Session 3: Joint Models for Longitudinal and Survival Data

M.-H. Chen

Overview

In this session, we will discuss Bayesian methods for modeling longitudinal and survival data.

- Joint models for survival and longitudinal data have recently become quite popular in cancer and AIDS clinical trials, where a longitudinal biologic marker such as CD4 count or immune response to a vaccine can be an important predictor of survival.

- Joint models for survival and longitudinal data are also commonly used in quality of life studies, where it is on interest to examine the relationship between a patient's quality of life and a time-to-event.

- Often in clinical trials where the primary endpoint is time to an event, patients are also monitored longitudinally with respect to one or more biologic endpoints throughout the follow-up period. This may be done by taking immunologic or virologic measures in the case of infectious diseases or perhaps with a questionnaire assessing

the quality of life after receiving a particular treatment.

- Often these longitudinal measures are incomplete or may be prone to measurement error. These measurements are also important because they may be predictive of survival.

- Therefore methods which can model both the longitudinal and the survival components jointly are becoming increasingly essential in most cancer and AIDS clinical trials.

- In this part of the session, we will give a detailed development of joint models for longitudinal and survival data, and discuss frequentist as well as Bayesian techniques for fitting such models. We will discuss various approaches to model development, and lay out the computational implementation in detail. Examples from AIDS and cancer vaccine trials will be presented.

M.-H. Chen

3.1 Introduction

♠ **Joint Modeling in AIDS Studies**

- In clinical trials of therapies for diseases associated with human immunodeficiency virus (HIV), immunologic and virologic markers are measured repeatedly over time on each patient. The interval lengths vary between data collection times and missing data is quite common.

- These markers are prone to measurement error and high within patient variability due to biological fluctuations. Modeling these covariates over time is preferable to using the raw data as noted by many.

- In addition, models provide estimates for time points where data are not available.

M.-H. Chen

- Many HIV clinical trials focus on the opportunistic infections (OI) associated with HIV disease where the survival endpoint is the time to development of the OI. In these trials, immunologic and virologic markers might be utilized as time-varying predictor variables.

- The most common measure used to assess immunological health of an HIV patient is the CD4+ lymphocyte count, or CD4 count for short.

- Higher CD4 counts indicate a stronger immune system that is more prepared to resist infection. Lower CD4 counts indicate a higher risk of an OI.

- Viral load is a measure of the amount of virus in the blood plasma. A lower viral load is preferable and may indicate successful treatment of the disease. A patient's success on treatment is often evaluated by these two markers.

M.-H. Chen

- When a patient begins a successful treatment regimen, the viral load may drop drastically and fall below a detectable level. The CD4 count may take longer to respond or may not respond at all.

- As viral load decreases, we may expect the CD4 count to increase as the immune system has time to recover. However, CD4 count is slower to respond than viral load.

- Because of this complex relationship between the immunologic and virologic markers, we may want a multivariate model for the longitudinal covariates.

M.-H. Chen

♠ Joint Modeling in Cancer Vaccine Trials

- In cancer vaccine (immunotherapy) trials, vaccinations are given to patients to raise the patient's antibody levels against the tumor cells.

- In these studies, the time-to-event endpoint is often time to disease progression or time to death. A successful vaccine activates the patient's immune system against future tumor growth.

- In this case, a patient's antibody production increases, indicating an increase in the bodies' immune strength. Therefore, measuring these antibodies helps the clinician to evaluate the immunity level.

- The primary measures of antibody response for many cancers are the IgG and IgM antibody titres. The levels of these markers are conjectured to be associated with the clinical outcome and are therefore monitored during follow-up.

- These markers are prone to measurement error; therefore, the raw data should not be used as covariates in a survival analysis. A method which jointly models the longitudinal marker as well as the survival outcome is therefore necessary.

♠ **Joint Modeling in Health-Related Quality of Life Studies**

- The collection of quality of life (QOL) data in clinical trials has become increasingly common, particularly when the survival benefit of a treatment is anticipated to be small or modest. In fact, one might argue that for a patient, quality of life is at times an even more important factor in treatment decisions than any modest survival benefit.

- Although this type of data provides much useful information for the decision-making process of both patient and physician, the challenges encountered in the collection and analysis of QOL data make it hard to provide meaningful statements about QOL differences by treatment.

M.-H. Chen

- A QOL survey instrument is typically administered to study participants at a number of prespecified time points during treatment and follow-up.

- Complete QOL data for patients at all of the specified collection times is frequently unavailable due to adverse events such as treatment toxicities or disease progression.

- Patients who are very ill when they report to the clinic may be less likely to complete the QOL instrument, and clinic personnel may feel that it is unethical to ask a patient to complete such a form when the patient feels so poorly.

- Therefore it is quite plausible that the missingness of QOL data is related to the patient's QOL at the assessment time, and strong evidence that QOL data is generally not missing at random has become accepted in the literature.

M.-H. Chen

- It is well known that such nonignorable missingness often leads to serious biases and must be taken into account at the time of analysis.

- These considerations lead to the development of joint models for longitudinal and survival data, where the longitudinal measure is QOL, and the survival component of the model acts as a type of nonignorable missing data mechanism.

- One such issue is that missingness of QOL data is often not monotone, yielding observations with nonignorable intermittent missingness. Another important issue is that a subject's QOL data is frequently subject to informative censoring by a terminal event such as death or disease progression.

M.-H. Chen

♠ **Recent Work on Bayesian Joint Modeling of Survival and Longitudinal Data**

- Brown, E.R., Ibrahim, J.G., and DeGruttola, V. (2005), "A Flexible B-Spline Model for Multiple Longitudinal Biomarkers and Survival," *Biometrics*, **61**, 65-74.

- Chen, M.-H., Ibrahim, J.G., and Sinha, D. (2004), "A New Joint Model for Longitudinal and Survival Data with a Cure Fraction," *Journal of Multivariate Analysis*, **91**, 18-34.

- Chi, Y. and Ibrahim, J.G. (2004), "A New Class of Joint Models for Longitudinal and Survival Data Accomodating Zero and Non-Zero Cure Fractions: A Case Study of an International Breast Cancer Study Group Trial," *Technical Report*, Department of Biostatistics, University of North Carolina at Chapel Hill.

- Ibrahim, J.G., Chen, M.-H., and Sinha, D. (2004), "Bayesian Methods for Joint Modeling of Longitudinal and Survival Data with Applications to Cancer Vaccine Studies," *Statistica Sinica*, **14**, 863-883.

- Brown, E.R. and Ibrahim, J.G. (2003), "A Bayesian Semiparametric Joint Hierarchical Model for Longitudinal and Survival Data," *Biometrics*, **59**, 221-228.

- Brown, E.R. and Ibrahim, J.G. (2003), "Bayesian Approaches to Joint Cure Rate and Longitudinal Models with Applications to Cancer Vaccine Trials," *Biometrics*, **59**, 686-693.

- Wang, Y. and Taylor, J. M. G. (2001), "Jointly modelling longitudinal and event time data, with applications to AIDS studies," *Journal of the American Statistical Association* **96**, 895-905.

M.-H. Chen

3.2 Methods for Longitudinal and Survival Data

- Often in time-to-event studies, patients are monitored throughout the study period with biologic measurements taken to evaluate health status. Statistical packages are widely available to perform survival analyses with time-dependent covariates.

- However, if the covariates are measured with error, the analysis becomes more complex. Simply including the raw measurements in the survival analysis leads to bias.

M.-H. Chen

♠ Partial Likelihood Models

- The hazard function of survival time with time-dependent covariates $\mathcal{X}^*(t)$ is generally expressed as

$$h(t|\mathcal{X}^*(t)) = \lim_{\delta \to 0} \frac{1}{\delta} P(t \leq T \leq t + \delta | T \geq t, \mathcal{X}^*(t)),$$

  where $\mathcal{X}^*(t)$ is the covariate history up to time t and T is the true survival time.

- In the presence of right censoring we only observe $Y = \min(T, C)$, where C is a potential censoring time and the failure indicator ν, which is equal to 1 if the individual is observed to fail $(T \leq C)$, and 0 otherwise.

- Therefore, we can write the hazard for those who fail as

$$\lim_{\delta \to 0} \frac{1}{\delta} P(t \leq Y \leq t + \delta, \nu = 1 | Y \geq t, \mathcal{X}^*(t)).$$

M.-H. Chen

- The proportional hazards model relates the hazard to time-dependent covariates,

$$h(t|\mathcal{X}^*(t)) = h_0(t)\varphi(\mathcal{X}^*(t), \boldsymbol{\beta}),$$

where $\varphi(\mathcal{X}^*(t), \boldsymbol{\beta})$ is a function of the covariate history specified up to an unknown parameter or vector of parameters $\boldsymbol{\beta}$.

- Leaving the underlying baseline hazard $h_0(t)$ unspecified, one approach for estimating $\boldsymbol{\beta}$ is to maximize the Cox's partial likelihood

$$\prod_{i=1}^{n} \left[\varphi(\mathcal{X}_i^*(y_i), \boldsymbol{\beta}) / \sum_{j=1}^{n} \varphi(\mathcal{X}_j^*(y_i), \boldsymbol{\beta}) I(y_j \geq y_i) \right]^{\nu_i}, \qquad (1)$$

where $I(y_j \geq y_i)$ is the indicator function so that $I(y_j \geq y_i) = 1$ if $y_j \geq y_i$ and 0 otherwise, $\mathcal{X}_i^*(t)$ is the covariate history for the ith case, y_i is the observed survival time, and ν_i is the censoring indicator, where $\nu_i = 1$ if y_i is a failure time and $\nu_i = 0$ otherwise, for $i = 1, 2, \ldots, n$.

M.-H. Chen

- In our present setting, the true covariate history, $\mathcal{X}^*(t)$, is not available. However, we may have observations, $X(t)$, representing some function of the true covariate, $X^*(t)$, which we refer to here as the *trajectory function*.

- Tsiatis, DeGruttola, and Wulfsohn (1995) present a computationally straightforward and easy-to-implement approach which reduces the bias in a model with time-varying covariates measured with error.

- They use asymptotic approximations to show consistency of estimates for modeling the longitudinal data separately, then plugging the estimates into a Cox proportional hazards model. Estimation and inference for the survival model are carried out using the partial likelihood theory.

- In the case where the trajectory function $X^*(t)$ and $X(t)$ have the same dimension, they specify the longitudinal model as

$$X(t) = X^*(t) + \epsilon(t),$$

  where $\epsilon(t)$ is measurement error with $E(\epsilon(t)) = 0$, $\mathrm{Var}(\epsilon(t)) = \sigma^2$ and $\mathrm{Cov}(\epsilon(t_1), \epsilon(t_2)) = 0$, $t_1 \neq t_2$, and $X^*(t)$ is the trajectory function.

- Letting $\mathcal{X}(t) = \{X(t_1), X(t_2), \ldots, X(t_j); \ t_j \leq t\}$ denote the history of the observed covariate up to time t leads to the hazard

$$h(t|\mathcal{X}(t)) = \int h(t|\mathcal{X}^*(t), \mathcal{X}(t)) dP(\mathcal{X}^*(t)|\mathcal{X}(t), Y \geq t).$$

- Further assumptions that neither the measurement error nor the timing of the visits prior to time t are prognostic yield

$$h(t|\mathcal{X}^*(t), \mathcal{X}(t)) = h(t|\mathcal{X}^*(t)) = h_0(t)\varphi(\mathcal{X}^*(t), \boldsymbol{\beta}).$$

M.-H. Chen

- Combining the previous two expressions results in

$$h(t|\mathcal{X}(t)) = h_0(t)E[\varphi(\mathcal{X}^*(t), \boldsymbol{\beta})|X(t_1), ..., X(t_j), t_j \leq t, Y \geq t]. \quad (2)$$

- Denote the conditional expectation in (2) by $E(t, \boldsymbol{\beta})$. If $E(t, \boldsymbol{\beta})$ were known, we could estimate $\boldsymbol{\beta}$ by maximizing Cox's partial likelihood,

$$\prod_{i=1}^{n} \left[E_i(y_i, \boldsymbol{\beta}) / \sum_{j=1}^{n} E_j(y_i, \boldsymbol{\beta}) I(y_j \geq y_i) \right]^{\nu_i}, \quad (3)$$

  where $E_i(t, \boldsymbol{\beta}) = E[\varphi(\mathcal{X}_i^*(t), \boldsymbol{\beta})|\mathcal{X}_i(t), Y_i \geq t]$, $\mathcal{X}_i^*(t)$ and $\mathcal{X}_i(t)$ denote the histories of true and observed covariates, and Y_i is the observed survival time for $i = 1, 2, \ldots, n$.

- Consider the case when the hazard is only a function of the univariate current value, $X(t)$. For the relative risk formulation of the original Cox model, $\varphi(x, \beta) = e^{x\beta}$. Thus, to compute $E(t, \beta)$ in (2), we must estimate $E[\exp\{\beta\mathcal{X}^*(t)\}|\mathcal{X}(t), Y \geq t]$, which is the moment generating function of the conditional distribution $[\mathcal{X}^*(t)|\mathcal{X}(t), Y \geq t]$.

M.-H. Chen

- Assuming a normal approximation to this conditional distribution, we are led to the moment generating function

$$\exp\{\beta\mu(t|\mathcal{X}(t)) + \beta^2\sigma^2(t|\mathcal{X}(t))/2\},$$

  where

$$\mu(t|\mathcal{X}(t)) = E\{X^*(t)|\mathcal{X}(t), Y \geq t\}$$

  and

$$\sigma^2(t|\mathcal{X}(t)) = \text{Var}\{X^*(t)|\mathcal{X}(t), Y \geq t\}.$$

- At each event time, a new model for the covariate is fit given all the covariate data up to that event time. The fitted value for that time is then plugged into the model.

- For models that are more complicated than the Cox model or the additive relative risk model, Tsiatis, DeGruttola, and Wulfsohn (1995) recommend using the first-order approximation:

$$E(t,\boldsymbol{\beta}) = E[\varphi(X^*(t),\boldsymbol{\beta})|\mathcal{X}(t), Y \geq t)] \approx \varphi(E(X^*(t)|\mathcal{X}(t), Y \geq t),\boldsymbol{\beta}).$$

M.-H. Chen

- This approximation allows any type of model to be used to fit the longitudinal covariates. The conditional expected values of the covariates $X^*(t)$ are simply plugged into (3) to get the maximum likelihood estimates of the regression parameter β in the proportional hazards model.

- It is not clear how appropriate this approximation is or how it can be validated.

- Use of this approximation gives the investigator a method which can be easily implemented using existing software. The analyst simply fits the longitudinal data using an appropriate model, then includes the fitted values from this model in a Cox proportional hazards model.

- Variance estimation follows from calculating the observed information from the partial likelihood function. However, alternative formulations would be more desirable because it would be easier to validate model assumptions and would make more efficient use of the data.

M.-H. Chen

♠ Joint Likelihood Models

- Although the two-stage procedure discussed in the previous section allows for an easy analysis of the data with existing software packages and reduces bias over using the raw covariate data $X(t)$ directly in a Cox model, a modeling approach that makes more efficient use of the data by modeling the outcomes jointly may be more desirable.

- One approach is based on a model for survival conditional on the observed longitudinal covariate with the joint likelihood equal to $f_{Y|X} f_X$. This model is used when time to event is the primary outcome of interest and longitudinal measurements may help predict the outcome.

- The second approach sets the model up as $f_{X|Y} f_Y$. This second approach is more often used in longitudinal studies where one might want to account for time to loss of follow-up.

M.-H. Chen

- DeGruttola and Tu (1994) propose such an approach and extend the general random effects model to the analysis of longitudinal data with informative censoring.

- A similar model for informatively censored longitudinal data was proposed by Schluchter (1992) and Schluchter, Greene, and Beck (2001).

- DeGruttola and Tu (1994) jointly model survival times and disease progression using normally distributed random effects.

- Assuming that these two outcomes are independent given the random effects, the joint likelihood is easily specified.

- The maximum likelihood estimates of the unknown parameters are obtained using the EM algorithm.

- Their model can be described as follows. Consider a sample of n subjects, indexed by i, each of whom has m_i observations of a marker of disease progression. Let X_i be an $m_i \times 1$ vector, whose elements X_{ij} are the observed values of the marker for the ith person on the jth occasion of measurement for $i = 1, 2, \ldots, n$; $j = 1, 2, \ldots, m_i$.

- Let $y_i = \min(t_i, c_i)$, where t_i and c_i denote the survival and censoring times for the ith subject, respectively. Then, the mixed effects model of the disease progression marker is

$$X_i = X_i^* + \epsilon_i \quad \text{and} \quad X_i^* = T_i \alpha + Z_i b_i,$$

where α is a $p \times 1$ vector of unknown parameters, T_i is a known full-rank $m_i \times p$ design matrix, $b_i \sim N(0, \Psi)$ denotes a $k \times 1$ vector of unknown individual effects, the b_i's are i.i.d., Z_i is a known full-rank $m_i \times k$ design matrix, $\epsilon_i \sim N_{m_i}(0, \sigma^2 I_i)$ is a vector of residuals, and I_i is an $m_i \times m_i$ identity matrix.

M.-H. Chen

- The following normal mixed effects model is used for the survival times (or a monotone transformation of survival, as appropriate)

$$y_i = \boldsymbol{w}_i'\boldsymbol{\zeta} + \boldsymbol{\lambda}'\boldsymbol{b}_i + r_i,$$

where $\boldsymbol{\zeta}$ is $q \times 1$ vector of unknown parameters, $\boldsymbol{w}_i = (w_{i1}, w_{i2}, \ldots, w_{iq})'$ is a $q \times 1$ known design matrix, $\boldsymbol{\lambda}$ is a $k \times 1$ vector of unknown parameters, $r_i \sim N(0, \omega^2)$, $\omega > 0$, and y_i is the survival time or some monotonic transformation of survival time such as the log of survival time.

- The longitudinal marker and survival times are independent conditional on the random effects; therefore, the complete data log-likelihood, i.e., the likelihood that would apply if $\boldsymbol{b}_i$ and y_i were observed, is written as

$$l_c = \sum_{i=1}^{n} \log[\phi(X_i|\boldsymbol{b}_i, \boldsymbol{\alpha}, \sigma^2)\phi(\boldsymbol{b}_i|\Psi)\phi(y_i|\boldsymbol{b}_i, \boldsymbol{\zeta}, \omega^2)], \qquad (4)$$

where $\phi(.|.)$ denotes the appropriate normal probability density function.

M.-H. Chen

- Estimation of the unknown parameters is accomplished using the EM algorithm, a technique which iterates between solving for the expected values of functions of the unobserved data (random effects and errors in this case) given the observed data and the maximum likelihood estimates of the parameters until convergence.

- The covariance matrix of the estimates of the parameters of interest (α, ζ) at convergence of the EM algorithm can be obtained by using Louis's formula (Louis, 1982).

- It is not clear how robust this model is to departures from parametric assumptions, especially in the assumption of normality for monotonic transformations of survival times.

- The link between failure times and longitudinal outcomes is not obvious or easily interpreted since it is made through the random effects alone. There is no parameter linking failure time to the longitudinal covariates directly.

M.-H. Chen

- Wulfsohn and Tsiatis (1997) also propose a joint likelihood model. They assume a proportional hazards model for survival conditional on the longitudinal marker.

- A random effects model is used to model the covariate and measurement error. Maximum likelihood estimates of all parameters in the joint model are obtained using the EM algorithm with numeric quadrature and a Newton-Raphson approximation performed at each iteration. The model is then applied to data from an HIV clinical trial.

- Denote by t_i the true survival time for individual i. Suppose we observe $y_i = \min(t_i, c_i)$, where c_i corresponds to a potential censoring time, and the censoring indicator ν_i, which takes 1 if the failure is observed and 0 otherwise.

- Assume the covariate for individual i is measured at time $\boldsymbol{t}_i = (t_{ij} : t_{ij} \leq y_i)$, where t_{ij} is the time from randomization for $j = 1, 2, \ldots, m_i$.

M.-H. Chen

- The random effects model for the longitudinal covariate is

$$X_{ij} = X_{ij}^* + \epsilon_{ij} \tag{5}$$

and

$$X_{ij}^* = b_{0i} + b_{1i}t_{ij}, \tag{6}$$

for $j = 1, 2, \ldots, m_i$, where

$$\boldsymbol{b}_i = (b_{0i}, b_{1i})' \sim N_2(\boldsymbol{b}, \Psi), \quad \epsilon_i \sim N_{m_i}(0, \sigma^2 I_i),$$

$\boldsymbol{b} = (b_0, b_1)'$, $\Psi = (\Psi_{jj^*})$ is a 2×2 unknown covariance matrix, and I_i is an $m_i \times m_i$ identity matrix, for $i = 1, 2, \ldots, n$.

- The hazard function for the proportional hazards model is given by

$$h(t|\boldsymbol{b}_i, X_i, \boldsymbol{t}_i) = h(t|\boldsymbol{b}_i) = h_0(t) \exp\{\beta(b_{0i} + b_{1i}t)\}.$$

M.-H. Chen

- The complete data log-likelihood is given by

$$\sum_{i=1}^{n} \left[\sum_{i=1}^{m_i} \log f(X_{ij}|\boldsymbol{b}_i, \sigma^2) + \log f(\boldsymbol{b}_i|\boldsymbol{b}, \Psi) + \log f(y_i, \nu_i|\boldsymbol{b}_i, h_0, \beta) \right],$$

  where $f(X_{ij}|\boldsymbol{b}_i, \sigma^2)$ and $f(\boldsymbol{b}_i|\boldsymbol{b}, \Psi)$ are the densities of $N(b_{0i} + b_{1i}t_{ij}, \sigma^2)$ and $N(\boldsymbol{b}, \Psi)$ distributions, respectively, and

$$f(y_i, \nu_i|\boldsymbol{b}_i, h_0, \beta) = [h_0(y_i) \exp\{\beta(b_{0i} + b_{1i}y_i)\}]^{\nu_i}$$

$$\times \exp\left[-\int_0^{y_i} h_0(u) \exp\{\beta(b_{0i} + b_{1i}u)\} du \right].$$

- The EM algorithm is used to obtain the parameter estimates.

- To estimate the variance of $\hat{\beta}$ at EM convergence, Wulfsohn and Tsiatis (1997) define the profile score $S_\beta(\hat{\boldsymbol{\theta}}_{-\beta}(\hat{\beta}))$ to be the derivative of the log-likelihood with respect to β evaluated at $\hat{\beta}$.

M.-H. Chen

- The remaining parameters, $\boldsymbol{\theta}_{-\beta}$, are estimated using restricted maximum likelihood estimates which are calculated by using a separate EM algorithm applied to the likelihood and keeping $\hat{\beta}$ fixed.

- They estimate the variance by calculating this score over several values of β, then they fit a line to these estimates and take the negative inverse of the slope of this line as the estimate of the variance.

- This involves implementing the EM algorithm several times to get estimates for the other parameters for each value of β to estimate this line. In their model and application, this linear approximation appeared valid.

- Some advantages of this approach are that it makes more efficient use of the data than the two-stage approach suggested by Tsiatis, DeGruttola, and Wulfsohn (1995), it uses the full likelihood in estimation, and makes a direct link between the survival and longitudinal covariate.

M.-H. Chen

- However, the parametric form of the trajectory function, X_{ij}^*, may be inappropriate in some settings.

- Also, the EM algorithm is slow to converge; therefore, it may not be feasible to extend the model of the trajectory function to the multivariate case. It took two hours to obtain estimates for a dataset with 137 individuals with 24 failures and approximately six covariate measurements after using reasonable starting values for the parameters. There is no mention of the amount of time required to get an estimate for the variance of $\hat{\beta}$.

- Although more complex models can be fit, it is not clear how computationally intensive they would be, or whether the assumption of linearity of $S_\beta(\hat{\boldsymbol{\theta}}_{-\beta}(\hat{\beta}))$ would still be reasonable.

♠ **Example 3.1: AIDS Data**

- Consider a double-blind placebo controlled trial of patients with advanced HIV. Of the total of 281 patients, 144 were randomized to received zidovudine (ZDV), and the remaining 137 patients were given a placebo.

- Measurements of CD4 count were taken prior to treatment and approximately every four weeks while on therapy. The study lasted for 18 weeks.

- The goal of the study was to understand the CD4 trajectories of the patients and to evaluate the strength of the relationship between the CD4 trajectory and survival.

- The unobserved trajectory here is assumed to be a true "biological marker" according to the definition of Prentice (1989) in that the treatment effect on survival is expressed through its effect on the marker, which then affects survival.

M.-H. Chen

- Thus, given the marker trajectory, there is no treatment effect on survival. The data was fit to both this model and the two-stage model proposed by Tsiatis, DeGruttola, and Wulfsohn (1995) in which the unobservable marker trajectory estimates are re-evaluated for each risk time.

- Table 3.1 compares the estimates of the longitudinal covariate from the two models for placebo group results. In Table 3.1, TSM and JM denote the two-stage model and the joint model, respectively. Estimates for the 8th and 16th event times are shown for the two-stage model.

- The parameter which describes the strength of the relationship between CD4 and survival, β, was estimated as -0.3029 compared to the two-stage model which estimated it as -0.284.

- The slope of the fitted line to the profile score for β is estimated to be -41.980, giving an estimate of the standard error of 0.154 in contrast to 0.144 in the two-stage model.

M.-H. Chen

TABLE 3.1. Parameter Estimates for AIDS Data.

Model	Event	b_0	b_1	σ^2
TSM	8th	4.23	-0.0050	0.301
	16th	4.27	-0.0045	0.305
JM		4.17	-0.0047	0.396
Model	Event	Ψ_{11}	Ψ_{12}	Ψ_{22}
TSM	8th	1.18	0.0016	0.000032
	16th	1.15	0.0024	0.000029
JM		1.11	0.0027	0.000014

- They explain that the increase in the variance estimate is due to the random effects being influenced by the uncertainty in the estimated growth curve parameters, therefore incorporating more variability.

- Also, they explain the increase in the estimate of standard error for pure measurement error in joint estimation as due to overfitting in the two-stage model.

M.-H. Chen

♠ Bayesian Methods for Joint Modeling of Longitudinal and Survival Data

- The methods discussed earlier for joint modeling have all been based on a frequentist approach. However, it may be advantageous to take a Bayesian approach to solving this problem.

- In the Bayesian paradigm, asymptotic approximations are not necessary, model assessment is more straightforward, computational implementation is typically much easier, and historical data can be easily incorporated into the inference procedure.

- The Bayesian methods we will review use an approach to building the model as Ibrahim, Chen, and Sinha (2004, Statistica Sinica).

- For their longitudinal model, they assume that two covariates $(X_1(t), X_2(t))$ are observed which both measure the true unobservable univariate antibody measure $(X^*(t))$.

M.-H. Chen

- The model is specified as

$$X_{i1}(t) = X_i^*(t) + \epsilon_{i1}(t) \tag{7}$$

  and

$$X_{i2}(t) = \alpha_0 + \alpha_1 X_i^*(t) + \epsilon_{i2}(t), \tag{8}$$

  where $\epsilon_i(t) = (\epsilon_{i1}(t), \epsilon_{i2}(t))'$,

$$\epsilon_i(t) \sim N_2 \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \right).$$

- Further motivation of the model in (7) and (8) is given in Example 3.2.

- $X_i^*(t)$ is modeled via a trajectory function, $g_{\gamma_i}(t)$, which may be either linear or quadratic in nature.

- The survival times (y_i's) are modeled using a proportional hazards model assuming the random errors of the longitudinal component are not prognostic of the survival time.

M.-H. Chen

- The hazard function is given by

$$h(t|\mathcal{X}_i^*(t^*), \mathcal{X}_i(t^*), z_i) = h(t|\mathcal{X}_i^*(t^*), z_i) = h_0(t)\exp\{\beta_1 X_i^*(t^*) + z_i'\boldsymbol{\beta_2}\},$$

  where $\mathcal{X}_i^*(t^*)$ and $\mathcal{X}_i(t^*)$ denote the histories of X_i^* and X_i up to time t^*, and z_i denotes a $p \times 1$ vector of baseline covariates for subject i.

- We assume that the z_i's are not measured with error.

- In addition, $\boldsymbol{\beta_2}$ is a $p \times 1$ vector of regression coefficients corresponding to z_i. Setting $X_i^*(t) = g_{\boldsymbol{\gamma}_i}(t)$, the quadratic parametric trajectory function, for example, takes on the form $\gamma_{i0} + \gamma_{i1}t + \gamma_{i2}t^2$.

- Let $t_{i1}, t_{i2}, \ldots, t_{im_i}$ denote the times at which the measurements X_{ij} are taken, and let $g_{\boldsymbol{\gamma}_i}(t_{ij})$ denote the trajectory function evaluated at t_{ij}.

- Conditional on the subject-specific trajectory parameters, $\boldsymbol{\gamma}_i = (\gamma_{i0}, \gamma_{i1}, \gamma_{i2})'$, the observed trajectories and survival times are independent.

M.-H. Chen

- Let $\mathbf{X}_{i1} = (x_{i11}, x_{i21}, \ldots, x_{im_i1})'$ and $\mathbf{X}_{i2} = (x_{i12}, x_{i22}, \ldots, x_{im_i2})'$, and let $X_1 = (X_{11}, X_{21}, \ldots, X_{n1})$, and $X_2 = (X_{12}, X_{22}, \ldots, X_{n2})$.

- Also let y_i denote the event time for the ith subject, which may be right censored, and let $\boldsymbol{y} = (y_{1,2}, \ldots, y_n)'$ denote the vector of event times.

- Further, let $\boldsymbol{\nu} = (\nu_1, \nu_2, \ldots, \nu_n)'$ denote the vector of censoring indicators, where $\nu_i = 1$ indicates a failure and $\nu_i = 0$ indicates a right censored observation, and $\boldsymbol{z} = (\boldsymbol{z}_1', \boldsymbol{z}_2', \ldots, \boldsymbol{z}_n')'$.

- We take $h_0(t)$ to be a constant λ_j over the time intervals $I_j = (s_{j-1}, s_j]$, for $j = 1, 2, \ldots, J$, where $s_0 = 0 < s_1 < \ldots < s_J < s_{J+1} = \infty$.

- 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 (X_1, X_2), and L_2 is the likelihood function for $\boldsymbol{y}$.

M.-H. Chen

- 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

$$L(\boldsymbol{\theta}) = L_1(\boldsymbol{\gamma}, \alpha_0, \alpha_1, \Sigma | X_1, X_2) L_2(\boldsymbol{\lambda}, \beta_1, \boldsymbol{\beta}_2, \boldsymbol{\gamma} | \boldsymbol{y}, \boldsymbol{\nu}, \boldsymbol{z}).$$

- The contribution of the longitudinal component to the likelihood is

$$L_1(\boldsymbol{\gamma}, \alpha_0, \alpha_1, \Sigma | X_1, X_2)$$

$$\propto \prod_{i=1}^{n} \left[|\Sigma|^{-m_i/2} \exp\left\{ -\frac{1}{2} \sum_{j=1}^{m_i} \left((x_{ij1} - g_{\boldsymbol{\gamma}_i}(t_{ij}), x_{ij2} - (\alpha_0 + \alpha_1 g_{\boldsymbol{\gamma}_i}(t_{ij})) \right) \right. \right.$$

$$\left. \left. \times \Sigma^{-1} (x_{ij1} - g_{\boldsymbol{\gamma}_i}(t_{ij}), x_{ij2} - (\alpha_0 + \alpha_1 g_{\boldsymbol{\gamma}_i}(t_{ij}))' \right) \right\} \right],$$

- The contribution of the survival component to the likelihood is

$$L_2(\boldsymbol{y}, \boldsymbol{\nu}, \boldsymbol{z} | \boldsymbol{\lambda}, \beta_1, \boldsymbol{\beta}_2, \boldsymbol{\gamma}) \propto \left\{ -\sum_{j=1}^{J} \sum_{i=1}^{n} \lambda_j B_{ij} \right\} \left(\prod_{j=1}^{J} \lambda_j^{d_j} \right)$$

$$\times \exp\left\{ \sum_{j=1}^{J} \sum_{i=1}^{n} \nu_i (\beta_1 g_{\boldsymbol{\gamma}_i}(t_{ij}^*) + \boldsymbol{\beta}_2' \boldsymbol{z}_i) \right\}, \qquad (9)$$

M.-H. Chen

where d_j is the number of failures in the kth time interval, t_{ij}^* is the the most recent time at which a measurement was taken, and
$M = \sum_{i=1}^{n} m_i$.

- In (9), 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 $j_{i1} = \max\{l : A_{il}^* \leq c_{j-1}\}$ and $j_{i2} = \max\{l : A_{il}^* \leq c_j\}$, where A_{il}^* is the rescaled A_{il} so that A_{il}^* has the same unit as t_i, then if $j_{i1} = j_{i2}$, $B_{ij} = (c_j - c_{j-1}) \exp\{\beta_1 g_{\gamma_i}(A_{ij_{i1}}) + z_i'\beta_2\}$, and if $j_{i1} < j_{i2}$,
$B_{ij} = (A_{i,j_{i1}+1}^* - c_{j-1}) \exp\{\beta_1 g_{\gamma_i}(A_{i,j_{i1}+1}) + z_i'\beta_2\} + \sum_{l=j_{i1}+1}^{j_{i2}}(A_{i,l+1}^* - A_{il}^*) \exp\{\beta_1 g_{\gamma_i}(A_{il}) + z_i'\beta_2\} + (c_j - A_{ij_{i2}}^*) \exp\{\beta_1 g_{\gamma_i}(A_{ij_{i2}}) + z_i'\beta_2\}$.

(iii) If $c_{j-1} < t_i \leq c_j$, using j_{i1} and j_{i2} given in (ii), then if $j_{i1} = j_{i2}$ or $t_i \leq A_{i,j_{i1}+1}^*$, when $j_{i1} < j_{i2}$,
$B_{ij} = (t_i - c_{j-1}) \exp\{\beta_1 g_{\gamma_i}(A_{ij_{i1}}) + z_i'\beta_2\}$, and otherwise, we define
$B_{ij} = (A_{i,j_{i1}+1}^* - c_{j-1}) \exp\{\beta_1 g_{\gamma_i}(A_{i,j_{i1}+1}) + z_i'\beta_2\} + \sum_{l=j_{i1}+1}^{k_i}(A_{i,l+1}^* - A_{il}^*) \exp\{\beta_1 g_{\gamma_i}(A_{il}) + z_i'\beta_2\} + (t_i - A_{ik_i}^*) \exp\{\beta_1 g_{\gamma_i}(A_{ik_i}) + z_i'\beta_2\}$,
where $j_{i1} + 1 \leq k_i \leq j_{i2}$ is chosen so that $A_{ik_i}^* < t_i \leq A_{i,k_i+1}^*$.

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

M.-H. Chen

- In the likelihood function (9), we have invoked an approximation, that is,

$$\int_0^{y_i} h_0(u) \exp\left\{\beta_1 g_{\gamma_i}(u) + \boldsymbol{\beta}_2' \boldsymbol{z}_i\right\} \, du \; \approx \; \sum_{i=1}^{n} \sum_{j=1}^{J} \lambda_j B_{ij}.$$

- The priors are all chosen to be noninformative with $\lambda_j \sim \mathcal{G}(a_j, b_j)$, $j = 1, 2, \ldots, J$, $\beta_1 \sim N(\zeta_1, v_1^2)$, $\beta_2 \sim N(\zeta_2, v_2^2)$ (for $p = 1$), $\alpha_0 \sim N(\zeta_3, v_3^2)$, $\alpha_1 \sim N(\zeta_4, v_4^2)$, $\Sigma^{-1} \sim W_2(n_0, Q_0)$, 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*.

M.-H. Chen

- In Example 3.2, we take $n_0 = 3$ and $Q_0^{-1} = 0.001 I_2$, where I_2 is the 2-dimensional identity matrix, and $\boldsymbol{\gamma}_i \overset{iid}{\sim} N_3(\boldsymbol{\mu}_0, \Sigma_0)$, where $\boldsymbol{\mu}_0 = (\mu_{01}, \mu_{02}, \mu_{03})$, and Σ_0 represents the variation in the parameters.

- Additional normal and inverse Wishart priors are placed on $\boldsymbol{\mu}_0$ and Σ_0, respectively. That is, $\mu_{01} \sim N(\xi_{01}, v_{01}^2)$, $\mu_{02} \sim N(\xi_{02}, v_{02}^2)$, $\mu_{03} \sim N(\xi_{03}, v_{03}^2)$, $\Sigma_0^{-1} \sim W_3(n_0^*, Q_0^*)$, and n_0^* and Q_0^* are chosen to make the prior noninformative.

M.-H. Chen

♠ **Example 3.2: Cancer Vaccine Data E2696**

- We consider a phase II melanoma clinical trial conducted by the Eastern Cooperative Oncology Group (ECOG), labeled here as E2696.

- Two treatment arms are used in our analysis here. 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 total sample size of $n = 70$ patients.

- The survival endpoint is relapse-free survival (RFS), measured in months. There were a total of 27 completely observed RFS times. The median RFS based on $n = 70$ patients is 17.71 months.

- IgG and IgM antibody titre measurements were taken at the five time points, 0, 4, 6, 12, and 52 weeks. Natural logarithms of IgG and IgM were used in all analyses.

M.-H. Chen

- Our primary goal in this analysis is to characterize the degree of association between RFS and the IgG and IgM antibody titre measurements using our proposed model.

- Both the IgG and IgM antibody measures are related to the true unobservable univariate antibody level X^*. From a biological perspective, a change in X_1 often implies a change in X^* over a suitable window of time on average, and hence X_1 is deemed more important as a source of information on X^* compared to X_2.

- This is why X_1 is assumed to be on an unbiased scale for X^*. Thus the forms of (7) and (8) are based on sound biological considerations.

M.-H. Chen

- Based on fitting a Cox model, the hazard ratio is 2.61 in favor of treatment A with a P-value of 0.02, and the 95% confidence interval is $(1.17, 5.82)$.

- Noninformative priors were used for all of the models in the analyses below. For example, for the quadratic trajectory model with $p = 1$ using $J = 8$, we take $a_j = b_j = 0$, $j = 1, 2, \ldots, 8$, $v_j^2 = v_{0j}^2 = 100$, $j = 1, 2, 3, 4$, $\xi_1 = \xi_2 = \xi_3 = \xi_4 = \xi_{01} = \xi_{02} = \xi_{03} = 0$, $a_{01} = a_{02} = 0$,

$$
b_{01} = b_{02} = 0.001, \text{ and } \Sigma_0^{-1} \sim W_3 \left(4, \begin{pmatrix} 1.0 & 0 & 0 \\ 0 & 0.1 & 0 \\ 0 & 0 & 0.1 \end{pmatrix} \right).
$$

M.-H. Chen

TABLE 3.2: Summary of transformed IgG and IgM Measures for E2696

TRT	Week	log(IgG)					log(IgM)				
		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
	# 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
	# of Missing	7	2	0	6	16	7	2	0	6	16

M.-H. Chen

TABLE 3.3. Parameter Estimates for Cancer Vaccine Data E2696

Parameter	Mean	SD	95% HPD	Parameter	Mean	SD	95% HPD
β_1	-0.26	0.20	(-0.65, 0.12)	σ_{011}	1.09	0.38	(0.36, 1.87)
β_2	1.01	0.43	(0.15, 1.85)	σ_{012}	0.26	0.09	(-0.15, 0.21)
ρ	0.60	0.05	(0.51, 0.69)	σ_{013}	-0.07	0.11	(-0.30, 0.14)
σ_1^2	4.12	0.42	(3.32, 4.96)	σ_{022}	0.03	0.02	(0.004, 0.06)
σ_2^2	6.79	0.57	(5.73, 7.96)	σ_{023}	-0.01	0.02	(-0.04, 0.02)
λ_1	0.06	0.04	(0.004, 0.13)	σ_{033}	0.06	0.06	(0.01, 0.17)
λ_2	0.05	0.04	(0.002, 0.13)	α_0	3.64	0.35	(2.96, 4.33)
λ_3	0.05	0.05	(0.002, 0.13)	α_1	-0.03	0.11	(-0.26, 0.18)
λ_4	0.05	0.05	(0.002, 0.13)	μ_{01}	2.76	0.18	(2.42, 3.11)
λ_5	0.11	0.10	(0.006, 0.29)	μ_{02}	1.80	0.52	(0.77, 2.78)
λ_6	0.02	0.02	(0.001, 0.05)	μ_{03}	-0.34	0.49	(-1.23, 0.66)
λ_7	0.03	0.03	(0.000, 0.09)	λ_8	0.26	0.51	(0.000, 0.88)

M.-H. Chen

FIGURE 3.1. Marginal posterior density for β_1 for cancer vaccine data.

M.-H. Chen

- The posterior mean and standard deviation of β_1 is -0.26 and 0.20, respectively. A plot of the marginal posterior density of β_1 is also given in Figure 3.1.

- The posterior estimate of β_1 is negative and far from 0, implying that there is an association between relapse free survival and antibody titre count.

- That is, the magnitude of β_1 implies that increased antibody titres counts are moderately associated with longer relapse-free survival.

- This phenomenon is also confirmed by the plots of the posterior hazards in Figures 3.2 and 3.3. In Figures 3.2 and 3.3, • and ∘ correspond to treatments A and B, respectively.

- As the antibody titre counts increase, there is a decrease in the posterior hazard estimate for each individual. This phenomenon is observed for both the IgG and IgM antibody measurements, and therefore gives further evidence to the association between antibody titres and relapse-free survival.
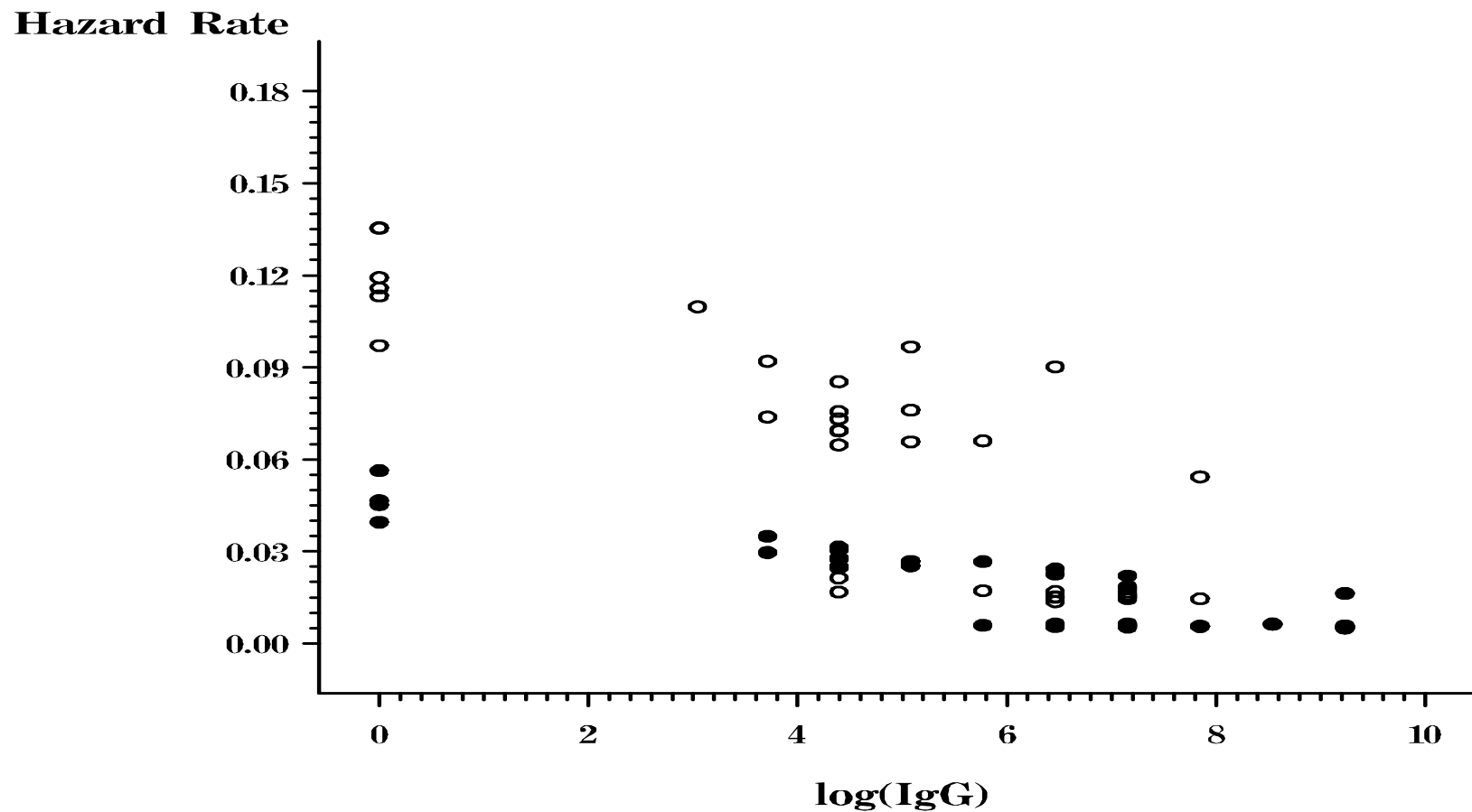
FIGURE 3.2. Estimated hazard rate as a function of log(IgG) taken at time point of peak measurement.

M.-H. Chen
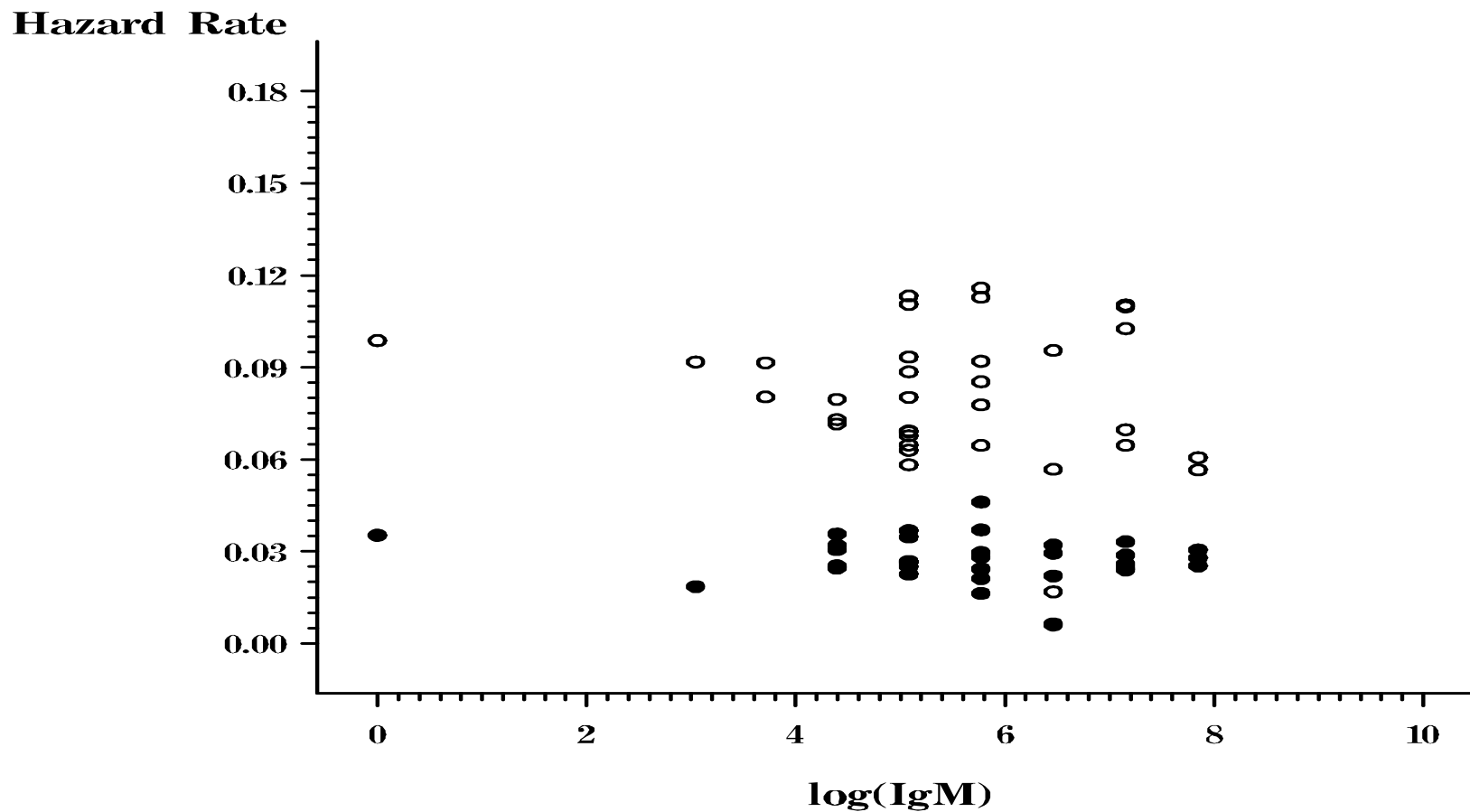
Hazard Rate

FIGURE 3.3. Estimated hazard rate as a function of log(IgM) taken at time point of peak measurement.

M.-H. Chen

- The treatment coefficient β_2 has posterior mean and standard deviation of 1.01 and 0.43, respectively, with 95% HPD interval $(0.15, 1.85)$.

- Thus it is clear that there is also an important treatment difference between A and B. Specifically, we see that treatment A (IFN + GMK) is superior to treatment B (GMK alone).

- The estimates for β_2 in Table 3.3 confirm this result. Figure 3.1 also confirms this, as we see that, for both the IgG and IgM antibody titres, the posterior hazard estimates are consistently smaller for treatment A then they are for treatment B, indicating a superior treatment A effect.

- A sensitivity analysis was conducted with respect to the choice of J. The posterior estimates of β_1, β_2, and the α_j's were also very robust with respect to the choice of J, yielding very similar estimates to those of Table 3.2 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.

3.3. A New Model for Longitudinal and Survival Data with a Cure Fraction

♠ The method we will review uses an approach to building the model as Chen, Ibrahim, and Sinha (2004, JMV).

♠ **E1694 Data**

- Data were from an Intergroup trial of the Eastern Cooperative Oncology Group, the Southwest Oncology Group and Cancer and Leukemia Group B.

- E1694 (Kirkwood et al., 2001) was designed to determine if GMK was superior to IFN with respect to RFS and Overall Survival (OS), with a secondary goal to determine the association of pre-existing and vaccine-induced IgM and IgG antibodies with RFS and OS.

M.-H. Chen

- GMK abbreviates GM2-KLH/QS-21, a vaccine that has recently been developed for treating melanoma.

- IFN denotes Interferon Alpha-2b, which is a chemotherapy and has a significant impact on RFS or OS.

- One of the major drawbacks of IFN, and chemotherapies in general, is that they are highly toxic.

- IgM and IgG measures were taken at baseline, 6, 9, 12, 18 and 24 months, respectively. In this analysis, we consider $n = 667$ patients and use RFS as the time-to-event variable in all analyses.

- The minimum RFS time was 0.0245 years and the maximum RFS time was 4.309 years. Once a patient relapsed, they dropped out of the study, and hence no longitudinal measures were collected after that time.

- Two covariates included in the analysis are age (z_{i1}) and gender (z_{i2}) $(z_{i2} = 0$ if male and 1 if female). We also let $x_i = 1$ denote the GMK arm and $x_i = 0$ the IFN arm.

M.-H. Chen

- Table 3.4 gives a detailed summary of RFS time, censoring status, covariates, and treatment.

TABLE 3.4. Summary of E1694 Data

	Median	IQR		Frequency
Follow-up	1.44	1.64	Censored	404
Time to Relapse	0.64	0.78	Relapse	263
Age	50.59	18.96		

	Frequency			Frequency
Sex	Male	428	Female	239
Treatment	GMK	344	IFN	323

- Compared to the E2696 study, the E1694 study was much better designed as the two treatment arms are completely separated. Therefore, with the E2696 study, we are able to make a direct comparison between GMK and IFN.

M.-H. Chen

- Figure 3.4 shows the Kaplan-Meier survival curve of the E1694 data. The Log-Rank test p-value is 0.056.
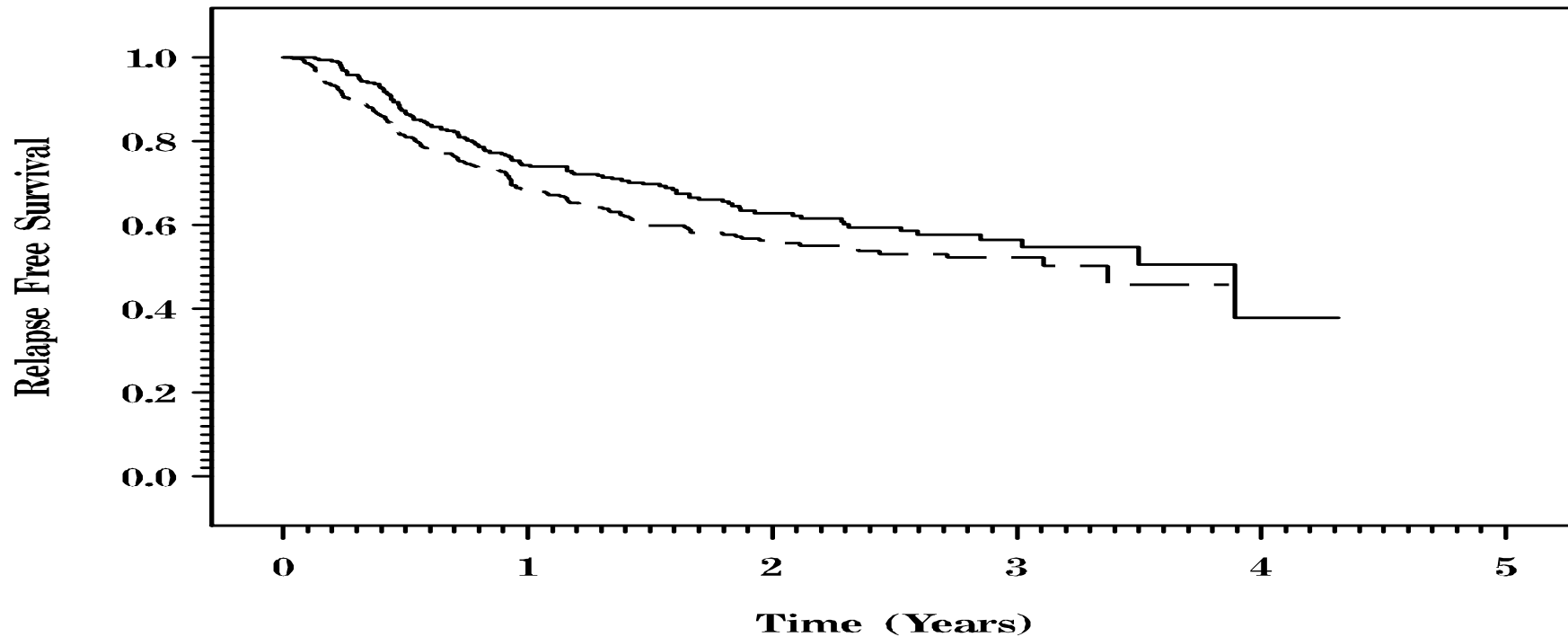

FIGURE 3.4. KM RFS plots for E1694, where the solid (top) curve and the dashed (bottom) curve corresponds to the IFN arm and the GMK arm, respectively.

M.-H. Chen

- Figure 3.4 shows a plateau in the Kaplan-Meier plot, motivating us to consider a cure rate model for these data.

- 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 3.5 gives a detailed summary of the $(\log(\mathrm{IgG}+1), \log(\mathrm{IgM}+1)))$ measures along with summaries of missing values.

- The trajectories of measures are shown in Figures 3.5 and 3.6.

M.-H. Chen

TABLE 3.5. Summary of Transformed IgG and IgM Measures

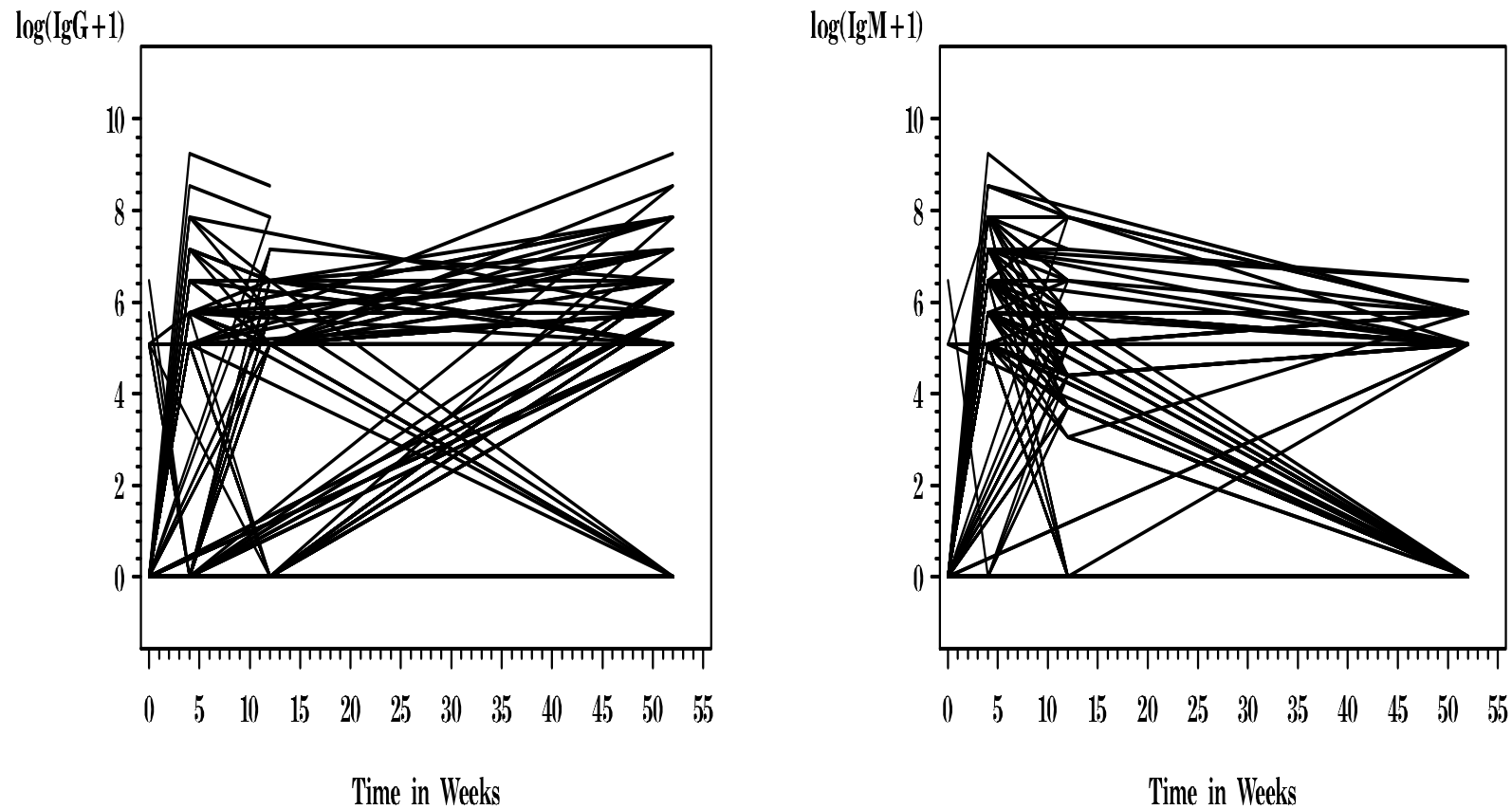| TRT	Week	log(IgG + 1)				log(IgM + 1)			
		0	4	12	52	0	4	12	52
GMK	Median	0.00	0.00	0.00	5.08	0.00	5.77	4.39	0.00
	Mean	0.23	1.87	2.05	3.42	0.07	5.30	3.17	1.76
	IQR	0.00	5.08	5.08	5.77	0.00	1.38	5.08	5.08
	# of Missing	108	90	144	185	19	84	85	222
IFN	Median	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Mean	0.04	0.08	0.42	0.00	0.03	0.07	0.09	0.12
	IQR	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	# of Missing	46	24	67	191	17	23	76	192

M.-H. Chen

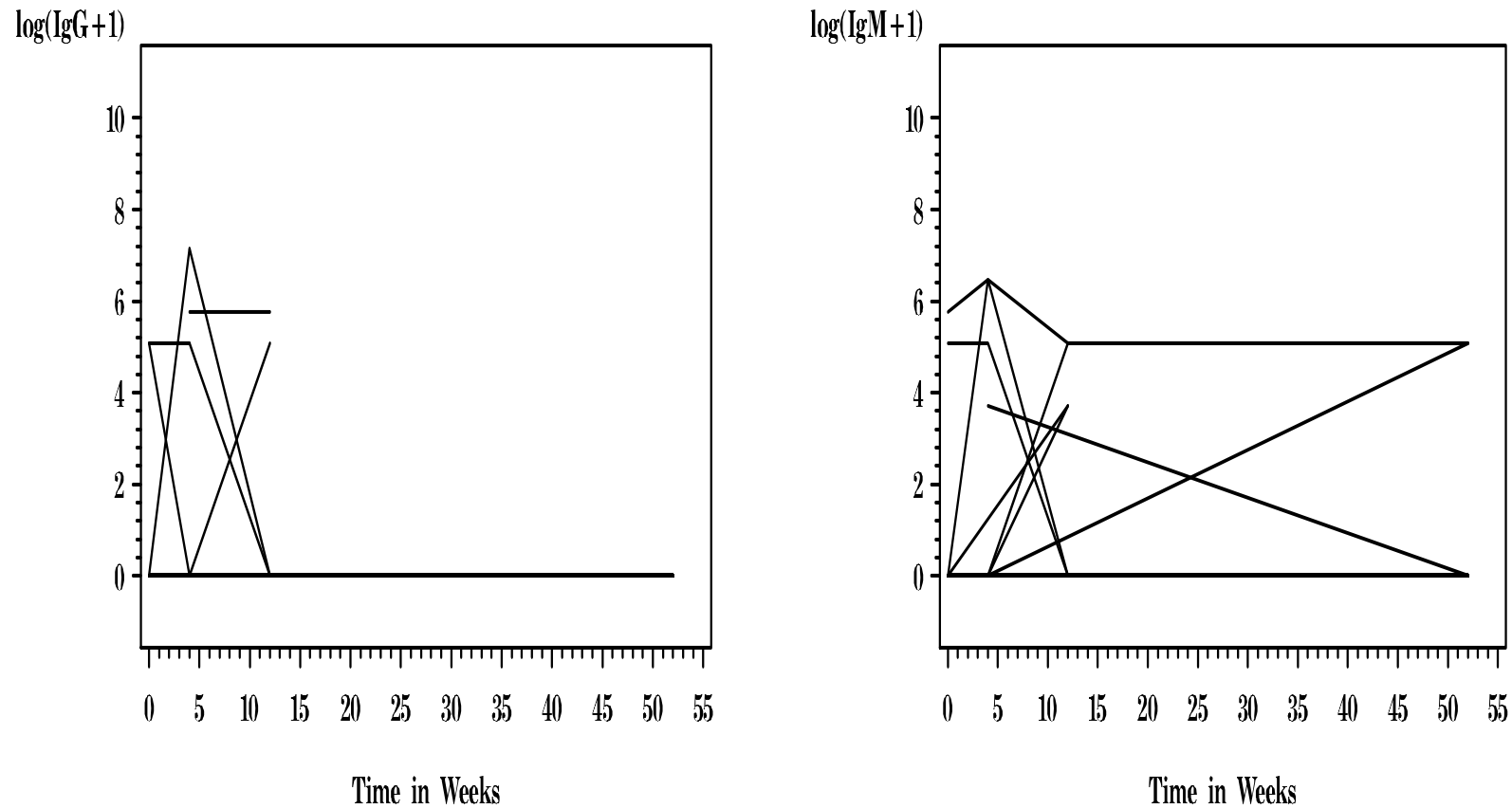FIGURE 3.5. Trajectory plots for log(IgG+1) (left) and log(IgM+1) (right) for the GMK arm.

M.-H. Chen

FIGURE 3.6. Trajectory plots for log(IgG+1) (left) and log(IgM+1) (right) for the IFN arm.

M.-H. Chen

- From Table 3.5, we can see that most IgG and IgM measures were 0 for the IFN arm.

- From Figure 3.5, we see evidence of a quadratic trend in the IgG trajectory with the peak IgG and IgM titers occurring at approximately 4 weeks for GMK.

- These numerical and graphical summaries also shed light on how to model longitudinal measures. In particular, the mean trajectory functions and the variability of these measures should be treatment-dependent.

- In order to study the relationship of these observable covariates to RFS or OS, we need to develop a Bayesian model for joint modeling of the survival data and the longitudinal IgG and IgM measurements.

M.-H. Chen

♠ Survival Component of the Model

- Let N_i denote the number of *metastatis-competent* tumor cells for subject i (in short, MCT).

- N_i's, $i = 1, \ldots, n$, are unobserved latent variables. Further, we assume that the N_i's are independent and each has a Poisson distribution with mean θ_i.

- Let $\boldsymbol{\alpha}_i$ be the vector of unobservable random effects characterizing the induced immunological response of the ith patient. Later, we model the patient specific random immune response $\boldsymbol{\alpha}_i$ dependent on the treatment covariate x_i.

- Assume that given $\boldsymbol{\alpha}_i$, the treatment x_i does not influence N_i, the number of MCTs.

M.-H. Chen

- We further assume that

$$\theta_i(x_i, z_i) \equiv \theta_i(z_i, \beta, \eta, \alpha_i) = \exp(z_i'\beta + \alpha_i'\eta),$$

  where η is a vector of regression parameters corresponding to α_i, and β is the vector of regression coefficients corresponding to the fixed covariates z_i, which includes an intercept but does not include x_i.

- The case $\eta = 0$ implies that the patient-specific immune response is not associated with the number of MCTs in the body.

- Let Z_{ij} denote the random time for the jth MCT to produce detectable metastatic disease in the ith subject.

M.-H. Chen

- Let $S(t|\boldsymbol{\lambda})$ denote the survival function of the Z_{ij}'s, which depends on the vector of parameters $\boldsymbol{\lambda}$, $F(t|\boldsymbol{\lambda}) = 1 - S(t|\boldsymbol{\lambda})$, and $f(t|\boldsymbol{\lambda}) = \frac{d}{dt} F(t|\boldsymbol{\lambda})$.

- Assume that $S(t)$ does not depend on x_i and $\boldsymbol{\alpha}_i$.

- The time to relapse of cancer, T_i, can be defined by $T_i = \min\{Z_{ij}, 1 \le j \le N_i\}$, and $T_i = \infty$ with probability 1 when $N_i = 0$.

- It follows that the conditional survival function is given by

$$S_{ip}(t|\boldsymbol{\beta}, \boldsymbol{z}_i, \boldsymbol{\eta}, \boldsymbol{\alpha}_i, \boldsymbol{\lambda}) = P(T_i > t|\boldsymbol{\beta}, \boldsymbol{z}_i, \boldsymbol{\eta}, \boldsymbol{\alpha}_i, \boldsymbol{\lambda})$$
$$= \exp[-\{\theta_i(\boldsymbol{z}_i, \boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\alpha}_i)F(t|\boldsymbol{\lambda})\}].$$

♠ Longitudinal Component of the Model

- Let $\Psi_i(t)$ for the ith patient denote the *trajectory function,* which is modeled as a known parametric function of t and $\boldsymbol{\alpha}_i$.

- For subject i, $\Psi_i(t)$ represents the true, unobservable *univariate* immune response level at time t.

- We take $\Psi_i(t)$ to be a linear or quadratic trajectory in t. Thus, the trajectory functions for subject i are of the form:

  Linear:
  $$\Psi_i(t) = \alpha_{i0} + \alpha_{i1}t,$$

  Quadratic:
  $$\Psi_i(t) = \alpha_{i0} + \alpha_{i1}t + \alpha_{i2}t^2.$$

- Let $\boldsymbol{\alpha}_i = (\alpha_{i0}, \alpha_{i1})'$ or $\boldsymbol{\alpha}_i = (\alpha_{i0}, \alpha_{i1}, \alpha_{i2})'$ to be unknown patient specific parameters.

M.-H. Chen

- Assume that the $\boldsymbol{\alpha}_i$'s vary within patients such that

$$\boldsymbol{\alpha}_i \sim N(\boldsymbol{\mu}(x_i), \Sigma(x_i)),$$

where $\boldsymbol{\mu}(x_i)$ is the mean of $\boldsymbol{\alpha}_i$, $\Sigma(x_i)$ is the covariance matrix of $\boldsymbol{\alpha}_i$, and both of these quantities depend on the treatment covariate x_i.

- We see that the model for $\Psi_i(t)$ is quite general and results in trajectory functions with different intercepts, slopes, and curvatures for each patient in a treatment group.

- To incorporate the effect of treatment x_i on $\boldsymbol{\mu}(x_i)$ and $\Sigma(x_i)$, we assume that

$$E(\alpha_{ij}|x_i, \boldsymbol{\mu}) = \mu_j x_i \text{ and } \text{Var}(\alpha_{ij}|x_i, \Sigma) = x_i \Sigma_1 + (1 - x_i)\Sigma_2.$$

M.-H. Chen

- Given that the patient i is at observed at time t, that is $i \in \mathcal{R}(t)$, where $\mathcal{R}(t)$ is the risk set at time t, let $\mathbf{Y}_i(t)$ denote the observed multivariate longitudinal response for subject i at time t.

- We assume that components of $\mathbf{Y}_i(t)$ are measurements of $\Psi_i(t)$ at different scales taken at time t. So, we take

$$\mathbf{Y}_i(t) = \mathbf{g}(\Psi_i(t)) + \epsilon_i(t),$$

  where $\mathbf{g}$ is a known function.

- At time point t, the random measurement error, $\epsilon_i(t) \sim N(0, \Sigma_0)$ are independent of $\Psi_i(t)$ and they are also independent of $\epsilon_i(t')$ for $t \neq t'$.

- The vector of known functions $\mathbf{g}$ reflects the fact that each observed antibody level measures the patients immune level against cancer in its own scale. For simplicity, we will use the linear relationship

$$g_j(\Psi_i(t)) = \phi_{0j} + \phi_{1j}x_i t + \alpha_{i0} + \alpha_{i1}t$$

M.-H. Chen

or the quadratic relationship

$$g_j(\boldsymbol{\Psi}_i(t)) = \phi_{0j} + \phi_{1j}x_it + \phi_{2j}x_it^2 + \alpha_{i0} + \alpha_{i1}t + \alpha_{i2}t^2$$

for j^{th} component of $\mathbf{Y}_i$, $j = 1, \ldots m$.

- Note that the terms $\phi_{0j} + \phi_{1j}x_it$ and $\phi_{0j} + \phi_{1j}x_it + \phi_{2j}x_it^2$ can be viewed as the fixed components of the trajectory function.

- We write

$$\boldsymbol{\Phi}_i(t) = (\boldsymbol{\Phi}_{i1}(t), \ldots, \boldsymbol{\Phi}_{im}(t))' = (\phi_{01} + \phi_{11}x_it, \ldots, \phi_{0m} + \phi_{1m}x_it)'$$

or

$$\boldsymbol{\Phi}_i(t) = (\boldsymbol{\Phi}_{i1}(t), \ldots, \boldsymbol{\Phi}_{im}(t))'$$
$$= (\phi_{01} + \phi_{11}x_it + \phi_{21}x_it^2, \ldots, \phi_{0m} + \phi_{1m}x_it + \phi_{2m}x_it^2)'.$$

- For the E1694 data, we have $m = 2$. The IgG and IgM measurements are respectively Y_{i1} and Y_{i2}.

M.-H. Chen

♠ The Main Features of the Proposed Joint Model

- Our model here characterizes the history of the three processes process $(\mathbf{Y}_i(t), \boldsymbol{\Phi}_i(t), \Psi_i(t))$ given $T_i \in \mathcal{R}(t)$ and x_i.

- The process $\Psi_i(t)$ is shared by the longitudinal components and the survival component to account for the correlation between them. Thus, the survival component of the model is connected to the longitudinal component of the model through $\boldsymbol{\eta}$ and $\boldsymbol{\eta} = 0$ implies that the longitudinal and survival measures are not associated.

- Each longitudinal component has its own treatment dependent process $\boldsymbol{\Phi}_{ij}(t)$, which is not shared with the survival component, to account for the sole contribution of the trajectory function to each longitudinal component in addition to the common process $\Psi_i(t)$.

M.-H. Chen

- The proposed model does not require that all longitudinal responses be observed at a given time point for the ith patient. This is an important feature of the model as in the E1694 data shown in Table 3.5, only one of the IgG and IgM measurements is available for many patients.

- The mean trajectory functions and the variances of α_i vary with treatment. This property of the model is most suitable for the E1694 data since we see in Table 3.5 that most of the longitudinal measures in the IFN arm are zero.

M.-H. Chen

♠ Likelihoods

- Let $a_{i1}, \ldots, a_{im_i}$ denote the times at which the antibody measurements are taken.

- Let $\Psi_i(a_{ij})$ denote the trajectory function evaluated at a_{ij}.

- Let the IgG and IgM antibody titers for subject i be denoted by $\boldsymbol{y}_{il} = (y_{i1l}, \ldots, y_{im_i l})'$, let $\boldsymbol{y}_l = (\boldsymbol{y}'_{1l}, \ldots, \boldsymbol{y}'_{nl})'$ for $l = 1, 2$, respectively, and let $\boldsymbol{y} = (\boldsymbol{y}'_1, \boldsymbol{y}'_2)'$.

- Let $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)'$ and $Z = (\boldsymbol{z}'_1, \boldsymbol{z}'_2, \ldots, \boldsymbol{z}'_n)'$.

- 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. 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.

M.-H. Chen

- Let $D = (\boldsymbol{y}, \boldsymbol{x}, Z, \boldsymbol{t}, \boldsymbol{\delta})$ denote the observed data.

- We consider a piecewise exponential model for $F(t|\boldsymbol{\lambda})$. Specifically, we construct a finite partition of the time axis, $0 < s_1 < \ldots < s_J$, with $s_J > t_i$ for all $i = 1, 2, \ldots, n$. Thus, we have the J intervals $(0, s_1]$, $(s_1, s_2]$, $\ldots$, $(s_{J-1}, s_J]$. We thus assume that the hazard for $F(t|\boldsymbol{\lambda})$ is equal to λ_j for the jth interval, $j = 1, 2, \ldots, J$, leading to

$$F(t|\boldsymbol{\lambda}) = 1 - \exp\left\{-\lambda_j(t - s_{j-1}) - \sum_{g=1}^{j-1} \lambda_g(s_g - s_{g-1})\right\}.$$

M.-H. Chen

- The likelihood function can thus be written as

$$L(\boldsymbol{\theta}|D) = L_1(\boldsymbol{\phi}, \Sigma_0, \Psi|\boldsymbol{y}, \boldsymbol{x})L_2(\boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\lambda}, \boldsymbol{\alpha}|\boldsymbol{t}, \boldsymbol{\delta}, \boldsymbol{z}),$$

where

$$L_1(\boldsymbol{\phi}, \Sigma_0, \Psi|D) = \prod_{i=1}^{n}\prod_{j=1}^{2}\prod_{k:a_k \leq t_i} |\Sigma_0|^{-1/2}$$

$$\times \exp\left\{-\frac{1}{2}(y_{ikj} - g_{ij}(a_k))'\Sigma_0^{-1}(y_{ikj} - g_{ij}(a_k))\right\}$$

and

$$L_2(\boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\alpha}|\boldsymbol{t}, \boldsymbol{\delta}, \boldsymbol{z}) = \prod_{i=1}^{n} (\theta_i f(t_i|\boldsymbol{\lambda}))^{\delta_i} \exp(-\theta_i F(t_i|\boldsymbol{\lambda})),$$

$\boldsymbol{\theta} = (\boldsymbol{\phi}, \Sigma_0, \Psi, \boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\lambda}, \boldsymbol{\alpha})$, $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1', \boldsymbol{\alpha}_2' \ldots, \boldsymbol{\alpha}_n')'$, and
$f(t_i|\boldsymbol{\lambda}) = \frac{d}{dt} F(t|\boldsymbol{\lambda})\Big|_{t=t_i}$.

M.-H. Chen

♠ Priors

- We specify the joint prior as

$$\pi(\Sigma_0, \phi, \mu, \Sigma_1, \Sigma_2, \beta, \eta, \lambda) = \pi(\Sigma_0)\pi(\phi)\pi(\mu)\pi(\Sigma_1)\pi(\Sigma_2)\pi(\beta)\pi(\eta)\pi(\lambda).$$

- We take Σ_0 to be of the form $\Sigma_0 = \text{Diag}(\sigma_{01}^2, \sigma_{02}^2)$, and take σ_{01}^2 and σ_{02}^2 to have independent inverse gamma priors.

- For ϕ, μ, β, and η, we take normal priors with diagonal variance-covariance matrices.

- The conditional prior for $\alpha|\mu, \Sigma_1, \Sigma_2$ is specified as follows:

$$\pi(\alpha|\mu, \Sigma_1, \Sigma_2) \propto \prod_{i=1}^{n} \left[|x_i\Sigma_1 + (1 - x_i)\Sigma_2|^{-1/2} \right.$$

$$\left. \times \exp\left\{ -\frac{1}{2}(\alpha_i - \mu(x_i))'[x_i\Sigma_1 + (1 - x_i)\Sigma_2]^{-1}(\alpha_i - \mu(x_i)) \right\} \right].$$

M.-H. Chen

- For Σ_1 and Σ_2, we assume $\Sigma_k = \text{Diag}(\sigma_{k0}^2, \sigma_{k1}^2)$ for the linear trajectory and $\Sigma_k = \text{Diag}(\sigma_{k0}^2, \sigma_{k1}^2, \sigma_{k2}^2)$ for the quadratic trajectory, where $k = 1, 2$. Then, we specify an inverse gamma prior for each variance component.

- Finally, we take independent gamma priors for $\boldsymbol{\lambda}$ as follows:

$$\pi(\boldsymbol{\lambda}) \propto \prod_{j=1}^{J} \lambda_j^{\zeta_0 - 1} \exp(-\tau_0 \lambda_j),$$

  where ζ_0 and τ_0 are pre-specified hyperparameters.

♠ Computational Development

To develop an efficient computational algorithm, we introduce the latent variables $N = (N_1, N_2, \ldots, N_n)'$ into the likelihood function for the survival component so that

$$L_2(\beta, \eta, \alpha, N | t, \delta, z) = \prod_{i=1}^{n} S(t_i|\lambda)^{N_i - \delta_i} \left(N_i f(t_i|\lambda)\right)^{\delta_i} \frac{\theta_i^{N_i}}{N_i!} \exp(-\theta_i),$$

where $S(t_i|\lambda) = 1 - F(t_i|\lambda)$. Note that $N_i = 0, 1, \ldots, \infty$ if $\delta_i = 0$ while $N_i = 1, 2, \ldots, \infty$ if $\delta_i = 1$. It can be shown that

$$\sum_{N} L_2(\beta, \eta, \alpha, N | t, \delta, z) = L_2(\beta, \eta, \alpha | t, \delta, z).$$

Using the collapsed Gibbs technique, we sample from the following conditional posterior distributions in turn:

M.-H. Chen

(1) $[\alpha|\phi, \Sigma_0, \beta, \eta, \lambda, \mu, \Sigma_1, \Sigma_2, D]$, which is log-concave.

(2) $[\phi|\Sigma_0, \alpha, D]$, which is a multivariate normal with mean and variance-covariance matrix depending on the form of the trajectory function.

(3) $[\Sigma_0|\phi, \alpha, D]$, which consists of two conditionally independent inverse gamma distributions.

(4) $[\mu|\alpha, \Sigma_1, \Sigma_2, D]$ is a multivariate normal, and its dimension, mean, and covariance matrix depends on the form of the trajectory.

(5) $[\Sigma_1, \Sigma_2|\alpha, \mu, D]$, which is the product of 4 (or 6) independent inverse gamma distributions depending on the linear (or quadratic) trajectory.

(6) $[\beta|\eta, \lambda, \alpha, D]$ has the density

$$\pi(\beta|\eta, \lambda, \alpha, D) \propto L_2(\beta, \eta, \alpha|t, \delta, z)\pi(\beta).$$

Note that this conditional distribution does not depend on N and it can be shown that $\pi(\beta|\eta, \lambda, \alpha, D)$ is log-concave.

M.-H. Chen

(7) $[\boldsymbol{\eta}|\boldsymbol{\beta}, \boldsymbol{\lambda}, \boldsymbol{\alpha}, D]$, which is similar to $\pi(\boldsymbol{\beta}|\boldsymbol{\eta}, \boldsymbol{\lambda}, \boldsymbol{\alpha}, D)$. Thus, the density of this conditional distribution is also log-concave.

(8) For each latent variable,

$$N_i|\boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\lambda}, D \sim \mathcal{P}(S(t_i|\boldsymbol{\lambda})\theta_i) + \delta_i,$$

where $\mathcal{P}(S(t_i|\boldsymbol{\lambda})\theta_i)$ denotes the Poisson distribution with mean $S(t_i|\boldsymbol{\lambda})\theta_i$.

(9) Given $\boldsymbol{N}$, the conditional posterior density for $\boldsymbol{\lambda}$ is

$$\pi(\boldsymbol{\lambda}|\boldsymbol{N}, \boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\alpha}, D) \propto \prod_{i=1}^{n} S(t_i|\boldsymbol{\lambda})^{N_i - \delta_i} f(t_i|\boldsymbol{\lambda})^{\delta_i} \pi(\boldsymbol{\lambda}).$$

With independent gamma priors for the λ_j's, it is easy to see that $[\boldsymbol{\lambda}|\boldsymbol{N}, \boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\alpha}, D]$ consists of J independent gamma distributions.

Thus, for (2) — (5), (8) and (9), the generation is straightforward, while for (1), (6) and (7), we can use the adaptive rejection algorithm of Gilks and Wild (1992), since the corresponding conditional posterior densities are log-concave.

M.-H. Chen

♠ Model Assessment via LPML

- To assess the goodness-of-fit of the proposed model, we consider the Conditional Predictive Ordinate (CPO).

- The CPO statistic is defined as

$$\text{CPO}_i = \int f(\boldsymbol{y}_{i1}, \boldsymbol{y}_{i2}, t_i, \delta_i | \boldsymbol{\theta}, x_i, \boldsymbol{z}_i) \pi(\boldsymbol{\theta} | D^{(-i)}) \, d\boldsymbol{\theta},$$

  where $\boldsymbol{\theta}$ is the vector of all model parameters, $D^{(-i)}$ denotes the data with the ith subject deleted, and $\pi(\boldsymbol{\theta} | D^{(-i)})$ is the posterior density of $\boldsymbol{\theta}$ based on the data $D^{(-i)}$.

- The LPML is defined as $\text{LPML} = \sum_{i=1}^{n} \log(\text{CPO}_i)$.

- We choose the model with the largest LPML value.

- For the E1694 data, we will use LPML to compare the linear and quadratic trajectory models.

- LPML is particularly suitable for cure rate models with noninformative priors.

M.-H. Chen

♠ Data Analysis

- We consider analyses with the linear and quadratic trajectories.

- In both analyses, a piecewise exponential model is used for $F(t|\boldsymbol{\lambda})$. We take $J = 10$ and the intervals $(s_{j-1}, s_j]$, $j = 1, 2, \ldots, J$ were chosen so that at least one failure falls in each interval.

- We take an improper uniform prior, i.e., $\pi(\boldsymbol{\lambda}) \propto 1$, for $\boldsymbol{\lambda}$. $N(0, 100)$ priors are specified for the location parameters β_j, η_j, ϕ_{j1}, ϕ_{j2}, and μ_j for $j = 0, 1, 2$.

- For the scale parameters σ_{01}^2 and σ_{02}^2, we independently take $\sigma_{0k}^2 \sim \mathcal{IG}(a = 1, b = 0.01)$ with density proportional to $(\sigma_{0k}^2)^{-(a+1)} \exp(-b/\sigma_{0k}^2)$. Similarly we take $\sigma_{kj}^2 \sim \mathcal{IG}(a = 1, b = 0.1)$ independently for $k = 1, 2$ and $j = 0, 1, 2$.

M.-H. Chen

- The table below shows posterior estimates of the parameters based on the linear as well as the quadratic trajectory model.

- The quadratic trajectory model give a better fit to the data than the linear trajectory model, as measured by the LPML statistic, as well as the posterior estimates for ϕ_{21} and ϕ_{22}, whose 95% Highest Posterior Density (HPD) intervals do not include zero.

- The posterior estimates of η_2 and μ_2 do not include zero, giving further evidence of the appropriateness of a quadratic trajectory model.

- The parameters that link the longitudinal model to the survival model are the η_j's, whose 95% HPD intervals all do *not* contain zero, indicating an important association between (IgG, IgM) and time-to-relapse.

M.-H. Chen

Parameter	Linear Trajectory			Quadratic Trajectory		
	Mean	SD	95% HPD Interval	Mean	SD	95% HPD Interval
η_0	-0.815	0.245	(-1.301, -0.343)	0.398	0.198	(0.021, 0.793)
η_1	0.173	0.046	(0.083, 0.265)	-0.326	0.108	(-0.530, -0.106)
η_2	-	-	-	1.369	0.769	(0.245, 2.891)
β_0	-0.479	0.122	(-0.723, -0.242)	-0.426	0.124	(-0.672, -0.186)
β_1	0.125	0.068	(-0.012, 0.256)	0.155	0.069	(0.020, 0.287)
β_2	-0.055	0.069	(-0.189, 0.080)	-0.026	0.070	(-0.164, 0.108)
σ_{01}^2	1.921	0.082	(1.760, 2.080)	1.919	0.078	(1.769, 2.074)
σ_{02}^2	3.591	0.131	(3.334, 3.847)	2.862	0.103	(2.666, 3.068)
ϕ_{01}	-0.163	0.042	(-0.246, -0.081)	-0.086	0.042	(-0.169, -0.005)
ϕ_{11}	-3.402	1.620	(-6.792, -0.585)	-2.103	2.509	(-6.941, 2.614)
ϕ_{21}	-	-	-	-7.679	1.303	(-10.21, -5.348)
ϕ_{02}	0.401	0.053	(0.297, 0.506)	0.162	0.049	(0.065, 0.256)
ϕ_{12}	-4.536	1.620	(-7.835, -1.647)	8.119	2.518	(3.318, 12.945)
ϕ_{22}	-	-	-	-19.95	1.329	(-22.53, -17.57)
μ_0	1.365	0.081	(1.209, 1.526)	0.957	0.079	(0.801, 1.120)
μ_1	7.877	1.668	(4.877, 11.259)	9.659	2.446	(5.180, 14.291)
μ_2	-	-	-	2.447	1.081	(0.869, 4.415)
σ_{10}^2	0.723	0.111	(0.508, 0.942)	0.818	0.108	(0.617, 1.042)
σ_{11}^2	18.303	4.480	(10.348, 27.432)	6.515	1.313	(4.052, 9.118)
σ_{12}^2	-	-	-	0.099	0.081	(0.011, 0.247)
σ_{20}^2	0.002	0.0003	(0.0016, 0.0028)	0.002	0.0003	(0.0016, 0.0028)
σ_{21}^2	0.002	0.0003	(0.0016, 0.0028)	0.002	0.0003	(0.0016, 0.0028)
σ_{22}^2	-	-	-	0.002	0.0003	(0.0016, 0.0028)
LPML	-7994.932			-7736.51		