in Bayesian framework to mimic maximum likelihood estimations in a frequentist approach in multilevel modeling of clustered grouped survival data can be easily applied with the use of the software OpenBUGS.

Introduction

In practice, monitoring the study subjects continuously is mostly expensive or even impossible; therefore, people are examined periodically at some regular pre-scheduled visits. Hence, grouped or discrete clustered failure time data are collected. Marginal approach in analyzing discrete clustered failure time data was proposed by Ross and Moore (1) that the marginal hazard of failure for individual subjects within a cluster is specified by a linear log odds survival model and the dependence structure is based on a gamma frailty model.

* Corresponding Author: Nooshin Akbari-Sharak, Postal Address: Department of Epidemiology and Biostatistics, School of Public Health, Tehran University of Medical Sciences, Tehran, Iran. Email: noooshin_akbari@yahoo.com

Please cite this article in press as: Rahimiforoushani A, Akbari-Sharak N, Kharazi-Fard M. Multilevel modeling of clustered grouped survival data of dental implant failures; a Bayesian approach. *J Biostat Epidemiol. 2017; 3(1): 7-12*

The dependency parameter value in discrete model provides vague information in the dependency level but intracluster association remains unknown. These methods are not suitable for data with large cluster size since estimation process is too complicated in practice. Random effects approach in generalized linear mixed model uses maximum likelihood and residual maximum likelihood to estimate the regression and dependence parameters which were proposed by Lam and Ip (2). Logistic regression and grouped proportional hazards model are considered that regression parameters in these models can be, respectively, interpreted as natural logarithm of the odds ratio and the natural logarithm of the relative risks. Estimation of intracluster correlation between the natural logarithm of the failure times from a same cluster is computable when grouped proportional hazards model used. The suggested models assumed that individuals in a same cluster share invisible random effects. As Lam and Ip stated, expanding models into multilevel modeling of this type of data, using REML and ML methods to accommodate more complicated dependence structure is only possible theoretically and impossible in practice because the dimension of design matrix would be extremely large and makes the computation difficult particularly in the estimation of the variance of the dependence parameters. Numerical estimations will be unstable and empirical and asymptotic properties of estimators may not be close to each other for a fixed moderate cluster size and each estimator's performance might be unknown. The aim of this article is to show the application of Monte Carlo Markov Chain (MCMC) and non-informative prior in Bayesian framework to mimic maximum likelihood estimations with the most common method in multilevel modeling of grouped survival data. The models considered in this article are extensions of the shared random effect models extended into multilevel modeling which are suggested by Lam and Ip (2) that allow a very general class of random effects models to accommodate situations with more complicated random effects structures. In clinical, it is important to predict future individual event using failure times and the covariate information from other observations in the same cluster, this class of generalized linear mixed models are preferred to marginal models because no prediction can be made with marginal models. The estimation will do by MCMC which is a popular tool for analyzing complex hierarchical data and regarding to advancements in computational technologies, it has found its way into the medical, public health, and dental arena.

Methods

A sample of 1011 persons visited at clinic dentistry of Tehran University of Medical Sciences between the years 2001 and 2012 for dental implant and 2368 implants were placed for them in total. Personal information of patients was gathered by direct questioning and information about implant was recorded by the respective doctor. The clinical status of dental implants was evaluated in three periods. Years 2001-2004 is the first period, 2005-2008 is the second, and 2009-2012 is the third period. The clinical statue of implants was determined by Boozer criterions which are the lack of complaints about permanent pains or the feeling of external object or Distzy, lack of infection around the implant area with suppuration, lack of looseness, and lack of permanent radiolucency around implant. If any implant failed, it gets the code 1 and if not, it got the code 0. If a failure happens within a period, that period and the one after that gets the code 1 and if it does not fail all three periods get the code 0. Suppose that T_{ij} is the failure time related to j^{th} implant from i^{th} person ($i = 1, 2\ldots, n$; $j = 1, 2\ldots, n_i$) measured in a continuous scale and x_{ij} is the p-dimensional vector of observed covariates associated with T_{ij}. Conditioned on random effect of e_i, T_{ij}'s are mutually independent. Assume time axis is divided into mutually independent intervals where with $a_s = 0$ and $a_d = \infty$. Furthermore, assume that $s_{ij} = s$ be the event that j^{th} implant from i^{th} person fails in I_s (i.e., $a_{s-1}<T_{ij}\leq a_s$). The conditional discrete hazard function λ is defined to be:

$$\lambda(s \mid x_{ij}, e_i) = P(S_{ij} = s|S_{ij}\geq s, x_{ij}, e_i) \qquad (1)$$

For modeling the conditional discrete hazard function (1) we use the binary regression with the general form λ with η and g is a

transformation that maps the whole real line (−∞, +∞), onto the unit interval l(0,1). γ_s (s = 1,…d) are a sequence of baseline parameters that show each interval has unique intercept and β is the vector of regression parameter. Two choices of g are of particular interest in this paper, the first case being the grouped proportional hazard regression model:

$$g(\eta) = 1-\exp[-\exp(\eta)] \quad (2)$$

And the second being the logistic regression:

$$g(\eta) = \exp(\eta)/[1+\exp(\eta)] \quad (3)$$

The covariate vector can be time-dependent by assuming that X_{ij} is constant in each interval. Distribution of random effects which determines the dependency structure of data is required for complete specification of model, therefore it is assumed that random effect e_i is consistent over time and normally distributed with a mean of 0 and a variance of σ_0^2. Overall, the choice of the distribution of the random effect e_i does not have strong impact on the estimate of β. As σ_0^2 characterizes the strength of association among individuals within the cluster, it is generally called the dependence parameter. Assume that failure time T_{ij} is observed to have failed in or censored right after the interval. Failure indicator is defined in this way:

$$y_{ijs} = \begin{cases} 1 & \text{if subject j of cluster i fails in } I_s^* \\ 0 & \text{if subject j of cluster i survives through } I_s^* \end{cases}$$

Therefore, the complete data log-likelihood function L can be expressed as:

$$L = \prod_{i=1}^{n}\prod_{j=1}^{n_i}\prod_{s=1}^{s_{ij}^*}[g(\eta_{ijs})^{y_{ijs}}\{1-g(\eta_{ijs})\}^{(1-y_{ijs})}] \quad (4)$$

In which L is the conditional log-likelihood of Y taking the random effect e as fixed. Assume that $m_i = \sum_{j=1}^{n_i} S_{ij}^*$ and $m = \sum_{i=1}^{n} m_i$.

We can express the log-likelihood function in the form of matrix by considering η_{ijl} as m dimensional vector.

$$\eta = X\beta^* + Ze$$

In which, $\beta^{*T} = \{\gamma_1, \ldots, \gamma_d, \beta_1^*, \ldots, \beta_p^*\}$, $e^T = \{e_1,\ldots,e_n\}$ and X and Z are design matrices.

Hence, by the usual Newton–Raphson iterative procedure for maximizing L with the set of initial values β_0^* and e_0 of β* and e. The estimations of β* and e can be obtained via updating the following equation:

$$\begin{pmatrix}\beta^* \\ \hat{e}\end{pmatrix} = \begin{pmatrix}\beta_0 \\ u_0\end{pmatrix} - v^{-1}\begin{pmatrix}0 \\ \theta^{-1}u_0\end{pmatrix} + v^{-1}(x \;\; z)^T\frac{\partial 1}{\partial \eta}$$

That

$$v = \begin{pmatrix}v_{11} & v_{12} \\ v_{21} & v_{22}\end{pmatrix} = \begin{pmatrix}x^T \\ z^T\end{pmatrix}\left[\frac{-\partial^2 1}{\partial\eta\,\partial\eta^T}\right](x \;\; z) + \begin{pmatrix}0 & 0 \\ 0 & \theta^{-1}I\end{pmatrix}$$

$$v^{-1} = \begin{pmatrix}v_{11} & v_{12} \\ v_{21} & v_{22}\end{pmatrix}$$

Matrix I is an identity matrix. The asymptotic variance-covariance matrix of the estimator β* is given by A11. The maximum likelihood estimator of θ is obtained by the following equation:

$$\hat{\theta}_{ML} = \hat{e}^T\hat{e}/(n - R_{ML})$$

And the asymptotic variance of $\hat{\theta}_{ML}$ can be estimated by the following equation:

$$2\hat{\theta}_{ML}^2\left[n - 2R_{ML} + \hat{\theta}_{ML}^{-2}\,tr\,(v_{22}^{-2})\right]^{-1}$$

Where, $R_{ML} = \hat{\theta}_{ML}^{-1}\,tr\,v_{22}^{-1}$ and the residual maximum likelihood estimator of θ is obtained in the same way:

$$\hat{\theta}_{REML} = \hat{e}^T\hat{e}/(n - R_{REML})$$

And the asymptotic variance of $\hat{\theta}_{REML}$ is like the following:

$$2\hat{\theta}_{REML}^2\left[n - 2R_{REML} + \hat{\theta}_{ML}^{-2}\,tr\,(A_{22}^{2})\right]^{-1}$$

Where, $R_{REML} = \hat{\theta}_{REML}^{-1}\,tr\,A_{22}$

In this article, based on likelihood of the observed data, Bayesian approach and imposing non-informative prior distribution for each of the parameters of interest we estimate the parameters. Variables $\gamma_1,\ldots,\gamma_d$, $\beta_1,\ldots,\beta_p$ are independently distributed according to $N(0,\varphi_1)$ and $1/\sigma_0^2$ are independently distributed according to a gamma distribution of $1/\varphi_1$ and $1/\varphi_2$ where φ_1 and φ_2 are chosen to be very large in practice, say 1000.

Under the assumption of exponential distribution for T, the correlation coefficient

logarithmic failure times from the same clusters is independent of independent variables and is like this (3):

$$corr(\log(T_{ij}), \log(T_{ij}) = \frac{\sigma_0^2}{\sigma_0^2 + \frac{\pi^2}{3}} \qquad j \neq 1$$

Where, σ_0^2 is the random effect variance of e_i.

Empirical results showed that these correlation coefficients are good approximation when the failure times follow a general proportional hazards model and are highly robust to the misspecification of the dependence structure of the random effects (2, 4).

Results

Both the grouped proportional hazards model and logistic regression model are considered here. A two-level model was considered that the implant or an implant tooth as level 1 unit and person is considered as level 2 unit. The sex variable (male $x_1 = 0$, female $x_1 = 1$) complication (there is a problem $x_2 = 0$, there is no problem while surgery $x_2 = 1$), biomaterial (the chemicals have been used $x_3 = 0$, the chemicals have not been used $x_3 = 1$), the place of tooth (upper jaw $x_4 = 0$, lower jaw $x_4 = 1$), and kind of tooth (molar tooth $x_5 = 1$, $x_6 = 0$, $x_7 = 0$, $x_8 = 0$, premolar tooth $x_6 = 1$, $x_7 = 0$, $x_8 = 0$, canine tooth $x_5 = 0$, $x_6 = 0$, $x_7 = 1$, $x_8 = 0$, and primary tooth $x_5 = 0$, $x_6 = 0$, $x_7 = 0$, $x_8 = 1$) were included as covariates. The graphical presentation of the model which is drawn in OpenBUGS software (OpenBUGS Foundation, Helsinki, Finland) is like this figure 1.

Where:

- Mu[i]: is the indicator of changing average related to each implant which is dependent on variables and other existing factors in model.
- Q [I, int[i]]: each person's failure possibility indicator.
- Y[i]: is the indicator of failure status for each person.
- Gender[i]: gender variable, complication: the variable of the existence of problem while doing surgery, location 1: the variable of tooth place, location1: Molar teeth location 2: Premolar tooth. Location 3: canine tooth and loc4 shows the primary tooth.
- V[j]: shows the effect of each person that sigma 2. Child is the related variance and tau. Child is its accuracy.
- Beta[j]: regression coefficients related to the parameters.
- Gam[j]: shows the effect of each interval.

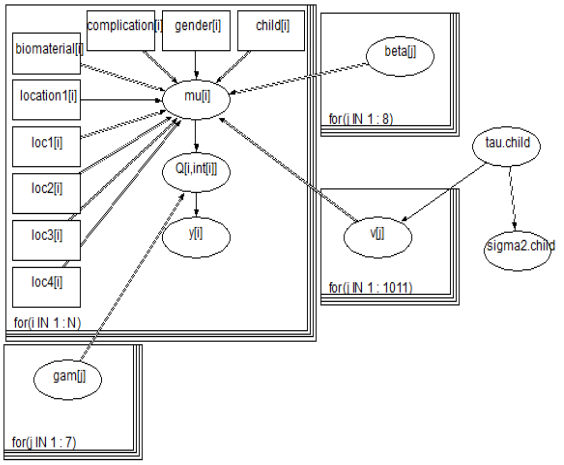

Figure 1. Graphical presentation of the estimated survival models in the analysis of the implant teeth data

In the estimation of the parameters, the first 5000 simulations were treated as burn-ins and discarded while the estimation was based on the next 10,00,000 simulations. Parameter estimates of the two models are shown in tables 1 and 2.

Results of the grouped proportional hazard model revealed that clustering effect among the log failure times of implants from the same person was fairly strong (correlation = 0.99). Complication and biomaterial variables have no effect on the implant failure, and there is no difference in the failure times related to the molar, premolar, canine, primary, and incisor since 95% credible interval (CI) included 0. The CI related to the gender and place of teeth not including 0 so these variables are significant in the model.

The estimates of the baseline parameters (γ_1, γ_2, and γ_3) were increasing indicating increasing hazard rates from interval 1-6. The simulation error related to all parameters is below 0.05, and it shows the convergence. Results of the logistic regression models were similar to those of the grouped proportional hazard model.

Table 1. Parameter estimates using OpenBUGS and grouped proportional hazards model results from 10,00,000 simulations after 5000 burn-ins

Variable	Grouped Proportional Hazards Model			
	Median	Confidence interval	SD	McErorr
Sex	−10.22	(−26.25, −1.296)	6.472	0.1972
Complication	−10.44	(−26.46, 1.661)	7.039	0.2187
Biomaterial	−8.337	(−22.03, 3.876)	6.477	0.197
Location 1	−13.75	(−29.49, −5.18)	6.111	0.1757
Location 1	−7.54	(−21.03, 1.251)	5.653	0.1664
Location 2	−2.113	(−12.24, 5.288)	4.333	0.1196
Location 3	−14.14	(−41.82, 1.959)	11.01	0.324
Location 4	2.773	(−14.12, 21.01)	8.639	0.2669
γ_1	−53.23	(−92.0, −28.94)	16.22	0.2662
γ_2	−26.83	(−45.02, −14.27)	7.566	0.2368
γ_3	−25.5	(−43.49, 13.2)	7.46	0.2336
σ^2_{person}	467.0	(−115.7, 1432.0)	338.0	10.41
$\text{corr}_{\text{person}}$	0.993	(0.9724, 0.9977)	0.006	2.086E-4

SD: Standard deviation

Discussion

Recently in dental research, different approaches have been proposed for handling clustered survival data with exact failure times (5-11). Different models and various Software like (SAS, S-plus) have been recommended for processing the analysis. In this study, the OpenBUGS software is used for analyzing the data in dentistry. The general multi-level formulas permit that we can study more incidental regression cases.

Bayesian inference has several advantages over the usual approaches, particularly in the flexibility of model building for complex data. Bayesian inference enables us to make exact inference for any sample size without resorting to asymptotic calculations, where usual approach relies a lot on asymptotic approximation and always there is the issue of whether the sample size is large enough for the asymptotic approximation to be valid (12). In this paper, Bayesian analysis performed with the use of non-informative priors and mimics usual approach.

The choice of non-informative prior for the random parameters are not straightforward, and there exists no standard choice for non-informative priors (13-15).

In this paper, the choice of the inverse gamma as the non-informative prior is mainly for mathematical convenient purpose. Moreover, as remarked by Turner et al. (14) that an inverse gamma prior for the variance of the random effects often leads to improved bias and coverage properties when compared with a locally uniform prior.

Table 2. Parameter estimates using OpenBUGS and logistic regression model results from 10,00,000 simulations after 5000 burn-ins

Variable	Logistic regression model			
	Median	Confidence interval	SD	McErorr
Sex	−10.66	(−28.22, −0.3781)	7.167	0.1895
Complication	−10.51	(−29.66, 1.151)	7.944	0.2063
Biomaterial	−8.323	(−25.41, 3.999)	7.263	0.1771
Location 1	−14.19	(−30.06, −5.418)	6.383	0.1669
Location 1	−7.635	(−22.61, 1.18)	6.064	0.1433
Location 2	−2.053	(−13.6, 5.509)	4.718	0.09597
Location 3	−13.23	(−38.35, 1.086)	10.16	0.2372
Location 4	3.949	(−14.15, 24.32)	9.448	0.1776
γ_1	−54.24	(−93.61, 28.05)	16.81	0.2888
γ_2	−26.88	(−47.52, 12.83)	9.012	0.2625
γ_3	−25.34	(−45.68, −11.63)	8.851	0.2568
σ^2_{person}	519.8	(−124.6, 1478.0)	368.0	11.16

SD: Standard deviation

Conclusion

Comparing the results obtained from this analysis with those reported previously by treating the multiple failure times from the same person as independent observations by ignoring the clustering effects, the estimated survival rates obtained from this analysis were higher. Since the correlation of the log failure times of the different implants from the same person was fairly strong (correlation = 0.99). The analysis without considering this correlation was not appropriate, and the statistical inference may not be valid.

Conflict of Interests

Authors have no conflict of interests.

Acknowledgments

Thanks to Dr. Kharazi-Fard and Clinic Dentistry of Tehran University of Medical Sciences, for their permission for us to use the dataset for this paper. The work described in this paper was extracted of a Master Thesis from Tehran University of Medical Science (Project No. 240.655).

References

1. Ross EA, Moore D. Modeling clustered, discrete, or grouped time survival data with covariates. Biometrics 1999; 55(3): 813-9.
2. Lam KF, Ip D. REML and ML estimation for clustered grouped survival data. Stat Med 2003; 22(12): 2025-34.
3. Lindeboom M, Van Den Berg GJ. Heterogeneity in models for bivariate survival: The importance of the mixing distribution. J R Stat Soc Series B 1994; 56(1): 49-60.
4. Lam KF, Lee YW. Merits of modelling multivariate survival data using random effects proportional odds model. Biom J 2004; 46(3): 331-41.
5. Gilthorpe MS, Mayhew MT, Bulman JS. Multilevel survival analysis of amalgam restorations amongst RAF personnel. Community Dent Health 2002; 19(1): 3-11.
6. Chuang SK, Wei LJ, Douglass CW, Dodson TB. Risk factors for dental implant failure: A strategy for the analysis of clustered failure-time observations. J Dent Res 2002; 81(8): 572-7.
7. Chuang SK, Tian L, Wei LJ, Dodson TB. Predicting dental implant survival by use of the marginal approach of the semi-parametric survival methods for clustered observations. J Dent Res 2002; 81(12): 851-5.
8. Harkanen T, Virtanen JI, Arjas E. Caries on permanent teeth: A non-parametric bayesian analysis. Scandinavian Journal of Statistics 2000; 27(4): 577-88.
9. Hannigan A, O'Mullane DM, Barry D, Schafer F, Roberts AJ. A re-analysis of a caries clinical trial by survival analysis. J Dent Res 2001; 80(2): 427-31.
10. Harkane T, Larmas MA, Virtanen JI, Arjas E. Applying modern survival analysis methods to longitudinal dental caries studies. J Dent Res 2002; 81(2): 144-8.
11. Bogaerts K, Leroy R, Lesaffre E, Declerck D. Modelling tooth emergence data based on multivariate interval-censored data. Stat Med 2002; 21(24): 3775-87.
12. Ibrahim JG, Chen MH, Sinha D. Bayesian survival analysis. Berlin, Germany: Springer Science & Business Media; 2001.
13. Kass RE, Wasserman L. The selection of prior distributions by formal rules. J Am Stat Assoc 1996; 91(435): 1343-70.
14. Turner RM, Omar RZ, Thompson SG. Bayesian methods of analysis for cluster randomized trials with binary outcome data. Stat Med 2001; 20(3): 453-72.
15. Spiegelhalter DJ. Bayesian methods for cluster randomized trials with continuous responses. Stat Med 2001; 20(3): 435-52.