THE LATENT GROUP-BASED TRAJECTORY MODEL: DEVELOPMENT OF DISCRIMINATION MEASURES AND JOINT MODELING WITH SUBDISTRIBUTIONS

by

Nilesh Shah

B.S. Computer Engineering,

Case Western Reserve University, 2001

Submitted to the Graduate Faculty of

the Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Doctor of Philiosphy

University of Pittsburgh

2012

UNIVERSITY OF PITTSBURGH

GRADUATE SCHOOL OF PUBLIC HEALTH

This dissertation was presented

by

Nilesh Shah

It was defended on

July 18, 2012

and approved by

Joyce Chung-Chou H. Chang, PhD, Associate Professor, Department of Medicine

School of Medicine, University of Pittsburgh

Lisa Weissfeld, PhD, Professor, Department of Biostatistics

Graduate School of Public Health, University of Pittsburgh

Jong Hyeon Jeong, PhD, Associate Professor, Department of Biostatistics

Graduate School of Public Health, University of Pittsburgh

Eydie Moses-Kolko, MD, Assistant Professor, Department of Psychiatry

School of Medicine, University of Pittsburgh

Dissertation Director: Joyce Chung-Chou H. Chang, PhD, Associate Professor,

Department of Medicine

School of Medicine, University of Pittsburgh

THE LATENT GROUP-BASED TRAJECTORY MODEL: DEVELOPMENT OF DISCRIMINATION MEASURES AND JOINT MODELING WITH SUBDISTRIBUTIONS

Nilesh Shah, PhD

University of Pittsburgh, 2012

In clinical research, patient care decisions are often easier to make if patients are classified into a manageable number of groups based on homogeneous risk patterns. Investigators can use latent group-based trajectory models to estimate the posterior probabilities that an individual will be classified into a particular group of risk patterns. Although this method is increasingly used in clinical research, there is currently no measure that can be used to determine whether an individual's group assignment has a high level of discrimination. We propose a discrimination index and provide confidence intervals of the probability of the assigned group for each individual. We also propose a modified form of entropy to measure discrimination. Additionally, when analyzing research involving disease processes, many researchers are interested in estimating the effect of longitudinally measured biomarkers on the event time outcomes in the presence of competing risks. We propose a method to estimate this effect under the joint modeling framework. The proposed joint model involves three submodels: the first one models the latent risk trajectory groups; the second one models the longitudinal pattern of biomarkers conditional on a specific risk group; and the third one models the subdistribution function conditional on a specific risk group.

These methods are significant to public health research since they enable researchers to more confidently assign individual patients to risk groups based on their clinical measurements.

The joint model also enables researchers to discover these distinct risk patterns more accurately by using patients' longitudinal data together with event time outcomes, while also adjusting for competing events.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

I would like to thank my advisor, Dr. Chang, and my committee members for all of their help. This dissertation would not be possible without their efforts and help. I would like to thank Allison Hipwell and the Pittsburgh Girls Study for making their data available to me. I would also like to thank the Department of Critical Care Medicine and John Kellum for providing me with an opportunity to work with the BioMaRK data. I would like to thank Dr. Howard Rockette for supporting me with his T32 grant. Finally, I would like to thank all my professors, collaborators, and fellow students for their help.

1.0 INTRODUCTION

The advent of personalized medicine has made statistical techniques like latent group-based trajectory[1] modeling more popular in recent years. These modeling techniques can separate populations into distinct risk patterns based on longitudinal data. Once separated, researchers are then able to evaluate each set of patients according to which group they are assigned. These modeling techniques are increasingly being implemented, but there are not many ways to evaluate these models or to determine whether individuals are assigned to their groups with a high degree of confidence. In this dissertation, we analyzed data from The Pittsburgh Girls Study. The investigators were interested in examining longitudinal trajectories of conduct disorder score in pre-adolescent girls. Using group-based trajectory modeling, we uncovered distinct behavioral groups and identify girls whose conduct gets worse or better over time. If individuals are to be evaluated and treated according to which behavioral group they belong to, investigators should be confident of the individuals' group assignment. There was previously no method to determine how confident this group assignment is. In this dissertation we develop methods to determine individual level discrimination. These measures serve as both individual patent evaluation and overall model adequacy checks.

Another data issue arises when we wish to uncover distinct behavioral subgroups based on both longitudinal data and event-time outcomes. Increasingly, medical studies have collected both longitudinal measurements with survival outcomes. One example of this is the Biological Markers of Recovery for the Kidney (BioMaRK) study. BioMaRK aims to study the relationship between inflammatory biomarker levels and recovery from Acute Kidney Injury (AKI). The investigators were interested in learned if there are distinct behavior groups in this data. Instead of using the traditional group-based model, our approach to this problem required modeling both longitudinal and survival data simultaneously. Traditionally

joint models are used to simultaneously model longitudinal and survival data. However, no one has accounted for a latent class joint model that simultaneously models longitudinal trajectories with competing risks subdistributions. One issue in the BioMaRK study is the fact that investigators wished to study recovery from AKI. During the course of the study, patients dropped out due to death. This is non-informative dropout that must be taken into account. Our model discovers latent group behavior based on both longitudinal and survival data. When examining data from BioMaRK, we modeled recovery from AKI as the main event of interest and used death as a competing event.

In the next section of this document, we review models and concepts essential to the dissertation. Then we develop measures of discrimination for latent group-based models. We also develop a measure called modified entropy to evaluate individual discrimination. We devise simulations and test these measures, then apply then to the Pittsburgh Girls Study. Then we develop a latent class joint model based on longitudinal trajectories and competing risks survival. We devise simulations then apply the model to the BioMaRK study. We then conclude the dissertation with discussion and mention of future work.

2.0 REVIEW OF MODELS

2.0.1 Group based modeling of development

The main latent class trajectory models used in this paper are based on Daniel Nagin's group-based models[1]. These models are used to model longitudinal data, with the models being able to separate the population into latent behavioral groups, or developmental trajectories. The motivation behind developing these models is that oftentimes, investigators believe that there may be different developmental patterns in a population over time. Using growth curves to model longitudinal data provides a mean trajectory with variation around that trajectory, but it cannot uncover these distinct behavioral patters over time. The main group-based model laid out by Nagin is:

$$P(Y_i) = \sum_{j}^{J} \pi_j P^j(Y_i) \tag{2.1}$$

where $P(Y_i)$ is the unconditional probability of observing individual i's longitudinal measurements Y. J represents to total number of behavioral groups, and π represents the probability of a randomly chosen individual belonging to group j. The group-based model is a variation of a finite mixture model. The longitudinal measurements, Y_i, can be assumed to follow any distribution the analyst wishes. Bobby Jones, et al. have implemented group-based modeling in SAS with the procedure PROC TRAJ [2]. Currently, PROC TRAJ supports modeling longitudinal data as normal (or tobit), binomial, or Poisson [2]. The model allows the longitudinal trajectories to be modeled as polynomial functions of time.

PROC TRAJ uses the quasi-Newton maximization method to maximize the likelihood and obtain parameter estimates [1]. Once estimates are obtained, one may calculate posterior probabilities for each individual. These are conditional probabilities for each individual, and

they represent the probability that an individual belongs to group j given his longitudinal measurements. Posterior probability is calculated as:

$$\hat{P}(j|Y_i) = \frac{\hat{P}(Y_i)\hat{\pi}_j}{\sum_j^J \hat{P}(Y_i)|j)\hat{\pi}_j} \tag{2.2}$$

Model selection is an important issue regarding group-based models. Since the analysts do not know how many latent groups are truly in the data, they must run multiple models with a differing number of groups. The Bayesian Information Criteria (BIC) is most commonly used to asses model performance [3]. BIC is defined as:

$$log(L) - 0.5 * k * log(N) \tag{2.3}$$

where L is the likelihood, k is the number of parameters estimated by the model, and N is the sample size. While the BIC itself can be used as a model selection criteria, Jones and Nagin also use an approximation of the Bayes Factor for model selection, with criteria outlined.

2.0.2 Survival analysis with subdistributions

When analyzing time to event outcomes, medical data often contain data with competing risks. Modeling these competing risks allows investigators to account for non-informative dropout in the survival process [4]. A main feature of competing risks analysis is that the model treats observations experiencing the competing event as still at risk to experience the main event. One standard way to account for competing risks is through the cause specific hazard function. The drawback of using the cause specific hazard approach is that it does not provide an interpretation regarding the probability of occurrence for the main event. Instead, Fine and Gray [4] devised a model to directly model the cumulative incidence function, or subdistribution of an event. They defined the subdistribution of the main event as:

$$F_1(t; Z) = 1 - \exp[-\int_0^t \lambda_{10}(s) \exp Z^T(s)\beta_0 ds] \tag{2.4}$$

where λ_{10} is the baseline hazard for the main event and Z is a set of covariates. The Fine and Gray model is a semiparametric approach since they assume the baseline hazard, λ_{10}, is not specified by a distribution.

The Fine and Gray model allows clinicians to interpret results more easily since they can directly interpret results as in terms of probabilities of experiencing events.

2.0.3 Parametric cumulative incidence function with subdistributions

Jeong and Fine [5] later developed a parametric from for regression on the cumulative incidence function. The parametric models are more amenable to maximum likelihood and allow the analyst to further extrapolate the probabilities of longer term events, something that is not appropriate in the semiparametric approach [5]. The assumed the baseline hazard follows a Gompertz distribution. The general form of the parametric form of subdistribution for the main event is:

$$F_1(t; Z) = 1 - \{1 + \alpha_k \exp(Z^T \beta_k) \mu_k(t)\}^{-1/\alpha_k} \tag{2.5}$$

where μ_k represents the baseline subdistribution hazard.

2.0.4 Joint Modeling

Oftentimes, investigators wish to model both longitudinal data and event-time outcomes. Previously, it was acceptable to model each of these outcomes separately. However, depending on the study, it may be entirely reasonable to assume that the longitudinal and survival processes are associated with each other. If this is the case, then modeling the outcomes separately will lead to biased parameter estimates. To account for the association between the longitudinal and survival process, researchers began developing joint models. Typically, joint models link the longitudinal and survival processes through a shared parameter, usually modeled as a random effect. These models usually used a mixed model for the longitudinal process and a proportional hazards survival model.

The joint model by Tsiatis and Davidian [6] is:

$$\prod_{i=1}^{n} \int \Big[\lambda_0(V_i) exp\{\gamma X_i(V_i) + \eta^T Z_i\}\Big]^{\Delta_i} exp\Big[- \int_0^{V_i} \lambda_0(u) exp\{\gamma X_i(u) + \eta^T Z_i\} du\Big] \tag{2.6}$$

$$\times \frac{1}{(2\pi\sigma^2)^{m_i/2}} exp\Bigg[- \sum_{j=1}^{m_i} \frac{\{W_i(t_{ij}) - X_i(t_{ij})\}^2}{2\sigma^2}\Bigg] p(\alpha_i | Z_i; \delta) d\alpha_i \tag{2.7}$$

which describes a longitudinal process and a survival process linked by random effects. These joint models can eliminate the bias caused by association between the survival and longitudinal processes, but model estimation and convergence created new problems. Typically, researchers would have to use Gaussian quadrature approximate the integral over the random effects. Latent class joint models can use group-based longitudinal trajectories and longitudinal data. The latent class joint models link the longitudinal and survival processes through group membership. Using the conditional independence assumption, we can assume that the longitudinal and survival processes are independent given group membership. That is, the longitudinal and survival processes are linked through the groups instead of through shared random effects. Lin, et al [7] specified a latent class joint model to study longitudinal trajectories of prostate specific antigen with onset of prostate cancer. They modeled used the conditional independence function to specify a likelihood that is the product of the multinomial group membership, the longitudinal process, and the survival process. Their log-likelihood was:

$$\sum_{i=1}^{n} log \sum_{k=1}^{K} [c_{ik} = 1 | X_i][y_i | X_i, c_{ik} = 1][N_i, Y_i | X_i, c_{ik} = 1] \tag{2.8}$$

where $[c_{ik} = 1|X_i]$ represents the multinomial group membership, $][y_i|X_i, c_{ik} = 1]$ represents the longitudinal process, and $[N_i, Y_i|X_i, c_{ik} = 1]$ represents the survival process.

3.0 MEASURES OF DISCRIMINATION FOR LATENT GROUP-BASED TRAJECTORY MODELS

3.1 INTRODUCTION

Latent group-based trajectory models[1] have increasing been used to identify distinct trajectory patterns in longitudinal data. One of the main advantages of latent group-based trajectory models is that they allow for the discovery of subgroup behaviors for a population with unobserved heterogeneity across time. For example, if a treatment is administered to a population of patients, there may be distinct response subgroups within that population. One group may respond favorably to the treatment while another group responds negatively, and maybe a third group does not respond at all. Discovering these unobserved patterns of behavior could be extremely useful. Clinicians may target more individual-based therapies based on a patient's profile if they believe a patient falls into one of these subgroups. It is worth noting that a latent group-based trajectory model is a special case of growth mixture models [8]. Latent group-based trajectory models assume a linear model to fit longitudinal trajectories over time, while growth mixture models assume a linear mixed model with possibly random intercepts and random slopes to fit the trajectories. These random components allow individuals within the same trajectory group to vary around the mean group trajectory.

To assess how well a model fits with the data, several goodness-of-fit measures have been proposed for the latent group-based trajectory models. Nagin[1] developed two measures to assess the model adequacy: use of the average posterior probability of assignment (APPA) and use of the odds of correct classification (OCC). Berkhof et al.[9] proposed a discrepancy measure to assess the model fit. Lindsay and Roeder[10] developed gradient-based diagnostic

measures for continuous outcomes and residual plots for discrete outcomes. Agresti[11] proposed the use of G^2 statistic to measure the absolute model fit.

In the analysis using latent group-based trajectory modeling, we need to first assume a certain number of latent groups and then estimate the best trajectory curve for each group via appropriate inference procedures. Several model selection techniques are available to decide what the number of latent groups should be chosen to best fit the data. Note that the likelihood ratio test does not have the usual large sample chi-square distribution properties due to the class probability parameter being at the border of its admissible space[6]. Commonly used model selection techniques include Akaike information criteria (AIC), Bayesian information criteria (BIC), and the bootstrap likelihood ratio test (BLRT). Simulations by Nylund et al.[8] showed that BIC and BLRT outperformed AIC and suggested that BIC and BLRT need to be compared together.

Although there are several methods to test the goodness-of-fit for latent group-based trajectory models, currently, there is no measure of discrimination to check the confidence of an individual being assigned to a certain group. This can be particularly troubling if treatment regimens are determined by an individual's group assignment. For example, in the two-group scenario, an individual may be in group one with a probability of 0.98 and in group two with a probability of 0.02. This individual's group assignment has a high level of discrimination and is assigned to group one. However, another individual may have a group one probability of 0.52 and a group two probability of 0.48. This individual is also assigned to group one even though his or her group membership has a poor level of discrimination. Recognizing this ambiguity may play a large role in how clinicians decide to treat individuals. In this paper, we propose two measures to evaluate discrimination, and they can be used alongside the goodness-of-fit techniques to evaluate latent group-based trajectory models.

In Section 3.2, we introduce the notation and revisit the latent-group trajectory models. In Section 3.3, we introduce the first discrimination measure by modifying entropy, and the second discrimination measure and its corresponding variance estimator based on the posterior probabilities of group membership. In Section 3.4, we conduct simulations to assess the performance of our discrimination measures.

In Section 3.5, we apply the proposed measures to a longitudinal study for the development of conduct disorders among young girls. We present conclusions in Section 3.6.

3.2 NOTATION AND MODEL

The latent group-based trajectory model[1] is a mixture of two components. The group membership is modeled via a multinomial regression and the longitudinal trajectories conditional on a given group membership are modeled via a linear model. The general form of the model can be specified as follows:

$$\sum_{j=1}^{J}\sum_{i=1}^{n_J} Pr(\pi_i = j|Z_i = z_i)Pr(Y_i = y_i|\pi_i = j, W_i = w_i), \tag{3.1}$$

where Z represents time-independent covariates and W represents time-independent or time-dependent covariates. For subject i, the first term of (1) represents the probability of belonging to group membership j,

$$Pr(\pi_i = j|Z_i = z_i) = \frac{\exp(\theta_j + \lambda_j^T z_i)}{\sum_{l=1}^{J}\exp(\theta_l + \lambda_l^T z_i)}. \tag{3.2}$$

The second term of (1) represents the probability of the longitudinal outcomes Y_i given the group membership j,

$$Pr(Y_i = y_i|\pi_i = j, W_i = w_i) = \frac{1}{(2\pi)^{\frac{t}{2}}|\Sigma_j|^{\frac{1}{2}}}\exp\left\{-\frac{1}{2}(y_i - \mu_j)^T\Sigma_j^{-1}(y_i - \mu_j)\right\}. \tag{3.3}$$

Mean μ_j can be specified as a polynomial function of time with the form

$$\mu_j = \beta_0 + \beta_1^T t + \beta_2^T t^2 + \beta_3^T t^3 + ... \tag{3.4}$$

The mean trajectories may also depend on covariates W. When there are only two latent groups involved, the log likelihood can be simplified as the form:

$$\log(L) = \sum_{i=1}^{n}\log\left\{\pi_i MVN(\mu_1, \Sigma_1) + (1 - \pi_i)MVN(\mu_2, \Sigma_2)\right\}, \tag{3.5}$$

where $MVN(\mu, \Sigma)$ is a multivariate normal density with mean vector μ and variance-covariance matrix Σ. The five parameters that need to be estimated from this model are

$\pi_i, \mu_1, \Sigma_1, \mu_2$, and Σ_2. Maximization of the log likelihood function can be done by using the quasi-Newton procedures. The estimated parameters are necessary for calculating the posterior probability of an individual being in a particular group.

3.3 DISCRIMINATION STATISTICS

Once parameter estimators are obtained, posterior probabilities can be calculated. Using the Bayes rule, the posterior probability of individual i belonging to group j given his or her longitudinal trajectory is

$$\hat{P}(\pi_i = j|Y_i) = \frac{\hat{P}(Y_i|\pi_i = j)\hat{\pi}_j}{\sum_j^J \hat{P}(Y_i|\pi_i = j)\hat{\pi}_j}. \tag{3.6}$$

For example, if there are 2 groups ($j = 2$), each individual will have a probability of being in group 1 and a probability of being in group 2. The group assignment depends on the largest of the two posterior probabilities. As mentioned above, the level of discrimination plays no part in group assignment. Therefore, individuals whose posterior probabilities are highly ambiguous are still assigned to groups just as individuals whose posterior probabilities are highly discriminated are.

Entropy is a statistic used to measure the amount of information or the degree of classification uncertainty in various fields including latent-class analysis. Individual-level entropy[12] is defined as

$$EN = -\sum_J \hat{p}_j \log(\hat{p}_j), \tag{3.7}$$

where p_j is an individual's posterior probability of being in group j. Larger value of entropy indicates higher level of uncertainty in discrimination. Therefore, subjects who are poorly discriminated should have higher values of entropy than subjects who are well discriminated. One important caveat is that entropy is based on all posterior probabilities, but for group assignment, we are most interested in the gap between the highest posterior probability and the second highest posterior probability. For example, in a four group scenario, if a subject's posterior probabilities are 0.25 for each group, discrimination will be poor and entropy high.

However, if the posterior probabilities are 0.4, 0.2, 0.2, and 0.2, entropy will still be relatively high even though we may confidently be able to assign the subject to the group with posterior probability 0.4. Since it is essentially the leading two posterior probabilities that determine discrimination status, we propose a modification of the entropy measure by considering only these leading two posterior probabilities. The modified entropy has the form

$$EN_m = -\left\{\hat{p}_{max}\log(\hat{p}_{max}) + \hat{p}_2\log(\hat{p}_2)\right\}, \tag{3.8}$$

where $\hat{p}_{max}$ and $\hat{p}_2$ are the largest and the second largest posterior probabilities, respectively.

To build up our second discrimination measure, we will first construct a confidence interval around the maximum posterior probability. An individual's discrimination will then be determined by whether the confidence interval of the maximum posterior probability contains the value of $(\hat{p}_{max} + \hat{p}_2)/2$. Another way to represent this is

$$\hat{p}_{max} - z_{\alpha/2}\frac{sd}{\sqrt{n}} < \frac{p_{max} + p_2}{2} < \hat{p}_{max} + z_{\alpha/2}\frac{sd}{\sqrt{n}}, \tag{3.9}$$

where sd represents the standard deviation of the $\hat{p}_{max}$ estimate.

We considered several methods to estimate the variance of posterior probabilities of group membership. One technique was calculating the distribution of the order statistic of the posterior probabilities. However, this would necessitate knowing how the probabilities are distributed. Another method considered was using the bootstrap technique[13], but it is very computationally intensive. We will adopt the method proposed by Menses et al.[14] to estimate the variance of the posterior probabilities. They derived the variance estimator using the delta method,

$$Var\left\{P(\pi_i = j|Y_i)\right\} \approx \left\{\frac{\partial P(\pi_i = j|Y_i)}{\partial y}\right\}^T Var(Y)\left\{\frac{\partial P(\pi_i = j|Y_i)}{\partial y}\right\}. \tag{3.10}$$

For simplicity, we define

$$S_j = \frac{\partial \hat{P}(\pi_i = j|Y_i)}{\partial y}, \tag{3.11}$$

and A as the denominator of the posterior probability, $A = \sum_{j=1}^{J}\hat{P}(Y_i|\pi_i = j)\hat{\pi}_j$. Therefore, S_j can be rewritten as the form $S_j = \left\{\hat{\pi}_j MVN(\hat{\mu}_j, \hat{\Sigma}_j)\right\}\left\{-\hat{\Sigma}_j(y - \hat{\mu}_j)\right\}$. Finally, the variance estimator of the posterior probability can be simplified as

$$Var\left\{\hat{P}(pi_i = j|Y_i)\right\} \approx \sum_j^J \left\{\frac{(S_j A - \sum_j^J S_j)\hat{\pi}_j MVN(\hat{\mu}_j, \hat{\Sigma}_j)}{A^2}\hat{\Sigma}_j\right\}. \tag{3.12}$$

3.4 SIMULATIONS

We simulated 500 datasets with a sample size of 500 to test the performance of the parameter estimates. For each dataset, we assumed three trajectories across three time points. The longitudinal trajectories are shown in Figure 1.

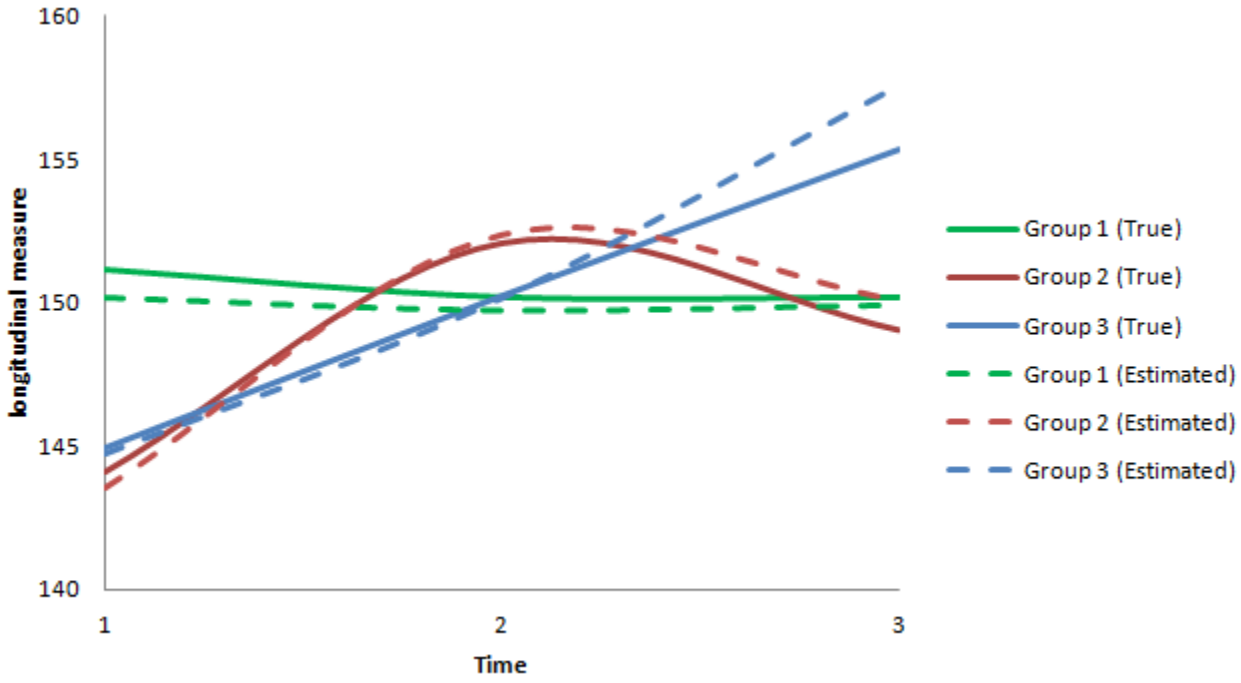

Figure 1: Three Group Trajectories

Table 1 summarizes the data generated according to simulated parameters along with the estimates. The results in Table 1 show that the estimated parameters are close to the generated parameters. The three-group trajectory model also estimated the trajectories close to the underlying setting. The two group model and four group models were also fit, and evaluation of the BIC showed that indeed the three group model is optimal.

To demonstrate the performance of our discrimination measures, we generated a single data set with N=150 and with three distinct groups of longitudinal trajectories (150,150,150), (145,152,148), and (145,150,155) over three time points.

Table 1: Posterior group membership probabilities with longitudinal trajectories.

	Simulated		Estimated	
Group j	Probability of group membership (π_j)	Trajectories at (t_1, t_2, t_3)	Probability of group membership ($\hat{\pi}_j$)	Trajectories at (t_1, t_2, t_3)
1	0.333	(150,150,150)	0.335	(150,150,150)
2	0.333	(145,152,148)	0.332	(145,152,148)
3	0.333	(145,150,155)	0.333	(145,150,155)

Once the likelihood was maximized and parameter estimates were obtained, we calculated the posterior probabilities of group membership for each individual.

The model poorly discriminated 21 out of 150 (14 percent) subjects according to the discriminant index, our second discrimination measure. Figure 2 depicts the density plots of entropy by the discrimination status. In fact, well discriminated subjects had a wide range of entropy, while poorly discriminated subjects tended to be at the upper end of the scale. Entropy ranged from 0.041 to 1.016 for the well discriminated subjects and from 0.685 to 1.094 for the poorly discriminated subjects. The overlap between discrimination and entropy takes place mainly when the posterior probability for one group is very low and nearly split between the two remaining groups. Since entropy is a measure of information, the measure provides information indicating that a subject most likely did not belong to the group with very low posterior probability. However, it has no way of discriminating between the other two groups. Two examples of this are shown in Table 2.

Figure 2 shows how entropy differs by discrimination status. It is evident that subjects who are poorly discriminated have higher levels of entropy than the well discriminated subjects, even though there is some overlap. Figure 2 also shows the distributions of entropy, with the solid line representing the density for well discriminated subjects and the dashed line for poorly discriminated subjects. Our modified entropy measure performs better. As expected, there is more agreement between discrimination status and modified entropy.

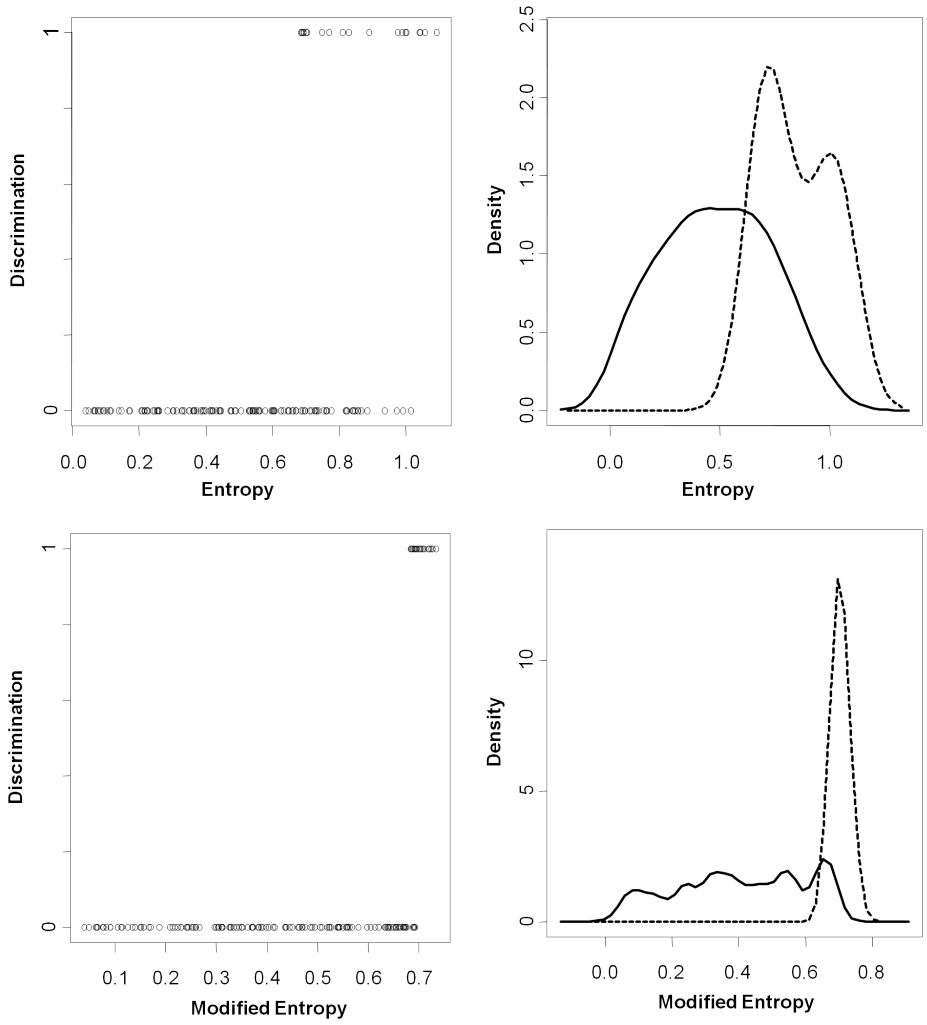

Figure 2: Three group entropy and modified entropy. The dashed line represents poorly discriminated subjects, and the solid line represents acceptably discriminated subjects.

3.5 EXAMPLE: THE PITTSBURGH GIRLS STUDY

The Pittsburgh Girls Study (PGS)[15] is a longitudinal study to follow an urban population sample of girls in Pittsburgh, Pennsylvania. The first assessment wave consisted of girls 5-8 years old. One objective of the study was to test developmental models for conduct

Table 2: Entropy and discrimination for two simulated subjects. This shows how entropy and discrimination can differ.

ID	$\hat{p}_1$	$\hat{p}_2$	$\hat{p}_3$	Poor discrimination (0=no, 1=yes)	Entropy
115	0.26	0.21	0.53	0	1.02
145	0.01	0.47	0.52	1	0.75

disorder (CD). We fit latent group-based trajectory models to uncover distinct longitudinal trajectories of CD severity scores and then applied the discrimination index to show how many subjects are well and poorly discriminated into these groups. There were separate cohorts depending on the starting age of the child. We examined the cohort that entered as five year olds (Cohort 5). Note that CD severity scores were based on the yearly self-reported information. Higher scores indicate more severe conduct problems while lower scores indicate fewer problems.

Cohort 5 consists of 588 subjects followed yearly from age 5-14. We analyze self reported data, which were collected from age 7 onwards. Only complete data cases are used, which reduced the dataset to N=471. There was no obvious pattern to the missing observations, and therefore they are assumed to be missing completely at random. We fit a latent-class longitudinal trajectory model with three groups, which were depicted in Figure 3. The number of groups was chosen based on BIC and clinical input to maintain a manageable number of groups.

Figure 3 shows the model results for the Pittsburgh Girls Study data. The trajectories showed one group (black, 90%) that made up the majority of the cohort. These girls had a consistently low CD score over time.

The group most interesting to researchers was denoted by the green line, and they made up 6.4% of the cohort. This is the group of girls whose conduct got worse over time. This group consists of girls whose conduct worsens as they age, and may be a signal to researchers that this group requires early intervention. It may also allow researchers to focus their

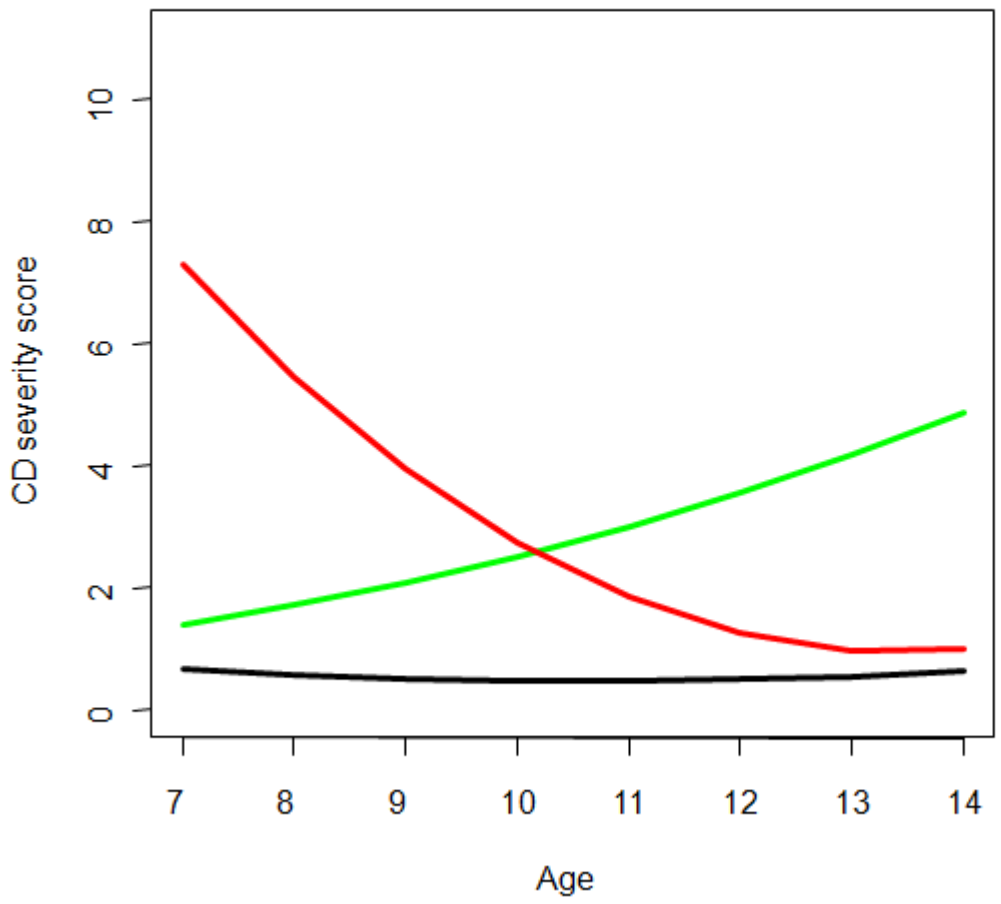

Figure 3: Trajectory plot for The Pittsburgh Girls Study. Group membership and descriptions can be seen in Table 3.

efforts on this particular group to discover why they are getting worse over time. The results showed overlaps in the trajectories, which may indicate a high level uncertainty in group assignments. However, application of the discriminant index showed that only 5 of the 471 (1.1%) subjects were poorly discriminated. Overall, the subjects were very well discriminated into their groups. This may be due to the fact that one group contained a large proportion of the subjects.

As Table 3 shows, poor discrimination rates were higher in the green group. Modified entropy also behaved as expected, with poorly discriminated subjects having higher modified entropy than acceptably discriminated subjects.

Table 3: Pittsburgh Girls Study results. The table shows group membership, percentage of poor discrimination, and the range of modified entropy.

Group	(n, %)	Poorly discriminated (n, %) within the group	Range of modified entropy among those not poorly discriminated	Range of modified entropy among those poorly discriminated
Black	(423, 89.9%)	(3, 0.7%)	0.00-0.61	0.68-0.69
Green	(29, 6.1%)	(2, 6.7%)	0.00-0.55	0.66-0.67
Red	(19, 4.0%)	(0, 0.0%)	0.00-0.55	NA

3.6 CONCLUSIONS

The discrimination index and modified entropy are useful tools for evaluating latent group-based trajectory models. The discrimination index based on the delta methods can successfully identify subjects whose discrimination is too poor to confidently be assigned to a particular latent group. While entropy or modified entropy can help us measure the amount of uncertainty in discrimination, this index can place confidence intervals and help us develop cut-off rules in order to identify which subjects are poorly discriminated. This can be very important, especially if interventions differ by group assignment. The index serves two purposes: first, to determine which individuals are poorly discriminated into their groups, and second, as a general test to evaluate the latent group-based trajectory model. Applying the method to The Pittsburgh Girls study showed that overall, the level of discrimination is very good. The discrimination index also identified which subjects were poorly discriminated during group assignment. The discriminant index provides a formal statistical test to determine an individual's group membership status, and should be used in tandem with goodness-of-fit methods to evaluate latent group-based trajectory models.

4.0 JOINT MODELING OF LATENT GROUP-BASED TRAJECTORY MODELS WITH SUBDISTRIBUTIONS

4.1 INTRODUCTION

Joint modeling can be used to simultaneously model longitudinal and survival data. However, many joint modeling techniques assume a homogeneous subject trajectory across time. Many joint modeling techniques also assume that the survival outcome is due to the cause of interest. There are scenarios where we need to account for latent class trajectories in the longitudinal data and we also need to account for competing risks in survival analysis. An example of ICU data by Deslandes and Chevret involved examining SOFA (Sequential Organ Failure Assessment) scores over time [16]. Their primary endpoint was 28 day survival, and their secondary endpoint was SOFA score measured over time. The authors used joint modeling to estimate the treatment effect on SOFA score. There are other applications where there may be different sub-classes of behavior within the population. A joint modeling approach has also been used (Lin, et al 2002) to examine diagnosis of prostate cancer with longitudinal biomarker measurements in a highly heterogeneous population.

Longitudinal data can be analyzed using mixed-effects models, but the analysis can be complicated due to informed dropout. For example, following people across time when modeling disease processes will result in some subjects dying or dropping out of the study for various reasons. A two-stage approach, where one estimated the longitudinal effects and uses them as covariates in a survival model, had previously been used [17], but this approach leads to biased estimates. Joint models typically look at longitudinal and survival data, treating the longitudinal data as a mixed-effects model and the survival data as Cox proportional hazards. As more researchers started using joint modeling, different applications necessitated

the modification of the models. Lin, et al [7] used a joint modeling approach to study whether a biomarker is related to onset of prostate cancer. The particular biomarker was comprised of highly heterogeneous trajectories, making the usual mixed model approach less than ideal. The authors used latent class analysis to discover the unobserved heterogeneity in the data, then fit each longitudinal group individually as a joint model. Other approaches have used different methods of survival analysis. For example, Deslandes and Chevret [7] used joint models to look at a treatment effect on SOFA scores over time. However, in ICU data, being discharged alive can often lead to informative censoring, so they used a competing risks survival approach to deal with this issue. However, no one has yet to model longitudinal data as group based trajectories and survival data as competing risks. Here we present a joint model framework that simultaneously models group based trajectories with competing risks survival outcomes.

In Section 4.2, we introduce model notation for the latent-group trajectory model, the competing risks survival model. In Section 4.3, we revisit the conditional independence assumption and present the latent class joint model. In Section 4.4, we conduct simulations to assess the performance of the joint model. In Section 4.5, we apply the proposed model to a study investigating biomarkers that may be associated with recovery from acute renal failure. We present conclusions in section 4.6.

4.2 NOTATION AND MODEL

4.2.1 Longitudinal Process

The latent group-based trajectory model [1] is a mixture of two components. The group membership is modeled via a multinomial regression and the longitudinal trajectories conditional on a given group membership are modeled via a linear model. The general form of the model can be specified as follows:

$$\sum_{j=1}^{J}\sum_{i=1}^{n_J} Pr(\pi_i = j|Z_i = z_i)Pr(Y_i = y_i|\pi_i = j, W_i = w_i), \tag{4.1}$$

where Z represents time-independent covariates for the group membership and W represents time-independent or time-dependent covariates for the longitudinal trajectory. π_i represents group membership. For subject i in group j, the first term of (4.1) represents the probability of belonging to group j,

$$\pi_i^j = Pr(\pi_i = j|Z_i = z_i) = \frac{\exp(\theta_j + \lambda_j^T z_i)}{\sum_{l=1}^{J} \exp(\theta_l + \lambda_l^T z_i)}. \tag{4.2}$$

The second term of (4.1) represents the probability of the longitudinal outcomes Y_i given the group membership j,

$$P^j(Y_i) = Pr(Y_i = y_i|\pi_i = j, W_i = w_i) = \frac{1}{(2\pi)^{\frac{t}{2}}|\Sigma_j|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(y_i - \mu_j)^T \Sigma_j^{-1}(y_i - \mu_j)\right\}. \tag{4.3}$$

Mean values for the longitudinal outcomes for group j, $\mu_j = E[Y]$, depend on covariates W and are specified as a polynomial function of time, (e.g. cubic) with the form

$$\mu_j = \beta_0 + \beta^T w_i + \beta_1^T t + \beta_2^T t^2 + \beta_3^T t^3. \tag{4.4}$$

4.2.2 Survival Process

The cumulative incidence function under the proportional subdistribution hazards assumption has the form [4]

$$F_k(t; Z) = 1 - \exp\{-\exp(Z^T \beta_k)\mu_k(t)\} \tag{4.5}$$

where $\mu_k(t) = \log_k\{\int_0^t \lambda_{k0}(s)ds\}$ is the log baseline cumulative subdistribution hazard function and k is cause of death. This is the Fine and Gray cumulative incidence shown by Jeong and Fine (2007). With this form, Fine and Gray use a semi-parametric baseline survival, while Jeong and Fine [5] assume a Gompertz distribution, i.e., $\mu^G(t; \rho, \tau) = \tau\{\exp(\rho t) - 1\}/\rho$. In this study, we assume a Gompertz parameter form as defined in Jeong and Fine [5].

When $\rho < 0$ and $t \to \infty$, there is a set of subjects who would never experience event k. This set of subjects is called the "cured" subjects and the cure fraction can be obtained by

$$\lim_{t\to\infty} F(t; Z) = 1 - \exp\{\tau_k \exp(Z^T \beta_k)/\rho_k\}. \tag{4.6}$$

Jeong and Fine [5] specify a parametric survival model with likelihood:

$$\prod_{i=1}^{n}\left[\Big\{\prod_{k=1}^{n_K} f_k(t_i,\psi_k;z_i)^{\delta_{ki}}\Big\}\Big\{1-\sum_{k=1}^{n_K}F_k(t_i,\psi_k;z_i)\Big\}^{1-\sum_{k=1}^{n_K}\delta_{ki}}\right] \tag{4.7}$$

where $\psi_k = (\beta_k, \rho_k, \tau_k)$, and δ is the censoring indicator.

We define $f_k(t_i, \psi_k; z_i) = dF_k(t_i, \psi_k; z_i)/dt$, therefore:

$$f_k(t_i,\psi_k;z_i) = \exp\{-\exp(Z^T\beta_k)\tau_k\exp(Z^T\beta_k)\{\exp(\rho_k t)-1\}\}\tau_k\exp(z^T\beta_k)\exp(\rho_k t) \tag{4.8}$$

4.2.3 Joint Model

The formulation of latent class joint model begins with the conditional independence assumption:

$$[Y_i, T_i|Z, j] = [Y_i|W, j][T_i|Z, j]. \tag{4.9}$$

The conditional independence assumption assumes that the longitudinal and survival processes (Y_i and T_i) are independent given the group membership j. The longitudinal and survival processes are linked through the latent classes. $[Y_i, T_i|Z, j]$ represents the joint longitudinal and survival process given a set of baseline covariates, Z and the latent group j. $[Y_i|Z, j]$ is the longitudinal process given a set of baseline covariates Z and group j. The survival process is $[T_i|Z, j]$, given a set of covariates Z and group j.

Extending this, Lin, et al. [7] show the framework for the log-likelihood of the joint model::

$$\sum_{i}^{n} log \sum_{j=1}^{J}\Big[f(j|Z)f(y_i|Z,j)f(t_i|Z,j)\Big] \tag{4.10}$$

For our joint model, from Equation 4.1 we set the latent group process as:$f(j|Z) = Pr(\pi_i = j|Z_i = z_i)$. Then we set:

$$f(y_i|Z, j) = Pr(Y_i = y_i|\pi_i = j, W_i = w_i) \tag{4.11}$$

where $Y_i \sim MVN(u_j, \Sigma)$ for the longitudinal process, also from Equation 4.1.

The survival process, from Equation 4.7, becomes:

$$f(t_i|Z,j) = \Big\{\prod_{k=1}^{n_K} f_k(t_i,\psi_k;z_i)^{\delta_{ki}}|\pi_i = j\Big\}\Big\{\Big\{1-\sum_{k=1}^{n_K}F_k(t_i,\psi_k;z_i)\Big\}^{1-\sum_{k=1}^{n_K}\delta_{ki}}|\pi_i = j\Big\} \tag{4.12}$$

Using the conditional independence assumption, we can combine the group-based longitudinal process and the competing risks survival process in a joint model.

$$\prod_{i}^{n}\Bigg[\sum_{j=1}^{J} Pr(\pi_i = j|Z_i = z_i)Pr(Y_i = y_i|\pi_i = j, W_i = w_i) \times$$
$$\Big\{\prod_{k=1}^{n_K} f_k(t_i, \psi_k; z_i)^{\delta_{ki}}|\pi_i = j\Big\}\Big\{1 - \sum_{k=1}^{n_K} F_k(t_i, \psi_k; z_i)\Big\}^{1-\sum_{k=1}^{n_K}\delta_{ki}}|\pi_i = j\Bigg] \tag{4.13}$$

Since we model the group membership as a multinomial and the longitudinal data as a multivariate normal to represent the vector of group means across time, Equation 4.13 is equivalent to:

$$\prod_{i}^{n}\Bigg[\sum_{j=1}^{J}\Big[\frac{exp(\theta_j + \lambda_j^T z_i)}{\sum_{l=1}^{J} exp(\theta_l + \lambda_l^T z_i)}\{MVN(\mu, \Sigma)|\pi_i = j\} \times$$
$$\Big\{(\prod_{k=1}^{n_K} f_k(t_i, \psi_{jk}; z_i)^{\delta_{ki}})(1 - \sum_{k=1}^{n_K} F_k(t_i, \psi_{jk}; z_i))^{1-\sum_{k=1}^{n_K}\delta_{ki}})|\pi_i = j\Big\}\Big]\Bigg] \tag{4.14}$$

and $\psi_{jk} = (\beta_k, \rho_{jk}, \tau_{jk})$.

4.3 SIMULATIONS

We simulated longitudinal and survival data for one-group, two-group, three-group, and four-group models. The longitudinal data was created for three time points and generated based on a multivariate normal distribution for each individual. Survival time was generated from a Gompertz proportional subdistribution hazards model and the corresponding parameter values were set based on the Jeong and Fine results. Approximately 75% of the observations are censored, 15% experience the main event, and 10% experience the competing event.

We generated longitudinal data for each individual based on a multivariate normal distribution. Each individual's longitudinal trajectory will follow $Y_i \sim MVN(\mu_j, \Sigma)$ where μ_j is a vector of group means at each time point and Σ is a matrix that represents the variance at each time point. For the two-group simulations, we generated one group to be flat across three time points, with each group mean generated with mean 150 and variance 8. Our second group increased over time. We generated group means of 150, 160, 170 across three

time points with variance 8. For the two group model, we generated 50% in each group. A sample of simulated data is shown below in Table 4. Figure 1 depicts these two trajectories over time.

To generate the proportion of subjects experiencing the main event, we look at the cumulative incidence function as $t \to \infty$ to represent the proportion of subjects experiencing the event. The proportion of subjects experiencing the main event can be shown[5] as:

$$F_{1j}(\infty; \psi_{kj}, Z) = 1 - \exp\{\tau_{1j} \exp(Z^T \beta_{1j})/\rho_{1j}\} \tag{4.15}$$

Where j represents the latent class and k represents the event type. Therefore, we generated a proportion of subjects to experience the main event with probability $F_{1j}(\infty; \psi_{1j}, Z)$ and competing event with probability $F_{2j} = 1 - F_{1j}(\infty; \psi_{kj}, Z)$. We denote the main event as $\delta = 1$ and the competing event as $\delta = 2$.

A treatment variable was generated from a Bernoulli random variable with $p = 0.5$.

To obtain the survival time using the inverse transformation method, we generate a random uniform variable and derive survival time T for subject i follows:

$$U(0,1) = \frac{F_{kj}(t, Z)}{F_{kj}(\infty, Z)}, \tag{4.16}$$

where U is a random uniform variable. Expanding the equation yields:

$$U(0,1) = \frac{1 - \exp\{-\exp(Z^T \beta_{kj})\tau_{kj}\{\exp(\rho_{kj} t) - 1\}/\rho_k\}}{1 - \exp\{-\tau_{kj} \exp(Z^T \beta_{kj})\}}, \tag{4.17}$$

Solving for t,

$$T_i = \frac{\log[1 - \frac{\rho_{kj} log[1 - U * A]}{\tau_{kj} \exp(Z^T \beta_{kj})}}{\rho_{kj}}, \tag{4.18}$$

where $A = 1 - \exp\{-\tau_{kj} \exp(Z^T \beta_{kj})/\rho_{kj}\}$.

To generate roughly 75% censoring, we used the following procedure to calculate censoring time:

$$C_i = -\frac{log(U(0,1))}{0.1}. \tag{4.19}$$

We let $X_i = min(T_i, C_i)$, which indicates the observed event time experienced by subject i. The censoring indicator, δ, is defined as:

$$\delta = I(C < t) = 0 \tag{4.20}$$

Table 4: Simulated longitudinal and competing risks data

Y_1	Y_2	Y_3	treatment	event time	δ
147.3421	164.2374	169.8981	0	0.041	1
147.3421	164.2374	169.8981	0	1.048	0
150.7401	147.9986	149.0599	1	0.6939	0
145.2746	157.6325	168.6481	1	7.444	2
150.0821	157.3202	172.5691	1	2.837	1
153.6890	161.3567	170.4538	0	57.684	0
148.9496	163.9706	171.9572	1	11.331	0
149.4707	152.2447	148.7789	0	9.790	0
149.0810	147.0524	146.7202	1	7.907	0
150.0501	157.8537	169.5550	1	4.247	0

We present two simulation scenarios.One in which the latent groups are driven by the longitudinal data, and one where they are driven by the survival process In the first scenario, the the two groups have different longitudinal processes but the same survival process.If the estimation method works, we expect the method identifies two subgroups from the population. We generated survival data similar to the dataset used in Jeong and Fine [6]. Table 5 shows the true and estimated parameter values under scenario 1.

Table 5: Two group simulation

parameter	simulated value	estimated value	se	bias
θ	0	0.002	0.062	0.002
b_{01}	150	150.004	0.114	0.004
b_{11}	0	0.0004	0.09	0.0004
b_{02}	150	150.007	0.11	0.007
b_{12}	10	9.997	0.09	0.003
Σ	8	8.02	0.19	0.02
τ_{11}	0.08	0.081	0.014	0.001
ρ_{11}	-0.25	-0.253	0.04	0.003
β_1	-0.54	-0.533	0.15	0.007
τ_{21}	0.01	0.0099	0.002	0.0001
β_2	-0.1	-0.066	0.204	0.034
τ_{12}	0.08	0.081	0.014	0.001
ρ_{12}	-0.25	-0.254	0.042	0.004
τ_{22}	0.01	0.01	0.002	0.002

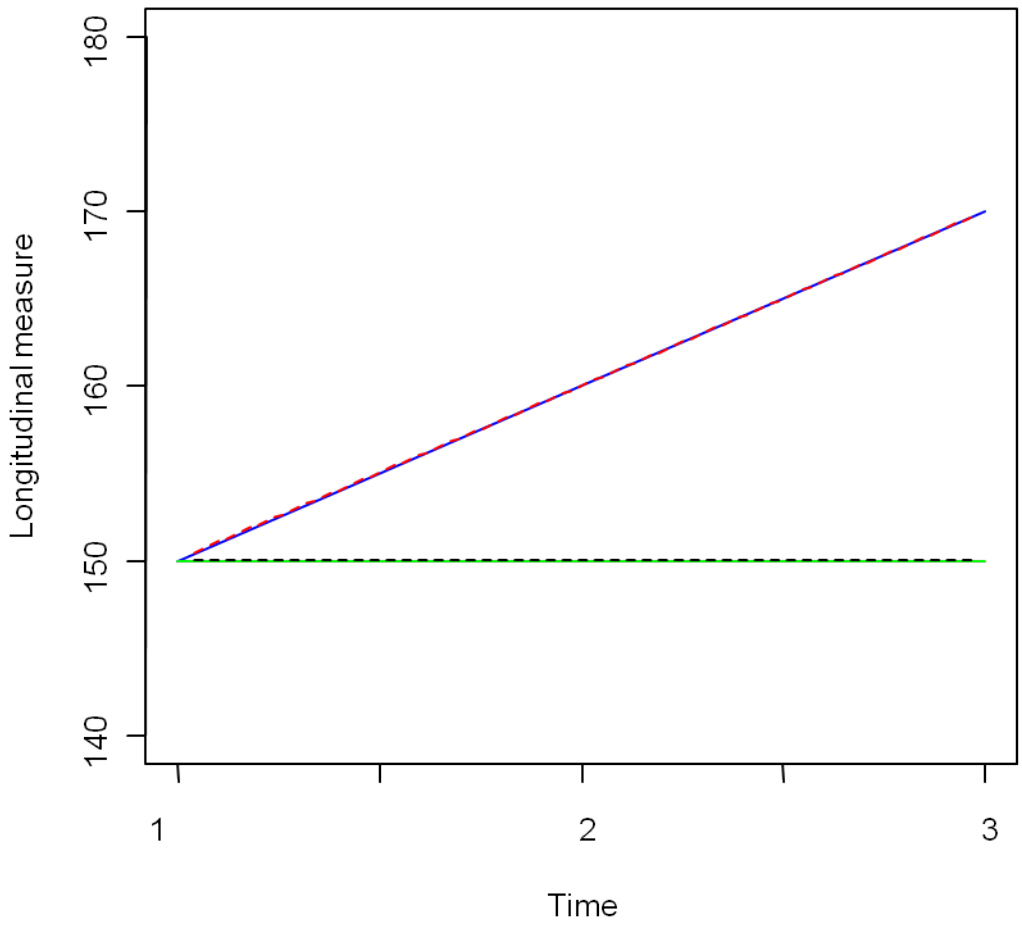

Figure 4: Simulation Scenario 1: Cumulative incidence functions.

In Table 5, b_{01} is the intercept of the longitudinal trajectory for group one and b_{11} is the slope. b_{02} is the intercept and b_{12} the slope for the longitudinal trajectory of group two. The model appears to do a good job estimating the group membership, the longitudinal trajectories, and the survival parameters. Figure 4 shows the cumulative incidence for the main event, and Figure 5 shows the longitudinal trajectories.

We then created another two group scenario where all subjects had the same longitudinal trajectory with two different survival patterns. Parameter estimates are shown in Table 6.

Figure 6 shows the simulated and estimated cumulative incidence functions. Here, the group membership estimate shows some bias. We generated a 50-50 split in the groups, but we get a probability of group membership of 58% for group 1 and 42% for group 2. Even though the membership parameter is a bit biased, we still get a much better picture of the data than we would with the two staged procedure. The two-stage procedure models the latent classes based solely on the longitudinal data, then stratifies the survival process by group membership. The two-stage procedure yields only one longitudinal trajectory. The two-stage approach then assumes there is only one latent group, and therefore one survival process instead of two. The results from the two-staged approach are shown in Tables 7 and 8. The two-stage longitudinal plot is shown in Figure 7 and the two-stage cumulative incidence function is shown in Figure 8.

We then created three and four-group simulations. The three group simulation is shown in Table 9 and the four-group simulation is shown in Table 10. These simulations assume most of the variation is from the longitudinal trajectories. Overall, the models perform well except for the group membership parameter when all of the differing behavior in the groups is due to the survival process.

Table 6: Two group simulation, varying survival

parameter	simulated value	estimated value	se	bias
θ	0	0.336	0.539	0.336
b_{01}	150	149.99	0.405	0.01
b_{11}	0	-0.0002	0.334	0.0002
b_{02}	150	150.002	0.299	0.002
b_{12}	00	0.001	0.254	0.001
Σ	8	7.89	0.28	0.11
τ_{11}	0.05	0.058	0.069	0.008
ρ_{11}	-0.30	-0.293	0.04	0.007
β_1	-0.54	-0.591	0.271	0.051
τ_{21}	0.03	0.035	0.014	0.005
β_2	-0.1	-0.12	0.244	0.02
τ_{12}	0.08	0.079	0.05	0.001
ρ_{12}	-0.25	-0.259	0.123	0.009
τ_{22}	0.01	0.009	0.007	0.001

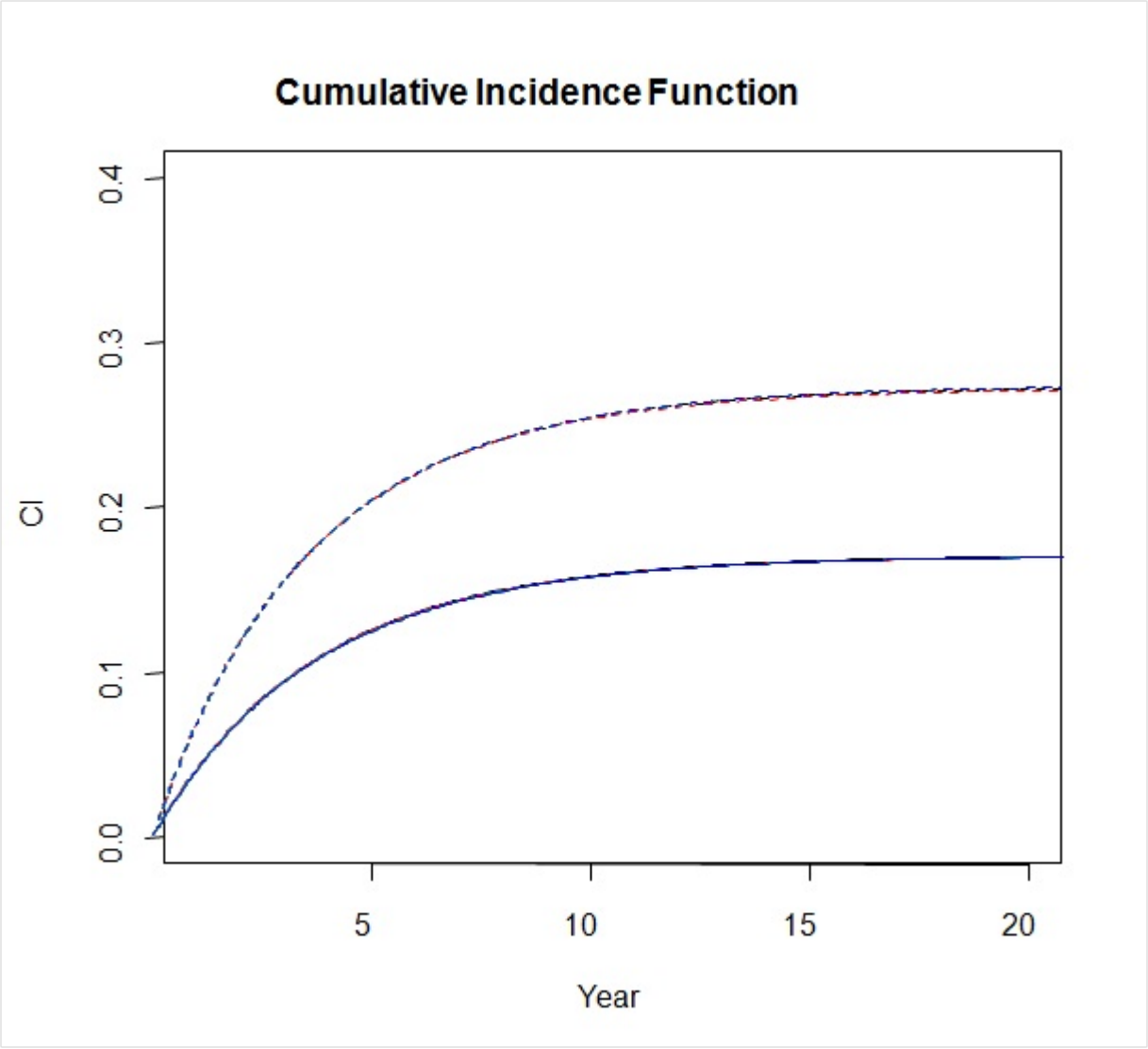

Figure 5: Simulation Scenario 1: Cumulative incidence.

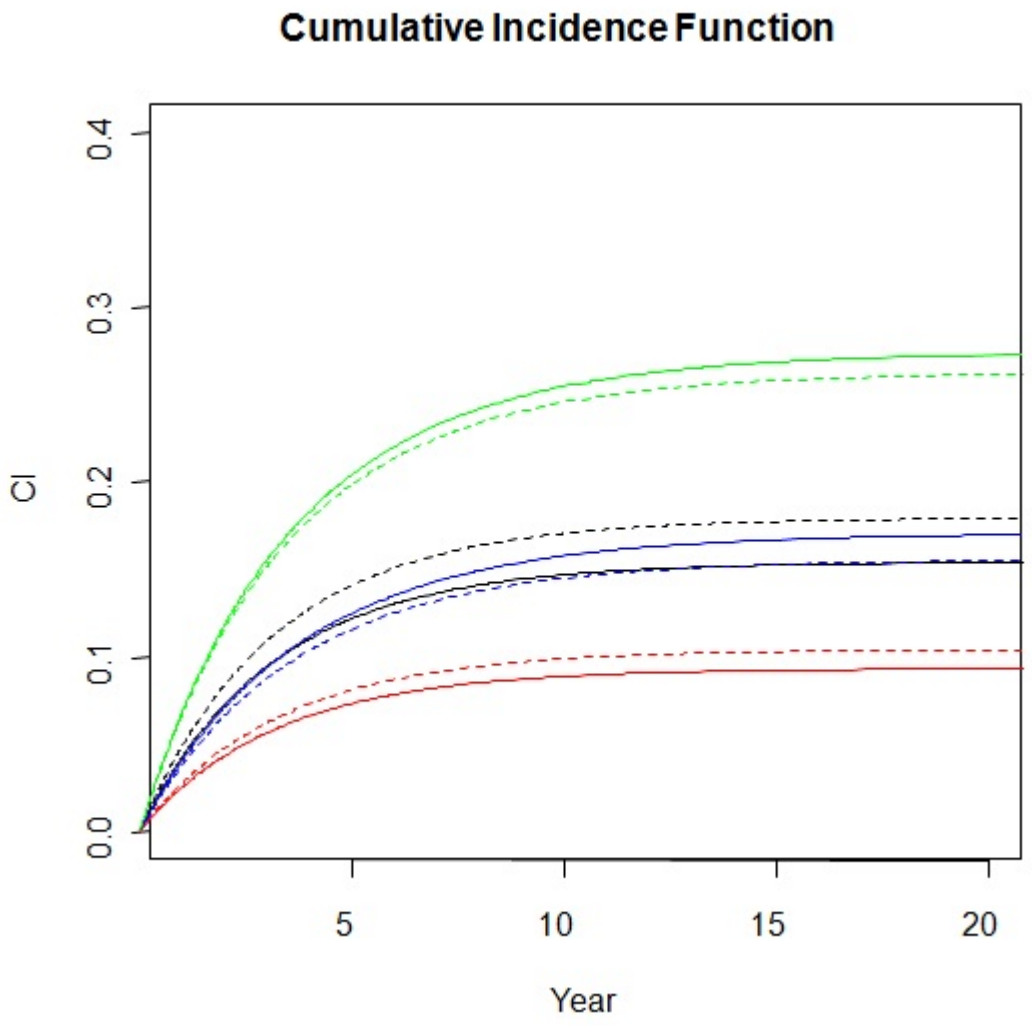

Figure 6: Simulation Scenario 2: Cumulative incidence functions Red is group=1 treatment=1. Black is group=1 treatment=0. Blue is group 2 treatment 1. Green is group=2 treatment=0. Solid represents simulated distribution and dotted is estimated.

Table 7: Two group simulation, varying survival: Two stage results- longitudinal trajectories

parameter	simulated value	estimated value
b_0	150	150
b_1	0	-0.0003
Σ	8	7.99

Table 8: Two group simulation, varying survival: Two stage results- survival outcomes

parameter	estimated value
τ_1	0.658
ρ_1	-0.274
β_1	-0.541
β_2	0.182
τ_2	-0.089

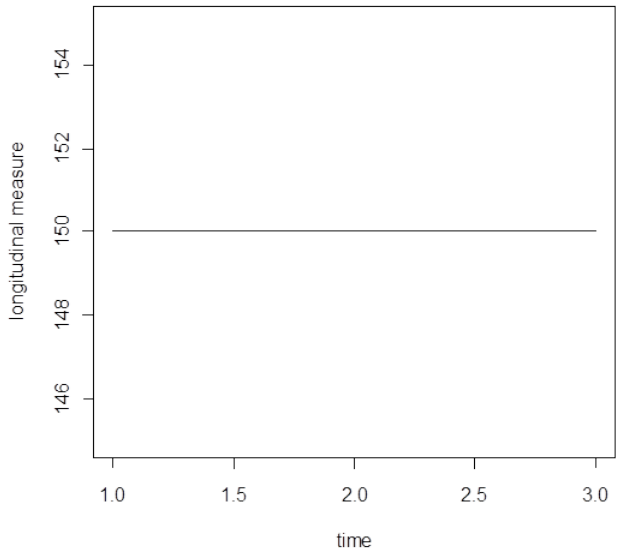

Figure 7: Two stage scenario 2: longitudinal trajectory

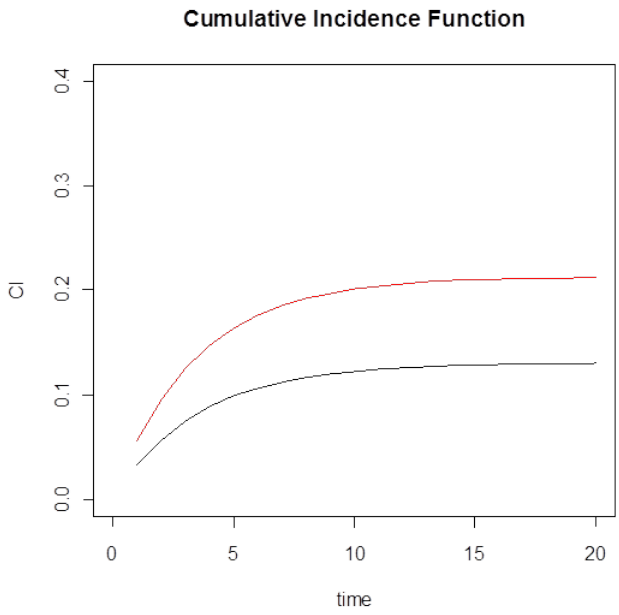

Figure 8: Two stage scenario 2: Cumulative incidence functions.

Table 9: Three group simulation

parameter	simulated value	estimated value	se	bias
θ_1	0	0.007380429	0.013660083	0.007380429
θ_2	0	0.007251138	0.013720109	0.007251138
b_{01}	150	149.995523	0.122708888	0.004477049
b_{11}	0	0.006767094	0.096443233	0.006767094
b_{02}	150	149.9983398	0.123857727	0.001660229
b_{12}	10	10.00891278	0.093364758	0.008912777
b_{03}	150	150.0067443	0.124174321	0.006744286
b_{13}	-10	-9.99833542	0.096617366	0.00166458
Σ	8	7.997340859	0.165837733	0.002659141
τ_{11}	0.08	0.080214294	0.014670442	0.000214294
ρ_{11}	-0.25	-0.252296777	0.047264329	0.002296777
β_1	-0.54	-0.51605992	0.130987059	0.02394008
τ_{21}	0.01	0.009877407	0.001933054	0.000122593
β_2	-0.10	-0.076655961	0.174349748	0.023344039
τ_{12}	0.08	0.080088833	0.014930994	8.88E-05
ρ_{12}	-0.25	-0.253980648	0.045521539	0.003980648
τ_{22}	0.01	0.010019783	0.002018567	1.98E-05
τ_{13}	0.08	0.080282881	0.015406932	0.000282881
ρ_{13}	-0.25	-0.253726204	0.048262471	0.003726204
τ_{23}	0.01	0.009937464	0.002076849	6.25E-05

Table 10: Four group simulation

parameter	simulated value	estimated value	se	bias
θ_1	0	0.014000978	0.020429434	0.014000978
θ_2	0	0.014315229	0.020196247	0.014315229
θ_3	0	0.013939894	0.020243941	0.013939894
b_{01}	150	150.0046316	0.129163406	0.004631618
b_{11}	0	0.002427609	0.108466652	0.002427609
b_{02}	150	150.0015145	0.123987557	0.001514543
b_{12}	10	10.0044659	0.102332277	0.004465903
b_{03}	150	150.0149517	0.134645141	0.014951692
b_{13}	-10	-10.00776779	0.102705067	0.007767786
b_{04}	150	150.0197878	0.132423656	0.019787751
b_{14}	50	49.99808612	0.103308052	0.001913883
Σ	8	8.018753658	0.163834362	0.018753658
τ_{11}	0.08	0.078639747	0.016839896	0.001360253
ρ_{11}	-0.25	-0.249539219	0.052761025	0.000460781
β_1	-0.54	-0.518097768	0.131984547	0.021902232
τ_{21}	0.01	0.009894038	0.002062416	0.000105962
β_2	-0.1	-0.086333667	0.17326322	0.013666333
τ_{12}	0.08	0.0815269	0.017099368	0.0015269
ρ_{12}	-0.25	-0.257555018	0.05485795	0.007555018
τ_{22}	0.01	0.009801578	0.002185425	0.000198422
τ_{13}	0.08	0.080739008	0.016922616	0.000739008
ρ_{13}	-0.25	-0.254859542	0.05551281	0.004859542
τ_{23}	0.01	0.010071575	0.002132759	7.16E-05
τ_{14}	0.08	0.080912189	0.016942175	0.000912189
ρ_{14}	-0.25	-0.254893844	0.052001886	0.004893844
τ_{24}	0.01	0.009855543	0.002278104	0.000144457

We then tested whether the BIC [3] is still a valid criteria for model selection. We generated 500 datasets, each with sample size 1200. We generated data using the same parameters from scenario 1 of the two group simulation study. Each data set was truly a two group data set. We then estimated the parameters and calculated the BIC for one group, two group, three group, and four group models. Using the BIC Factor[2] criteria, we selected the best model fit. 94% of the time, the correct two group model was selected by the BIC Factor. It appears the BIC factor remains a valid model selection tool for the latent class joint model.

4.4 EXAMPLE: THE BIOMARK STUDY

BioMaRK is a study that aims to test associations between a panel of inflammatory biomarkers and recovery from acute renal failure. The entire cohort consists of 819 subjects whose biomarker measurements were taken at day 1 and day 8. A subset of these patients (n=104) had daily biomarker measurements taken from day 1 through day 8. The biomarker of interest in this study is interleukin-6 (IL-6). Our event-time outcome of interest was recovery from acute kidney injury, and we treated death as a competing risk. We analyzed the data and examined if there are distinct behavioral patterns in in these data regarding both longitudinal and survival data. In the parent trial [18], subjects were randomized to either high-intensity renal replacement therapy or low-intensity renal replacement therapy. We included treatment as the only covariate in the model. It is important to model the subdistributions in this study, since subjects drop out of the study due to death.

4.4.1 The joint model approach

We ran the joint model for 1, 2, and 3 groups, and used the BIC factor to determine that the two-group model was the best. Table 11 shows the model-fitting results. Using the BIC Factor criteria laid out by Jones, where a BIC Factor ≤ 2 is not important, we choose the next lowest group. For our data, the two group model is optimal.

Table 11: Model fitting for BioMaRK data

groups	parameters	LL	BIC	BIC factor
1	8	-1556.99	-1575.66	NA
2	14	-1413.21	-1445.72	259.69
3	20	-1430.88	-1477.32	-63.21

We fit linear longitudinal trajectories for simplicity. Parameter estimates from the two group model are shown in Table 12 below. The longitudinal trajectories are shown in Figure 9. The cumulative incidence function for the main event, recovery is shown in Figure 11. The cumulative incidence function for the competing event, death, is shown in Figure 11. The group membership parameter indicates that 68% of the population is in the black group and 32% in the red group. The red group has a higher baseline IL-6 and decreases over time, while the black group remains mostly stable over time. Though the red group decreases over time, they do not reach IL6 levels of the black group on day 8. The cumulative incidence function for recovery shows the black group with a higher probability of recovery than the red group. This result is consistent with what is known about inflammatory markers; higher inflammatory biomarker levels are typically associated with worse outcomes.The λ parameter indicates that there is no difference in the treatment assignments across the the two groups (p=0.11). The treatment effect for recovery is non-significant in the black group (p=0.66) and the red group (p=0.46), though the low intensity treatment performs better than the high intensity treatment. The subdistribution for the competing event shows the red group with higher probabilities of death. Again, this is consistent with higher levels of biomarkers leading to worse outcomes. The treatment effect for death is also non-significant for the black group (p=0.50) and the red group (p=0.86).

4.4.2 The two-stage approach

We can also analyze the data using the two-stage approach. Here we modeled the longitudinal trajectories using SAS PROC TRAJ to obtain the trajectories and group memberships. Then we stratified by groups and calculated the subdistributions for each group.

Modeling the longitudinal trajectories yields similar results to the joint model. From Figure 10, we see one relatively flat trajectory with group membership of 70% and a decreasing group with membership 30%.Figure shows the longitudinal trajectories. Plotting the subdistributions again gives us similar results to the joint model. Figure 11 shows the subdistributions for each group. The treatment effect for recovery is non-significant in the black (p=0.65) and red (p=0.25) groups.

Table 12: Joint and two-stage BioMaRK results.

	Joint Model		Two Stage	
parameter	estimate	std error	estimate	std error
λ (treatment across groups)	0.75	0.46	0.76	0.47
θ (Group membership)	-0.734	0.246	-0.857	-
b_{01} (Group 1 intercept)	4.58	0.11	4.61	0.11
b_{11} (Group 1 slope)	-0.05	0.02	-0.05	0.02
b_{02} (Group 2 intercept)	7.18	0.19	7.25	0.19
b_{12} (Group 2 slope)	-0.23	0.03	-0.24	0.03
Σ	1.01	0.06	1.01	0.03
τ_{11}	0.04	0.01	0.03	0.01
ρ_{11}	-0.04	0.01	-0.04	0.01
β_{11} (Group 1 treatment effect recovery)	-0.15	0.35	0.01	0.33
τ_{21}	0.007	0.002	0.008	0.002
β_{21} (Group 1 treatment effect death)	0.301	0.451	0.082	0.417
τ_{12}	0.027	0.015	0.03	0.015
ρ_{12}	-0.034	0.018	-0.032	0.018
τ_{22}	0.012	0.005	0.009	0.004
β_{12} (Group 2 treatment effect recovery)	-0.461	0.611	-0.647	0.567
β_{22} (Group 2 treatment effect death)	0.104	0.603	0.228	0.574

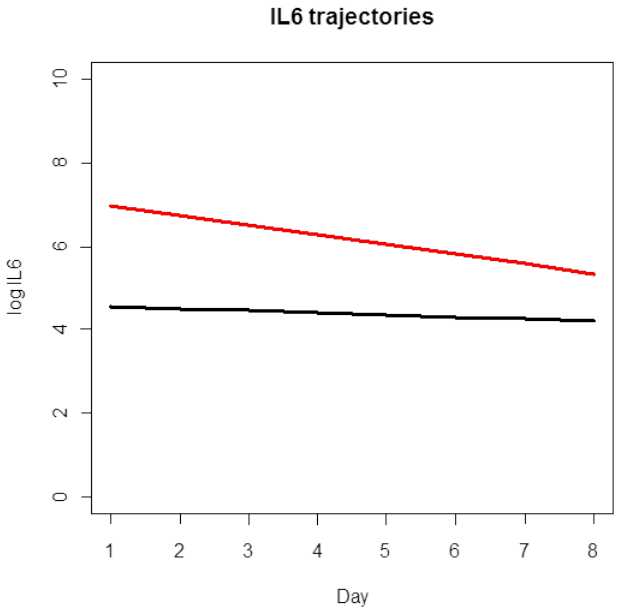

Figure 9: BioMaRK: Joint model longitudinal group trajectories

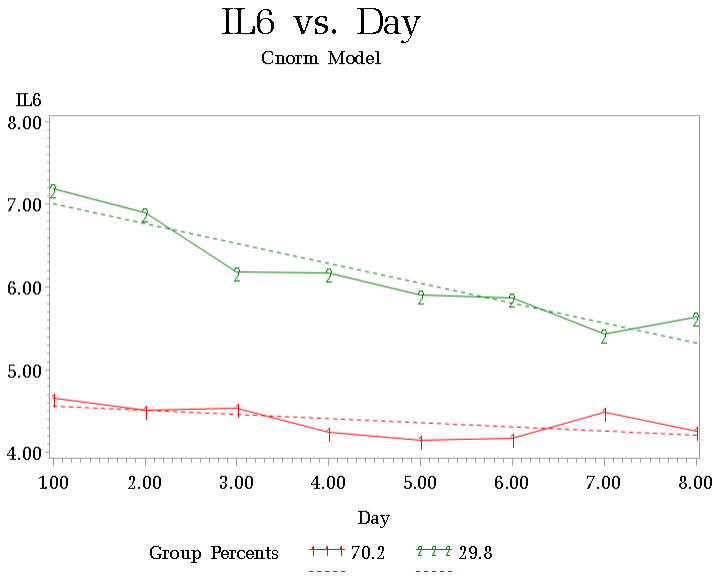

Figure 10: BioMaRK: Two-Stage longitudinal trajectories

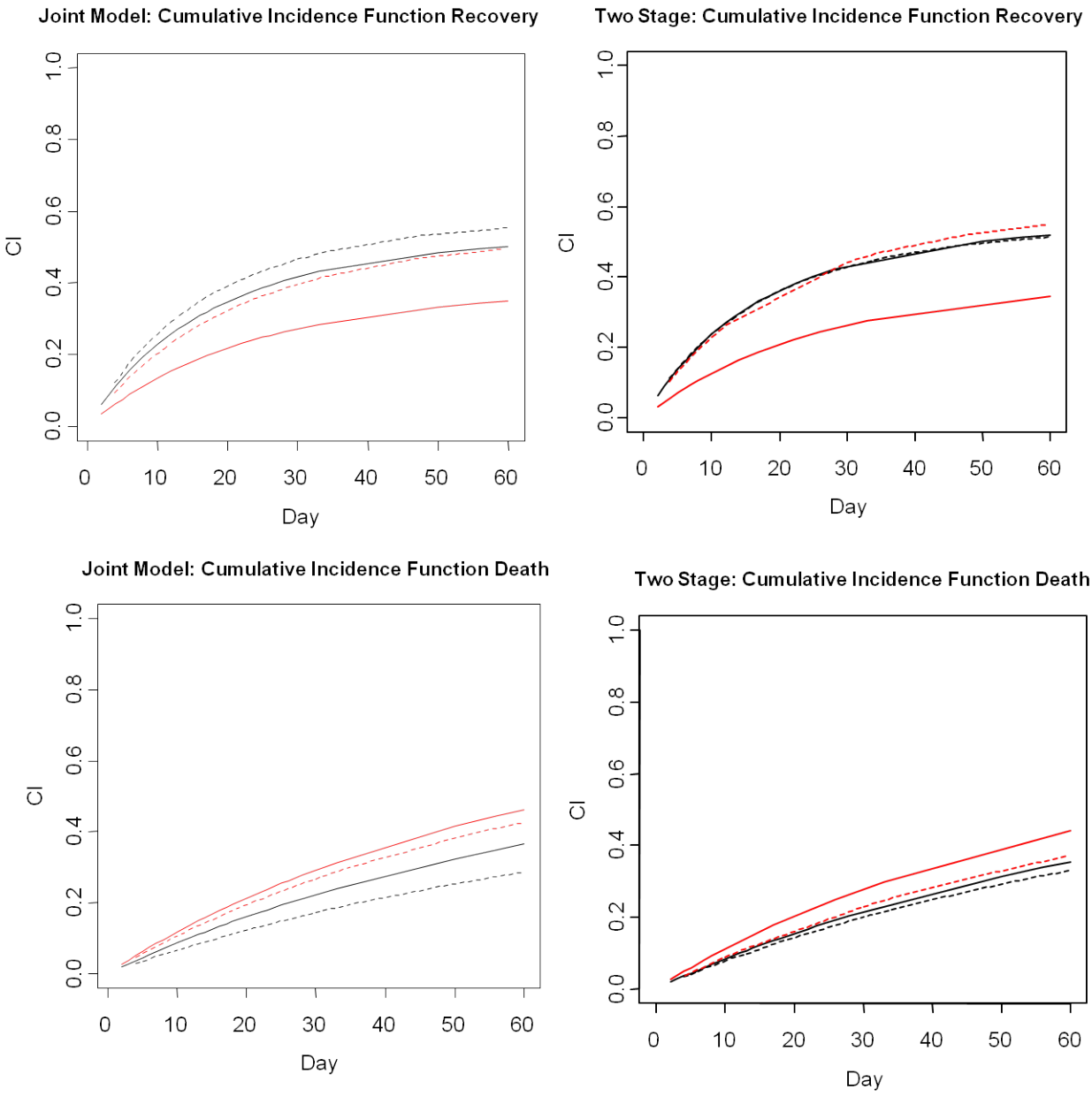

Figure 11: BioMaRK: Cumulative incidence functions for recovery and death.

4.5 CONCLUSIONS

The latent class joint model allows us to discover distinct behavioral groups and simultaneously model the longitudinal process and competing risks survival process. This allows investigators to uncover behavioral groups based on both longitudinal and survival data. Using the conditional independence assumption, we can construct a likelihood by multiplying the multinomial group distribution with the longitudinal and survival processes. We modeled the longitudinal trajectories assuming a multivariate normal, and the survival process as subdistributions. The simulation studies showed that the group membership parameter may be biased in the extreme case where all of the group separation is caused by the survival process. We applied the joint model to data from the BioMaRK study, and showed that in the data set, there are two distinct behavioral groups. One group has a higher baseline with decreasing levels of IL6 over time, and another group has fairly constant levels over time. We found that the group with consistently lower levels of IL6 was more likely to recover and less likely to die. The main advantage of the latent class joint model is that researchers can discover latent groups using both longitudinal and survival data. This can give researchers a more accurate picture of group behavior. If the latent groups are driven by differing survival processes, a two-stage approach will not reflect this. However, in the BioMaRK data, the joint model and two-stage approach yield very similar results. Even if the two approaches produce similar results, the joint model can be used as a sensitivity analysis to show that the two-stage is not overly biased. The biggest disadvantage is model complexity and computation. Estimation can be quite cumbersome and issues with model convergence can be problematic. If the latent groups are more driven by the survival process, it is much easier for the analyst to use existing software, such as PROC TRAJ in SAS, to determine group membership. The analyst can then stratify by group membership and run different survival models. Future work in the are could involve studying how the model performs based on if the group membership is driven more by the longitudinal or survival process. Also, developing a standard package to run these models could make analysts more likely to use them. Even if the two-stage analysis does not differ much from the joint model analysis, one can use the joint model as a sensitivity check.

5.0 OVERALL CONCLUSIONS

Latent group based modeling can be powerful tool to investigators. If they believe that their data contains distinct behavioral groups, using group based modeling can give researchers a better idea of the true nature of their data. These models can be powerful tools in discovering patterns researchers may not have known about. Group-based modeling not only allows researchers to separate individuals into a manageable number of risk groups, but can also identify why the groups are different, leading to potential discoveries. One of the main issues we addressed is the relative lack of diagnostic statistics for these models. Our discrimination index gives an indicator of whether an individual is assigned to a group with a high or low degree of confidence. Modified entropy can also be used as a continuous measure. The percent of subjects that are well discriminated can also be used as an overall model diagnostic. If clinicians are interested in treating individuals based on their group assignments, the overall discrimination rate will tell them whether or not their model is reliable. Researchers are also becoming more interested in modeling both longitudinal trajectories with survival outcomes. While much work has been done developing classic joint models, we developed a latent class joint model that can be used to model latent groups based on both longitudinal data and competing risks survival data. These models are becoming increasingly useful, especially with more work being done with biomarker measurements. The competing risks models are useful especially when researchers use clinical outcomes such as recovery from a disease. Future work includes developing an R package for latent group based trajectory modeling with discrimination index. Currently, most group-based longitudinal trajectories are modeled using Mplus or PROC TRAJ in SAS. Developing more user friendly software packages can make ensure that these models are more widely used.

APPENDIX A

GROUP-BASED MODELING R CODE

```
library(mvtnorm)

data2<-function(n,means1,means2,v) {

library(mvtnorm)



tp<-dim(means1)[1]
tp2<-tp+1
nmeans1<-as.vector(t(means1))
nmeans2<-as.vector(t(means2))

corr<-diag(tp)
va<-v*corr
#x<-matrix(NA,nrow=n,ncol=tp)
x<-matrix(NA,nrow=n,ncol=tp2)
#q1<-rmvnorm(n/2,nmeans1,va)
#q2<-rmvnorm(n/2,nmeans2,va)
#n.2<-n/2
#n.3<-(n/2)+1
```

```
#x[1:n.2,1:tp]<-q1
#x[n.3:n,1:tp]<-q2
z<-rbinom(n,1,0.5)
for(i in 1:n) {
if(z[i]==1) {x[i,1:tp]= rmvnorm(1,nmeans1,va)} else{x[i,1:tp]= rmvnorm(1,nmeans2,va)}
}
x[,tp2]<-rbinom(n,1,0.5)
return(x)
}

data3<-function(n,means1,means2,means3,v) {

library(mvtnorm)

tp<-dim(means1)[1]
nmeans1<-as.vector(t(means1))
nmeans2<-as.vector(t(means2))
nmeans3<-as.vector(t(means3))
corr<-diag(tp)
va<-v*corr
x<-matrix(NA,nrow=n,ncol=tp)
q1<-rmvnorm(n/3,nmeans1,va)
q2<-rmvnorm(n/3,nmeans2,va)
q3<-rmvnorm(n/3,nmeans3,va)

n.2<-n/3
n.3<-(n/3)+1
n.4<-2*n/3
n.5<-(2*n/3)+1
x[1:n.2,1:tp]<-q1
```

```
x[n.3:n.4,1:tp]<-q2
x[n.5:n,1:tp]<-q3
return(x)
}

data4<-function(n,means1,means2,means3,means4,v) {

library(mvtnorm)
tp<-dim(means1)[1]
nmeans1<-as.vector(t(means1))
nmeans2<-as.vector(t(means2))
nmeans3<-as.vector(t(means3))
nmeans4<-as.vector(t(means4))

corr<-diag(tp)
va<-v*corr


x<-matrix(NA,nrow=n,ncol=tp)
q1<-rmvnorm(n/4,nmeans1,va)
q2<-rmvnorm(n/4,nmeans2,va)
q3<-rmvnorm(n/4,nmeans3,va)
q4<-rmvnorm(n/4,nmeans4,va)

n.2<-n/4
n.3<-(n/4)+1
n.4<-2*n.2
n.5<-n.4+1
n.6<-3*n/4
n.7<-n.6+1
```

```
x[1:n.2,1:tp]<-q1
x[n.3:n.4,1:tp]<-q2
x[n.5:n.6,1:tp]<-q3
x[n.7:n,1:tp]<-q4
return(x)
}

# Likelihood functions
#1 grp
fr1<-function(parm,dat)
{
mu<-parm[1:3]
sig<-matrix(c(parm[4],0,0,0,parm[4],0,0,0,parm[4]),nrow=3,ncol=3)

deny<-dmvnorm(dat,mu,sig)
LL<-sum(log(deny))
return(-LL)
}



#2 grps
fr2<-function(parm,dat)
{

t2<-0:2
t<-t(t(t2))
l<-length(t)
tp<-(t-2)/2
# u1<-1+tp
# u2<-u1+1
```

```
# u3<-u2+tp-1
theta<-parm[1]
b01<-parm[2]
b11<-parm[3]
b02<-parm[4]
b12<-parm[5]
iden<-diag(l)
sig<-parm[6]*iden


deny <- (1/(1+exp(theta))*dmvnorm(dat,b01+(b11*t),sig) +  exp(theta)/(1+exp(theta))
*dmvnorm(dat,b02+(b12*t),sig))
LL<-sum(log(deny))
return(-LL)
}

#2 grps trt cov
fr2c<-function(parm,dat)
{

t2<-0:2
t<-t(t(t2))
l<-length(t)
tp<-(t-2)/2
# u1<-1+tp
# u2<-u1+1
# u3<-u2+tp-1
theta<-parm[1]
b01<-parm[2]
b11<-parm[3]
```

```
b02<-parm[4]
b12<-parm[5]
iden<-diag(l)
sig<-parm[6]*iden
lambda<-parm[7]

d<-1+exp(theta + lambda*dat[,4])

deny <- (1/d)*(dmvnorm(dat[,1:3],b01+(b11*t),sig)) +  (exp(theta +
lambda*dat[,4])/d)*(dmvnorm(dat[,1:3],b02+(b12*t),sig))
LL<-sum(log(deny))
return(-LL)
}




#3 grps
fr3<-function(parm,dat)
{
l<-length(parm)
tp<-(l-3)/3
u1<-2+tp
u2<-u1+1
u3<-u2+tp-1
u4<-u3+1
u5<-u4+tp-1
theta2<-parm[1]
theta3<-parm[2]
mu1<-parm[3:u1]
```

```
mu2<-parm[u2:u3]
mu3<-parm[u4:u5]
iden<-diag(tp)
sig<-parm[l]*iden


d<-1+exp(theta2)+exp(theta3)
deny <- ((1/d)*dmvnorm(dat,mu1,sig) + (exp(theta2)/d)*dmvnorm(dat,mu2,sig)+
 (exp(theta3)/d)*dmvnorm(dat,mu3,sig))
LL<-sum(log(deny))
return(-LL)
}



#4 grps
fr4<-function(parm,dat)
{
t2<-0:6
t<-t(t(t2))
l<-length(t)
# tp<-(l-4)/4
u1<-3+tp
u2<-u1+1
u3<-u2+tp-1
u4<-u3+1
u5<-u4+tp-1
u6<-u5+1
u7<-u6+tp-1
```

```
theta2<-parm[1]
theta3<-parm[2]
theta4<-parm[3]
b01<-parm[4]
b11<-parm[5]
b02<-parm[6]
b12<-parm[7]
b03<-parm[8]
b13<-parm[9]
b04<-parm[10]
b14<-parm[11]
iden<-diag(1)
sig<-parm[12]*iden

d<-1+exp(theta2)+exp(theta3)+exp(theta4)
deny <- ((1/d)*dmvnorm(dat,b01+b11*t,sig) + (exp(theta2)/d)*dmvnorm(dat,b02+b12*t,sig)+
(exp(theta3)/d)*dmvnorm(dat,b03+b13*t,sig)+(exp(theta4)/d)*dmvnorm(dat,b04+b14*t,sig)
)
LL<-sum(log(deny))
return(-LL)
}

fr32<-function(parm,dat)
{
t2.1<-dim(dat)[2]
t2<-1:t2.1-1
t<-t(t(t2))
t.2<-t^2
l<-length(t)
```

```
theta2<-parm[1]
theta3<-parm[2]
b01<-parm[3]
b11<-parm[4]
b21<-parm[5]
b02<-parm[6]
b12<-parm[7]
b22<-parm[8]
b03<-parm[9]
b13<-parm[10]
b23<-parm[11]
iden<-diag(1)
sig<-parm[12]*iden

d<-1+exp(theta2)+exp(theta3)
deny <- ((1/d)*dmvnorm(dat,b01+b11*t+b21*t.2,sig) +
 (exp(theta2)/d)*dmvnorm(dat,b02+b12*t+b22*t.2,sig)+
(exp(theta3)/d)*dmvnorm(dat,b03+b13*t+b23*t.2,sig)
)
LL<-sum(log(deny))
return(-LL)
}




#4 grps
fr42<-function(parm,dat)
{
```

```
t2.1<-dim(dat)[2]
t2<-1:t2.1-1
t<-t(t(t2))
t.2<-t^2
l<-length(t)

theta2<-parm[1]
theta3<-parm[2]
theta4<-parm[3]
b01<-parm[4]
b11<-parm[5]
b21<-parm[6]
b02<-parm[7]
b12<-parm[8]
b22<-parm[9]
b03<-parm[10]
b13<-parm[11]
b23<-parm[12]
b04<-parm[13]
b14<-parm[14]
b24<-parm[15]
iden<-diag(l)
sig<-parm[16]*iden

d<-1+exp(theta2)+exp(theta3)+exp(theta4)
deny <- ((1/d)*dmvnorm(dat,b01+b11*t+b21*t.2,sig) +
 (exp(theta2)/d)*dmvnorm(dat,b02+b12*t+b22*t.2,sig)+
(exp(theta3)/d)*dmvnorm(dat,b03+b13*t+b23*t.2,sig)+
(exp(theta4)/d)*dmvnorm(dat,b04+b14*t+b24*t.2,sig)
)
```

```
LL<-sum(log(deny))
return(-LL)
}




#_____________________________________________________________________

# Posterior probabilites for 2,3,4 groups. x=data,e=parameters,j=groups

post.p<-function(x,e,j) {

if(j==2)
{



tp<-dim(x)[2]
l<-length(e$par)
u1<-1+tp
u2<-u1+1
u3<-u2+tp-1
iden<-diag(tp)
#2 grp
e2.var1<-e$par[l]*iden


py.1.2<-dmvnorm(x,e$par[2:u1],e2.var1)
py.2.2<-dmvnorm(x,e$par[u2:u3],e2.var1)
```

```
p<-1/(1+exp(e$par[1]))
g2p1.y<-py.1.2*p/(py.1.2*p + py.2.2*(1-p))
g2p2.y<-py.2.2*(1-p)/(py.1.2*p + py.2.2*(1-p))

n<-dim(x)[1]

post.x<-matrix(NA,nrow=n,ncol=4)
post.x[,1]<-g2p1.y
post.x[,2]<-g2p2.y

for(i in 1:n)
{
if (post.x[i,1]>post.x[i,2]) post.x[i,3]=1 else post.x[i,3]=2
}
# post.x[1:600,4]=1
# post.x[601:1200,4]=2
}

if(j==3)
{
tp<-dim(x)[2]
l<-length(e$par)
u1<-2+tp
u2<-u1+1
u3<-u2+tp-1
u4<-u3+1
u5<-u4+tp-1
iden<-diag(tp)
```

```
#3 grp
e3.var1<-e$par[l]*iden


mu1<-e$par[3]+e$par[4]*t+e$par[5]*t.2
mu2<-e$par[6]+e$par[7]*t+e$par[8]*t.2
mu3<-e$par[9]+e$par[10]*t+e$par[11]*t.2

py.1.3<-dmvnorm(x,mu1,e3.var1)
py.2.3<-dmvnorm(x,mu2,e3.var1)
py.3.3<-dmvnorm(x,mu3,e3.var1)


p1<-1/(1+exp(opt$par[1])+exp(opt$par[2]))
p2<-exp(e$par[1])/(1+exp(e$par[1])+exp(e$par[2]))
p3<-exp(e$par[2])/(1+exp(e$par[1])+exp(e$par[2]))

g3p1.y<-py.1.3*p1/(py.1.3*p1 + py.2.3*p2 + py.3.3*p3)
g3p2.y<-py.2.3*p2/(py.1.3*p1 + py.2.3*p2 + py.3.3*p3)
g3p3.y<-py.3.3*p3/(py.1.3*p1 + py.2.3*p2 + py.3.3*p3)

n<-dim(x)[1]

post.x<-matrix(NA,nrow=n,ncol=5)
post.x[,1]<-g3p1.y
post.x[,2]<-g3p2.y
post.x[,3]<-g3p3.y


for(i in 1:n)
```

```
{
if (max(post.x[i,1:3])==post.x[i,1]) post.x[i,4]=1
if (max(post.x[i,1:3])==post.x[i,2]) post.x[i,4]=2
if (max(post.x[i,1:3])==post.x[i,3]) post.x[i,4]=3
}
# post.x[1:400,5]=1
# post.x[401:800,5]=2
# post.x[801:1200,5]=3
}

if(j==4)
{

tp<-dim(x)[2]
# l<-length(e$par)
u1<-3+tp
u2<-u1+1
u3<-u2+tp-1
u4<-u3+1
u5<-u4+tp-1
u6<-u5+1
u7<-u6+tp-1

time<-1:tp-1
t<-t(t(time))
t.2<-t^2
mu1<-e$par[4]+e$par[5]*t+e$par[6]*t.2
mu2<-e$par[7]+e$par[8]*t+e$par[9]*t.2
mu3<-e$par[10]+e$par[11]*t+e$par[12]*t.2
mu4<-e$par[13]+e$par[14]*t+e$par[15]*t.2
```

```
iden<-diag(tp)

#4 grp
e4.var1<-e$par[16]*iden


py.1.4<-dmvnorm(x,mu1,e4.var1)
py.2.4<-dmvnorm(x,mu2,e4.var1)
py.3.4<-dmvnorm(x,mu3,e4.var1)
py.4.4<-dmvnorm(x,mu4,e4.var1)

p1<-1/(1+exp(e$par[1])+exp(e$par[2])+exp(e$par[3]))
p2<-exp(e$par[1])/(1+exp(e$par[1])+exp(e$par[2])+exp(e$par[3]))
p3<-exp(e$par[2])/(1+exp(e$par[1])+exp(e$par[2])+exp(e$par[3]))
p4<-exp(e$par[3])/(1+exp(e$par[1])+exp(e$par[2])+exp(e$par[3]))


g4p1.y<-py.1.4*p1/(py.1.4*p1 + py.2.4*p2 + py.3.4*p3 + py.4.4*p4)
g4p2.y<-py.2.4*p2/(py.1.4*p1 + py.2.4*p2 + py.3.4*p3 + py.4.4*p4)
g4p3.y<-py.3.4*p3/(py.1.4*p1 + py.2.4*p2 + py.3.4*p3 + py.4.4*p4)
g4p4.y<-py.4.4*p4/(py.1.4*p1 + py.2.4*p2 + py.3.4*p3 + py.4.4*p4)

n<-dim(x)[1]

post.x<-matrix(NA,nrow=n,ncol=5)
post.x[,1]<-g4p1.y
post.x[,2]<-g4p2.y
post.x[,3]<-g4p3.y
```

```
post.x[,4]<-g4p4.y

for(i in 1:n)
{
if (max(post.x[i,1:4])==post.x[i,1]) post.x[i,5]=1
if (max(post.x[i,1:4])==post.x[i,2]) post.x[i,5]=2
if (max(post.x[i,1:4])==post.x[i,3]) post.x[i,5]=3
if (max(post.x[i,1:4])==post.x[i,4]) post.x[i,5]=4

}
}
return(post.x)
}
#_____________________________________________________________________

#BIC
bic<-function(e,num_parm,N)
{
return(-e$value-(.5*num_parm*log(N)))
}
#_____________________________________________________________

mis.rat<-function(post) {

count<-0
n<-dim(post)[1]
n.col<-dim(post)[2]
for(i in 1:n)
{
if(post[i,n.col-1]!=post[i,n.col]) count=count+1
```

```
}

return(c(count, count/n))
}

# 2 group scores
scores<-function(post.mat) {
return(mean(abs(post.mat[,1]-0.5)/0.5))
}

# 3 group scores
scores3<-function(post.mat) {
w1<-.8
w2<-.2
p2<-0
b<-array(dim=3)
d<-dim(post.mat)[1]
s<-array(dim=d)
for(i in 1:d) {
b<-post.mat[i,1:3]
p1<-max(b)
p2<-median(b)
p3<-min(b)
# s[i]<-((w1*(p1-p2) + w2*(p2-p3)))/w1
s[i]<-(p1-(1/3))/(2/3)
}

return(mean(s))
}
```

```
# 4 group scores
scores4<-function(post.mat) {
b<-array(dim=4)
d<-dim(post.mat)[1]
s<-array(dim=d)
for(i in 1:d) {
b<-post.mat[i,1:4]
p1<-max(b)
p2<-median(b)
p3<-min(b)
s[i]<-(p1-(1/4))/(3/4)
}

return(mean(s))
}



# 2 group simulation
msim2<-function(N) {
sdff<-array(1200)
for(i in 1:N)
{
x<-data2(1200,150,150,150,150,150,155,15,13)
opt<-optim(c(0.5,150,150,150,150,150,155,15,13),fr2,dat=x, method="BFGS")
bic<-bic(opt,5,length(x[,1]))
post.mat<-post.p(x,opt,2)
mis<-mis.rat(post.mat)
sdff[i]<-scores(post.mat)
}
```

```
return(sdff)
}


# 3 group simulation
msim3<-function(N) {
sdff<-array(1200)
for(i in 1:N)
{
x<-data3(1200,160,155,150,150,150,150,140,145,150,12,12,12)
opt<-optim(c(0.3,0.3,160,155,150,150,150,150,140,145,150,12,12,12),
           fr3,dat=x, method="BFGS")
bic<-bic(opt,8,length(x[,1]))
post.mat<-post.p(x,opt,3)
mis<-mis.rat(post.mat)
sdff[i]<-scores3(post.mat)
}
return(sdff)
}

# 4 group simulation
msim4<-function(N) {
sdff<-array(1200)
for(i in 1:N)
{
x<-data4(1200,150,150,150,145,155,150,140,150,155,155,150,150,12,12,12,12)
opt<-optim(c(0.25,0.25,0.25,150,150,150,145,155,150,140,150,155,155,150,150,12,12,12,12)
            fr4,dat=x, method="BFGS")
bic<-bic(opt,11,length(x[,1]))
post.mat<-post.p(x,opt,4)
```

```
mis<-mis.rat(post.mat)
sdff[i]<-scores4(post.mat)
}
return(sdff)
}
```

APPENDIX B

DISCRIMINATION INDEX R CODE

```
var.est2<-function(x,e,j) {

#2 grp
e2.var1<-matrix(c(e$par[8],0,0,0,e$par[8],0,0,0,e$par[8]),nrow=3,ncol=3)
e2.var2<-matrix(c(e$par[9],0,0,0,e$par[9],0,0,0,e$par[9]),nrow=3,ncol=3)

py.1.2<-as.matrix(dmvnorm(x,e$par[2:4],e2.var1))
py.2.2<-as.matrix(dmvnorm(x,e$par[5:7],e2.var2))

n<-dim(x)[1]
x.t<-t(x)
x_mu<-matrix(NA,nrow=3*n,ncol=1)
x_mu2<-matrix(NA,nrow=3*n,ncol=1)
x_mu1.t<-matrix(NA,nrow=3,ncol=n)

mu1_hat<-as.matrix(e$par[2:4])
mu2_hat<-as.matrix(e$par[5:7])


l<-1
```

```
for(i in 1:n)
{

k<-l+2
x_mu[l:k,]<-x[i,]-mu1_hat
l<-l+3
}

l<-1
for(i in 1:n)
{
k<-l+2
x_mu2[l:k,]<-x[i,]-mu2_hat
l<-l+3
}

var.delt<-matrix(NA,nrow=n,ncol=1)
#var.delt1<-matrix(NA,nrow=n,ncol=3)
#var.delt2<-matrix(NA,nrow=n,ncol=3)


l<-1
for(i in 1:n)
{



k<-l+2
Sp1<-(t(solve(e2.var1)))%*%(x_mu[k:l,])
Sp2<-(solve(e2.var1))%*%(x_mu[k:l,])
Sp3<-(t(solve(e2.var2)))%*%(x_mu2[k:l,])
```

```
Sp4<-(solve(e2.var2))%*%(x_mu2[k:l,])
S1<-(e$par[1]*py.1.2[i,])*(-.5*(Sp1+Sp2))
S2<-(1-e$par[1])*py.2.2[i,]*(-.5*(Sp3+Sp4))
A<-(py.1.2[i,]*e$par[1] + py.2.2[i,]*(1-e$par[1]))

var.delt1<-t(((S1*A-((S1+S2)*e$par[1]*py.1.2[i,]))/A^2))
%*%e2.var1%*%((S1*A-((S1+S2)*e$par[1]*py.1.2[i,]))/A^2)
var.delt2<-t(((S2*A-((S1+S2)*(1- e$par[1])*py.2.2[i,]))/A^2))
%*%e2.var2%*%((S2*A-((S1+S2)
          *(1e$par[1])*py.2.2[i,]))/A^2)
var.delt[i,]<-var.delt1+var.delt2
l<-l+3
}

return(var.delt)

}

#-------------------------------------------------------------------
n<-200
rat<-array(dim=250)

for(z in 1:250) {
x<-data2(n,150,150,150,145,152,148,15,15)
e<-optim(c(0.5,150,150,150,145,152,148,15,15),fr2,dat=x, method="BFGS")
v<-var.est2(x,e,2)
ci<-1.96*sqrt(v)/sqrt(n)

#-----------------------------------------
```

```
# Max probability matrix
post.mat<-post.p(x,e,2)
maxmat<-matrix(NA,nrow=n,ncol=5)
maxmat[,1:4]<-post.mat

emat<-matrix(NA,nrow=n,ncol=5)

for(i in 1:n)
{
emat[i,1]<-max(post.mat[i,1],post.mat[i,2])
emat[i,2]<-ci[i,]
emat[i,3]<-emat[i,1]-emat[i,2]
emat[i,4]<-emat[i,1]+emat[i,2]
if(emat[i,3]<.5) {emat[i,5]<-1} else {emat[i,5]<-0}
}

tb1<-table(emat[,5])

emat2<-matrix(NA,nrow=n,ncol=7)

for(i in 1:n)
{
emat2[i,1]<-max(post.mat[i,1],post.mat[i,2])
emat2[i,2]<-min(post.mat[i,1],post.mat[i,2])
emat2[i,3]<-emat2[i,1]-emat[i,2]
emat2[i,4]<-emat2[i,1]+emat[i,2]
emat2[i,5]<-emat2[i,2]-emat[i,2]
emat2[i,6]<-emat2[i,2]+emat[i,2]

if(emat2[i,3]<emat2[i,6]) {emat2[i,7]<-1} else {emat2[i,7]<-0}
```

```
}

tb2<-table(emat2[,7])

rat[z]<-tb1[2]/(tb1[1]+tb1[2])
}




for(i in 1:250)
{
if(is.na(rat[i])==TRUE) {rat[i]<-0}
}



# 3 group delta method
#_____________________________________________________________________

var.est3<-function(x,e,j) {

#3 grp
# e2.var1<-matrix(c(e$par[12],0,0,0,e$par[12],0,0,0,e$par[12]),nrow=3,ncol=3)
# e2.var2<-matrix(c(e$par[13],0,0,0,e$par[13],0,0,0,e$par[13]),nrow=3,ncol=3)
# e2.var3<-matrix(c(e$par[14],0,0,0,e$par[14],0,0,0,e$par[14]),nrow=3,ncol=3)


# py.1.2<-as.matrix(dmvnorm(x,e$par[3:5],e2.var1))
```

```
# py.2.2<-as.matrix(dmvnorm(x,e$par[6:8],e2.var2))
# py.3.2<-as.matrix(dmvnorm(x,e$par[9:11],e2.var3))

l<-length(e$par)
tp<-dim(x)[2]
iden<-diag(tp)

e2.var1<-e$par[l]*iden
time<-1:tp-1
t<-t(t(time))
t.2<-t^2
mu1<-e$par[3]+e$par[4]*t+e$par[5]*t.2
mu2<-e$par[6]+e$par[7]*t+e$par[8]*t.2
mu3<-e$par[9]+e$par[10]*t+e$par[11]*t.2




py.1.2<-as.matrix(dmvnorm(x,mu1,e2.var1))
py.2.2<-as.matrix(dmvnorm(x,mu2,e2.var1))
py.3.2<-as.matrix(dmvnorm(x,mu3,e2.var1))




n<-dim(x)[1]
x.t<-t(x)
x_mu<-matrix(NA,nrow=tp*n,ncol=1)
x_mu2<-matrix(NA,nrow=tp*n,ncol=1)
x_mu3<-matrix(NA,nrow=tp*n,ncol=1)
```

```
x_mu1.t<-matrix(NA,nrow=3,ncol=n)

mu1_hat<-mu1
mu2_hat<-mu2
mu3_hat<-mu3


# l<-1
# for(i in 1:n)
# {

# k<-l+2
# x_mu[l:k,]<-x[i,]-mu1_hat
# l<-l+3
# }

# l<-1
# for(i in 1:n)
# {
# k<-l+2
# x_mu2[l:k,]<-x[i,]-mu2_hat
# l<-l+3
# }

# l<-1
# for(i in 1:n)
# {
# k<-l+2
# x_mu3[l:k,]<-x[i,]-mu3_hat
```

```
# l<-l+3
# }

l<-1
for(i in 1:n)
{

k<-tp+l-1
x_mu[l:k,]<-x[i,]-mu1_hat
l<-l+tp
}

l<-1
for(i in 1:n)
{
k<-tp+l-1
x_mu2[l:k,]<-x[i,]-mu2_hat
l<-l+tp
}

l<-1
for(i in 1:n)
{
k<-tp+l-1
x_mu3[l:k,]<-x[i,]-mu3_hat
l<-l+tp
}
```

```
var.delt<-matrix(NA,nrow=n,ncol=1)
#var.delt1<-matrix(NA,nrow=n,ncol=3)
#var.delt2<-matrix(NA,nrow=n,ncol=3)

p1<-1/(1+exp(e$par[1])+exp(e$par[2]))
p2<-exp(e$par[1])/(1+exp(e$par[1])+exp(e$par[2]))
p3<-exp(e$par[2])/(1+exp(e$par[1])+exp(e$par[2]))



l<-1
for(i in 1:n)
{


k<-tp+l-1
Sp1<-(t(solve(e2.var1)))%*%(x_mu[k:l,])
Sp2<-(solve(e2.var1))%*%(x_mu[k:l,])
Sp3<-(t(solve(e2.var1)))%*%(x_mu2[k:l,])
Sp4<-(solve(e2.var1))%*%(x_mu2[k:l,])
Sp5<-(t(solve(e2.var1)))%*%(x_mu3[k:l,])
Sp6<-(solve(e2.var1))%*%(x_mu3[k:l,])

S1<-(p1*py.1.2[i,])*(-.5*(Sp1+Sp2))
S2<-(p2*py.2.2[i,])*(-.5*(Sp3+Sp4))
S3<-(p3*py.3.2[i,])*(-.5*(Sp5+Sp6))
A<-(py.1.2[i,]*p1 + py.2.2[i,]*p2+ py.3.2[i,]*p3)

var.delt1<-t(((S1*A-((S1+S2+S3)*p1*py.1.2[i,]))/A^2))%*%e2.var1%*%
```

```
((S1*A-((S1+S2+S3)*p1*py.1.2[i,]))/A^2)
var.delt2<-t(((S2*A-((S1+S2+S3)*p2*py.2.2[i,]))/A^2))%*%e2.var1%*%
((S2*A-((S1+S2+S3)*p2*py.2.2[i,]))/A^2)
var.delt3<-t(((S3*A-((S1+S2+S3)*p3*py.3.2[i,]))/A^2))%*%e2.var1%*%
((S3*A-((S1+S2+S3)*p3*py.3.2[i,]))/A^2)

var.delt[i,]<-var.delt1+var.delt2+var.delt3
l<-l+tp
}

return(var.delt)

}



n<-150
rat<-array(dim=100)

for(z in 1:100) {

x<-data3(n,150,150,150,155,150,145,145,150,155,8,8,8)
e<-optim(c(0.3,0.3,150,150,150,155,150,145,145,150,155,8,8,8),fr3,dat=x, method="BFGS")
v<-var.est3(c5data,e,3)
ci<-1.96*sqrt(v)/sqrt(n)

# Max probability matrix 3
post.mat1<-post.p(c5data,e,3)
post.mat<-post.mat1[,1:4]
maxmat<-matrix(NA,nrow=n,ncol=5)
maxmat[,1:4]<-post.mat
```

```
emat<-matrix(NA,nrow=n,ncol=5)

for(i in 1:n)
{
emat[i,1]<-max(post.mat[i,1],post.mat[i,2],post.mat[i,3])
emat[i,2]<-ci[i,]
emat[i,3]<-emat[i,1]-emat[i,2]
emat[i,4]<-emat[i,1]+emat[i,2]
if(emat[i,3]<.333) {emat[i,5]<-1} else {emat[i,5]<-0}

}

ent_mat3<-matrix(NA,nrow=n,ncol=2)
emat2<-matrix(NA,nrow=n,ncol=5)
for(i in 1:n)
{
emat2[i,1]<-max(post.mat[i,1],post.mat[i,2],post.mat[i,3])
# emat2[i,2]<-median(post.mat[i,1:3])
med<-median(post.mat[i,1:3])
emat2[i,2]<-(emat2[i,1]+med)/2


emat2[i,3]<-emat2[i,1]-emat[i,2]
emat2[i,4]<-emat2[i,1]+emat[i,2]
# emat2[i,5]<-emat2[i,2]-emat[i,2]
# emat2[i,6]<-emat2[i,2]+emat[i,2]
if(emat2[i,3]<emat2[i,2]) {emat2[i,5]<-1} else {emat2[i,5]<-0}
ent_mat3[i,1]<-emat2[i,1]
ent_mat3[i,2]<-med
```

```
}
# Use a 2 sample proportion test instead of looking at the overlapping CIs.
tb1<-table(emat2[,5])
rat[z]<-tb1[2]/(tb1[1]+tb1[2])
}

# Max probability matrix 3
post.mat1<-post.p(c5data,e,3)
post.mat<-post.mat1[,1:3]
maxmat<-matrix(NA,nrow=n,ncol=4)
maxmat[,1:3]<-post.mat
post.mat2<-data.frame(post.mat)

emat<-matrix(NA,nrow=n,ncol=6)

for(i in 1:n)
{
emat[i,1]<-max(post.mat[i,1],post.mat[i,2],post.mat[i,3])
emat[i,2]<-median(c(post.mat[i,1],post.mat[i,2],post.mat[i,3]))


emat[i,3]<-(emat[i,1]+emat[i,2])/2
emat[i,4]<-emat[i,1]-ci[i]
emat[i,5]<-emat[i,1]+ci[i]
if(emat[i,4]<emat[i,3]) {emat[i,6]<-1} else {emat[i,6]<-0}
}

tb1<-table(emat[,6])
```

```
#_____________________________________________________________________________
var.est4<-function(x,e,j) {

#4 grp
tp<-dim(x)[2]
l<-length(e$par)
u1<-3+tp
u2<-u1+1
u3<-u2+tp-1
u4<-u3+1
u5<-u4+tp-1
u6<-u5+1
u7<-u6+tp-1

iden<-diag(tp)

#4 grp
e2.var1<-e$par[l]*iden
time<-1:tp-1
t<-t(t(time))
t.2<-t^2
mu1<-e$par[4]+e$par[5]*t+e$par[6]*t.2
mu2<-e$par[7]+e$par[8]*t+e$par[9]*t.2
mu3<-e$par[10]+e$par[11]*t+e$par[12]*t.2
mu4<-e$par[13]+e$par[14]*t+e$par[15]*t.2
```

```
py.1.2<-as.matrix(dmvnorm(x,mu1,e2.var1))
py.2.2<-as.matrix(dmvnorm(x,mu2,e2.var1))
py.3.2<-as.matrix(dmvnorm(x,mu3,e2.var1))
py.4.2<-as.matrix(dmvnorm(x,mu4,e2.var1))



n<-dim(x)[1]
x.t<-t(x)
x_mu<-matrix(NA,nrow=tp*n,ncol=1)
x_mu2<-matrix(NA,nrow=tp*n,ncol=1)
x_mu3<-matrix(NA,nrow=tp*n,ncol=1)
x_mu4<-matrix(NA,nrow=tp*n,ncol=1)


x_mu1.t<-matrix(NA,nrow=tp,ncol=n)
mu1_hat<-mu1
mu2_hat<-mu2
mu3_hat<-mu3
mu4_hat<-mu4


l<-1
for(i in 1:n)
{

k<-tp+l-1
x_mu[l:k,]<-x[i,]-mu1_hat
```

```
l<-l+tp
}

l<-1
for(i in 1:n)
{
k<-tp+l-1
x_mu2[l:k,]<-x[i,]-mu2_hat
l<-l+tp
}

l<-1
for(i in 1:n)
{
k<-tp+l-1
x_mu3[l:k,]<-x[i,]-mu3_hat
l<-l+tp
}

l<-1
for(i in 1:n)
{
k<-tp+l-1
x_mu4[l:k,]<-x[i,]-mu4_hat
l<-l+tp
}


var.delt<-matrix(NA,nrow=n,ncol=1)
#var.delt1<-matrix(NA,nrow=n,ncol=tp)
```

```
#var.delt2<-matrix(NA,nrow=n,ncol=tp)

p1<-1/(1+exp(e$par[1])+exp(e$par[2])+exp(e$par[3]))
p2<-exp(e$par[1])/(1+exp(e$par[1])+exp(e$par[2])+exp(e$par[3]))
p3<-exp(e$par[2])/(1+exp(e$par[1])+exp(e$par[2])+exp(e$par[3]))
p4<-exp(e$par[3])/(1+exp(e$par[1])+exp(e$par[2])+exp(e$par[3]))



l<-1
for(i in 1:n)
{


k<-tp+l-1
Sp1<-(t(solve(e2.var1)))%*%(x_mu[k:l,])
Sp2<-(solve(e2.var1))%*%(x_mu[k:l,])
Sp3<-(t(solve(e2.var1)))%*%(x_mu2[k:l,])
Sp4<-(solve(e2.var1))%*%(x_mu2[k:l,])
Sp5<-(t(solve(e2.var1)))%*%(x_mu3[k:l,])
Sp6<-(solve(e2.var1))%*%(x_mu3[k:l,])
Sp7<-(t(solve(e2.var1)))%*%(x_mu4[k:l,])
Sp8<-(solve(e2.var1))%*%(x_mu4[k:l,])


S1<-p1*py.1.2[i,]*(-.5*(Sp1+Sp2))
S2<-p2*py.2.2[i,]*(-.5*(Sp3+Sp4))
S3<-p3*py.3.2[i,]*(-.5*(Sp5+Sp6))
S4<-p4*py.4.2[i,]*(-.5*(Sp7+Sp8))
```

```
A<-(py.1.2[i,]*p1 + py.2.2[i,]*p2+ py.3.2[i,]*p3+ py.4.2[i,]*p4)

var.delt1<-
         (((S1*A-((S1+S2+S3+S4)*p1*py.1.2[i,]))/A^2))%*%
        e2.var1%*%((S1*A-((S1+S2+S3+S4)*p1*py.1.2[i,]))/A^2)
var.delt2<-
           (((S2*A-((S1+S2+S3+S4)*p2*py.2.2[i,]))/A^2))%*%
          e2.var1%*%((S2*A-((S1+S2+S3+S4)*p2*py.2.2[i,]))/A^2)
var.delt3<-
          t(((S3*A-((S1+S2+S3+S4)*p3*py.3.2[i,]))/A^2))%*%
        e2.var1%*%((S3*A-((S1+S2+S3+S4)*p3*py.3.2[i,]))/A^2)
var.delt4<-
            t(((S4*A-((S1+S2+S3+S4)*p4*py.4.2[i,]))/A^2))%*%
           e2.var1%*%((S4*A-((S1+S2+S3+S4)*p4*py.4.2[i,]))/A^2)



var.delt[i,]<-var.delt1+var.delt2+var.delt3+var.delt4
l<-l+tp
}

return(var.delt)

}

n<-200
rat<-array(dim=100)

for(z in 1:100) {

x<-data4(n,150,150,150,150,145,140,140,140,140,140,145,150,12,12,12,12)
```

```
e<-optim(c(0.25,0.25,0.25,150,150,150,150,145,140,140,140,140,140,
145,150,12,12,12,12),fr4,dat=x, method="BFGS")
v<-var.est4(c5data,e,4)
n<-dim(c5data)[1]
ci<-1.96*sqrt(v)/sqrt(n)

# Max probability matrix 4
post.mat1<-post.p(c5data,e,4)
post.mat<-post.mat1[,1:4]
maxmat<-matrix(NA,nrow=n,ncol=5)
maxmat[,1:4]<-post.mat
post.mat2<-data.frame(post.mat)
o_p<-matrix(NA,nrow=n, ncol=4)
for(i in 1:n)
{
o_p[i,]<-rank(post.mat[i,])

}

emat<-matrix(NA,nrow=n,ncol=6)

for(i in 1:n)
{
emat[i,1]<-max(post.mat[i,1],post.mat[i,2],post.mat[i,3],post.mat[i,4])



for(j in 1:4)
{
if(o_p[i,j]==3) {emat[i,2]<-post.mat[i,j]}
}
```

```
emat[i,3]<-(emat[i,1]+emat[i,2])/2
emat[i,4]<-emat[i,1]-ci[i]
emat[i,5]<-emat[i,1]+ci[i]
if(emat[i,4]<emat[i,3]) {emat[i,6]<-1} else {emat[i,6]<-0}
}

tb1<-table(emat[,6])
```

APPENDIX C

JOINT MODEL R CODE

```
bic<-function(e,num_parm,N)
{
return(-e$value-(.5*num_parm*log(N)))
}
#_____________________________________________

#2 group longitudinal daata
data2<-function(n,means1,means2,v) {

library(mvtnorm)



tp<-dim(means1)[1]
nmeans1<-as.vector(t(means1))
nmeans2<-as.vector(t(means2))

corr<-diag(tp)
va<-v*corr
```

```
x<-matrix(NA,nrow=n,ncol=tp)
#q1<-rmvnorm(n/2,nmeans1,va)
#q2<-rmvnorm(n/2,nmeans2,va)
#n.2<-n/2
#n.3<-(n/2)+1
#x[1:n.2,1:tp]<-q1
#x[n.3:n,1:tp]<-q2
z<-rbinom(n,1,0.5)
for(i in 1:n) {
if(z[i]==1) {x[i,]= rmvnorm(1,nmeans1,va)} else{x[i,]= rmvnorm(1,nmeans2,va)}
}
return(x)
}
#_____________________________________________________
cdata<- function(n) {

data<-matrix(nrow=n, ncol=6)
for(i in 1:n) {



rho1<- -0.25
tau1<-0.08
rho2<- -0.00772
#rho2<- 0.000152
tau2<-0.01
beta1<- -0.54
beta2<- -0.10

trt<-rbinom(1,1,0.5)
linpred_1<-trt*beta1
```

```
linpred_2<-trt*beta2


f1inf<- 1-exp((tau1/rho1)*exp(linpred_1))
f2inf<-1-f1inf

e<-rbinom(1,1,f1inf)

if(e==1){
u<-runif(1,0,1);
t<-log(1-(log(1-u*f1inf)*rho1)/(tau1*exp(linpred_1)))/rho1}
else {u<-runif(1,0,1);t<-log(max(1e-15,1-(log(1-u*f2inf)*rho2)/
(tau2*exp(linpred_2))))/rho2};

e2<-e
if(e==0) {e2=2}

data[i,1]<-trt
data[i,2]<-t
data[i,3]<-e2

if(data[i,3]==1) {(data[i,4]=1) & (data[i,5]=0)}
else {(data[i,4]=0) & (data[i,5]=1)};


cen<-runif(1,0,1)
C<- -log(cen)/0.1

data[i,2]<-min(t,C)
if(C<t){data[i,3]<-0}
```

```
if(data[i,3]==0) {data[i,6]=1;data[i,5]=0;data[i,4]=0} else{data[i,6]=0};
}
return(data)
}
#_____________________________________________________________

data1<-function(n,nmean,v)
{
tp<-dim(nmean)[1]
nmean1<-as.vector(t(nmean))
corr<-diag(tp)
va<-v*corr
q<-rmvnorm(n,nmean1,va)
return(q)
}
_____________________________________________________________

cdata2gr<- function(n,p) {

data<-matrix(nrow=n, ncol=6)
for(i in 1:n) {

b<-rbinom(1,1,p)
if(b==0) {
rho1<- -0.25
tau1<-0.08
rho2<- -0.00772
#rho2<- 0.000152
tau2<-0.01
```

```
beta1<- -0.54
beta2<- -0.10 }

else {
rho1<- -0.30
tau1<-0.05
#rho2<- 0.000152
tau2<-0.03
rho2<-(tau2)/(log(1-exp(tau1/rho1)))
beta1<- -0.54
beta2<- -0.10 };

trt<-rbinom(1,1,0.5)
linpred_1<-trt*beta1
linpred_2<-trt*beta2


f1inf<- 1-exp((tau1/rho1)*exp(linpred_1))
f2inf<-1-f1inf

e<-rbinom(1,1,f1inf)

if(e==1){
u<-runif(1,0,1);
t<-log(1-(log(1-u*f1inf)*rho1)/(tau1*exp(linpred_1)))/rho1}
else {u<-runif(1,0,1);t<-log(max(1e-15,1-(log(1-u*f2inf)*rho2)/
(tau2*exp(linpred_2))))/rho2};

e2<-e
if(e==0) {e2=2}
```

```
data[i,1]<-trt
data[i,2]<-t
data[i,3]<-e2

if(data[i,3]==1) {(data[i,4]=1) & (data[i,5]=0)}
else {(data[i,4]=0) & (data[i,5]=1)};



cen<-runif(1,0,1)
C<- -log(cen)/0.1

data[i,2]<-min(t,C)
if(C<t){data[i,3]<-0}

if(data[i,3]==0) {data[i,6]=1;data[i,5]=0;data[i,4]=0} else{data[i,6]=0};
}
return(data)
}
#_____________________________________________________________________
#4 grps
frjm4<-function(parm,dat) {

t2<-0:2
t<-t(t(t2))
l<-length(t)
tp<-(t-2)/2
theta1<-parm[1]
theta2<-parm[2]
theta3<-parm[3]
```

```
b01<-parm[4]
b11<-parm[5]
b02<-parm[6]
b12<-parm[7]
b03<-parm[8]
b13<-parm[9]
b04<-parm[10]
b14<-parm[11]
iden<-diag(1)
sig<-parm[12]*iden
minn<-1e-15
rhop21<- -0.00772
taup11<-parm[13]
rhop11<-parm[14]
betap1<-parm[15]
taup21<-parm[16]
betap2<-parm[17]
taup12<-parm[18]
rhop12<-parm[19]
taup22<-parm[20]
rhop22<- -0.00772
taup13<-parm[21]
rhop13<-parm[22]
taup23<-parm[23]
rhop23<- -0.00772
taup14<-parm[24]
rhop14<-parm[25]
taup24<-parm[26]
rhop24<- -0.00772
```

```
f11<- taup11*exp(dat[,4]*betap1)*exp(rhop11*dat[,5])*
          exp(-exp(dat[,4]*betap1)*(taup11/rhop11)*(exp(rhop11*dat[,5])-1))
f21<- taup21*exp(dat[,4]*betap2)*exp(rhop21*dat[,5])*
         exp(-exp(dat[,4]*betap2)*(taup21/rhop21)*(exp(rhop21*dat[,5])-1))
f12<- taup12*exp(dat[,4]*betap1)*exp(rhop12*dat[,5])*
           exp(-exp(dat[,4]*betap1)*(taup12/rhop12)*(exp(rhop12*dat[,5])-1))
f22<- taup22*exp(dat[,4]*betap2)*exp(rhop22*dat[,5])*
           exp(-exp(dat[,4]*betap2)*(taup22/rhop22)*(exp(rhop22*dat[,5])-1))
f13<- taup13*exp(dat[,4]*betap1)*exp(rhop13*dat[,5])*
           exp(-exp(dat[,4]*betap1)*(taup13/rhop13)*(exp(rhop13*dat[,5])-1))
f23<- taup23*exp(dat[,4]*betap2)*exp(rhop23*dat[,5])*
            exp(-exp(dat[,4]*betap2)*(taup23/rhop23)*(exp(rhop23*dat[,5])-1))
f14<- taup14*exp(dat[,4]*betap1)*exp(rhop14*dat[,5])*
           exp(-exp(dat[,4]*betap1)*(taup14/rhop14)*(exp(rhop14*dat[,5])-1))
f24<- taup24*exp(dat[,4]*betap2)*exp(rhop24*dat[,5])*
          exp(-exp(dat[,4]*betap2)*(taup24/rhop24)*(exp(rhop24*dat[,5])-1))
```

```
F11<-1-(exp(-exp(dat[,4]*betap1)*(taup11/rhop11)*(exp(rhop11*dat[,5])-1)))
F21<-1-(exp(-exp(dat[,4]*betap2)*(taup21/rhop21)*(exp(rhop21*dat[,5])-1)))
F12<-1-(exp(-exp(dat[,4]*betap1)*(taup12/rhop12)*(exp(rhop12*dat[,5])-1)))
F22<-1-(exp(-exp(dat[,4]*betap2)*(taup22/rhop22)*(exp(rhop22*dat[,5])-1)))
F13<-1-(exp(-exp(dat[,4]*betap1)*(taup13/rhop13)*(exp(rhop13*dat[,5])-1)))
F23<-1-(exp(-exp(dat[,4]*betap2)*(taup23/rhop23)*(exp(rhop23*dat[,5])-1)))
```

```
F14<-1-(exp(-exp(dat[,4]*betap1)*(taup14/rhop14)*(exp(rhop14*dat[,5])-1)))
F24<-1-(exp(-exp(dat[,4]*betap2)*(taup24/rhop24)*(exp(rhop24*dat[,5])-1)))



F1<-1-F11-F21
F2<-1-F12-F22
F3<-1-F13-F23
F4<-1-F14-F24



deny<-(1/(1+exp(theta1)+exp(theta2)+exp(theta3))*dmvnorm(dat[,1:3],b01+(b11*t),sig)*
(f11^dat[,7])*(f21^dat[,8])*(F1^dat[,9])+
exp(theta1)/(1+exp(theta1)+exp(theta2)+exp(theta3))*dmvnorm(dat[,1:3],b02+(b12*t),sig)*
(f12^dat[,7])*(f22^dat[,8])*(F2^dat[,9])+
exp(theta2)/(1+exp(theta1)+exp(theta2)+exp(theta3))*dmvnorm(dat[,1:3],b03+(b13*t),sig)*
(f13^dat[,7])*(f23^dat[,8])*(F3^dat[,9])+
exp(theta3)/(1+exp(theta1)+exp(theta2)+exp(theta3))*dmvnorm(dat[,1:3],b04+(b14*t),sig)*
(f14^dat[,7])*(f24^dat[,8])*(F4^dat[,9]))
LL<-sum(log(deny))
return(-LL)
}


#_____________________________________________________________________
#3 grps
frjm3<-function(parm,dat) {
```

```
t2<-0:2
t<-t(t(t2))
l<-length(t)
tp<-(t-2)/2
theta1<-parm[1]
theta2<-parm[2]
b01<-parm[3]
b11<-parm[4]
b02<-parm[5]
b12<-parm[6]
b03<-parm[7]
b13<-parm[8]
iden<-diag(l)
sig<-parm[9]*iden
minn<-1e-15
rhop21<- -0.00772
taup11<-parm[10]
rhop11<-parm[11]
betap1<-parm[12]
taup21<-parm[13]
betap2<-parm[14]
taup12<-parm[15]
rhop12<-parm[16]
taup22<-parm[17]
rhop22<- -0.00772
taup13<-parm[18]
rhop13<-parm[19]
taup23<-parm[20]
rhop23<- -0.00772
```

```
f11<- taup11*exp(dat[,4]*betap1)*exp(rhop11*dat[,5])*
exp(-exp(dat[,4]*betap1)*(taup11/rhop11)*(exp(rhop11*dat[,5])-1))
f21<- taup21*exp(dat[,4]*betap2)*exp(rhop21*dat[,5])*
exp(-exp(dat[,4]*betap2)*(taup21/rhop21)*(exp(rhop21*dat[,5])-1))
f12<- taup12*exp(dat[,4]*betap1)*exp(rhop12*dat[,5])*
exp(-exp(dat[,4]*betap1)*(taup12/rhop12)*(exp(rhop12*dat[,5])-1))
f22<- taup22*exp(dat[,4]*betap2)*exp(rhop22*dat[,5])*
exp(-exp(dat[,4]*betap2)*(taup22/rhop22)*(exp(rhop22*dat[,5])-1))
f13<- taup13*exp(dat[,4]*betap1)*exp(rhop13*dat[,5])*
exp(-exp(dat[,4]*betap1)*(taup13/rhop13)*(exp(rhop13*dat[,5])-1))
f23<- taup23*exp(dat[,4]*betap2)*exp(rhop23*dat[,5])*
exp(-exp(dat[,4]*betap2)*(taup23/rhop23)*(exp(rhop23*dat[,5])-1))

F11<-1-(exp(-exp(dat[,4]*betap1)*(taup11/rhop11)*(exp(rhop11*dat[,5])-1)))
F21<-1-(exp(-exp(dat[,4]*betap2)*(taup21/rhop21)*(exp(rhop21*dat[,5])-1)))
F12<-1-(exp(-exp(dat[,4]*betap1)*(taup12/rhop12)*(exp(rhop12*dat[,5])-1)))
F22<-1-(exp(-exp(dat[,4]*betap2)*(taup22/rhop22)*(exp(rhop22*dat[,5])-1)))
F13<-1-(exp(-exp(dat[,4]*betap1)*(taup13/rhop13)*(exp(rhop13*dat[,5])-1)))
F23<-1-(exp(-exp(dat[,4]*betap2)*(taup23/rhop23)*(exp(rhop23*dat[,5])-1)))


F1<-1-F11-F21
F2<-1-F12-F22
F3<-1-F13-F23
```

```
deny<-(1/(1+exp(theta1)+exp(theta2))*dmvnorm(dat[,1:3],b01+(b11*t),sig)*
(f11^dat[,7])*(f21^dat[,8])*(F1^dat[,9])+
 exp(theta1)/(1+exp(theta1)+exp(theta2))*dmvnorm(dat[,1:3],b02+
(b12*t),sig)*(f12^dat[,7])*(f22^dat[,8])*(F2^dat[,9])+ exp(theta2)/(1+exp(theta1)+
exp(theta2))*dmvnorm(dat[,1:3],b03+(b13*t),sig)*
            (f13^dat[,7])*(f23^dat[,8])*(F3^dat[,9]))

LL<-sum(log(deny))
return(-LL)
}
#_____________________________________________________________________________


#2 grps
frjm2<-function(parm,dat) {

t2<-0:2
t<-t(t(t2))
l<-length(t)
tp<-(t-2)/2
# u1<-1+tp
# u2<-u1+1
# u3<-u2+tp-1
theta<-parm[1]
b01<-parm[2]
b11<-parm[3]
b02<-parm[4]
b12<-parm[5]
iden<-diag(l)
```

```
sig<-parm[6]*iden
minn<-1e-15
rhop21<- -0.00772
taup11<-parm[7]
rhop11<-parm[8]
betap1<-parm[9]
taup21<-parm[10]
# rhop21<-parm[11]
betap2<-parm[11]
taup12<-parm[12]
rhop12<-parm[13]
taup22<-parm[14]
rhop22<- -0.00772


f11<- taup11*exp(dat[,4]*betap1)*exp(rhop11*dat[,5])*
exp(-exp(dat[,4]*betap1)*(taup11/rhop11)*(exp(rhop11*dat[,5])-1))
f21<- taup21*exp(dat[,4]*betap2)*exp(rhop21*dat[,5])*
exp(-exp(dat[,4]*betap2)*(taup21/rhop21)*(exp(rhop21*dat[,5])-1))
f12<- taup12*exp(dat[,4]*betap1)*exp(rhop12*dat[,5])*
exp(-exp(dat[,4]*betap1)*(taup12/rhop12)*(exp(rhop12*dat[,5])-1))
f22<- taup22*exp(dat[,4]*betap2)*exp(rhop22*dat[,5])*
exp(-exp(dat[,4]*betap2)*(taup22/rhop22)*(exp(rhop22*dat[,5])-1))

F11<-1-(exp(-exp(dat[,4]*betap1)*(taup11/rhop11)*(exp(rhop11*dat[,5])-1)))
F21<-1-(exp(-exp(dat[,4]*betap2)*(taup21/rhop21)*(exp(rhop21*dat[,5])-1)))
F12<-1-(exp(-exp(dat[,4]*betap1)*(taup12/rhop12)*(exp(rhop12*dat[,5])-1)))
F22<-1-(exp(-exp(dat[,4]*betap2)*(taup22/rhop22)*(exp(rhop22*dat[,5])-1)))

F1<-1-F11-F21
```

```
F2<-1-F12-F22

# for(i in 1:1000) {
# f11[i]<-max(1e-15, f11[i])
# }

# for(i in 1:1000) {
# f21[i]<-max(1e-15, f21[i])
# }

# for(i in 1:1000) {
# f12[i]<-max(1e-15, f12[i])
# }

# for(i in 1:1000) {
# f22[i]<-max(1e-15, f22[i])
# }

# for(i in 1:1000) {
# F1[i]<-max(1e-15, F1[i])
# }

# for(i in 1:1000) {
# F2[i]<-max(1e-15, F2[i])
# }


deny<-(1/(1+exp(theta))*dmvnorm(dat[,1:3],b01+(b11*t),sig)*
          (f11^dat[,7])*(f21^dat[,8])*(F1^dat[,9])+ exp(theta)/(1+exp(theta))*
         dmvnorm(dat[,1:3],b02+(b12*t),sig)*
```

```
        (f12^dat[,7])*(f22^dat[,8])*(F2^dat[,9]))

LL<-sum(log(deny))
return(-LL)
}
#_____________________________________________________________________________


#1 grp
frjm1<-function(parm,dat) {

t2<-0:2
t<-t(t(t2))
l<-length(t)
tp<-(t-2)/2
b01<-parm[1]
b11<-parm[2]
iden<-diag(l)
sig<-parm[3]*iden
minn<-1e-15
rhop2<- -0.00772
taup1<-parm[4]
rhop1<-parm[5]
betap1<-parm[6]
taup2<-parm[7]
betap2<-parm[8]


f1<- taup1*exp(dat[,4]*betap1)*exp(rhop1*dat[,5])*exp(-exp(dat[,4]*betap1)*
(taup1/rhop1)*(exp(rhop1*dat[,5])-1))
```

```
f2<- taup2*exp(dat[,4]*betap2)*exp(rhop2*dat[,5])*exp(-exp(dat[,4]*betap2)*
(taup2/rhop2)*(exp(rhop2*dat[,5])-1))

F1<-1-(exp(-exp(dat[,4]*betap1)*(taup1/rhop1)*(exp(rhop1*dat[,5])-1)))
F2<-1-(exp(-exp(dat[,4]*betap2)*(taup2/rhop2)*(exp(rhop2*dat[,5])-1)))

F<-1-F1-F2


deny<-dmvnorm(dat[,1:3],b01+(b11*t),sig)*(f1^dat[,7])*(f2^dat[,8])*(F^dat[,9])

LL<-sum(log(deny))
return(-LL)
}
#_____________________________________________________________________________




res<-matrix(NA,nrow=500,ncol=26)
s<-system.time(
for(i in 1:500) {
long1<-as.matrix(c(150,150,150))
long2<-as.matrix(c(150,160,170))
long3<-as.matrix(c(150,140,130))
long4<-as.matrix(c(150,200,250))
data<-matrix(NA,nrow=1200,ncol=9)
#data[,1:3]<-data2(1000,long1,long2,8)
#data[,1:3]<-data3(1200,long1,long2,long3,8)
#data[,1:3]<-data1(1000,long1,8)
```

```
data[,4:9]<-cdata(1200)
#data[,4:9]<-cdata2gr(1000,.5)
data[,1:3]<-data4(1200,long1,long2,long3,long4,8)

# 2grp
#opt<-optim(c(0,150,0,150,10,8,0.08, -0.25, -0.54, 0.01, -0.10, 0.08, -0.25, 0.01),
frjm2,dat=data, method="Nelder-Mead",control=list(maxit=10000))
#opt<-optim(c(0,150,0,150,0,8,0.05, -0.3, -0.54, 0.03, -0.10, 0.08, -0.25, 0.01),
frjm2,dat=data, method="Nelder-Mead",control=list(maxit=10000))

#opt 1 grp
#opt<-optim(c(150,0,8,0.08, -0.25, -0.54, 0.01, -0.10),frjm1,dat=data,
method="Nelder-Mead",control=list(maxit=10000))

#opt 3 grp
#opt<-optim(c(0,0,150,0,150,10,150,-10,8,0.08, -0.25, -0.54, 0.01, -0.10, 0.08,
-0.25, 0.01, 0.08, -0.25, 0.01),frjm3,dat=data, method="Nelder-Mead",
control=list(maxit=10000))



#opt 4 grp
opt<-optim(c(0,0,0,150,0,150,10,150,-10,150,50,8,0.08, -0.25, -0.54,
0.01, -0.10, 0.08, -0.25, 0.01, 0.08, -0.25, 0.01, 0.08, -0.25, 0.01),frjm4,
dat=data, method="Nelder-Mead",control=list(maxit=10000))




res[i,1:26]<-opt$par
}
)#system time end
```

```
results<-matrix(NA,nrow=26,ncol=3,dimnames=list(c("theta1","theta2","theta3","b01",
"b11","b02","b12","b03","b13","b04","b14","sigma","tau11","rho11","beta1","tau21",
"beta2","tau12","rho12","tau22","tau13","rho13","tau23","tau14","rho14",
"tau24"),c("estimate","se","bias")) )


sv4<-c(0,0,0,150,0,150,10,150,-10,150,50,8,0.08, -0.25, -0.54, 0.01, -0.10,0.08,-0.25,
0.01,0.08,-0.25,0.01,0.08,-0.25,0.01)

for(i in 1:26) {
results[i,1]<-mean(res[,i],na.rm=TRUE)
results[i,2]<-sd(res[,i],na.rm=TRUE)
results[i,3]<-abs(results[i,1]-sv4[i])
}
write.csv(results, "C:\\Users\\Nilesh\\Desktop\\Nilesh\\Pitt\\Dissertation
\\jeongfineresults\\
jmm1000results4gptry.csv")

#-------------------------------------
#bic testing
bic.mat<-matrix(NA,nrow=5,ncol=5)
conv.mat<-matrix(NA,nrow=5,ncol=5)
s<-system.time(
for(i in 1:5) {
long1<-as.matrix(c(150,150,150))
long2<-as.matrix(c(150,160,170))
data<-matrix(NA,nrow=1000,ncol=9)
data[,1:3]<-data2(1000,long1,long2,8)
data[,4:9]<-cdata(1000)
```

```
#opt 1 grp
opt<-optim(c(150,0,8,0.08, -0.25, -0.54, 0.01, -0.10),frjm1,dat=data, method=
"Nelder-Mead",control=list(maxit=10000))
bic.mat[i,1]<-bic(opt,8,1000)
conv.mat[i,1]<-opt$convergence

# 2grp
opt<-optim(c(0,150,0,150,10,8,0.08, -0.25, -0.54, 0.01, -0.10, 0.08, -0.25, 0.01),
frjm2,dat=data, method="Nelder-Mead",control=list(maxit=10000))
bic.mat[i,2]<-bic(opt,14,1000)
conv.mat[i,2]<-opt$convergence

#opt 3 grp
opt<-optim(c(0,0,150,0,150,10,150,-10,8,0.08, -0.25, -0.54, 0.01, -0.10, 0.08, -0.25,
 0.01, 0.08, -0.25, 0.01),frjm3,dat=data, method="Nelder-Mead",control=
list(maxit=20000))
bic.mat[i,3]<-bic(opt,20,1000)
conv.mat[i,3]<-opt$convergence

#opt 4 grp
opt<-optim(c(0,0,0,150,0,150,10,150,-10,150,5,8,0.08, -0.25, -0.54, 0.01, -0.10,
0.08, -0.25, 0.01, 0.08, -0.25, 0.01, 0.08, -0.25, 0.01),
frjm4,dat=data, method="Nelder-Mead",control=list(maxit=20000))
bic.mat[i,4]<-bic(opt,26,1000)
conv.mat[i,4]<-opt$convergence

bic.mat[i,5]<-0
if(max(bic.mat[i,1:4])==bic.mat[i,1]) {bic.mat[i,5]=1}
```

```
if(max(bic.mat[i,1:4])==bic.mat[i,2]) {bic.mat[i,5]=2}
if(max(bic.mat[i,1:4])==bic.mat[i,3]) {bic.mat[i,5]=3}
if(max(bic.mat[i,1:4])==bic.mat[i,4]) {bic.mat[i,5]=4}

conv.mat[i,5]<-sum(conv.mat[i,1:4])
})
```

BIBLIOGRAPHY

[1] Nagin DS. *Group Based Modeling of Development*. Harvard University Press: Cambridge, MA., 2005.

[2] Jones B, Nagin D, Roeder K. A SAS Procedure Based on Mixture Models for Estimating Developmental Trajectories. *Sociological Methods and Research* 2001; **29(3)**:374-393.

[3] Schwarz, G. Estimating the dimension of a model. *Annals of Statistics* 1978; **6(2)**: 461-464.

[4] Fine, J., Gray, R., A Proportional Hazards Model for the Subdistribution of a Competing Risk, *Journal of the American Statistical Association*, **Volume 94**, N. 446,1999, pages 496-509.

[5] Jeong, J., Fine, J., Parametric Regression on Cumulative Incidence Function,*Biostatistics*, **8,2**, 2007, 184-196.

[6] Tsiatis, A., Davidian, M., Joint Modeling of Longitudinal and Time-to Event Data: An Overview,*Statistica Sinica*, **Volume 14**, 2004, Pages 809 834.

[7] Lin H., Turnbull, B., McCulloch, C., Slate, E., Latent Class Models for Joint Analysis of Longitudinal Biomarker and Event Process Data: Application to Longitudinal Prostate-Specific Antigen Readings and Prostate Cancer, *Journal of the American Statistical Association*, **Volume 97**, No. 457, 2002.

[8] Nylund K, Asparouhov T, Muthen B. Deciding on the number of classes in latent class analysis and growth mixture modeling: a monte carlo simulation study. *Structural Equation Modeling* 2007;**14(4),**: 535-569.

[9] Berkof J, Mechlen I, and Gelman A. A Bayesian approach to the selection and testing of latent class models. *Statistica Sinica* 2003; **13**: 423–442.

[10] Lindsay B, Roeder K. Residual diagnostics for mixture models. *Journal of the American Statistical Association* 1992;**87(419)**: 785-794.

[11] Agresti A. *Categorical Data Analysis.* Wiley: New York, 1990.

[12] Titterington DM, Smith AFM, Makov UE. *Statistical Analysis of Finite Mixture Distributions.* Wiley: New York,1985.

[13] Efron, B. Bootstrap methods: Another look at the jackknife. *The Annals of Statistics* 1979; **7(1),**: 1-26.

[14] Meneses J, Antle CE, Bartholomew MJ, and Lengerich RL. A simple algorithm for delta method variances for multinomial posterior Bayes probability estimates. *Communications in Statistics – Simulation and Computation* 1990; **19(3)**: 837-845.

[15] Keenan K, Hipwell A, Chung T, Stepp S, Stouthamer-Loeber M, Loeber R, McTigue K. The Pittsburgh Girls Study: overview and initial findings. *Journal of Clinical Child and Adolescent Psychology* 2010; **39(4)**: 506-21.

[16] Deslandes, E., Chevret, S., Joint Modeling of Multivariate Longitudinal Data and the Dropout Process in a Competing Risk Setting: Application to ICU Data, *BMC Medical Research Methodology*, 2010, **10(69)**.

[17] Wulfsohn, M., Tsiatis A, A Joint Model for Survival and Longitudinal Data Measured With Error, *Biometrics*, **No. 1**, 1997, pages 330-339.

[18] Palevsky PM, Zhang JH, O'Connor TZ, Chertow GM, Crowley ST, Choudhury D et al. Intensity of renal support in critically ill patients with acute kidney injury. *N.Engl.J.Med.* 2008; **359**:7-20.

[19] Biernacki C, Govaert G. Choosing models in model-based clustering and discriminant analysis. *Institut National de Recherche en Informatique et en Automatique* 1998; **3509**: 1-22.

[20] Collins L, Lanza S. *Latent Class and Latent Transition Analysis.* Wiley: New York, 2010.

[21] Henderson, R., Diggle, P., Dobson, A. Joint Modelling of Longitudinal Measurements and Event Time Data,*Biostatistics*, **1,4**, 2000, 465-480.