

Correlated frailty model: an advantageous approach for covariate analysis of tuberculosis data

C. Ponnuraja and P. Venkatesan
Tuberculosis Research Centre (ICMR), Chennai-600 031, India.
 cponnuraja@gmail.com, venkaticmr@gmail.com

Abstract

The demonstration of varying treatment effects among different subjects of patients is an important part of the analysis of clinical trials. But issues of censoring, truncation and inclusion criteria complicate the analysis of clinical trial data. Recent advances in proportional hazard methodologies provide regression diagnostics, improved point and interval estimates of the parameters of survival functions, handling of time dependent covariates in the analysis. This paper discusses the interactions between treatment and patient in the presence of censoring and to account heterogeneity using frailty model. The application of the frailty model with respect to pulmonary tuberculosis data are presented and discussed.

Keywords: Tuberculosis, Cox proportional hazard model, time-dependent covariates, gamma shared frailty model.

Introduction

Survival analysis application implicitly assumes a homogeneous population to be studied; i.e. all individuals are in principle subject under the same risk (Cox, 1972). In most of the clinical trial applications, the study population cannot be assumed to be homogeneous the effect of drug may be individual specific or group specific or each subjects has its own biological response to treatment. The natural course of a disease varies a lot from person to person so does the effect of treatment, or the influence of various risk factors (Aalen, 1988). In clinical trials are drawn conclusions about the average effect of treatment and do not say much about how the effect varies between patients. Therefore, there is a mixture of individuals with different hazards (Manton & Stallard, 1981). It is always not possible to obtain all relevant covariates related to the study on disease of interest, because of economical reasons and important covariates is still unknown. The heterogeneity may be explainable in terms of observed covariates though there will always be an unexplained residual. The heterogeneity is considered basically as unobserved, and is manifesting only indirectly (Aalen, 1988). We intend to account for the dependence in clustered event times in the lifetimes of patients in a clinical trial (Andersen *et al.*, 1999). A usual way to model dependence of clustered event times is through the introduction of a cluster-specific random effect. This random effect gives explanation the dependence in the sense that had we known the frailty, the events would be independent. A frailty is an unobserved random proportionality factor that modifies the hazard function of an individual, or of related individuals. Frailty models are extensions of the proportional hazards model, namely Cox model, the most popular model in survival analysis. The word 'frailty' was introduced by Vaupel *et al.* (1979) for univariate data and

later Clayton (1978) introduced for multivariate data. Generally the frailty in a model is considered to be constant for a particular individual, whereas in the medical context, one expects frailty to increase with age. Also the frailty term is included in the univariate frailty model mainly to describe heterogeneity.

Frailty models

This random effect for the i^{th} cluster v_i is incorporated conditionally into the proportional hazard function as

$$h(t/v_i) = v_i h_0(t) \exp(\beta x_{ij}) \quad \text{----- (1)}$$

which may be re-expressed as

$$h(t/v_i) = h_0(t) \exp(\beta x_{ij} + \eta_i) \quad \text{----- (2)}$$

showing v_i , actually behaves as an unknown covariate for the i^{th} cluster in the model. Using relationships between the survival and hazard function, the conditional survival function as

$$S(t/v_i) = \text{Exp}[v_i A_0(t) \exp(\beta x_{ij})] \quad \text{----- (3)}$$

and the conditional likelihood as

$$L(\gamma, \beta | v_i) = \prod_{i=1}^I \prod_{j=1}^{n_i} (h(t_{ij} | v_i)^{\delta_{ij}} S(t_{ij} | v_i)) \quad \text{----- (4)}$$

where there are i clusters, i^{th} one being of size n_i and γ and β represent baseline hazard and regression parameters, respectively. Substitution gives:

$$= \prod_{i=1}^I \prod_{j=1}^{n_i} \left(\frac{\phi}{\Phi} \right)^{\delta_{ij}} \prod_{i=1}^I \exp(-\Phi) (\Phi)^{\delta_i} \quad \text{----- (5)}$$

where

$$\phi = v_i \exp(\beta' X_i) \exp(\alpha' W_{ij}) \gamma_j^{r-1}$$

$$\Phi = v_i \exp(\beta' X_i) \sum_{j=1}^{n_i} \exp(\alpha' W_{ij}) \gamma_j^r = v_i \exp(\beta' X_i) e_i$$

$$L(\gamma, \beta | v_i) = \prod_{i=1}^I \prod_{j=1}^{n_i} [h_0(t) v_i \exp(\beta X_{ij})]^{\delta_{ij}} \exp[-v_i \Lambda_0(t) \exp(\beta X_{ij})]$$

$$= \prod_{i=1}^I \prod_{j=1}^{n_i} \left(\frac{\exp(\alpha W_{ij}) \gamma_j^{r-1}}{e_i} \right)^{\delta_{ij}} \prod_{i=1}^I (\exp(\beta X_i) e_i \tau)^{\delta_i} \frac{\Gamma(\delta_i + 1/\tau)}{\Gamma(1/\tau)} \left(\frac{1}{1 + \exp(\beta X_i) e_i \tau} \right)^{\delta_i + 1}$$

The marginal likelihood, $L(\gamma, \beta)$, is obtained through integration of the random effect distribution. A common assumption is that the random effect will follow a Gamma distribution with mean 1 and variance τ , i.e.,

$$f(v_i) = \frac{v_i^{1/\tau - 1} \exp(-v_i / \tau)}{\Gamma(1/\tau) \tau^{1/\tau}} \quad \text{----- (6)}$$

The marginal likelihood is then obtained to be

$$L(\gamma, \beta, \alpha, \tau) = \prod_{i=1}^I \prod_{j=1}^{n_i} \int_0^\infty L(\gamma, \beta, \alpha | v_i) dG(v_i) \quad \text{----- (7)}$$

Shared gamma frailty models

Frailty models have been used when groups of subjects have responses that are likely to be dependent in some general way. When multiple events have been observed on the same subjects Liang *et al.* (1995) discussed the use of frailty models with multivariate failure time data. If the value of the frailty is assumed to be constant within groups, the models are called shared frailty models. Also the model that allows for dependence between related individuals in a group is the shared frailty. Many authors have considered this concept (Clayton & Cuzick, 1985; Hougaard, 1986; Clayton, 1991; Andersen *et al.*, 1997; Parner, 1998; Hougaard, 2000; Vu *et al.*, 2001; Duchateau *et al.*, 2002; Rahgozar *et al.*, 2008). The available software for frailty model (Klein & Moeschberger, 1997) provide SAS macros that fit the Gamma frailty model and as an additional development made a set of programs by Jenkins (1997) for use in Stata that fit a frailty model.

In shared frailty model, the conditional hazard function of T_{ij} given the unobservable frailty random variable v_i of the i^{th} group and fixed observed covariate vector x_{ij} , is given by

$$h_{ij}(t | v_i, x_{ij}) = v_i h_0(t) \exp(x_{ij}^T \beta) \quad i = 1, \dots, n, \quad j = 1, \dots, n_i \quad \text{(8)}$$

where $h_0(t)$ an unknown baseline hazard function common to every subject and β is the vector of fixed effect parameters. The shared frailty variable v_i is assumed to be independent and identically distributed for groups of patients.

Gamma distribution is used as

$$f v_i(v_i) = \frac{1}{\Gamma \theta} \theta^\theta v_i^{\theta-1} \exp(-\theta v_i), \quad i = 1, \dots, n \quad \text{----- (9)}$$

Higher values of θ^{-1} signify larger variances for v_i , consequently greater heterogeneity among different groups of patients. The role of shared frailty model is more useful when we consider multivariate survival times.

The joint survival function for the k_i individuals within the i^{th} group is

$$S(t_{i1}, \dots, t_{ik_i}) = Pr(T_{i1} > t_{i1}, \dots, T_{ik_i} > t_{ik_i})$$

$$= \int_0^\infty \prod_{j=1}^{k_i} Pr(T_{ij} > t_{ij} | v_i) g(v_i) dv_i$$

$$= \left[1 + \frac{1}{\theta} \sum_{j=1}^{k_i} \Lambda_0(t_{ij}) \exp(\beta^t X_{ij}) \right]^{-\theta} \quad \text{----- (10)}$$

where the $\beta, \theta, \Lambda_0(t)$ are estimates

Cox proportional hazards models

The PH model describes the relationship between the hazard function of the risk of an event and a set of covariates. The Cox PH model is usually written in terms of the hazard model. The Cox (1972) model is

$$h(t, X) = h_0(t) e^{\sum_{i=1}^p \beta_i X_i} \quad \text{----- (11)}$$

where $h_0(t)$ is baseline hazard and β_i is parameter vector and X_i are covariates.

This model gives an expression for the hazard at time t for an individual with a given specification of a set of explanatory variables denoted by X . That is, X represents a collection of predictor variables that is being modeled to predict an individual's hazard. An important feature of this model is that the baseline hazard is a function of t , but does not involve the X 's. The baseline hazard function $h_0(t)$ is the hazard function when all covariates set to

zero and also the baseline hazard function is left unspecified so that the time-to-event random variable is not assumed to follow any particular distribution.

Time dependent covariate models

Time dependent covariates have been studied a number of authors (Crowley & Hu, 1977; Cox & Oakes, 1984; Andersen, 1986; Fisher & Lin 1999). A baseline Cox analysis ignores the change of updated covariate values usually yields smaller effect estimates than a time-dependent analysis using all temporal information available (Aydemir *et al.*, 1999). Also Altman and De Stavola (1994) called this is the time decay of the effects of entry values. One of the earliest applications of the use of time varying covariates in a biomedical setting may be found in Crowley (1977).

Let $x(t)$ denote the value of the covariate x measured at time t . Let $x_i(t_i)$ denote the value of the covariate for subject i at time t_i .

$$x_i(t_i) = [x_{i1}(t_i), x_{i2}(t_i), \dots, x_{ip}(t_i)] \quad \text{----- (12)}$$

The notation in the above equation is completely general in the sense that, if a particular covariate, x_k , is fixed then

$$x_{ik}(t_i) = x_{ik}(t = 0) = x_{ik}$$

and this has lead to use the time-dependent notation in equation(12) exclusively. The generalization of the proportional hazards regression function to include possibly multiple time varying covariates is

$$h(t, x(t), \beta) = h_0(t) \exp[x' \beta(t)] \quad \text{----- (13)}$$

and the generalization of the partial likelihood function

$$l_p(\beta) = \prod_{i=1}^n \left[\frac{e^{x_i \beta}}{\sum_{j \in R(t_i)} e^{x_j \beta}} \right] = \prod_{i=1}^n \left[\frac{e^{x_i'(t_i) \beta}}{\sum_{l \in R(t_i)} e^{x_l'(t_i) \beta}} \right]^{c_i} \quad \text{----- (14)}$$

Application to tuberculosis trial data

The data consists of 1236 tuberculosis patients admitted in randomized controlled clinical trial into three treatments including a control regimen, of 6 months duration each conducted at Tuberculosis Research

Centre (ICMR, 1994). The event of interest is time until the sputum conversion during treatment period. The number of events observed was 1061(85.8%). The covariates considered are: Age (in years), Sex (Male-1 and Female-0), Treatment group (Treatment-A, Treatment-B & Control), Weight at baseline (in Kg), Pre treatment sensitivity (Drug Susceptibility Pattern) (Any Drug Resistance-1 and All Drug Sensitivity -0) and Sputum Culture (Low Grade Positive-0 and High Grade Positive-1). Event is coded as 1 and censoring coded as 0.

Age was clustered into seven groups (≤ 20 yrs-1, 21 to 25 yrs-2, 26 to 30 yrs-3, 31 to 35 yrs-4, 36 to 40 yrs-5, 41 to 45 yrs-6 and $\geq 46=7$) for the applications of frailty model. The data base is used to compare empirically the survival difference between two groups of treatments under various model assumptions.

We consider the Cox model with six covariates. Table 1 gives the regression coefficients using the proportional hazards model. The data involves three group of patients treated with anti-tuberculosis regimens. The data also contains sputum culture grade which is a known prognostic indicator of response for tuberculosis patients. The basic question of interest here is comparison of the response experience of the three treatment groups adjusting for possible confounding and interaction effects of the other covariates for modeling purpose. We include only covariates which are depending on time: weight and age.

The Table 1 and 2 gives Cox PH model along with time dependent covariate model with different covariates according to their hierarchical order.

Table 1. Comparisons between Cox PH model and time dependent covariate model

Cox PH Models					
Variable	β	SE(β)	P	HR	95% CI
Treatment	0.176	0.083	0.035	1.192	1.01 - 1.40
Sp.Culture	0.007	0.113	NS	1.007	0.81 - 1.26
Sex	0.257	0.078	0.001	1.293	1.11 - 1.50
Age	-0.003	0.003	NS	0.997	0.99 - 1.00
Weight	0.011	0.005	0.022	1.011	1.00 - 1.02
Drug Supc	0.638	0.090	0.000	1.893	1.59 - 2.26
Treatmnt * Sp.Culture	-0.101	0.137	NS	0.904	0.69 - 1.18

Table 2. Time dependent covariate model (age)

T_Cov_	0.022	0.006	0.000	1.023	1.01 - 1.04
Treatment	0.183	0.083	0.035	1.201	1.02 - 1.41
Sp. Culture	0.026	0.113	NS	1.026	0.82 - 1.28
Sex	0.251	0.076	0.001	1.286	1.11 - 1.49
Weight	0.011	0.005	0.021	1.011	1.00 - 1.02
Drug Supc	0.641	0.090	0.000	1.899	1.59 - 2.26
Treatment * Sp.Culture	-0.123	0.137	NS	0.885	0.67 - 1.16

The standard regimen showed better response rates than the new split regimen. Age and weight as time dependent variable are considered in the extended Cox model. The results are based on age as a time dependent covariates, after adjusting all the covariates. We observed that the result of time dependent covariate is better than the results of Cox PH Model. The application of frailty model to the same data assuming individual heterogeneity is also considered.

The survival curves are presented (Fig.1). The adjusted survival curves differ significantly from unadjusted survival curve.

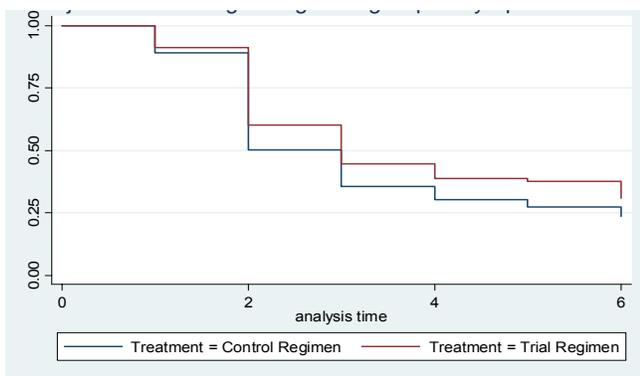


Fig. 1: Survival curve for treatment (adjusted for covariates)

From Table 3, we see that all the covariates are statistically significant except sputum culture. This leads to the conclusion that, the individual heterogeneity in sputum culture conversion differs significantly between the patients.

The Gamma shared frailty model is random effects models where the frailties are common or shared among groups of individuals based on age and are randomly distributed across groups.

Table 3. Frailty models (gamma)

	HR	SE	Z	P	95% CI
Treatment1	1.344	0.205	1.94	0.052	0.997 - 1.812
Sp. Culture	1.319	0.204	1.79	0.074	0.973 - 1.787
Sex	0.421	0.068	-5.34	0.000	0.306 - 0.579
Weight	1.025	0.011	2.41	0.016	1.005 - 1.047
Drug Suscept	2.764	0.478	5.88	0.000	1.969 - 3.880

$-2LL = 2004.4$

Table 4. Gamma shared frailty

	H R	SE	Z	P	95% CI
Treatment1	1.297	0.098	3.43	0.001	1.118 - 1.504
Sp. Culture	1.205	0.092	2.45	0.014	1.038 - 1.340
Sex	0.721	0.052	-4.54	0.000	0.625 - 0.830
Weight	1.016	0.005	3.48	0.001	1.007 - 1.025
DrugSuscept	2.738	0.248	11.1	0.000	2.293 - 3.269

$-2LL = 2305.5$

The age factor is clustered into different groups, the effects between two treatments on age are showing differences among each other and all other covariates influence the outcome factor (Table 4). The shared gamma frailty model is able to account the heterogeneity and shows that the shared gamma frailty model is better than the gamma frailty model.

Discussion

The term v varies from individual to individual and is not observable. Hence the distribution of v of the population $G(v)$ must be specified. Since the hazard function is non-negative, v must be restricted to non-negative values. The univariate frailty model assumes the random effect varies between individuals. The multivariate frailty model assumes unexplained heterogeneity shared by related individuals and frailty as common to several individuals. In this paper we focused only on univariate case. In most of the cases, a frailty model can only imply a positive correlation within group. On the other hand, there are some situations in which the association is negative. Secondly, the model constraints the observed factors to be the same within a group of clustered observations implying constant correlation between all individuals in cluster, and also to be the same during follow-up. This is unsatisfactory in many situations, because not always reflecting the reality. As a final point the dependence parameter and the population heterogeneity are determined at the same time, and can be confounded. This can lead to complexity in the interpretation when these conditions not met out in the medical context. The main objective is to find a surrogate marker for frailty to select the frail patients. Additionally, with clusters containing different subjects, the situation is quite different. In the shared frailty model, we assume that persons in the same cluster share the same frailty term and we have to predict the value of the frailty term for that cluster.

References

1. Aalen (1988) Heterogeneity in survival analysis. *Stat. in Medicine.* 7, 1121-1137.
2. Altman and De Stavola B (1994) Practical problems in fitting a proportional hazards model to data with updated measurements of covariates. *Stat. Med.* 13, 301-341.
3. Andersen PK (1986) Time-dependent covariates and Markov processes. In: Modern statistical methods in chronic disease epidemiology. Moolgavkar SH & Prentice RL (eds), Wiley, NY. pp: 82-103.
4. Andersen PK, Klein JP and Zhang MJ (1999) Testing for centre effects in multi-centre survival studies: a

- Monte Carlo comparison of fixed and random effects tests. *Stat. Med.* 18, 1489- 1500.
5. Andersen PK, Klein JP, Knudsen KM and Palacios RT (1997) Estimation of variance in Cox's regression model with shared gamma frailties. *Biometrics.* 53, 1475-1484.
 6. Aydemir O, Aydemir S and Dirschedl P (1999) Analysis of time-dependent covariates in failure time data. *Stat. Med.* 18, 2123-2134.
 7. Clayton D (1978) A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika.* 65,141-151.
 8. Clayton D (1991) A Monte Carlo method for Bayesian inference in frailty models. *Biometrics.* 47, 467-485.
 9. Clayton D and Cuzick J (1985) Multivariate generalization of the proportional hazards model (with Discussion). *J. Royal Stat. Soc.* 148, 82-117.
 10. Cox DR (1972) Regression model and life tables (with discussion). *J. Royal Stat. Soc.(B)*, 34, 187-220.
 11. Cox DR (1975) Partial likelihood. *Biometrika.* 62, 269-276.
 12. Cox DR and Oakes D (1984) Analysis of survival data. London Chapman & Hall.
 13. Crowley J and Hu M (1977) Covariate analysis of heart transplant survival data. *J. Am. Stat. Ass.* 78, 27-36.
 14. Duchateau L, Janssen P, Lindsey P, Legrand C, Nguti R and Sylvester R (2002) The shared frailty model and the power for heterogeneity tests in multicenter trials. *Comp. Stat. Data Analysis.* 40, 603-620.
 15. Fisher LD and Lin DY (1999) Time-dependent covariates in the cox proportional-hazards regression model. *Ann. Rev. Pub. Health.* 20,145-57.
 16. Hougaard P (1986) A class of multivariate failure time distributions. *Biometrika.* 73, 671-678.
 17. Hougaard P (2000) Analysis of multivariate survival data. Springer: NY. pp: 312-381.
 18. Jenkins SP (1997) Discrete time proportional hazard regression. *STATA Tech. Bull.* 39, 17-32.
 19. Klein JP and Moeschberger ML (1997) Survival Analysis Techniques for Censored and truncated data, Springer-Verlag: NY.
 20. Klein JP (1992) Semi-parametric estimation of random effects using the Cox model based on the EM algorithm. *Biometrics.* 48, 795-806.
 21. Liang KY, Self SG, Bandeen-Roche KJ and Zeger SL (1995) Some recent developments for regression analysis of multivariate failure time data. *Lifetime Data Anal.* 1, 403-415.
 22. Manton K and Stallard E (1981) Methods for evaluating the heterogeneity of aging processes in human populations using vital statistics data: explaining the black/white mortality crossover by a model of mortality selection. *Human Bio.* 53, 47-67.
 23. Murphy SA (1992) Consistency in a proportional hazards model incorporating a random effect. *Annals Stat.* 22, 712-731.
 24. Parner E (1998) Asymptotic theory for the correlated gamma frailty model. *Ann. Stat.* 26, 183-214.
 25. Rahgozar M, Faghihzadeh S, Rouchi GB and Peng Y (2008) The power of testing a semi-parametric shared gamma frailty parameter in failure time data. *Stat. Med.* 27, 4328-4339.
 26. Tuberculosis Research Centre, ICMR, Chennai, India (2004) Split-drug regimens for the treatment of patients with sputum smear-positive pulmonary tuberculosis- a unique approach. *Trop. Med. Int. Health.* 9, 551-58.
 27. Vaupel JW, Manton K and Stallard E (1979) The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography.* 16, 439-454.
 28. Vu H, Segal MR, Knuiman MW and James IR (2001) Asymptotic and small sample statistical properties of random frailty variance estimates for shared gamma frailty models. *Comm. Stat-Simulation Comp.* 30, 581-595.