

Proportional Hazards Mixed Models: A Review with Applications to Twin Models

Ronghui Xu¹

Abstract

We describe our recent work on mixed effects models for right-censored data. Vaida and Xu (2000) provided a general framework for handling random effects in proportional hazards (PH) regression, in a way similar to the linear, non-linear and generalized linear mixed effects models that allow random effects of arbitrary covariates. This general framework includes the frailty models as a special case. Maximum likelihood estimates of the regression parameters, the variance components and the baseline hazard, and empirical Bayes estimates of the random effects can be obtained via an MCEM algorithm. Variances of the parameter estimates are approximated using Louis' formula.

We show interesting applications of the PH mixed effects model (PHMM) to a US Vietnam Era Twin Registry study on alcohol abuse, with the primary goal of identifying genetic contributions to such events. The twin pairs in the registry consist of monozygotic and dizygotic twins. After model fitting and for interpretation purposes, the proportional hazards formulation is converted to a linear transformation model before the results on genetic contributions are reported. The model also allows examination of gene and covariate interactions, as well as the modelling of multivariate outcomes (comorbidities).

1 Clustered survival data and frailty models

The proportional hazards model, since its introduction in Cox (1972), has been widely used for independent, identically distributed right-censored time to event data. Correlated survival data, however, arise in various fields of applications. These include multicenter clinical trials, familial or genetic data, recurrent events, matched pairs design, etc. An example of correlated survival data from multicenter clinical trials is a lung cancer trial conducted by the Eastern Cooperative Oncology Group (ECOG), to compare two different chemotherapy regimens, where there was evidence that the treatment effect varied significantly across the 31 participating institutions (Gray, 1994, 1995). Figure 1 shows the estimated treatment effects in terms of the

¹ Department of Biostatistics, Harvard School of Public Health and Dana-Farber Cancer Institute, Boston, MA 02115; rxu@jimmy.harvard.edu

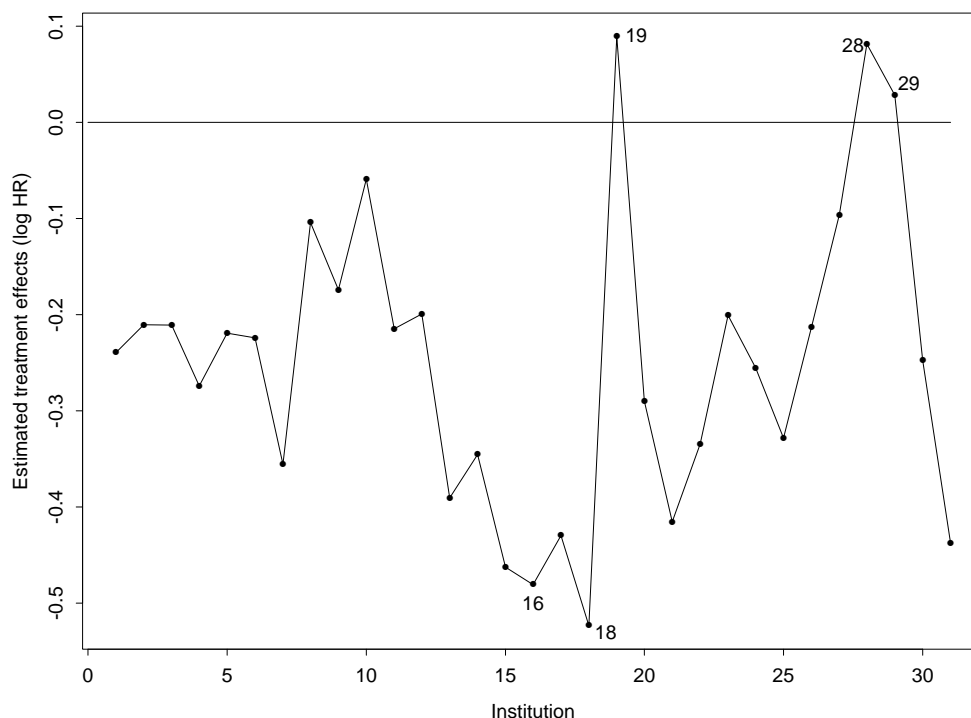


Figure 1: Estimated treatment effects by institution for the ECOG lung cancer data.

log hazard ratio under a proportional hazards model from the 31 institutions (Vaida and Xu, 2000). As seen in the figure, the log hazard ratios in some institutions such as number 16 and 18 are around -0.5, while in some other institutions such as number 19, 28 and 29 the log hazard ratios are almost zero.

Until recently, clustered survival data had been modeled mainly through frailty models, a name coined by Vaupel *et al.* (1972). A univariate frailty model has the form

$$\lambda_{ij}(t) = \lambda_0(t) \exp(\beta' \mathbf{z}_{ij} + b_i), \quad (1.1)$$

where $i = 1, \dots, m$ indicates the cluster, and $j = 1, \dots, n_i$ indicates the individual within the i th cluster, $\lambda_{ij}(t)$ its hazard function, $\lambda_0(t)$ the baseline hazard, and \mathbf{z}_{ij} a vector of covariates. The term b_i is the random effect or, log frailty, for cluster i . For interpretation under model (1.1), note that the random effect acts on the log hazard additively, with no interaction between any covariate and the clustering variable. Therefore it cannot be used to model the varying treatment effect as shown in the lung cancer data above.

Early work under model (1.1) was mainly motivated by unobserved heterogeneity and the fact that, unlike linear models, the Cox regression models are generally not nested, so that leaving out potential covariates from a Cox model can lead to nonproportionality of the hazards. See Xu (1996) for a review and discussion. The cluster sizes for modelling unobserved heterogeneity are equal to one. The early frailty work was then extended to handle multivariate survival data by allowing

cluster sizes greater than one; for reference to some of those see Vaida and Xu (2000) and Hougaard (2000). In the literature the distribution of $\exp(b_i)$ was often assumed to be Gamma for mathematical convenience, since it leads to conjugate priors. Another distribution that was sometimes used was log-normal.

The univariate frailty model was later extended to additive frailty models, of the form

$$\lambda_{ij}(t) = \lambda_0(t)(\mathbf{a}'_i \mathbf{w}_{ij}) \exp(\beta' \mathbf{z}_{ij}). \quad (1.2)$$

Here \mathbf{w} is a design vector of 0's and 1's, and \mathbf{a} is a vector of gamma frailties. Petersen (1998) summarizes the use of model (1.2) for genetic and familial data. Model (1.2), however, is still unable to incorporate random effects for arbitrary covariates, or equivalently, covariate by cluster interactions as exemplified in the lung cancer data. Furthermore, as will be shown in Section 3, the interpretation of the frailties is difficult, especially when one wishes to explain the results from model (1.2) in the context of established literature on genetic epidemiology.

2 Proportional hazards mixed effects model

Vaida and Xu (2000) proposed the proportional hazards model with general random effects

$$\lambda_{ij}(t) = \lambda_0(t) \exp(\beta' \mathbf{z}_{ij} + \mathbf{b}'_i \mathbf{w}_{ij}), \quad (2.1)$$

with the same notation as before, and \mathbf{w}_{ij} is a vector of covariates that have random effects. In contrast with the additive frailty models, in model (2.1) the random effects act on the same scale as the fixed effects, enabling covariate by cluster interactions. We call model (2.1) the proportional hazards mixed effects model, or PHMM. As an example, for the lung cancer data, if a single covariate $z = 0$ or 1 indicates one of the two treatment assignments, the following model can be used to capture cluster-specific treatment effect:

$$\lambda_{ij}(t) = \lambda_0(t) \exp\{b_{0i} + (\beta + b_{1i})z_{ij}\}.$$

The treatment effect from institution i is then $\beta + b_{1i}$.

For the distribution of the random effects in (2.1), it is assumed that

$$\mathbf{b}_i \stackrel{iid}{\sim} N(0, G). \quad (2.2)$$

The advantages of the normal assumption for random effects, as explained in Vaida and Xu (2000), are symmetry and scale-invariance. The normality assumption however, is not crucial and in the development below can be replaced by other parametric distribution for the random effects.

Model (2.1) is in parallel with the linear, non-linear and generalized linear mixed effects models. On the log hazard scale it is written

$$\log \lambda_{ij}(t) = \log \lambda_0(t) + \beta' \mathbf{z}_{ij} + \mathbf{b}'_i \mathbf{w}_{ij}.$$

This describes the conditional (log) hazard, or equivalently, the conditional distribution, of the event time T of an individual as a function of the covariates and the random effects. Mathematically, model (2.1) can be written in the equivalent form of a linear transformation model

$$g(T_{ij}) = -(\beta' \mathbf{z}_{ij} + \mathbf{b}'_i \mathbf{w}_{ij}) + e_{ij}, \quad (2.3)$$

where $g(t) = \log \int_0^t \lambda_0(s) ds$, and e has a fixed extreme value distribution with variance $\sigma^2 = 1.645$. In view of formulation (2.3), the PHMM then decomposes the variation in the transformed event time, into contributions from the fixed covariate effects, the random effects, and a fixed error distribution. This formulation is often useful for interpretation purposes of the PHMM, as will be seen in the applications of the next section.

Inference under model (2.1) is based on nonparametric maximum likelihood. The unknown vector of parameters is $\theta = (\beta, G, \lambda_0)$. The nonparametric likelihood discretizes the continuous function $\lambda_0(\cdot)$ to mass points on the observed event times (Johansen, 1993). The likelihood is then a function of finite dimensional parameters, albeit without a closed form due to the integration with respect to the distribution of the random effects. In Vaida and Xu (2000) the EM algorithm was used to find the maximum likelihood estimator $\hat{\theta}$, where the M-step implementation was similar to the standard Cox model, and the E-step involved Markov Chain Monte Carlo (MCMC) to approximate the conditional distribution of the random effects given data. Simulation studies have since been carried out to evaluate the performance of the estimators and the algorithm, and the results (unpublished) have been good. In particular, unlike some of the other methods and algorithms that have appeared in the literature (private communication), the maximum likelihood estimate obtained through the MCEM algorithm has been stable in all our experience with both simulated and real data sets.

Following the estimation of θ , Louis (1982) formula can be used to estimate the variance of $\hat{\theta}$. The magnitude of the random effects is also sometimes of interest, and can be obtained from the posterior, or “empirical Bayes”, distribution:

$$\hat{\mathbf{b}}_i = E(\mathbf{b}_i | \mathbf{y}_i; \hat{\theta}) \quad (2.4)$$

where \mathbf{y}_i denotes the observed data from cluster i . Note that this posterior distribution is a by-product of the EM algorithm at convergence; for details see Vaida and Xu (2000).

3 Application to twin data

The following contains material from a doctoral thesis written by I-Chao Liu in the Department of Epidemiology, Harvard School of Public Health.

3.1 A twin model

As mentioned in the first section, one of the applications of PHMM is genetic data. In particular, much work has been done in genetic epidemiology to make use of twin

data, as they provide a way to examine the relative contribution of genetic and environmental effects on diseases. Monozygotic (MZ) twins have the exact same genes, and any difference in disease within twin pairs should come from environmental factors. In contrast, dizygotic (DZ) twins only share half of their genes, and differences in disease could be the result of environment and genetics. More specifically, The total variance of twin resemblance on traits is partitioned into three parts: additive genetic (G), common environment (C), and unique environment (E). Additive genetic represents twin similarity arising from additive effects of alleles; common environment includes all non-genetic factors shared within a twin pair and leads to twin similarity; unique environmental factors includes all non-genetic factors unshared within a twin pair and leads to twin dissimilarity.

When the endpoint of a twin study is time to event with possible right-censoring, such as age of onset of a disease, the PHMM can be used. To model the dependence structure of the MZ and DZ twins, we need a vector of six random effects: $\mathbf{b} = (b_1, b_2, \dots, b_6)'$. For an MZ twin pair i , let b_{i1} denote the common genetic G and environmental C effect, and b_{i2} and b_{i3} denote unique environmental E effect for twin 1 and twin 2, respectively, so that

$$\lambda_{i1}(t) = \lambda_0(t) \exp(b_{i1} + b_{i2}), \quad (3.1)$$

$$\lambda_{i2}(t) = \lambda_0(t) \exp(b_{i1} + b_{i3}), \quad (3.2)$$

when no covariates are included. We can write

$$\text{var}(b_{i1}) = \sigma_G^2 + \sigma_C^2, \quad \text{var}(b_{i2}) = \text{var}(b_{i3}) = \sigma_E^2. \quad (3.3)$$

For a DZ twin pair i , let b_{i4} denote the common genetic (1/2) and environmental effect, and b_{i5} and b_{i6} denote unique genetic (1/2) and environmental effect for twin 1 and twin 2, respectively, so that

$$\lambda_{i1}(t) = \lambda_0(t) \exp(b_{i4} + b_{i5}), \quad (3.4)$$

$$\lambda_{i2}(t) = \lambda_0(t) \exp(b_{i4} + b_{i6}). \quad (3.5)$$

We can then write

$$\text{var}(b_{i4}) = \frac{\sigma_G^2}{2} + \sigma_C^2, \quad \text{var}(b_{i5}) = \text{var}(b_{i6}) = \frac{\sigma_G^2}{2} + \sigma_E^2. \quad (3.6)$$

Due to the construction of these random effects, i.e. to decompose the total variance, it is natural to assume that they are independent of each other. Therefore the variance matrix for \mathbf{b} is

$$G = \begin{pmatrix} \sigma_G^2 + \sigma_C^2 & & & & & \\ & \sigma_E^2 & & & & \\ & & \sigma_E^2 & & & \\ & & & \frac{\sigma_G^2}{2} + \sigma_C^2 & & \\ & & & & \frac{\sigma_G^2}{2} + \sigma_E^2 & \\ & & & & & \frac{\sigma_G^2}{2} + \sigma_E^2 \end{pmatrix}. \quad (3.7)$$

Table 1: Estimates from the VET Registry data.

Contribution	Estimates (SE)	Proportion of Total* (SE)
σ_G^2	1.08 (0.15)	36% (5%)
σ_C^2	0.27 (0.05)	9% (2%)
σ_E^2	0.02 (0.01)	55% (4%)

* Total is $\sigma_G^2 + \sigma_C^2 + \sigma_E^2 + 1.645$.

3.2 VET Registry data

The above twin model was applied to the Vietnam Era Twin (VET) Registry data. The VET Registry in the United States was set up to follow the male-male twin pairs born from 1939 to 1957, in which both twins had the experience of serving in the military during the Vietnam era (1965-1975). In 1992, trained interviewers from the Institute for Survey Research at Temple University invited approximately 5,000 twin pairs from the registry to participate in telephone-administrated psychiatric interviews based on Diagnostic interview schedule for DSM-III-R (DIS-III-R) (Robins *et al.*, 1989). After giving verbal informed consent, the twins were asked questions about several aspects of alcohol use, including frequency, amount, age at onset of various symptoms, duration of each symptom, and mood or perception changes associated with alcohol drinking. Of the 10,253 eligible individuals, 8,169 (80%) were successfully interviewed. Zygosity was determined by responses to questions relating to the similarity of physical appearance, along with blood group typing method. The 3,372 complete twin pairs of known zygosity (1,874 MZ pairs and 1,498 DZ pairs) were included in the analysis described below.

By applying standard DIS-III-R computer algorithms, individual symptoms and the diagnoses of alcohol abuse based on DSM-III-R were derived from the responses to questions. Using the age at onset of alcohol abuse as outcome, we fit the above twin model to the VET Registry data, first with no covariates. The estimates are given in Table 1, where in the parentheses are the standard errors. In order to interpret the results as genetic and environmental contributions to the actual age of onset of alcohol abuse, as opposed to the hazard of it, we make use of the transformation model formulation (2.3). Therefore the total variance is $\sigma_G^2 + \sigma_C^2 + (\sigma_E^2 + \sigma^2)$, and the genetic contribution σ_G^2 as a percentage of it is estimated to be 36%, while the common environmental contribution is estimated to be 9%. Such results turned out to be very comparable to what had been previously reported in the literature on genetic contributions to alcohol abuse. In the parentheses of the last column are again the standard errors of the estimated percentages.

3.3 Extended twin data applications

The twin model described above can be easily extended to incorporate covariates. It is also possible to incorporate covariate by gene interactions. In Liu's thesis, we considered the possibly different genetic and common environment contributions to the transition period from regular alcohol use to alcohol dependence, among early

and late regular alcohol users. The early users were defined as those who started regular alcohol use before the age of 40, while the late users after age 40. For such a dichotomized covariate, a vector of 12 random effects were used to model the covariate interaction with the genetic contribution. The variance matrix is similar to (3.7), albeit with two sets of the variance parameters, one for the early users and one for the late users. We found substantially larger genetic contributions among the early users than among the late users, while the common environmental contribution was much larger among the late users. Such findings have potential prevention or intervention implications, since for late users appropriate strategies might be considered to intervene on environmental determinants so as to reduce the risk of developing alcohol dependence.

Finally, the above twin model can be extended to handle multivariate outcomes in twin data, such as the ages of onset of multiple categories of alcohol dependence symptoms, or joint modelling of ages of onset of alcohol and tobacco abuses. In the case of bivariate outcomes, such as alcohol and tobacco abuses, again a vector of 12 random effects is used. But unlike the case with a binary covariate, the variance matrix is no longer diagonal, since we cannot assume the independence between the genetic contributions to alcohol and to tobacco abuses. More specifically, denote $(b_{1A}, b_{2A}, \dots, b_{6A})$ the random effects corresponding to alcohol abuse, and $(b_{1T}, b_{2T}, \dots, b_{6T})$ the random effects corresponding to tobacco abuse. Then the variance matrix G for the vector of random effects $\mathbf{b} = (b_{1A}, b_{1T}, b_{2A}, b_{2T}, \dots, b_{6A}, b_{6T})'$ is block-diagonal:

$$\text{diag} \left[\begin{pmatrix} \sigma_{GA}^2 + \sigma_{CA}^2 & \tau_1 \\ \tau_1 & \sigma_{GT}^2 + \sigma_{CT}^2 \end{pmatrix}, \begin{pmatrix} \sigma_{EA}^2 & \tau_2 \\ \tau_2 & \sigma_{ET}^2 \end{pmatrix}, \begin{pmatrix} \sigma_{EA}^2 & \tau_2 \\ \tau_2 & \sigma_{ET}^2 \end{pmatrix}, \right. \\ \left. \begin{pmatrix} \frac{\sigma_{GA}^2}{2} + \sigma_{CA}^2 & \tau_3 \\ \tau_3 & \frac{\sigma_{GT}^2}{2} + \sigma_{CT}^2 \end{pmatrix}, \begin{pmatrix} \frac{\sigma_{GA}^2}{2} + \sigma_{EA}^2 & \tau_4 \\ \tau_4 & \frac{\sigma_{GT}^2}{2} + \sigma_{ET}^2 \end{pmatrix}, \begin{pmatrix} \frac{\sigma_{GA}^2}{2} + \sigma_{EA}^2 & \tau_4 \\ \tau_4 & \frac{\sigma_{GT}^2}{2} + \sigma_{ET}^2 \end{pmatrix} \right],$$

where τ_1, \dots, τ_4 are the covariances of the pairwise random effects. Work on applying the above model to the VET Registry data is currently in progress.

4 Discussion

Following the lecture at the 2003 Methodology and Statistics conference in Ljubljana, Slovenia, on which this paper is based, there was discussion on the use of random effects PH model versus stratified PH model. O'Quigley and Stare (2002) showed that the stratified PH model is reasonably efficient as compared to a univariate frailty model. Note that the univariate frailty model (1.1) is nested in a stratified PH model, but the general PHMM is not. While the univariate frailty model is sometimes used to describe 'unobserved heterogeneity' as mentioned in Section 1, the purpose of the PHMM is to capture mixed, or cluster-specific, covariate effects. For example, the cluster-specific treatment effects as shown in figure ?? cannot be estimated from a stratified PH model. Other types of motivation for using the PHMM, as illustrated in Section 3, arise from situations where the variance components are of primary interest. Therefore the use of PHMM should come

from practical applications, when a stratified PH model cannot answer the questions under investigation.

The software for all the applications given above are stand alone programs written in C, and are available from the author.

Acknowledgment

The author would like to thank the 2003 Methodology and Statistics conference organizers for the opportunity to review the work on PHMM. The author would also like to thank Dr. Florin Vaida and Dr. I-Chao Liu for the joint work on which this review was based.

References

- [1] Cox, D.R. (1972): Regression models and life tables (with discussion). *Journal of the Royal Statistical Society, Series B*, **34**, 187-220.
- [2] Gray, R. (1995): A Bayesian analysis of institutional effects in a multicenter cancer clinical trial. *Biometrics*, **50**, 244-253.
- [3] Gray, R. (1995): Tests for variation over groups in survival data. *Journal of the American Statistical Association*, **90**, 198-203.
- [4] Hougaard, P. (2000): *Analysis of Multivariate Survival Data*. Springer-Verlag.
- [5] Johansen, S. (1993): An extension of Cox's regression model. *International Statistical Review*, **51**, 258-262.
- [6] Louis, T. (1982): Finding the observed information matrix with using the EM algorithm. *Journal of the Royal Statistical Society, Series B*, **44**, 190-200.
- [7] O'Quigley, J. and Stare, J. (2002): Proportional hazards models with frailties and random effects. *Statistics in Medicine*, **21**, 3219-3233.
- [8] Peterson, J. (1998): An additive frailty model for correlated life times. *Biometrics*, **54**, 646-661.
- [9] Robins, L., Helzer, J., Cottler, L., and Goldring, E. (1989): *Diagnostic Interview Schedule Version III Revised*. Rockville, MD: Nat. Inst. of Mental Health.
- [10] Vaida, F. and Xu, R. (2000): Proportional hazards model with random effects. *Statistics in Medicine*, **19**, 3309-3324.
- [11] Vaupel, J.W., Manton, K.G., and Stallard, E. (1979): The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography*, **16**, 439-454.
- [12] Xu, R. (1996): *Inference for the proportional hazards model*. Ph.D thesis of University of California, San Diego.