Statistica Sinica **14**(2004), 863-883

BAYESIAN METHODS FOR JOINT MODELING OF LONGITUDINAL AND SURVIVAL DATA WITH APPLICATIONS TO CANCER VACCINE TRIALS

Joseph G. Ibrahim, Ming-Hui Chen and Debajyoti Sinha

University of North Carolina, University of Connecticut and Medical University of South Carolina

Abstract: Vaccines have received a great deal of attention recently as potential therapies in cancer clinical trials. One reason for this is that they are much less toxic than chemotherapies and potentially less expensive. However, little is currently known about the biologic activity of vaccines and whether they are associated with clinical outcome. The antibody immune measures IgG and IgM have been proposed as potential useful measures in melanoma clinical trials because of their observed association with clinical outcome in pilot studies. To better understand the role of the IgG and IgM antibodies for a particular vaccine, we examine a case study in melanoma and investigate the association between clinical outcome and an individual's antibody (IgG and IgM titers) history over time. The Cox proportional hazards model is used to study the relationship between the antibody titers as a time varying covariate and survival. We develop a Bayesian joint model for multivariate longitudinal and survival data and give its biologic motivation. Various scientific features of the model are discussed and interpreted. In addition, we present a model assessment tool called the *multivariate L measure* that allows us to formally compare different models. A detailed analysis of a recent phase II melanoma vaccine clinical trial conducted by the Eastern Cooperative Oncology Group is presented.

Key words and phrases: Antibody IgG titers, antibody IgM titers, cancer, longitudinal data, melanoma, proportional hazards, random effects, survival model.

1. Introduction

Melanoma incidence is increasing at a rate that exceeds all solid tumors. Although education efforts have resulted in earlier detection of melanoma, patients who have deep primary melanoma ($> 4mm$) or melanoma metastatic to regional draining lymph nodes, classified as *high-risk melanoma* patients, continue to have high relapse and mortality rates of 60% to 75% (see Kirkwood, Ibrahim, Sondak, Richards, Flaherty, Ernstoff, Smith, Rao, Steele and Blum (2000)). Recently, several post-operative (adjuvant) chemotherapies have been proposed for this class of melanoma patients, and the one which seems to provide the most significant impact on relapse-free survival (RFS) and overall survival (OS) is Interferon

Alpha-2b (IFN). One of the major drawbacks of IFN, and chemotherapies in general, is that they are highly toxic. As a result, there has been a recent surge in research activity for finding effective vaccines for treating malignant melanoma. These vaccines are not nearly as toxic and have much milder side-effects, and could prove to be as efficacious as other chemotherapies.

The purpose of a vaccine therapy is that, after administration of a vaccine, more antibodies would be induced by the immune system and that would help eradicate the tumor cells. The primary measures of antibody response are the IgG and IgM antibody titers. These are continuous serology measurements which assess the degree of antibody activity. In vaccine clinical trials, it is then of importance to assess the amount of IgG and IgM antibodies induced after several vaccinations are given at various time points. It is of interest to assess the association of the antibody measurements with primary outcome measures, such as RFS. From several pilot studies conducted, it has been conjectured that the amount of antibody titers produced from certain vaccines for cancer are associated with clinical outcome. However, this hypothesis has never been formally assessed in any statistical or mathematical model. In this paper, we formally address this issue in the context of melanoma.

A vaccine that has recently been developed for treating melanoma is called GM2-KLH/QS-21, which we abbreviate here by GMK. GM2 is a ganglioside expressed by most melanoma cells. From previous pilot studies, patients with pre-existing natural antibody against GM2 or patients who developed antibodies against GM2 as a result of immunization, demonstrated an improved RFS compared to patients without GM2 antibodies. From these studies, GMK appears to enhance antibody response to GM2. However, the level of antibody response to GM2 induced by GMK is not directly observable. The IgG and IgM titers are measured intermittently over several time periods, and they can be viewed as independent measurements of a latent covariate process – an unobservable antibody response to GM2. The prognostic value of these covariates is of great interest in these studies, and the covariate process itself may be of interest, as it sheds light on the natural history of the disease. In order to study the relationship of these observable covariates to RFS or OS, we develop a Bayesian model for joint modeling of the survival data and the longitudinal IgG and IgM measurements. Since inference is based on the parameters that describe the covariate process and those that describe the risk of failure as a function of the covariate process at the same time, our method not only uses the observed covariate data, but also uses survival information to get estimates of the true latent covariate values at any time. We can expect, therefore, more precise and accurate estimates of the strength of the relationship between the latent covariate and the risk of failure.

We consider a case study of a recent Eastern Cooperative Oncology Group (ECOG) phase II clinical trial, E2696, that used the GMK vaccine. Two treatment arms are used in our analysis. The treatment arms consist of a combination of interferon (IFN) and the ganglioside vaccine (GMK), which we label as A (IFN + GMK). The other treatment arm consists of GMK alone, labeled as B. There were 35 patients on each treatment arm, resulting in a sample size of $n = 70$ patients. There were a total of 27 completely observed RFS times. The median RFS based on $n = 70$ patients was 11.93 months. IgG and IgM antibody titer measurements were taken at the five time points, 0, 4, 6, 12 and 52 weeks. In all of the analyses, the IgG and IgM measures were transformed to logarithms. Since many of the IgG and IgM measures were 0 before transformation, we first added a value of 1 to all IgG and IgM titer values, then took natural logarithms. Table 1 gives a detailed summary of the $(\log(\text{IgG}+1), \log(\text{IgM}+1))$ measures along with summaries of missing values. For ease of exposition, we denote this transformation by log(IgG) and log(IgM) throughout the paper.

Table 1. Summary of transformed IgG and IgM measures.

		log(IgG)					log(IgM)				
Treatment	Week	0	4	6	12	52	0	4	6	12	52
A	Median	0.00	3.38	4.39	3.71	7.15	0.00	5.08	5.08	3.71	4.73
	Mean	0.00	2.84	4.17	3.05	6.74	0.14	5.07	5.26	3.31	3.53
	SD	0.00	2.53	2.18	2.40	2.15	0.78	1.84	1.54	2.29	2.59
	Number of Missing	3	1	5	5	13	3	1	5	5	13
B	Median	0.00	3.04	3.71	3.04	6.46	0.00	5.08	5.08	4.39	3.71
	Mean	0.00	2.23	3.12	2.23	5.87	0.58	5.02	4.99	3.85	2.99
	SD	0.00	2.35	2.47	2.39	1.86	2.04	1.94	1.40	2.04	2.46
	Number of Missing	7	2	0	6	16	7	2	0	6	16

There are three main objectives in this study: (i) to formally examine and assess the association between the antibody titers and RFS for these data in order to determine the potential efficacy of GMK, (ii) to examine an appropriate form of a longitudinal model for describing the behavior of the longitudinal measures over time, and (iii) to examine the relationships between the (IgG, IgM) antibody titers through an appropriate statistical model. Characterizing the relationship between the (IgG, IgM) antibody titers may shed light on which is the better measure to use in assessing immune response, and whether these two measures act in concert over time.

Toward these goals, an exploratory analysis of these data revealed the following features.

1. By plotting the antibody titer counts vs. a model-based estimate of the hazard, given in Figure 1 (see Section 5), we see a decrease in the hazard estimate for each individual. This phenomenon is observed for both treatment arms as well as for both the IgG and IgM antibody measurements, and therefore gives us initial evidence that there is indeed some positive association between high antibody titers and longer relapse-free survival.

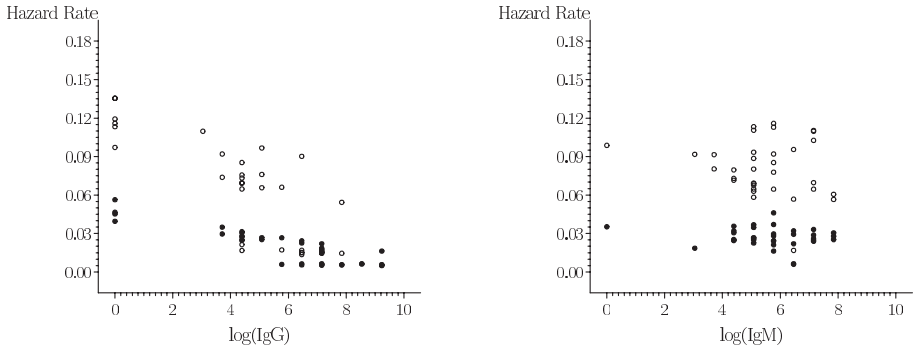

Figure 1. Estimated hazard rate as a function of IgG (left) or IgM (right) taken at time point of peak IgG, where • and ∘ correspond to treatments A and B, respectively.

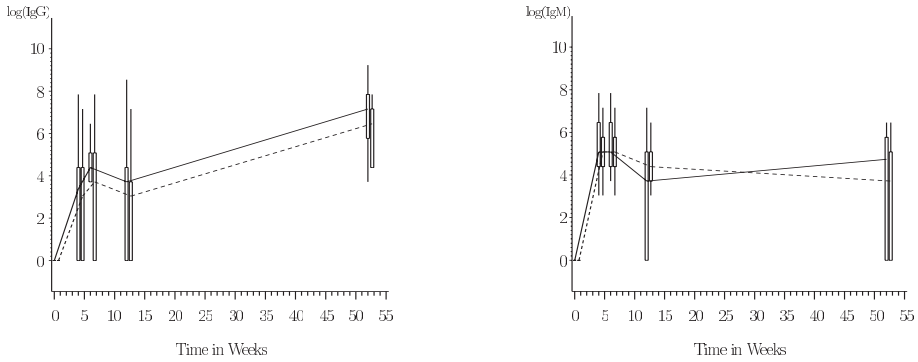

Figure 2. Boxplots for log(IgG) (left) and log(IgM) (right) where the solid and dashed lines correspond to treatments A and B, respectively.

2. Figure 2 shows box plots of the logarithm of the IgG and IgM antibody titer measurements, respectively, against time, at each of the five time points. From Figure 2, we see evidence of the quadratic trend in the IgG trajectory, which occurs in the period from $2 - 8$ weeks, with the peak IgG titer occurring at approximately 52 weeks. Figure 3 confirms that the peak (maximum) IgG titer occur at 52 weeks. Figure 2 also shows evidence of the quadratic trend in the IgM trajectory, which also occurs in the period from $2 - 8$ weeks, with the

peak IgM titer occurring at approximately 4 weeks. Figure 3 confirms that the peak (maximum) IgM titer occurs at 4 weeks. Figure 4 shows the individual patient trajectory plots by treatment arm for each individual patient for the IgG and IgM titers, respectively. Figure 4 confirms the need for a quadratic trajectory.

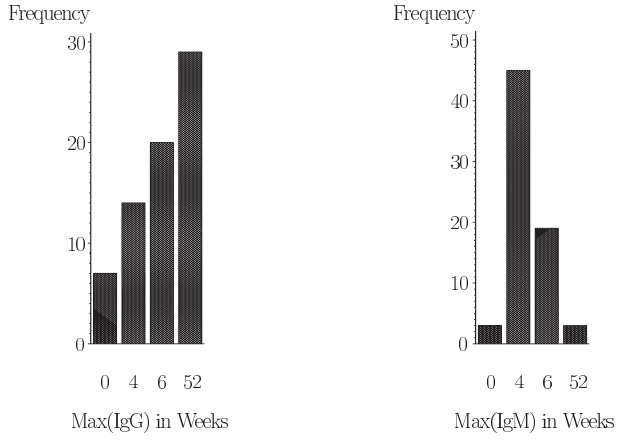

Figure 3. Histograms of Max(IgG) (left) and Max(IgM) (right).

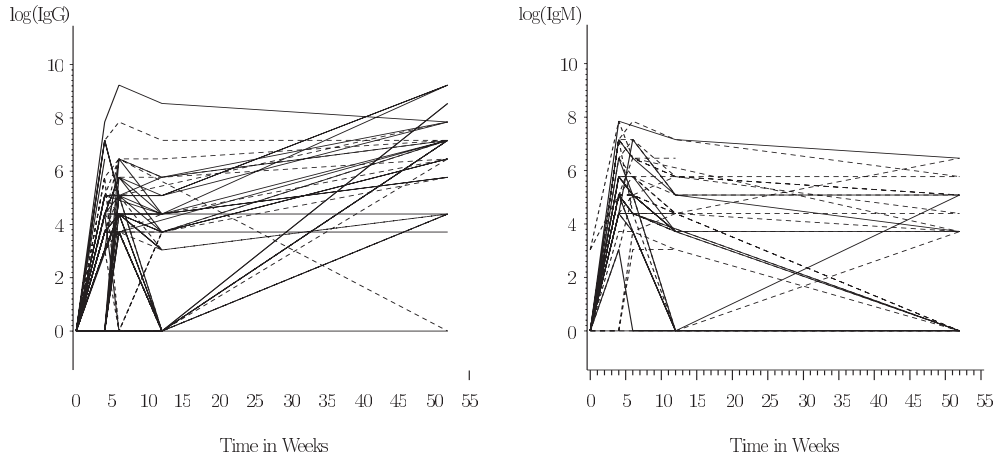

Figure 4. Trajectory plots for log(IgG) (left) and log(IgM) (right) where the solid and dashed lines correspond to Treatments A and B, respectively.

3. Table 1 shows a statistical summary of the IgG and IgM measures for each treatment arm for each time point, and it also shows the fraction of missing antibody titers at each time point. Table 1 indicates that, on average, the IgM titer is greater than that of the IgG titer in a wide window of time. This

implies that IgG and IgM are linked on the same scale biologically, and this should be taken into account when jointly modeling these two longitudinal measures.

There has been much previous work in joint modeling of survival and longitudinal data focused almost exclusively on AIDS studies, and in particular, jointly modeling of survival data and univariate or multivariate longitudinal CD4 counts. These articles include DeGruttola and Tu (1994), Tsiatis, DeGruttola and Wulfsohn (1995), Faucett and Thomas (1996), Wulfsohn and Tsiatis (1997), Taylor, Cumberland and Sy (1994) and Wang and Taylor (2001). Other approaches considering a multivariate longitudinal measure and related to our development here include Henderson, Diggle and Dobson (2000), Xu and Zeger (2001a, 2001b) and Song, Davidian and Tsiatis (2002).

The rest of this paper is organized as follows. In Section 2, we present the statistical model and provide a biological motivation for it. In Section 3, we present the likelihood and the priors as well as providing biological interpretations of the parameters. In Section 4, we provide model assessment tools that will enable us to judge the goodness of fit for the various models as we consider different forms of the trajectory, priors, and survival model. In Section 5, we provide a detailed analysis of the E2696 data and in Section 6, we present a discussion of our findings and examine their biological and scientific implications.

2. Model and Notation

For a clear focus and ease of exposition, we develop our model in the context of vaccines in cancer clinical trials and, in particular, for the E2696 study. The model, however, is quite general and can be used in many other contexts. The longitudinal component of the model can be described as follows. Let $X_i(t)$ denote the true, unobservable antibody level against GM2, and let $\boldsymbol{Y}_i(t) = (Y_{i1}(t), Y_{i2}(t))'$ denote the 2×1 vector of observed (IgG, IgM) immunological titer measures for subject i at risk at time t, $i = 1, \ldots, n$. We assume

$$Y_{i1}(t) = X_i(t) + \epsilon_{i1}(t), \tag{2.1}$$

$$Y_{i2}(t) = \alpha_0 + \alpha_1 X_i(t) + \epsilon_{i2}(t)\ , \tag{2.2}$$

$$\boldsymbol{\epsilon}_i(t) \sim N_2\left(\left(\begin{array}{c} 0 \\ 0 \end{array}\right), \Sigma = \left(\begin{array}{cc} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{array}\right)\right), \tag{2.3}$$

where $\boldsymbol{\epsilon}_i(t) = (\epsilon_{i1}(t), \epsilon_{i2}(t))'$ is independent of $X_i(t)$. We also model $X_i(t)$ through a known but arbitrary function $g_{\boldsymbol{\gamma}_i}(t)$ indexed by parameter vector $\boldsymbol{\gamma}_i$.

The form of (2.1) and (2.2) is based on sound biological considerations and is partially verified from Table 1. For melanoma, the IgG antibody titer is a biologically preferred measure of the unobserved level $X_i(t)$ of antibody response to GM2 compared to IgM. It has also been established by Kirkwood et al. (2000) that an IgG response typically implies a change in antibody response ($X_i(t)$) to GM2. In Table 1 we see that, on average, the IgM titer is greater than that of the IgG titer in a wide window of time. For this reason, we measure $X_i(t)$ on the same scale as the IgG response. The error term $\boldsymbol{\epsilon}_i(t)$ represents the fact that the (IgG, IgM) measures are not perfect measures of the unobserved immunity level, but rather, in some sense, an approximation to the true underlying immunity level. Since the (IgG, IgM) are serologic measures on the same individual conducted on the same assay, they are correlated conditional on $X_i(t)$, and ρ represents this unknown correlation. The unobserved immunity level of a patient is related to his/her serologic measurements $(Y_1(t), Y_2(t))$, and the correlation between the serologic measurements is primarily due to their relationship to the common latent factor - immunity level. The argument for this is analogous to the assumption that the frailty distribution has mean 1 in the frailty model. In our model, the $X_i(t)$'s can be viewed as time varying frailties, that is, latent variables measuring subject specific morbidity, which are also related to the serological immune response measures. Since the (IgG, IgM) measures have different unknown dispersion, given $X_i(t)$, we assume different error variances (σ_1^2, σ_2^2) for (IgG, IgM). A more general version of this model can be considered by assuming within subject autocorrelation over time. An empirical investigation of this showed that the autocorrelations are nearly zero for these data, and such a model is perhaps not needed here.

For subject i, write $X_i(t) \equiv g_{\boldsymbol{\gamma}_i}(t)$, where $\boldsymbol{\gamma}_i$ is a vector of random effects for subject i. We call $g_{\boldsymbol{\gamma}_i}(t)$ the *trajectory function* throughout. The name "trajectory" is meaningful, as $g_{\boldsymbol{\gamma}_i}(t)$ actually measures the trajectory of immune response to GM2. The function g explains the behavior of the unobserved covariate process over time, and thus is a critical component of the model for understanding the biology of the vaccine and the disease. The exploratory analyses for E2696 shown by Figure 2 indicated that a quadratic trajectory might be suitable for these data and, in this case, $\boldsymbol{\gamma}_i = (\gamma_{1i}, \gamma_{2i}, \gamma_{3i})$ and $X_i(t) = \gamma_{1i} + \gamma_{2i}t + \gamma_{3i}t^2$. One can also incorporate interactions and higher order terms in $X_i(t)$. In this paper, we develop the model assuming a general structure for $X_i(t) = g_{\boldsymbol{\gamma}_i}(t)$, and consider specific forms for $g_{\boldsymbol{\gamma}_i}(t)$ in Section 5.

The random effects $\boldsymbol{\gamma}_i$ play a critical role in the model development. They represent the subject specific trajectory parameters. A common distribution that is specified for the $\boldsymbol{\gamma}_i$'s is a multivariate normal distribution, as in the usual

assumption in a mixed model. In the case of a quadratic trajectory model, we assume in our data analysis that

$$\boldsymbol{\gamma}_i \sim N_3(\boldsymbol{\mu}_0, \Sigma_0), \tag{2.4}$$

where $\boldsymbol{\mu}_0 = (\mu_{01}, \mu_{02}, \mu_{03})$ represents the population mean of the individual-specific trajectory parameters, and $\Sigma_0 = (\sigma_{0jk})$ represents the variation in the parameters that define the individual trajectories across the population of patients. The form of Σ_0 plays a crucial role in the analysis and, in Section 5, we consider two main analyses. The first analysis is based on an unstructured Σ_0 in which we take a Wishart prior for Σ_0^{-1},

$$\Sigma_0^{-1} \sim W_3(n_0^*, Q_0^*), \tag{2.5}$$

where n_0^* and Q_0^* are chosen to make the prior noninformative. An unstructured form of Σ_0 is the most biologically meaningful specification since it allows general correlations between the subject-specific trajectory parameters (the $\boldsymbol{\gamma}_i$'s). In Section 5, our main analyses will be based on an unstructured Σ_0. If the associations between the $\boldsymbol{\gamma}_i$'s is not too strong it may be reasonable to consider simplified versions of Σ_0, such as a diagonal Σ_0. This more parsimonious structure for Σ_0 has less biological meaning than an unstructured Σ_0 but may be empirically attractive since it may result in a more efficient parameter estimates if indeed the associations between the γ_i's are weak. For comparison to the unstructured Σ_0, we carry out a second analysis using a diagonal form

$$\Sigma_0 = \begin{pmatrix} \sigma_{01}^2 & 0 & 0 \\ 0 & \sigma_{02}^2 & 0 \\ 0 & 0 & \sigma_{03}^2 \end{pmatrix}, \tag{2.6}$$

where the σ_{0j}^2 have independent inverse gamma distributions. In addition, we take $\mu_{01} \sim N(\xi_{01}, v_{01}^2)$, $\mu_{02} \sim N(\xi_{02}, v_{02}^2)$, $\mu_{03} \sim N(\xi_{03}, v_{03}^2)$, $\sigma_{01}^2 \sim IG(e_{01}, f_{01})$, $\sigma_{02}^2 \sim IG(e_{02}, f_{02})$ and $\sigma_{03}^2 \sim IG(e_{03}, f_{03})$. Here $IG(a, b)$ denotes the inverse gamma distribution with shape parameter a and scale parameter b. In the data analysis, The hyperparameters are all chosen so that the resulting priors for $(\boldsymbol{\mu}_0, \Sigma_0)$ in Section 5 are noninformative. The above specifications can be easily extended to more general trajectory models.

Several papers in the literature have questioned the adequacy and robustness of the normality assumption for the random effects $\boldsymbol{\gamma}_i$, and various alternative approaches, Bayesian and non-Bayesian, have been proposed in the literature. Brown and Ibrahim (2003) consider a nonparametric approach where they specify a Dirichlet process prior for the $\boldsymbol{\gamma}_i$'s. In any case, the normal distribution assumption for the $\boldsymbol{\gamma}_i$'s needs careful scrutiny and examination.

The survival component of the model is taken to have a proportional hazards structure. For the ith subject at risk at time t, let $\mathcal{X}_i(t)$ denote the history of $X_i(\cdot)$ up to time t, $\mathcal{Y}_i(t)$ denote the history of the observable covariates $(Y_{i1}(\cdot), Y_{i2}(\cdot))$ up to t, and let $\boldsymbol{z}_i$ denote a $p \times 1$ vector of baseline covariates for subject i, such as treatment, gender, age, and so forth. We assume that the $\boldsymbol{z}_i$'s are measured without error. We make the usual assumption that the hazard $h(t|\mathcal{X}_i(t), \mathcal{Y}_i(t), \boldsymbol{z}_i) = h(t|\mathcal{X}_i(t), \boldsymbol{z}_i)$, and model the hazard function for the ith subject as

$$h(t|\mathcal{X}_i(t), \mathcal{Y}_i(t), \boldsymbol{z}_i) = h_0(t) \exp\left\{\beta_1 X_i(t) + \boldsymbol{z}_i' \boldsymbol{\beta}_2\right\} \ , \tag{2.7}$$

where $h_0(t)$ is the baseline hazard function, β_1 is a scalar regression coefficient for the longitudinal covariate process, and $\boldsymbol{\beta}_2$ is a $p \times 1$ vector of regression coefficients for the baseline covariates. The random measurement errors are not prognostic of the survival time. That is, given the unobservable trajectory $X_i(t)$, the observable serologic measurements are not prognostic. The regression coefficient β_1 plays a critical role in the model, since it is the parameter that yields information on the association between the longitudinal measurements and survival. Inference about this parameter will help us address point 1 in the previous section. Values of β_1 near 0 imply a weak association, whereas values of β_1 far from 0 imply a strong association.

Another important aspect of the model is the form of the trajectory function. The assessment of the appropriate form of the trajectory function is critical for making predictions and for understanding the biology of the vaccine and the disease. There are several other important features of the joint model given by (2.1), (2.2) and (2.7). First, the model explains a good portion of the variability in survival time through the variability in the 'unobservable' immunity level for an individual. This relationship between risk of cancer and time-varying immunity level is of great clinical importance. Second, the form of the hazard function (2.7) assumes that the hazard at time t only depends on the current, true, unobservable antibody level $X_i(t)$. In the cancer vaccine context, this implies that the true antibody levels measured at time t are the only relevant quantities characterizing the survival time at time t. This seems to be a reasonable assumption for this cancer vaccine application since the course of the vaccine treatment is given over a long period of time, and the effects of the vaccine on survival are best assessed at the most recent vaccine injection.

To fully characterize the relationship between the survival time t_i and the covariate process $\boldsymbol{Y}_i(t)$, it is of interest to compute the conditional survival function of t_i given $\mathcal{X}_i(t)$, $\mathcal{Y}_i(t)$ and $\boldsymbol{z}_i$. This can be expressed as a posterior expectation, given by $E[h(t|\mathcal{X}_i(t), \mathcal{Y}_i(t), \boldsymbol{z}_i)|D]$, where D denotes the observed data. Gibbs sampling can then be used to calculate this quantity once samples from the joint

posterior distribution of the parameters are available. The major Gibbs steps are provided and the implementation of these Gibbs steps is discussed in detail in the Appendix.

3. Likelihood and Priors

To construct the likelihood function, we discretize the time axis into J intervals. Let $A_{i1}, \ldots, A_{im_i}$ denote the times at which the antibody measurements are taken, and let $g_{\boldsymbol{\gamma}_i}(A_{il})$ denote the trajectory function evaluated at A_{il}. For example, for a quadratic trajectory, we have $g_{\boldsymbol{\gamma}_i}(A_{il}) = \gamma_{1i} + \gamma_{2i}A_{il} + \gamma_{3i}A_{il}^2$ and $\boldsymbol{\gamma}_i = (\gamma_{1i}, \gamma_{2i}, \gamma_{3i})'$. Let the IgG and IgM antibody titers for subject i be denoted by $\boldsymbol{Y}_{i1} = (y_{i11}, \ldots, y_{im_i1})'$ and $\boldsymbol{Y}_{i2} = (y_{i12}, \ldots, y_{im_i2})'$, and let $\boldsymbol{Y}_1 = (\boldsymbol{Y}'_{11}, \ldots, \boldsymbol{Y}'_{n1})'$, and $\boldsymbol{Y}_2 = (\boldsymbol{Y}'_{12}, \ldots, \boldsymbol{Y}'_{n2})'$. Further, let t_i denote the event time for the ith subject, which may be right censored and let $\boldsymbol{t} = (t_1, \ldots, t_n)'$ denote the vector of event times. We take $h_0(t)$ to be a constant λ_j over the time intervals $I_j = (c_{j-1}, c_j]$, for $j = 1, \ldots, J$, where $c_0 = 0 < c_1 < \ldots < c_J < c_{J+1} = \infty$. Let $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_n)'$ denote the vector of censoring indicators, where $\delta_i = 1$ indicates a failure and $\delta_i = 0$ indicates a right censored observation. The likelihood function for the joint model involves two components, denoted by L_1 and L_2. The first component L_1 is the likelihood for $(\boldsymbol{Y}_1, \boldsymbol{Y}_2)$, and L_2 is the likelihood function for $\boldsymbol{t}$. Let $\boldsymbol{\theta}$ be a generic label for the vector of all the parameters in L_1 and L_2. The likelihood function of $\boldsymbol{\theta}$ is given by

$$L(\boldsymbol{\theta}) = L_1(\boldsymbol{\gamma}, \alpha_0, \alpha_1, \Sigma | \boldsymbol{Y}_1, \boldsymbol{Y}_2) L_2(\boldsymbol{\lambda}, \beta_1, \boldsymbol{\beta}_2, \boldsymbol{\gamma} | \boldsymbol{t}, \boldsymbol{\delta}, \boldsymbol{z}), \tag{3.1}$$

where

$$L_1(\boldsymbol{\gamma}, \alpha_0, \alpha_1, \Sigma | \boldsymbol{Y}_1, \boldsymbol{Y}_2) \propto \prod_{i=1}^{n} \Big[|\Sigma|^{-m_i/2} \exp\Big\{ -\frac{1}{2} \sum_{l=1}^{m_i} \Big((y_{il1} - g_{\boldsymbol{\gamma}_i}(A_{il}), y_{il2} - (\alpha_0 + \alpha_1 g_{\boldsymbol{\gamma}_i}(A_{il}))) \Sigma^{-1} (y_{il1} - g_{\boldsymbol{\gamma}_i}(A_{il}), y_{il2} - (\alpha_0 + \alpha_1 g_{\boldsymbol{\gamma}_i}(A_{il})))\Big)' \Big\} \Big], \tag{3.2}$$

$$L_2(\boldsymbol{\lambda}, \beta_1, \boldsymbol{\beta}_2, \boldsymbol{\gamma} | \boldsymbol{t}, \boldsymbol{\delta}, \boldsymbol{z}) \propto \exp\Big\{ -\sum_{j=1}^{J} \sum_{i=1}^{n} \lambda_j B_{ij} \Big\} \Big(\prod_{j=1}^{J} \lambda_j^{d_j} \Big) \exp\Big\{ \sum_{j=1}^{J} \sum_{i=1}^{n} \delta_i (\beta_1 g_{\boldsymbol{\gamma}_i}(t_{ij}^*) + \boldsymbol{z}_i' \boldsymbol{\beta}_2) \Big\}, \tag{3.3}$$

$\boldsymbol{z} = (\boldsymbol{z}'_1, \ldots, \boldsymbol{z}'_n)'$, $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_n)'$, $\boldsymbol{\alpha} = (\alpha_0, \alpha_1)'$, $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_J)'$, $\boldsymbol{\theta} = (\boldsymbol{\gamma}, \boldsymbol{\alpha}, \beta_1, \boldsymbol{\beta}_2, \sigma_1, \sigma_2, \boldsymbol{\lambda})$, d_j is the number of failures, and t_{ij}^* denotes the nearest past time point where antibody measurements are taken. We also assume here that the scheduling of the (IgG, IgM) measurements is not predictive of survival. In (3.3), the computational algorithm for B_{ij} proceeds as follows.

(i) If $t_i < c_{j-1}$, $B_{ij} = 0$.

(ii) If $t_i > c_j$, letting $j_{i1} = \max\{l : A^*_{il} \le c_{j-1}\}$ and $j_{i2} = \max\{l : A^*_{il} \le c_j\}$, where A^*_{il} is the rescaled A_{il} so that A^*_{il} has the same unit as t_i, then if $j_{i1} = j_{i2}$, $B_{ij} = (c_j - c_{j-1}) \exp\{\beta_1 g_{\boldsymbol{\gamma}_i}(A_{ij_{i1}}) + \boldsymbol{z}_i'\boldsymbol{\beta}_2\}$, and if $j_{i1} < j_{i2}$,

$$\begin{aligned} B_{ij} = {} & (A^*_{i,j_{i1}+1} - c_{j-1}) \exp\left\{\beta_1 g_{\boldsymbol{\gamma}_i}(A_{i,j_{i1}+1}) + \boldsymbol{z}_i'\boldsymbol{\beta}_2\right\} \\ & + \sum_{l=j_{i1}+1}^{j_{i2}} (A^*_{i,l+1} - A^*_{il}) \exp\left\{\beta_1 g_{\boldsymbol{\gamma}_i}(A_{il}) + \boldsymbol{z}_i'\boldsymbol{\beta}_2\right\} \\ & + (c_j - A^*_{ij_{i2}}) \exp\left\{\beta_1 g_{\boldsymbol{\gamma}_i}(A_{ij_{i2}}) + \boldsymbol{z}_i'\boldsymbol{\beta}_2\right\}. \end{aligned}$$

(iii) If $c_{j-1} < t_i \le c_j$, using j_{i1} and j_{i2} given in (ii), then if $j_{i1} = j_{i2}$ or $t_i \le A^*_{i,j_{i1}+1}$, when $j_{i1} < j_{i2}$, $B_{ij} = (t_i - c_{j-1}) \exp\left\{\beta_1 g_{\boldsymbol{\gamma}_i}(A_{ij_{i1}}) + \boldsymbol{z}_i'\boldsymbol{\beta}_2\right\}$, and otherwise, we define

$$\begin{aligned} B_{ij} = {} & (A^*_{i,j_{i1}+1} - c_{j-1}) \exp\left\{\beta_1 g_{\boldsymbol{\gamma}_i}(A_{i,j_{i1}+1}) + \boldsymbol{z}_i'\boldsymbol{\beta}_2\right\} \\ & + \sum_{l=j_{i1}+1}^{k_i} (A^*_{i,l+1} - A^*_{il}) \exp\left\{\beta_1 g_{\boldsymbol{\gamma}_i}(A_{il}) + \boldsymbol{z}_i'\boldsymbol{\beta}_2\right\} \\ & + (t_i - A^*_{ik_i}) \exp\left\{\beta_1 g_{\boldsymbol{\gamma}_i}(A_{ik_i}) + \boldsymbol{z}_i'\boldsymbol{\beta}_2\right\}, \end{aligned}$$

where $j_{i1} + 1 \le k_i \le j_{i2}$ is chosen so that $A^*_{ik_i} < t_i \le A^*_{i,k_i+1}$.

When j_{il} $(l = 1, 2)$ does not exist, we define $j_{il} = 1$, and the calculation of B_{ij} needs a minor adjustment.

In the likelihood function (3.3), we have invoked an approximation. The likelihood contribution for the ith subject in $L_2(\boldsymbol{\lambda}, \beta_1, \boldsymbol{\beta}_2, \boldsymbol{\gamma} | \boldsymbol{t}, \boldsymbol{\delta}, \boldsymbol{z})$ is given by $(h_0(t_i) \exp\{\beta_1 g_{\boldsymbol{\gamma}_i}(t_i) + \boldsymbol{z}_i'\boldsymbol{\beta}_2\})^{\delta_i} \exp\{-\int_0^{t_i} h_0(u) \exp\{\beta_1 g_{\boldsymbol{\gamma}_i}(u) + \boldsymbol{z}_i'\boldsymbol{\beta}_2\} du\}$. An approximation of this integral is needed primarily for computational convenience. We approximate the integral involving the term $g_{\boldsymbol{\gamma}_i}(u)$ by $\sum_{i=1}^n [\int_0^{t_i} h_0(u) \exp\{\beta_1 g_{\boldsymbol{\gamma}_i}(u) \boldsymbol{z}_i'\boldsymbol{\beta}_2\}] du \approx \sum_{i=1}^n \sum_{j=1}^J \lambda_j B_{ij}$. This approximation should work well as long as the trajectory of the latent antibody level affecting the hazard function does not change over time too rapidly compared to the scheduled serology measurements. If they do change too rapidly, we have too little information about the trajectory to use any kind of model or analysis for such an example. It is also more sensible to use this approximation in the likelihood instead of the previous one since, in our data, we have indirect information on the trajectory only at the t^*_{ij}'s, and to use the values of the trajectory at other time points will require interpolations which are impossible to validate through data. This type of approximation was also used by Tsiatis, DeGruttola and Wulfsohn (1995).

For ease of exposition, and without loss of generality, we assume $p = 1$ for the survival component of the model. The priors for the parameters in the survival component of the model are taken as: $\lambda_j \sim \text{gamma}(a_j, b_j)$, $\beta_1 \sim N(\xi_1, v_1^2)$, $\beta_2 \sim N(\xi_2, v_2^2)$, $\alpha_0 \sim N_2(\xi_3, v_3^2)$, $\alpha_1 \sim N(\xi_4, v_4^2)$, $\Sigma^{-1} \sim W_2(n_0, Q_0)$, where Q_0 is a 2×2 symmetric and positive definite matrix, $W_2(n_0, Q_0)$ denotes the Wishart distribution with degrees of freedom n_0 and mean matrix $n_0 Q_0$ and n_0 and Q_0 are prespecified *a priori*. For the E2696 analysis, we take $n_0 = 3$ and $Q_0^{-1} = 0.001 I_2$, where I_2 is the 2-dimensional identity matrix, so that the prior is sufficiently diffuse. In Section 5, noninformative priors are taken for all parameters. We can see that the above prior specification can be easily generalized for $p > 1$.

4. Model Choice via The L Measure

The choice of the parametric form of the trajectory function $g_{\boldsymbol{\gamma}_i}(t)$ is an important issue. Here we propose a multivariate criterion for model assessment called the *multivariate L measure*. It is motivated as follows. Consider an experiment that yields the data $\boldsymbol{w} = (\boldsymbol{w}_1, \ldots, \boldsymbol{w}_n)'$, where each $\boldsymbol{w}_i = (w_{i1}, w_{i2}, \ldots, w_{iq})'$ is a q vector of response variables. In our joint modeling application, we have $\boldsymbol{w}_i = (y_{i1}, y_{i2}, t_i)'$ so that $q = 3$. Denote the joint sampling density of the $\boldsymbol{w}_i$'s by $p(\boldsymbol{w}|\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is a vector of indexing parameters. We allow the $\boldsymbol{w}_i$'s to be fully observed or right censored. In our specific application, the (y_{i1}, y_{i2}) are fully observed and t_i is right censored. Let $\boldsymbol{v} = (\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n)'$ denote future values of a replicate experiment. That is, $\boldsymbol{v}$ is a future response vector with the same sampling density as $\boldsymbol{w}|\boldsymbol{\theta}$. The idea of using a future response matrix $\boldsymbol{v}$ in developing a criterion for assessing a model or comparing several models has been used by Ibrahim and Laud (1994), Laud and Ibrahim (1995), Ibrahim and Chen (1997) and Ibrahim, Chen and Sinha (2001).

Let $\boldsymbol{f}(\cdot) = (f_1(\cdot), \ldots, f_q(\cdot))'$ be known functions, and let the jth component of $\boldsymbol{w}_i$ be transformed to $w_{ij}^* = f_j(w_{ij})$, with $v_{ij}^* = f_j(v_{ij})$, $j = 1, \ldots, q$. For example, one may take the logarithms of the survival times t_i, and in this case $f_3(t_i) = \log(t_i) = w_{i3}^*$. One may also want to transform the the longitudinal measures to logarithms or apply some other transformation. In the analysis of the E2696 data, we transform the survival time and the (IgG, IgM) measures to logarithms.

The multivariate L measure can be written as

$$L_2(\boldsymbol{w}^*) = \sum_{i=1}^{n} \text{Cov}(\boldsymbol{v}_i^*|\boldsymbol{w}^*) + \nu \sum_{i=1}^{n} (\boldsymbol{\mu}_i - \boldsymbol{w}_i^*)(\boldsymbol{\mu}_i - \boldsymbol{w}_i^*)', \tag{4.1}$$

where $0 < \nu < 1$. Since $L_2(\boldsymbol{w}^*)$ is a matrix, we take its determinant as our assessment statistic, leading to $\hat{L} = \mid L_2(\boldsymbol{w}^*) \mid$. Allowing ν to vary between zero and

one gives the user a great deal of flexibility in the tradeoff between bias and variance. If $\boldsymbol{w}^*$ is fully observed, then (4.1) is straightforward to compute. However, if $\boldsymbol{w}^*$ contains right censored observations, then (4.1) is computed by taking the expectation of these censored observations with respect to the posterior predictive distribution of the censored observations. Let $\boldsymbol{w}^* = (\boldsymbol{w}^*_{obs}, \boldsymbol{w}^*_{cens})'$, where $\boldsymbol{w}^*_{obs}$ denotes the completely observed components of $\boldsymbol{w}^*$ and $\boldsymbol{w}^*_{cens}$ denotes the censored components. We assume that $\boldsymbol{w}^*_{cens}$ is a random quantity and $\boldsymbol{a}^* < \boldsymbol{w}^*_{cens}$, where $\boldsymbol{a}^*$ is known. For ease of exposition, we let $D = (n, \boldsymbol{w}_{obs}, \boldsymbol{a}^*)$ denote the observed data. Then (4.1) is modified as

$$L(\boldsymbol{w}^*_{obs}) = E_{\boldsymbol{w}_{cens}|D}[I(\boldsymbol{a}^* < \boldsymbol{w}^*_{cens})L_2(\boldsymbol{w}^*)], \tag{4.2}$$

where $I(\boldsymbol{a}^* < \boldsymbol{w}^*_{cens})$ is a generic indicator function taking the value 1 if $\boldsymbol{a}^* < \boldsymbol{w}^*_{cens}$ and 0 otherwise. When $\boldsymbol{a}^*$ and $\boldsymbol{w}^*_{cens}$ are vectors, then $\boldsymbol{a}^* < \boldsymbol{w}^*_{cens}$ means that the inequalities hold for each component of these vectors. Write (4.2) as $L(\boldsymbol{w}^*_{obs}) = \int \int_{\boldsymbol{w}^*_{cens} > \boldsymbol{a}^*} L_2(\boldsymbol{w}^*) p(\boldsymbol{w}^*_{cens}|\boldsymbol{\theta}) p(\boldsymbol{\theta}|D) d\boldsymbol{w}^*_{cens} d\boldsymbol{\theta}$, where $p(\boldsymbol{w}^*_{cens}|\boldsymbol{\theta})$ is the sampling density of $\boldsymbol{w}^*_{cens}$ and $p(\boldsymbol{\theta}|D)$ is the posterior density of $\boldsymbol{\theta}$ given the observed data D. It can be shown that (4.1) can be expressed as a posterior expectation, so that

$$L_2(\boldsymbol{w}^*) = \sum_{i=1}^{n} \left\{ E_{\boldsymbol{\theta}|D} \left(E \left[\boldsymbol{v}^*_i \boldsymbol{v}^{*'}_i | \boldsymbol{\theta} \right] \right) - \boldsymbol{\mu}_i \boldsymbol{\mu}'_i \right\} + \nu \sum_{i=1}^{n} (\boldsymbol{\mu}_i - \boldsymbol{w}^*_i)(\boldsymbol{\mu}_i - \boldsymbol{w}^*_i)', \tag{4.3}$$

where $\boldsymbol{\mu}_i = E_{\boldsymbol{\theta}|D}[E(\boldsymbol{v}^*_i|\boldsymbol{\theta})]$. Once the posterior samples of $\boldsymbol{\theta}$ are obtained, (4.3) and (4.2) can be easily evaluated.

5. Analysis of the E2696 Data

For the E2696 data, there were 8 patients with intermittent missingness, 15 patients that were administratively censored before 52 weeks, of which 6 patients had missing longitudinal measures, 28 patients were administratively censored after 12 months, 18 patients had missing data at timepoints before relapse, 23 patients relapsed before 52 weeks and 4 patients who relapsed after 52 weeks. Once a patient relapsed, they dropped out of the study, and hence no longitudinal measures were collected after that time. Also, the minimum RFS time was 0.59 months and the maximum RFS time was 23.95 months based on $n = 70$ patients.

We consider two major analyses, one in which Σ_0 is assumed unstructured with a Wishart prior given by (2.5), and one in which Σ_0 is assumed diagonal with prior given by (2.6). Noninformative priors were used for all of the models in the analyses below. For example, for the quadratic trajectory model with $p = 1$, $J = 8$, and unstructured Σ_0, we take, $a_j = b_j = 0$, $j = 1, \ldots, 8$, $(\xi_l, v^2_l) = (0, 100)$, $l = 1, \ldots, 4$, $\xi_{01} = \xi_{02} = \xi_{03} = 0$, $v^2_{01} = v^2_{02} = v^2_{03} = 100$, $\Sigma_0^{-1} \sim$

$W_3\left(4, \begin{pmatrix} 1.0 & 0 & 0 \\ 0 & 0.1 & 0 \\ 0 & 0 & 0.1 \end{pmatrix}\right)$. For the analysis with a diagonal Σ_0, we take $e_{01} = e_{02} = e_{03} = 0.001$ and $f_{01} = f_{02} = f_{03} = 0.001$.

Table 2. Posterior estimates for all parameters for unstructured Σ_0 and $J = 8$.

	Model with Quadratic Trajectory			Model with Linear Trajectory		
Parameter	Mean	SD	95% HPD	Mean	SD	95% HPD
β_1	-0.256	0.195	$(-0.648, 0.118)$	-0.228	0.182	$(-0.602, 0.118)$
β_2	1.009	0.433	$(0.154, 1.853)$	1.007	0.430	$(0.154, 1.839)$
ρ	0.603	0.048	$(0.507, 0.695)$	0.611	0.043	$(0.527, 0.695)$
σ_1^2	4.123	0.421	$(3.321, 4.959)$	4.261	0.399	$(3.500, 5.054)$
σ_2^2	6.787	0.574	$(5.726, 7.958)$	6.788	0.569	$(5.722, 7.929)$
α_0	3.638	0.353	$(2.964, 4.330)$	3.632	0.334	$(3.002, 4.299)$
α_1	-0.032	0.114	$(-0.256, 0.182)$	-0.029	0.104	$(-0.231, 0.173)$
μ_{01}	2.765	0.176	$(2.421, 3.112)$	2.758	0.191	$(2.390, 3.139)$
μ_{02}	1.802	0.517	$(0.772, 2.781)$	1.460	0.153	$(1.156, 1.764)$
μ_{03}	-0.344	0.488	$(-1.229, 0.659)$	–	–	–
σ_{011}	1.085	0.379	$(0.363, 1.866)$	1.221	0.371	$(0.546, 1.982)$
σ_{012}	0.026	0.090	$(-0.150, 0.206)$	0.035	0.113	$(-0.195, 0.252)$
σ_{013}	-0.069	0.111	$(-0.295, 0.145)$	–	–	–
σ_{022}	0.026	0.022	$(0.004, 0.065)$	0.058	0.045	$(0.007, 0.148)$
σ_{023}	-0.006	0.017	$(-0.040, 0.023)$	–	–	–
σ_{033}	0.064	0.056	$(0.007, 0.173)$	–	–	–
λ_1	0.055	0.039	$(0.0043, 0.130)$	0.054	0.037	$(0.0050, 0.128)$
λ_2	0.052	0.042	$(0.0022, 0.131)$	0.049	0.038	$(0.0020, 0.122)$
λ_3	0.053	0.045	$(0.0016, 0.135)$	0.049	0.039	$(0.0018, 0.124)$
λ_4	0.051	0.046	$(0.0020, 0.133)$	0.046	0.037	$(0.0031, 0.116)$
λ_5	0.111	0.100	$(0.0060, 0.292)$	0.098	0.080	$(0.0058, 0.251)$
λ_6	0.019	0.019	$(0.0006, 0.053)$	0.017	0.015	$(0.0004, 0.045)$
λ_7	0.032	0.034	$(0.0001, 0.092)$	0.028	0.027	$(0.0005, 0.078)$
λ_8	0.257	0.507	$(0.0003, 0.882)$	0.207	0.347	$(0.0003, 0.704)$

Table 2 shows posterior estimates of all the parameters for the linear and quadratic trajectory model using $J = 8$, assuming that Σ_0 is unstructured and $\rho \neq 0$. For both the linear and quadratic trajectory models the posterior mean of β_1 is negative and similar in magnitude for both models. The negative value of β_1 indicates that increased (IgG, IgM) levels are associated with longer relapse-free survival. For the quadratic trajectory model, the posterior mean and standard deviation (SD) of β_1 are -0.256 and 0.195, respectively. Figure 4 also gives the marginal posterior distribution of β_1 under the quadratic trajectory model. The posterior mode is -0.245. The 95% Highest Posterior Density (HPD) interval

for β_1 is $(-0.648, 0.118)$. Although the 95% HPD interval for β_1 includes 0, Figure 4 indicates that $P(\beta_1 < 0|D) = 0.85$, and thus 85% of the area under the density is to the left of 0. This confirms that there is indeed a moderate to strong association between the (IgG, IgM) measures and relapse-free survival. This is an interesting finding since it has long been conjectured that increased antibody response is perhaps associated with longer relapse-free survival in melanoma. This phenomenon is also confirmed in the posterior hazard plots given in Figure 1 where as the antibody titers increase, there is a decrease in the posterior hazard estimate for each individual. The posterior mean of ρ for the quadratic trajectory model is 0.603 with 95% HPD interval (0.507, 0.695). The estimate of ρ indicates a positive correlation between the (IgG, IgM) measures, implying that these measures act in concert, that is, a higher IgG implies a higher IgM, and the correlation between them is fairly strong. Also, the posterior means of σ_1^2 and σ_2^2 are quite different: 4.123 and 6.787, respectively. In Table 2 the parameter estimates are similar for the linear and quadratic trajectory models, indicating that they are fairly robust with respect to the choice of trajectory. The posterior estimate of μ_{03} is -0.344 with 95% HPD interval $(-1.229, 0.659)$. This includes 0, which perhaps partially explains why the linear and quadratic trajectories give similar parameter estimates in Table 2.

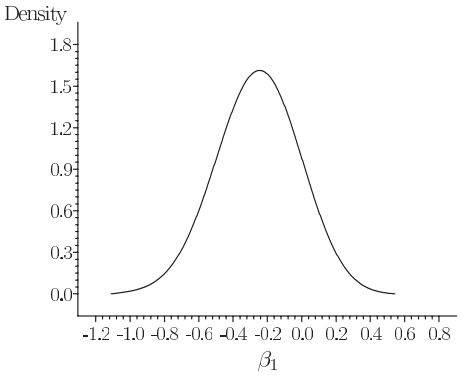

Figure 5. Marginal posterior density for β_1.

We next investigate the importance of modeling ρ in (2.3). Table 3 shows L measure values $(\hat{L}/10^4)$ based on several models assuming an unstructured Σ_0 and $J = 8$. These include models with linear and quadratic trajectories, models with $\rho = 0$ and $\rho \neq 0$, and models with $\beta_1 = 0$ and $\beta_1 \neq 0$. A value of $\beta_1 = 0$ corresponds to fitting a proportional hazards model to the survival data, "ignoring " the latent immune response information in the (IgG, IgM) antibody measures. Table 3 shows that for all values of ν, the quadratic trajectory model has a substantially smaller L measure value compared to the linear trajectory

model, regardless of the value of β_1 or ρ. Models with $\rho \neq 0$ have much smaller L measure values than the corresponding models which assume $\rho = 0$, for both the linear and quadratic trajectories. This clearly indicates the need for modeling a non-zero correlation between the error terms as given in (2.3). Moreover, models in which $\beta_1 \neq 0$ have smaller L measure values than models assuming $\beta_1 = 0$. This is appealing since it further demonstrates the importance of the association between the (IgG, IgM) titer levels and relapse-free survival.

Table 3. L measure values for unstructured Σ_0 using $J = 8$.

		Model with $\rho \neq 0$		Model with $\rho = 0$	
ν	Trajectory	$\beta_1 = 0$	$\beta_1 \neq 0$	$\beta_1 = 0$	$\beta_1 \neq 0$
0.2	linear	162.49	160.05	242.11	242.99
	quadratic	154.21	149.91	207.78	207.10
0.5	linear	277.01	270.28	422.92	418.55
	quadratic	266.12	256.55	366.38	359.77
1.0	linear	557.65	538.87	882.78	861.07
	quadratic	544.79	520.71	774.60	748.98

Table 4. L measure values for diagonal Σ_0 using $J = 8$.

		Model with $\rho \neq 0$		Model with $\rho = 0$	
ν	Trajectory	$\beta_1 = 0$	$\beta_1 \neq 0$	$\beta_1 = 0$	$\beta_1 \neq 0$
0.2	linear	170.76	167.46	246.76	246.69
	quadratic	171.19	166.75	216.84	214.06
0.5	linear	287.93	279.71	416.87	411.95
	quadratic	288.99	278.88	361.74	353.48
1.0	linear	570.53	549.27	833.10	814.15
	quadratic	573.41	548.44	711.17	688.04

Table 4 shows results based on an analysis assuming a diagonal form of Σ_0 in (2.6), with noninformative inverse gamma priors for σ_{0j}^2, $j = 1, 2, 3$. One striking feature here is that, according to the L measure, the fit of the diagonal Σ_0 model is much worse than that of the unstructured Σ_0 model. Somewhat similar trends in Table 4 are observed as in Table 3. However, we do see some differences. There is not much difference in the fits between the linear and the quadratic trajectory models when $\rho \neq 0$, and even for $\rho = 0$ the differences are not as dramatic as in Table 3. Also, when $\beta_1 = 0$ and $\rho \neq 0$, the linear trajectory model actually fits slightly better than the quadratic trajectory model. In summary, the unstructured Σ_0 model fits the data much better than the diagonal Σ_0 model, and it does a much better job of distinguishing between the various models under consideration, including linear vs. quadratic trajectory, $\beta_1 = 0$ vs. $\beta_1 \neq 0$ and

$\rho = 0$ vs. $\rho \neq 0$. Our results for unstructured Σ_0 and a diagonal Σ_0 are consistent with the findings of Heagerty and Kurland (2001).

Table 5. L measure values for several models with $\beta_1 \neq 0$ using $J = 4$.

		Unstructured Σ_0		Diagonal Σ_0	
ν	Trajectory	$\rho \neq 0$	$\rho = 0$	$\rho \neq 0$	$\rho = 0$
0.2	linear	204.50	310.16	213.88	315.59
	quadratic	193.67	257.93	219.20	272.04
0.5	linear	354.33	549.58	366.86	540.87
	quadratic	338.32	458.28	375.66	457.80
1.0	linear	724.79	1163.50	739.27	1096.85
	quadratic	700.37	976.10	757.19	907.87

We also conducted a sensitivity analysis on J. Table 5 shows L measure values for several models, including the unstructured and diagonal Σ_0 models. The L measure values based on $J = 4$ are much larger than the L measure values in Tables 3 and 4 based on $J = 8$, implying that $J = 8$ gives a better model fit in general, regardless of the structure of Σ_0. Models with $\rho \neq 0$ have much smaller L measure values than models which assume $\rho = 0$ for both the linear and quadratic trajectories, consistent with the results of Tables 3 and 4. In addition, the posterior estimates of β_1, β_2 and (α_0, α_1) were robust with respect to the choice of J, yielding similar estimates to those of Table 2 for several different values of J. As a result, the posterior hazard and trajectory function estimates were also robust with respect to the choice of J. In addition to the treatment covariate, we also considered other covariates, such as gender, age and weight, in the model given in (2.7). We found that, except for the treatment covariate, all of the other covariates under consideration were highly insignificant. However, with the other covariates included in (2.7), we still obtained L measure values for the linear and quadratic trajectories with $J = 4$ and $J = 8$ that are similar to those reported in Tables 3 and 4. In the above analyses, we chose the subintervals $(c_{j-1}, c_j]$ with equal numbers of failures or censored observations. Different constructions of $(c_{j-1}, c_j]$, such as those with approximately equal lengths subject to the restriction that at least one failure occurs in each interval, were also considered. The posterior estimates were fairly robust with respect to these different constructions.

In the E2696 data, many patients had zero (IgG, IgM) antibody titer measures throughout the course of vaccine treatment and, of course, most patients had zero (IgG, IgM) antibody titer measures at baseline. This was an ELISA assay, and the zero (IgG, IgM) measures are "exact" in the sense that they are not values representing truncation or censoring due to lower detection limits in

the assay. Moreover, it is well established fact that, in this patient population, approximately 3−5% of all patients have naturally occurring antibodies, that is, their baseline IgG and IgM measures will *not* be 0. Thus, in our data this phenomenon did in fact occur, and it has nothing to do with detection limits of the assay. Those patients that had IgG or IgM titers greater than 0 at baseline are those with naturally occurring antibodies.

To further examine robustness of the model, we carried out an analysis omitting the four cases with RFS values beyond twelve months. The resulting posterior estimates for most of the parameters are quite similar to those of Table 2, and thus, those four cases do not appear to be influential. The only noticeable change in the estimates is in β_1, where the posterior mean, the standard deviation, and the 95% HPD interval for β_1 change to -0.404, 0.290 and $(-0.979, 0.149)$, respectively. Still the 95% HPD interval contains 0 and we conclude that the omission of these four cases does not affect the interpretations of the results and the conclusions of this analysis.

In the computations, 50,000 Gibbs samples were used to compute all posterior estimates using a burn-in of 1,000 samples. The convergence of the Gibbs sampler was checked using several diagnostic procedures as recommended by Cowles and Carlin (1996), and we found that the Gibbs sampler converged before 1,000 iterations. To ease the computations, the L measure was computed by averaging the antibody titer measurements over the five time points. The computations of all posterior quantities and the multivariate L measure took about 30 minutes on a COMPAQ XP1000 unix workstation. The computer code was written in FORTRAN 77 using double precision accuracy. Software for fitting this model is readily available on the second authors website at http://merlot.stat.uconn.edu/∼mhchen.

6. Discussion

The results here suggest that vaccines such as GMK or treatment combinations involving GMK, may be as (or perhaps more) efficacious as standard chemotherapies in improving relapse free survival and survival, and have much less toxicity. However, since the sample size for this study was not large ($n = 70$), one has to interpret these results with great caution. One other important issue is whether the (IgG, IgM) measures are associated with overall survival (i.e., time from randomization to death). Since a cancer treatment is ultimately judged on its efficacy for overall survival, this is a critical issue that still needs a detailed investigation. There were only seven deaths on the combined treatment arms at the time of this E2696 analysis, and therefore not enough events had occurred for any meaningful analyses relating (IgG, IGM) to overall survival. This is a topic of current investigation. The results presented here for RFS, however, are

encouraging enough to at least warrant the investigation of a larger phase III trial in order to more accurately assess the efficacy of GMK and other related vaccines in this patient population. Such studies are currently being investigated in ECOG.

Acknowledgements

The authors wish to thank the Editor, an associate editor and a referee for several suggestions which have greatly improved the paper. Dr. Ibrahim's research was partially supported by NIH grants #CA 70101 and #CA 74015, Dr Chen's research was partially supported by NSF grant #DMS-9702172 and NIH grants #CA 70101 and #CA 74015, and Dr. Sinha's research was partially supported by NCI grant #R29-CA69222.

Appendix: Computational Development

Although, the joint posterior distribution of $\boldsymbol{\theta}$ does not have a closed form, the conditional posteriors do, and thus implementation of the Gibbs sampler is straightforward. Let D denote the data, and "rest" denote the remaining parameters.

We have the following.

(1) $[\lambda_j|D, \text{rest}] \sim \text{gamma}(d_j + a_j, B_{+j} + b_j)$, where $B_{+j} = \sum_{i=1}^{n} B_{ij}$.

(2) $[\Sigma^{-1}|D, \text{rest}] \sim W_2(n_0 + \sum_{i=1}^{n} m_i,\ [Q_0^{-1} + \sum_{i=1}^{n}\sum_{l=1}^{m_i} (y_{il1} - g_{\boldsymbol{\gamma}_i}(A_{il}), y_{il2} - \alpha_0 - \alpha_1 g_{\boldsymbol{\gamma}_i}(A_{il}))' (y_{il1} - g_{\boldsymbol{\gamma}_i}(A_{il}), y_{il2} - \alpha_0 - \alpha_1 g_{\boldsymbol{\gamma}_i}(A_{il}))]^{-1})$.

(3) $[\alpha_0|D, \text{rest}] \sim N(\mu_{\alpha_0}, \sigma^2_{\alpha_0})$, where $\mu_{\alpha_0} = ((m_3/v_3^2) + (\sigma_2^2(1-\rho^2))^{-1} \sum_{i=1}^{n}\sum_{l=1}^{m_i} (y_{il2} - \alpha_1 g_{\boldsymbol{\gamma}_i}(A_{il}) - (\rho\sigma_2/\sigma_1)(y_{il1} - g_{\boldsymbol{\gamma}_i}(A_{il}))))\sigma^2_{\alpha_0}$ and $\sigma^2_{\alpha_0} = ((v_3^2)^{-1} + \sum_{i=1}^{n} (m_i/\sigma_2^2(1-\rho^2)))^{-1}$.

(4) $[\alpha_1|D, \text{rest}] \sim N(\mu_{\alpha_1}, \sigma^2_{\alpha_1})$, where $\mu_{\alpha_1} = \sigma^2_{\alpha_1}((m_4/v_4^2) + (\sigma_2^2(1-\rho^2))^{-1} \sum_{i=1}^{n} \sum_{l=1}^{m_i} g_{\boldsymbol{\gamma}_i}(A_{il})(y_{il2} - \alpha_0 - ((\rho\sigma_2)/\sigma_1)(y_{il1} - g_{\boldsymbol{\gamma}_i}(A_{il}))))$ and $\sigma^2_{\alpha_1} = ((v_4^2)^{-1} + (\sigma_2^2(1-\rho^2))^{-1} \sum_{i=1}^{n}\sum_{l=1}^{m_i} [g_{\boldsymbol{\gamma}_i}(A_{il})]^2)^{-1}$.

(5) $[\boldsymbol{\mu}_0|D, \text{rest}]$ is a multivariate normal for the polynomial trajectory $g_{\boldsymbol{\gamma}_i}(A_{il})$, and the derivation of mean and variance-covariance matrix depends on the form of the trajectory.

(6) $[\Sigma_0|D, \text{rest}]$ is an inverse Wishart distribution for the polynomial trajectory $g_{\boldsymbol{\gamma}_i}(A_{il})$.

(7) The conditional posterior density of β_1 and $\boldsymbol{\beta}_2$ is log-concave in each component of β_1 and $\boldsymbol{\beta}_2$.

(8) The conditional posterior density of $\boldsymbol{\gamma}$ is log-concave in each component of $\boldsymbol{\gamma}$ for the polynomial trajectory $g_{\boldsymbol{\gamma}_i}(A_{il})$.

Thus, for (1)−(6), the generation is straightforward, while for (7) and (8) we can use an adaptive rejection algorithm, since the corresponding conditional posterior densities are log-concave.

References

Brown, E. R. and Ibrahim, J. G. (2003). A Bayesian semiparametric joint hierarchical model for longitudinal and survival data. *Biometrics* **59**, 221-228.

Cowles, M. K. and Carlin, B. P. (1996). Markov chain Monte Carlo convergence diagnostics: a comparative review. *J. Amer. Statist. Assoc.* **91**, 883-904.

DeGruttola, V. and Tu, X. M. (1994). Modeling progression of CD4-lymphocyte count and its relationship to survival time. *Biometrics* **50**, 1003-1014.

Faucett, C. J. and Thomas, C. C. (1996). Simultaneously modelling censored survival data and repeatedly measured covariates: a Gibbs sampling approach. *Statistics in Medicine* **15**, 1663-1685.

Heagerty, P. and Kurland, B. F. (2001). Misspecified maximum likelihood estimates and generalized linear mixed models. *Biometrika* **88**, 973-985.

Henderson, R., Diggle, P. and Dobson, A. (2000). Joint modelling of longitudinal measurements and event time data. *Biostatistics* **1**, 465-480.

Ibrahim, J. G. and Chen, M.-H. (1997). Predictive variable selection in the multivariate linear model. *Biometrics* **53**, 465-478.

Ibrahim, J. G., Chen, M.-H. and Sinha, D. (2001). Criterion-based methods for Bayesian model assessment. *Statist. Sinica* **11**, 419-443.

Ibrahim, J. G. and Laud, P. W. (1994). A predictive approach to the analysis of designed experiments. *Journal of the American Statistical Association* **89**, 309-319.

Kirkwood, J. M., Ibrahim, J. G., Sondak, V. K., Richards, J., Flaherty, L. E., Ernstoff, M. S., Smith, T. J., Rao, U., Steele, M. and Blum, R. H. (2000). The role of high- and low-dose interferon alfa-2b in high-risk melanoma: first analysis of intergroup trial E1690/S9111/C9190. *J. Clinical Oncology* **18**, 2444-2458.

Laud, P. W. and Ibrahim, J. G. (1995). Predictive model selection. *J. Roy. Statist. Soc. Ser. B* **57**, 247-262.

Song, X., Davidian, M. and Tsiatis, A. A. (2002). An estimator for the proportional hazards model with multiple longitudinal covariates measured with error. *Biostatistics* **3**, 511-528.

Taylor, J. M. G., Cumberland, W. G. and Sy, J. P. (1994). A stochastic model for analysis of longitudinal AIDS data. *J. Amer. Statist. Assoc.* **89**, 727-736.

Tsiatis, A. A., DeGruttola, V. and Wulfsohn, M. S. (1995). Modeling the relationship of survival to longitudinal data measured with error: applications to survival and CD4 counts in patients with AIDS. *J. Amer. Statist. Assoc.* **90**, 27-37.

Wang, Y. and Taylor, J. M. G. (2001). Jointly modelling longitudinal and event time data, with applications to AIDS studies. *J. Amer. Statist. Assoc.* **96**, 895-905.

Wulfsohn, M. S. and Tsiatis A. A. (1997). A joint model for survival and longitudinal data measured with error. *Biometrics* **53**, 330-339.

Xu, J. and Zeger, S. L. (2001a). Joint analysis of longitudinal data comprising repeated measures and times to events. *Appl. Statist.* **50**, 375-387.

Xu, J. and Zeger, S. L. (2001b). The evaluation of multiple surrogate endpoints. *Biometrics* **57**, 81-87.

Department of Biostatistics, UNC School of Public Health, University of North Carolina, Chapel Hill, NC 27599, U.S.A.

E-mail: ibrahim@bios.unc.edu

Department of Statistics, University of Connecticut, 215 Glenbrook Road, U-4120, Storrs, CT 06269-4120, U.S.A.

E-mail: mhchen@stat.uconn.edu

Department of Biostatistics and Epidemiology, Medical University of South Carolina, U.S.A.
E-mail: sinhad@musc.edu

(Received October 2002; accepted June 2003)