APPLIED BAYESIAN STATISTICAL ANALYSIS
(Proceeding of NSABSA 08, 23-24 May, 2008, 132-138)

Bayesian Separate and joint modeling for Controlled Clinical Trial data using BUGS

Venkatesan, P. and Ponnuraja, C.

Department of Statistics, Tuberculosis Research centre (ICMR),
Chennai-600 031, India.

ABSTRACT

Many clinical trials and other medical studies generate both longitudinal (repeated measurements) and survival (time to event) data. The existing methods are inappropriate when the longitudinal variable is correlated. Earlier articles proposed a joint model for longitudinal and survival data, obtaining maximum likelihood estimates via the EM algorithm based on Bayesian approach implementing via Markov Chain Monte Carlo (MCMC) methods. The longitudinal and survival responses are assumed independent given a linking latent bivariate Gaussian process and available covariates. We use the approach to jointly model the longitudinal and survival data from a clinical trial comparing treatments and also its interactions. The joint Bayesian approach appears to offer significantly improved and enhanced estimation of survival times and other parameters of interest like gender, age and weight. In spite of the complexity the model, we find it to be relatively straight forward to implement and understand using the WinBUGS software.

KEY WORDS: Bayesian Approach; Joint model; Markov chain Monte Carlo (MCMC); WinBUGS.

1. Introduction

Many clinical trials and other medical studies generate both longitudinal (repeated measurements) and survival (time to event) data. In HIV clinical trials, one measures the number of CD4 cells per ml^3 of blood (longitudinal) and time until death or disease progression (survival). The above two are obviously correlated (low CD4 is prognostic of poor survival outcome). Henderson et al. (2000) proposed an expectation-maximization (EM) algorithm to fit the joint model and also connect the longitudinal and survival processes with bivariate random effects following a latent bivariate Gaussian process, Wang and Taylor (2001) include the longitudinal marker as a time-dependent covariate in the (proportional hazards) survival model. Lin et al. (2002) employ a latent class model of logistic model for each subject's class membership of longitudinal and survival processes are independent given this membership though marginally dependent. Wulfsohn and Tsiatis (1997) propose a joint likelihood model. They assume a proportional hazards model for survival conditional on the longitudinal marker. Guo and Carlin (2004) focused the fully Bayesian version of the joint modeling approach implemented via MCMC methods via the WinBUGS (Spiegelhalter, et al. 2003) package.

2. Methods

The parameters r_1, r_2 and r_3 in the survival model measure the association between the two sub-models induced by the random intercepts, slopes and fitted longitudinal value at the event time respectively. Let Y_{ij} denote the j^{th} CD4 count measurement on the i^{th} patient in the trial, $j = 1,2,...n_i$ and $i = 1,2,...m$. We include four explanatory variables as main effect in our analysis: Treatment, Gender, Age and Weight. Our main goal is to analyze the association among Cd4 count, survival time, treatment group, gender, age, weight at baseline accounting for all relevant correlations and subject-specific random effects. We selected very vague prior distributions and we use proper priors, but with hyperparameter value chosen so that the prior will have minimal impact relative to the data. In the longitudinal sub-model we take multivariate normal and inverse gamma priors for the main effects vector $\beta_1 = (\beta_{11}, \beta_{12}, \beta_{13}, \beta_{14}, \beta_{15}, \beta_{16})$ and the error variance $\sigma_\varepsilon 2$, similarly vague normal and inverse gamma priors for $\beta_2 = (\beta_{21}, \beta_{22}, \beta_{23}, \beta_{24}, \beta_{25})$ and $\sigma_3 2$ in the sub model. The parameters common for both models, we take a inverse Wishart(Carlin and Louis, 2000), which is again vague but does provide at least some shrinkage of random effects towards 0, ensuring good identifiability of the main effects(Carlin and Louis, 2000). The association parameters we choose normal prior for $\gamma_1, \gamma_2, \gamma_3$ those are quite vague relative to these parameters' likely posterior magnitude. Our priors are chosen so that our

APPLIED BAYESIAN STATISTICAL ANALYSIS
(Proceeding of NSABSA 08, 23-24 May, 2008, 132-138)

Bayesian analysis reproduces a corresponding likelihood analysis, but where likelihoods are re standardized and interpreted as probability distribution on the parameter.

It appears that a patient's survival is related to two characteristics driving the patient's longitudinal data pattern, namely the initial CD4 level and the rate of CD4 decrease. This is clinically reasonable, since high CD4 count represents better health status; patients with CD4 counts that are low or in more rapid decline would be expected to have poorer survival.

3. Application to Tuberculosis Data

The data used in this application consists of 229 cases of tuberculosis with HIV infected patients, admitted in randomized controlled clinical trial into two anti tuberculosis treatments; patient was assigned randomly to receive either of six months treatment or of nine months treatment according to their sputum grade as well as cd4 counts classification. The event of interest is sputum smear conversion (positive into negative) during treatment period. There are covariates considered here, CD4 counts were measured at study entry, and again at 6^{th} and 12^{th} monthly visits.

1. Age (in years)
2. Sex (i) Male -1
   (ii) Female -0
3. Treatment group (i) TreatmentA-1
   (ii) Treatment - 2
4. Weight at baseline (in Kg)
5. CD4 counts at three time points(Baseline, 6^{th} month and 12^{th} month)

Time and status are also involved here. Event is coded as1 and censoring is coded as 0.

Our Bayesian results for both separate and joint models are computed entirely in WinBUGS. The fully Bayesian version of joint modeling approach implemented based on the sampling chains of 5,000 iterations each after 5000 burning. We followed the optimum of 35,000 iteration and the burning speed of 119 seconds approximately to all stages of every 5,000 iteration. By default, WinBUGS assumes the parameter vector contains both fixed and random effects and provides the components for the two sub-models (the terms in the log-likelihood arising from longitudinal and survival model components) to evaluate their relative contributions.

Table 1 Separate Analysis and Joint analysis - WinBUGS 10000^{th} iterations

	Separate Analysis		**Joint Analysis**	
				Longitudinal sub-models
Parameter	Posterior Mean	95% CI	Posterior Mean	95% CI
Intercept (β_{11})	7.943	3.44 - 2.56	7.929	3.40-12.53
Time (β_{12})	-0.128	-0.35 - 0.09	-0.139	-0.35- 0.06
Time*Treat(β_{13})	-0.002	-0.29 - 0.01	0.010	-0.27- 0.3
Gender (β_{14})	-0.283	-1.24 - 0.69	-0.283	-1.21- 0.67
Age (β_{15})	0.017	-0.07 - 0.10	0.017	-0.07-0.12
Weight (β_{16})	0.120	0.03 - 0.21	0.121	0.03- 0.21
				Survival sub-models
Parameter	Posterior Mean	95% CI	Posterior Mean	95% CI
Intercept (β_{21})	-0.673	-1.62 - 0.27	-0.663	-1.66-0.26
Treatment (β_{22})	-0.033	-0.34 - 0.27	-0.035	-0.35-0.28
Gender (β_{23})	0.008	-0.21 - 0.23	0.005	-0.21-0.23
Age (β_{24})	-0.011	-0.03 - 0.01	-0.01	-0.03-0.01
Weight (β_{25})	0.002	-0.02 - 0.02	0.002	-0.02-0.02
r_1	0.667	0.14 - 1.17	0.743	0.32-1.3
r_2	1.411	0.31 - 2.42	1.562	0.68-2.67
r_3	-0.673	-1.17 - (-1.53)	-0.750	-1.3-(-33)
touz	0.089	0.07 - 0.12	0.089	0.07-0.12

APPLIED BAYESIAN STATISTICAL ANALYSIS
(Proceeding of NSABSA 08, 23-24 May, 2008, 132-138)

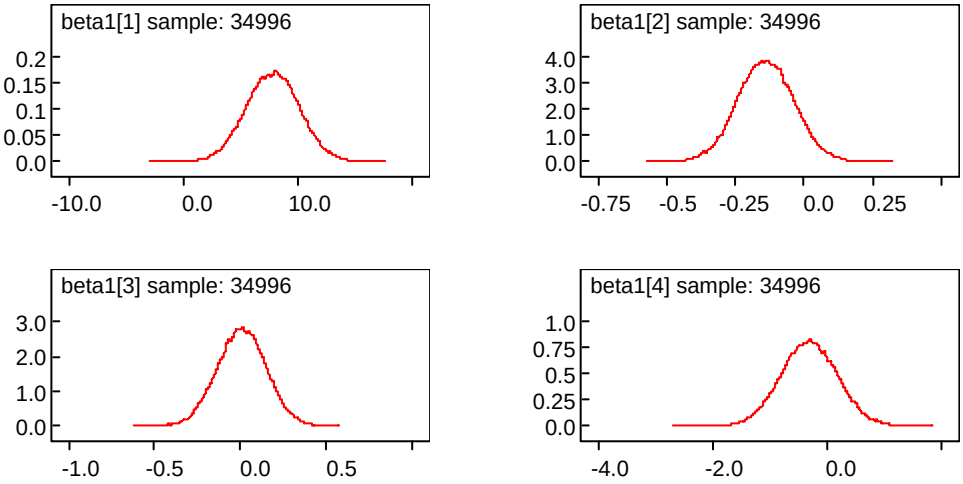
Table 2: Separate Analysis and Joint analysis - WinBUGS 50000th iterations

	Separate Analysis		Joint Analysis	
			Longitudinal sub-models	
Parameter	Posterior Mean	95% CI	Posterior Mean	95% CI
Intercept (β_{11})	7.691	3.08-12.26	7.852	3.36-12.41
Time (β_{12})	-0.054	-0.33- 0.31	-0.135	-0.34-0.07
Time*Treat(β_{13})	0.030	-0.26- 0.33	0.005	0.28-0.29
Gender (β_{14})	-0.313	-1.28- 0.67	-0.296	-1.27-0.68
Age (β_{15})	0.020	-0.07- 0.11	0.019	-0.07-0.11
Weight (β_{16})	0.122	0.03- 0.21	0.121	0.03-0.21
			Survival sub-models	
Intercept (β_{21})	-0.717	-1.77- 0.29	-0.666	-1.70-0.40
Treatment (β_{22})	-0.039	-0.36- 0.28	-0.035	-0.35-0.28
Gender (β_{23})	0.005	-0.21- 0.23	0.008	-0.21-0.23
Age (β_{24})	-0.010	-0.03- 0.01	-0.011	-0.03-0.01
Weight (β_{25})	0.002	-0.02- 0.02	0.002	-0.12-0.02
r1	0.715	-0.79- 1.26	0.700	0.25-1.33
r2	1.499	-1.69- 2.56	1.470	0.50-2.72
r3	-0.353	-1.26- 0.79	-0.703	-1.33-(-0.25)
touz	0.089	0.07-0.12	0.089	0.07-0.12

The posterior estimates of the regression coefficients β_1 and β_2 and their 95%confidence intervals are summarized in Table 1 and 2 on different iterations. Here the results from the separate and joint analyses are quite similar to each other. However, the posterior estimates of the association parameters in the joint analysis are significant and negative association from zero. Bayesian approach to longitudinal and survival responses of the joint model results differ much more noticeably from the separate model results also significantly increasing the survival time in each. Moreover, the joint model in reality reverses the separate models' findings, in the sense that the patient with good CD4 direction is now predicted to survive much longer than the patient the poor direction.

The joint models considered here in fit using the WinBUGS, thus avoiding the need for complex EM programming and facilitating the models' use in practice. The Weibull model with the shape parameter r following a gamma prior distribution for avoiding complexity by simply setting r=1. The estimation of the random effects via empirical Bayes, with associated standard errors was obtained by the delta method. Approximate 95% prediction intervals can be obtained by assuming asymptotic normality. The asymmetry of the posteriors suggests traditional confidence intervals based on asymptotic normality and approximate standard errors will not be very accurate.

Figure 1: Densities



APPLIED BAYESIAN STATISTICAL ANALYSIS
(Proceeding of NSABSA 08, 23-24 May, 2008, 132-138)

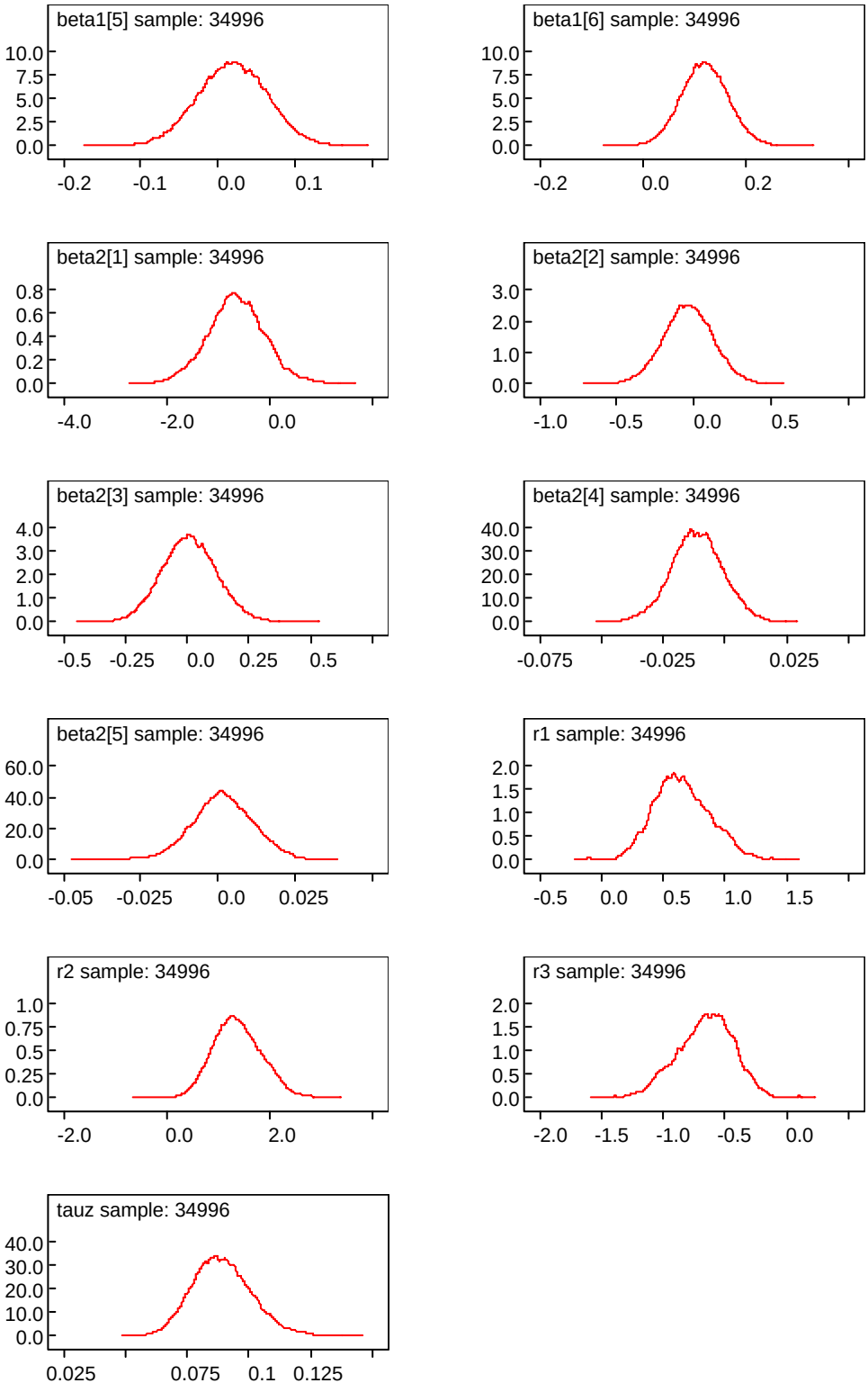
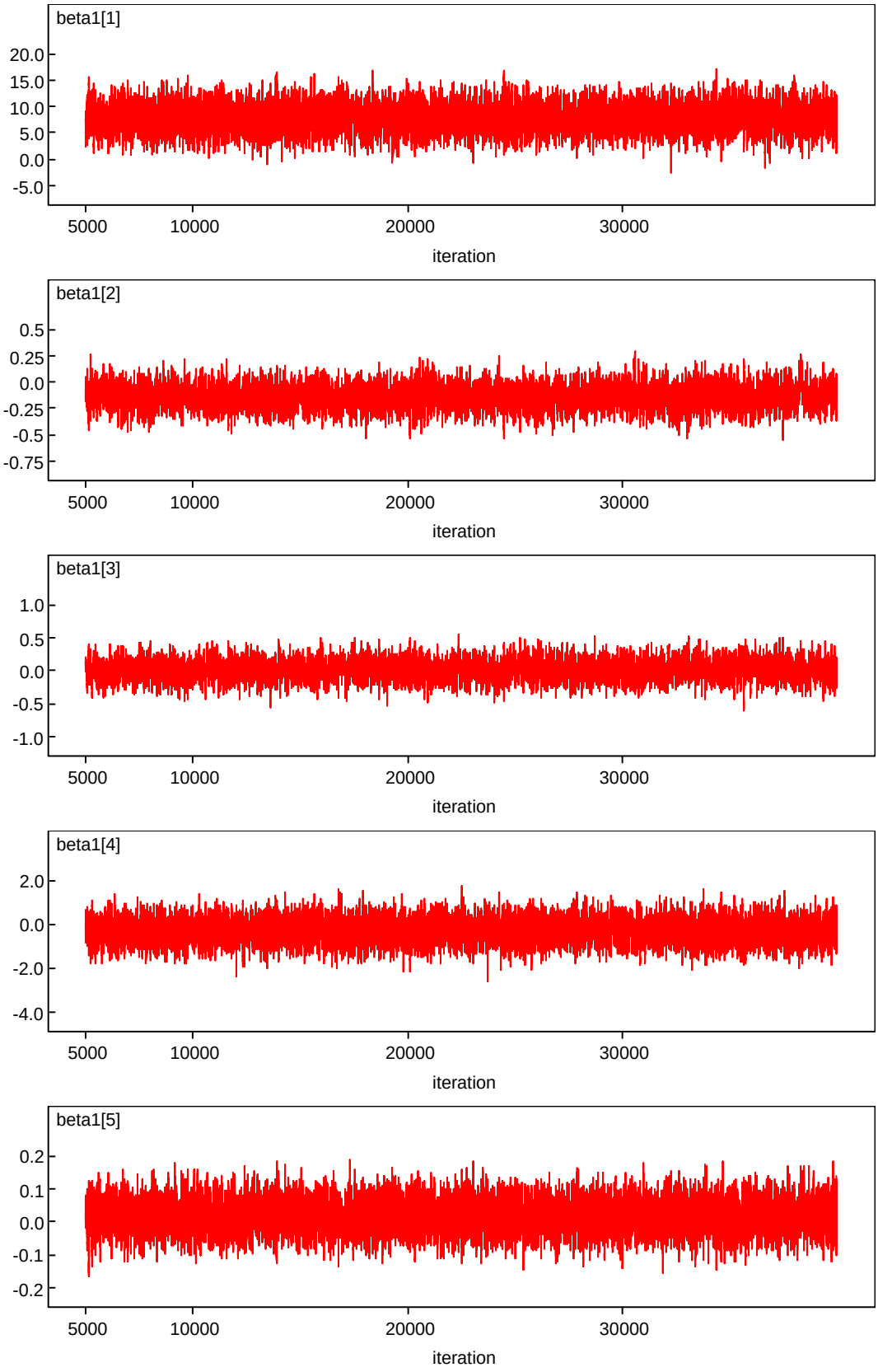
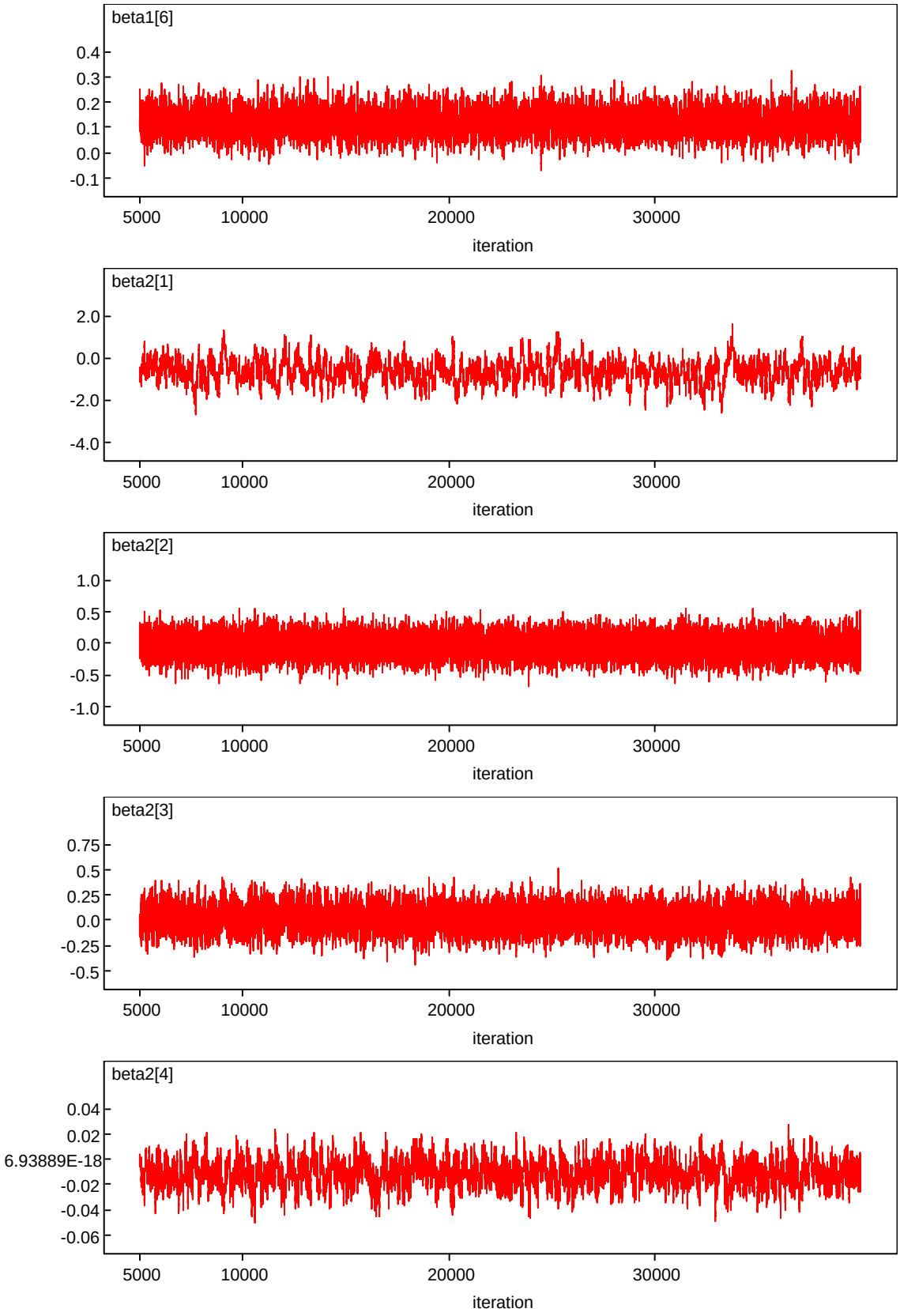


Figure 2: Time Series

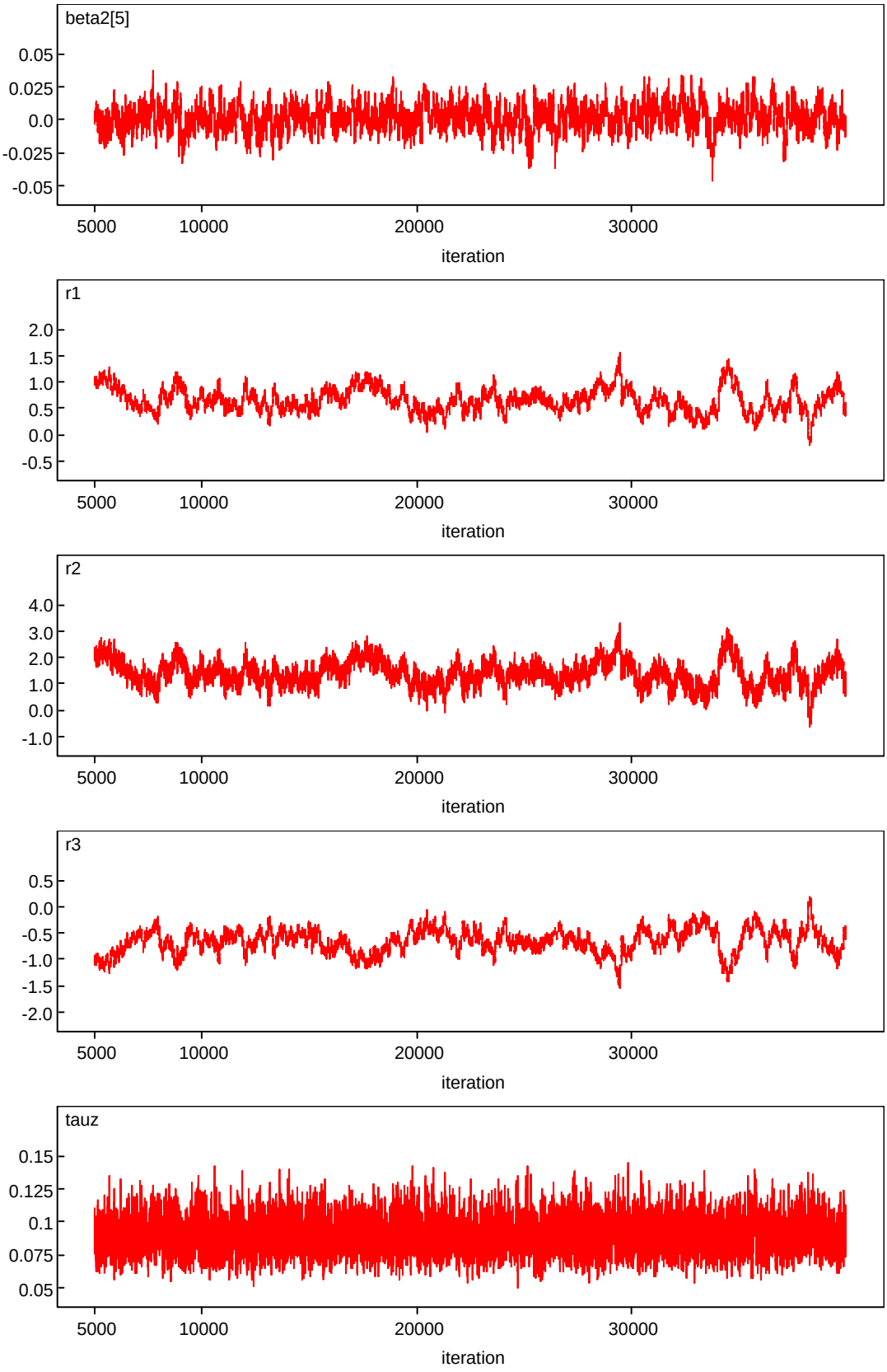
APPLIED BAYESIAN STATISTICAL ANALYSIS
(Proceeding of NSABSA 08, 23-24 May, 2008, 132-138)



APPLIED BAYESIAN STATISTICAL ANALYSIS
(Proceeding of NSABSA 08, 23-24 May, 2008, 132-138)



APPLIED BAYESIAN STATISTICAL ANALYSIS
(Proceeding of NSABSA 08, 23-24 May, 2008, 132-138)



APPLIED BAYESIAN STATISTICAL ANALYSIS
(Proceeding of NSABSA 08, 23-24 May, 2008, 132-138)

4. Summary

Bayesian approach seems both simpler *and* easier here. The Bayesian approach using MCMC for Separate and of Joint models using longitudinal and survival data is illustrated here. The joint models considered here in fit using the WinBUGS, thus avoiding the need for complex EM programming and facilitating the models' use in practice. The Weibull model with the shape parameter r following a gamma prior distribution for avoiding complexity by simply setting r=1. The estimation of the random effects via empirical Bayes, with associated standard errors was obtained by the delta method. Approximate 95% prediction intervals can be obtained by assuming asymptotic normality. The asymmetry of the posteriors suggests traditional confidence intervals based on asymptotic normality and approximate standard errors will not be very accurate.

However the joint analysis increases the estimated mean survival times moderately in both longitudinal sub-model and survival sub-model. This is due to the model's accounting for the correlations between the longitudinal and survival data and the hypothetical patient has covariate value normally associated with a good prognosis, this is reflected in the dramatically improved predicted survival times. Moreover the findings of the patients good CD4 trajectory is now predicted to survive much longer than the patients with bad trajectory based on of joint model but it is actually reverse the findings of separate models. However, the asymmetry of the posteriors which are similar to the likelihood, due to vague priors, suggests traditional confidence intervals based on asymptotic normality and approximate standard errors will not be very accurate.

References

1. Carlin, B. P., and Louis, T. A. (2000). Bayes and Empirical Bayes Methods for Data Analysis (2nd ed.), Boca Raton, FL: Chapman and Hall/CRC Press.

2. Guo, X. and Carlin, B.P. (2004). Separate and joint modelling of longitudinal and event time data using standard computer packages. *The American Staistician*, 58, 16-24.

3. Henderson, R., Diggle, P., and Dobson, A. (2000). Joint modelling of longitudinal measurements and event time data. *Biostatistics*, 1, 465-80.

4. Lin, H., Turnbull, B.W., McCulloch, C. E., and Slate, E. H. (2002). Latent Class Models for Joint Analysis of Longitudinal Biomarker and Event Process Data: Application to Longitudinal Prostate-Specific Antigen Readings and Prostate Cancer, *Journal of the American Statistical Association*, 97, 53–65.

5. Spiegelhalter, D., Thomas, A., Best, N., and Lunn, D. (2003). WinBUGS version14 user manual. *Technical Report.*

6. Wang, Y and Taylor, (2001). Jointly modeling longitudinal and event time data with application to acquired immunodeficiency syndrome. *Journal of the American Statistical Association*, 96, 895-905.

7. Wulfsohn, M. and Tsiatis, A. (1997). A joint model for survival and longitudinal data measured with error. *Biometrics*, 53, 330-39.