

**GROWTH MIXTURE MODELING TO IDENTIFY PATTERNS OF THE  
DEVELOPMENT OF ACUTE LIVER FAILURE IN CHILDREN**

by

Song Zhang

B.Med., Harbin University of Medicine, China, 1995

M.Sc., Wright State University, 2000

Submitted to the Graduate Faculty of  
Graduate School of Public Health in partial fulfillment  
of the requirements for the degree of  
Master of Science

University of Pittsburgh

2012

UNIVERSITY OF PITTSBURGH

Graduate School of Public Health

This thesis was presented

by

Song Zhang

It was defended on

June 28, 2012

and approved by

Committee Chair:

Sati Mazumdar, Ph.D., Professor, Department of Biostatistics  
Graduate School of Public Health, University of Pittsburgh

Committee Members:

Steven Belle, Ph.D., Professor, Department of Epidemiology  
Graduate School of Public Health, University of Pittsburgh

Joyce Chang, Ph.D., Associate Professor

School of Medicine, University of Pittsburgh

Ruosha Li, Ph.D., Assistant Professor, Department of Biostatistics  
Graduate School of Public Health, University of Pittsburgh

Copyright © by Song Zhang

2012

**Sati Mazumdar, PhD**

**GROWTH MIXTURE MODELING TO IDENTIFY PATTERNS OF THE  
DEVELOPMENT OF ACUTE LIVER FAILURE IN CHILDREN**

Song Zhang, M.S

University of Pittsburgh, 2012

**ABSTRACT**

Pediatric Acute Liver Failure (PALF) is a clinical syndrome in which the affected children lose hepatic function and become critically ill within days. The causes of PALF remain indeterminate for about half of the cases. Liver transplantation is a lifesaving procedure but has long term adverse effects. It is critical to advance clinical insight by distinguishing patients who die without liver transplantation from those who are able to survive without transplantation. The PALF study is a multicenter study for children under 18 years old who present with acute liver failure. The study collected clinical and laboratory data for the first 7 days or until one of the events: death, transplantation or discharge occurred within 7 days following study enrollment. Growth Mixture Modeling (GMM) was applied to detect the trajectory patterns of INR (International Normalized Ratio) for hepatic-based coagulation through the first 7 days.

Three subgroups were identified by INR trajectories with 10.3% classified as high-INR, 34.7% as middle-INR and 55.0% as low-INR. The children with an indeterminate diagnosis were more likely to be classified into the high-INR group ( $p < 0.0001$ ) than were children with a specific diagnosis. The mortality without liver transplantation within 21 days of study entry was similar between the children in the high-INR group (19%) and in the middle-INR group (17%),

( $p=0.70$ ). The percentage of participants having liver transplantation was significantly higher among the children in the high-INR group (61%) than those in the middle-INR group (46%), ( $p=0.01$ ).

INR is used as a biomarker for determining the need of liver transplantation. Children with an indeterminate diagnosis were more likely to be in the high-INR group, and more likely to undergo liver transplantation as compared to other children with a specified diagnosis. The results suggest that INR was not a strong indicator for death without liver transplantation. Further studies should attempt to reveal biological mechanisms among the indeterminate diagnosis patients. This study has public health significance for its design to better understand the mechanism and progression of the children with acute liver failure from a multi-center collaboration.

## TABLE OF CONTENTS

<b>1.0</b>	<b>INTRODUCTION.....</b>	<b>1</b>
1.1	PEDIATRIC ACUTE LIVER FAILURE (PALF).....	1
1.2	THE PEDIATRIC ACUTE LIVER FAILURE (PALF) STUDY.....	2
1.3	GROWTH MIXTURE MODELS IN PALF STUDY.....	2
<b>2.0</b>	<b>DATA AND METHODS.....</b>	<b>4</b>
2.1	GROWTH MIXTURE MODELING AND INFERENCE.....	4
2.2	STUDY DESIGN, SAMPLE AND OUTCOME.....	6
2.2.1	Study Design and Sample.....	6
2.2.2	Outcome Measure.....	8
2.3	GROWTH MIXTURE MODEL FOR INR MODELING.....	9
2.4	GROWTH MIXTURE MODEL SELECTION.....	11
2.5	MISSING DATA.....	12
<b>3.0</b>	<b>RESULTS.....</b>	<b>14</b>
3.1	STUDY SAMPLE.....	14
3.2	NUMBER OF LATENT CLASSES.....	15
3.3	CLASSIFICATIONS FROM A 3-CLASS MODEL.....	15
<b>4.0</b>	<b>DISCUSSION.....</b>	<b>17</b>

<b>APPENDIX A - TABLES.....</b>	<b>19</b>
<b>APPENDIX B - FIGURES.....</b>	<b>23</b>
<b>BIBLIOGRAPHY.....</b>	<b>26</b>

## LIST OF TABLES

Table 1. Characteristics of PALF Participants by Diagnosis.....	19
Table 2. Model Fit Criteria for Different Models.....	20
Table 3. Association of 3 Latent Class Groups from 3-class Model with Diagnosis .....	21
Table 4. Association of 3 Latent Class Groups from 3-class Model with 21-day Outcome .....	22



## LIST OF FIGURES

Figure 1. Means of INR and 95% Confidence Interval in the First 7 Days.....	23
Figure 2. Diagram of 3-class, 2-piecewise GMM without Covariate for INR .....	24
Figure 3. Sample Means And Estimated Means from the fitted 3-class Model .....	25

## **1.0 INTRODUCTION**

### **1.1 PEDIATRIC ACUTE LIVER FAILURE (PALF)**

Acute liver failure (ALF) is a dramatic clinical syndrome in which previously healthy people lose hepatic function and become critically ill within days. It is a life-threatening illness of multiple etiologies, unusual severity and a rapid clinical course. A variety of infectious, metabolic, cardiovascular and drug-induced causes for ALF have been identified (1). However, the cause of pediatric ALF remains unknown (termed indeterminate) for about half of the cases (1). Due to the unknown etiology for the indeterminate children, it is difficult to determine treatment strategies. This may be due to lack of insight to clinical development, including the probability of recovery without liver transplantation. Children with an indeterminate diagnosis have greater probability of receiving liver transplantation than children with a specified etiology (1). Although liver transplantation is a lifesaving procedure for children with PALF, given a shortage of available organs, many die prior to undergoing a liver transplant action. Of patients with ALF who received liver transplantation, long-term survival is diminished in comparison to those who received liver transplantation for Wilson's or other chronic cases (2). Also, as some children spontaneously recover while awaiting a liver transplantation, it is possible that some children undergo transplantation who would have recovered without transplantation. With the current shortage of donor livers, and long term adverse sequelae of liver transplantation, it is

critical to discover methods to identify those children who have a high likelihood to survive without transplantation and those who are unlikely to survive without transplantation.

## **1.2 THE PEDIATRIC ACUTE LIVER FAILURE (PALF) STUDY**

The Pediatric Acute Liver Failure (PALF) study is a multinational collaborative study of infants, children and adolescents less than 18 years old who presented with ALF at one of the participating centers (1). The PALF was formed to facilitate an improved understanding of the pathogenesis, treatment and outcome of ALF in children which would serve to identify factors that would help to predict the likelihood of death or need for liver transplantation (1). One aim of the PALF study is to develop better methods than currently exist to predict outcomes by identifying factors to predict spontaneous survival, i.e., survival without liver transplantation. The PALF study created a database of a cohort of 986 children that included demographic, clinical, and laboratory data and short-term outcomes. As an observational study, all measures and treatment were performed as standard of care at each of the participating centers. Clinical and laboratory data were collected for the first 7 days following study enrollment or until one of the events, death, transplantation, or discharge occurred within 21 days of study entry.

## **1.3 GROWTH MIXTURE MODELS IN PALF STUDY**

Since etiology is so often unknown in PALF, it is of interest to identify patients with distinctive clinical patterns which are associated with different prognostics and outcomes. Mixture modeling

explores the data structure for detecting the homogeneous subgroups from a heterogeneous population. In the case of a heterogeneous population, mixture modeling is able to determine the probability of each participant's belonging to a particular subgroup. Growth Mixture Models (GMM) developed by Muthen and his colleagues (3, 4, 5) provided a method to detect heterogeneity through growth trajectory patterns by finding distinctive patterns.

In this paper, we apply growth mixture modeling method to the PALF cohort data. We classify the PALF patients based on the patterns of changes in the International Normalized Ratio (INR), which is a biomarker of liver function measuring hepatic-based coagulopathy. Higher values of INR mean that blood is taking more time to coagulate or clot. INR is an important biomarker for deciding upon listing a patient for liver transplantation (6). The major aim in this paper is to group PALF patients based on their INR trajectory in the first 7 days after enrollment to the study and examine whether the subgroups distinguish the participants who survive without liver transplantation from the ones who die without such a procedure. In addition, categorizing participants into distinctive patterns of disease progression may lead to better understanding of indeterminate patients, or expanding definitions of existing diagnoses, and to enable more precise insight of PALF.

## 2.0 DATA AND METHODS

### 2.1 GROWTH MIXTURE MODELING AND INFERENCE

We use some notations from Muthen (3). In a longitudinal design the response variables  $y_i$  are continuous observed variables ( $y_{i1}, y_{i2}, \dots, y_{iT}$ ) for individual  $i$  with  $i = 1, \dots, N$ , at potential  $T$  time points  $t_{i1}, \dots, t_{iT}$ . In growth mixture modeling, we explore the relationship of the response variables to two different kinds of latent variables. The first kind is an  $M$ -dimensional vector of latent continuous growth variables  $\boldsymbol{\eta}_i = (\eta_{i1}, \dots, \eta_{iM})^T$  in individual  $i$  representing the random effects regarding intercepts and slopes. The second kind is a  $K$ -dimensional latent categorical variables (also called latent class variables),  $\mathbf{c}_i = (c_{i1}, \dots, c_{iK})^T$  where  $c_{ik}=1$  or zero depending on whether the individual  $i$  belongs to class  $k$  for  $k = 1, \dots, K$ .

In mixture model analysis, models with different numbers of latent classes are explored and compared in terms of fit for the best number of classes. For  $K$  latent classes, let  $\boldsymbol{\pi}_i = (\pi_{i1}, \pi_{i2}, \dots, \pi_{iK})^T$  be the probability vector associated with  $\mathbf{c}_i$ , where  $\pi_{ik}$ , equal to  $P(c_{ik}=1)$ , denotes the probability of the individual  $i$  belonging to the class  $k$  and  $\sum_{k=1}^K \pi_{ik} = 1$ .

For individual  $i$ , without any covariate, the logit model for  $\boldsymbol{\pi}_i$  is expressed as

$$\text{logit}(\boldsymbol{\pi}_i) = \left[ \log\left(\frac{\pi_{i1}}{\pi_{iK}}\right), \log\left(\frac{\pi_{i2}}{\pi_{iK}}\right), \dots, \log\left(\frac{\pi_{iK-1}}{\pi_{iK}}\right) \right]^T = \boldsymbol{\alpha}_e \quad (1)$$

where the class  $K$  is a reference class,  $\boldsymbol{\alpha}_c$  is a  $(K-1)$  parameter vector, and  $\boldsymbol{\pi}_i$  is modeled by a multinomial logit regression as unordered categorical outcomes.

The latent class variable is incorporated in the growth mixture model by letting each individual's intercept and slope vary. The distributions of these intercepts and slopes are determined by their class membership. The expression of  $y_i$  related to the continuous latent variable  $\boldsymbol{\eta}_i$  for individual  $i$  in the mixture model framework is

$$y_i = \boldsymbol{A}_y \boldsymbol{\eta}_i + \boldsymbol{\varepsilon}_i \quad (2)$$

where  $\boldsymbol{A}_y$  is a  $T \times M$  design matrix;  $\boldsymbol{\eta}_i$  is affected by the membership of the latent class variable which identifies the individual  $i$  into one of  $K$  classes,

$$\boldsymbol{\eta}_i = \boldsymbol{W} \boldsymbol{c}_i + \boldsymbol{\zeta}_i \quad (3)$$

where  $\boldsymbol{W}$  is an  $M \times K$  matrix containing columns  $\boldsymbol{\omega}_k$  as class-specific parameters for each class representing the mean of  $[\boldsymbol{\eta}_i | \boldsymbol{c}_i]$ , for  $k = 1, \dots, K$ , without any covariate adjustment. The distribution of residual vector  $\boldsymbol{\varepsilon}_i$  is  $N(\mathbf{0}, \boldsymbol{\Theta})$  with  $\boldsymbol{\Theta}$  a diagonal covariance matrix, and the distribution of residual vector  $\boldsymbol{\zeta}_i$  is  $N(\mathbf{0}, \boldsymbol{\Omega}_k)$  accounting for the class-specific feature, with assuming  $\boldsymbol{\varepsilon}_i$  and  $\boldsymbol{\zeta}_i$  uncorrelated to each other. Given latent class variable  $\boldsymbol{c}_i$ , the conditional distribution  $[y_i | \boldsymbol{c}_i]$  follows  $N_T(\boldsymbol{A}_y \boldsymbol{W} \boldsymbol{c}_i, \boldsymbol{A}_y \boldsymbol{\Omega}_k \boldsymbol{A}_y' + \boldsymbol{\Theta})$ .

Growth mixture modeling explains unobserved heterogeneity among subjects in their longitudinal progress using both random effects (7) and finite mixtures (8) by allowing separate sets of parameters for mixture components. In mixture modeling, parameters are estimated by the method of maximum likelihood (e.g., EM algorithm). The growth mixture modeling techniques were implemented with the Mplus program (9).

Methods for fitting GMM have been demonstrated elsewhere (3, 9). A brief description of model fitting is summarized here. GMM applies the EM algorithm by treating the continuous

latent variables  $\boldsymbol{\eta}_i$  and the categorical latent class variables  $c_i$  as missing data. We can write the complete-data log-likelihood for the individual  $i$  as follows ( $[x]$  denotes a density or distribution of a random variable vector  $x$ ).

$$\log L = \sum_{i=1}^N (\log[c_i] + \log[\boldsymbol{\eta}_i | c_i] + \log[y_i | \boldsymbol{\eta}_i]) \quad (4)$$

where

$$\sum_{i=1}^N (\log[c_i]) = \sum_{i=1}^N \sum_{k=1}^K c_{ik} \log \pi_{ik} \quad (5)$$

and  $[\boldsymbol{\eta}_i | c_i]$  is assumed to follow the normal distribution  $N_M(\mathbf{W}c_i, \boldsymbol{\Omega}_k)$  and  $[y_i | \boldsymbol{\eta}_i]$  is assumed to follow the normal distribution  $N_T(\boldsymbol{A}_y \boldsymbol{\eta}_i, \boldsymbol{\Theta})$ ,  $N_B(z)$  representing the normal distribution for a B-dimensional normal vector  $z$ .

In E-step, to maximize the expectation of the complete-data log likelihood in model (4) with respect to the missing data, given the response observations  $y_i$ ,  $c_i$  can be written as

$$P(c_{ik} = 1 | y_i) = \pi_{ik} N_T(\boldsymbol{A}_y \mathbf{W}c_i, \boldsymbol{A}_y \boldsymbol{\Omega}_k \boldsymbol{A}_y + \boldsymbol{\Theta}) / \sum_{k=1}^K \pi_{ik} N_T(\boldsymbol{A}_y \mathbf{W}c_i, \boldsymbol{A}_y \boldsymbol{\Omega}_k \boldsymbol{A}_y + \boldsymbol{\Theta}) \quad (6)$$

Model (6) represents the posterior probability  $p_{ik}$  for the latent class variables  $c_{ik}$ .

Repetition of the E-steps and M-steps continues until convergence is reached. To achieve the global maximum, several different starting values were employed in the final model.

## 2.2 STUDY DESIGN, SAMPLE AND OUTCOME

### 2.2.1 Study Design and Sample

The PALF Study Group (PALFSG) is a multi-center and multi-national collaborative study consisting of 24 sites from the United States (21 sites), Canada (1 site) and the United Kingdom

(2 sites), that developed a dataset to facilitate and improve understanding of the pathogenesis, treatment and outcome of acute liver failure in infants, children, and adolescents. Participants through 17 years of age were eligible for the PALF study if they met the following entry criteria : 1) no known evidence of chronic liver disease; 2) biochemical evidence of acute liver injury; 3) hepatic-based coagulopathy defined as a prothrombin time (PT) was between 15.0-19.9 seconds or INR was between 1.50-1.99 not corrected by vitamin K in the presence of clinical hepatic encephalopathy (HE) or PT at least 20 seconds or INR at least 2.0, regardless of the presence or absence of clinical HE. PT is an alternative test to INR for liver coagulopathy, but the result of the PT test depends on the method used so that there is variability among the laboratories of the participating centers. HE is defined as a disturbance in central nervous system function because of hepatic insufficiency. The screening for eligibility was performed prior to any plasma therapy and no more than 72 hours prior to enrollment in the PALF study. Following informed consent from a parent or legal guardian, demographic, clinical and laboratory information were recorded daily for the first 7 days from enrollment into the PALF study or until the occurrence of death, liver transplantation, or discharge within 7 days. A final diagnosis was determined for each participant by the primary investigator at the clinical site. The primary outcome was determined at 3 weeks (21 days) after entry into the study as the earliest of the following events: death without transplantation, transplantation and survival without transplantation.

As of July 2011, there were 986 participants in the PALF database. INR values were used in this paper as the longitudinal assessments of a patient's clinical severity, with higher values indicating worse hepatic coagulopathy. One site from the United Kingdom, which measured hepatic-based coagulopathy by PT instead of INR, was excluded from the analysis. The remaining 23 sites had 914 patients. There were another 30 patients for whom INR values were



not collected so these participants were excluded. This resulted in 884 participants with at least one INR measured in the first 7 days after study enrollment. In PALF, eligibility regarding INR or PT were performed between 72 hours prior to enrollment and up to enrollment, however, the actual INR/PT values used for eligibility were not recorded in the PALF database, which means that there are some participants whose INR levels were always below 1.5 following enrollment. Since this was an observational study, INR may not have been collected due to limitations in drawing blood or physician's judgment.

### **2.2.2 Outcome Measure**

INR is used to measure the speed of a particular pathway of coagulation, comparing it to the normal condition. With a heavily damaged liver, the INR increases with the synthesis of vitamin-K dependent coagulation factor getting impaired. The coagulopathy of participants in PALF cannot be corrected by vitamin K as that is an exclusion criterion. INR was designed to be collected daily for the first 7 days from enrollment if PALF participants stayed in the hospital before transplantation. As shown in Figure 1, the INR collected during the first 7 days averaged 3.34, 2.94, 2.72, 2.72, 2.59, 2.59, 2.49, depicting a trend of decline in the first 3 days followed by a relatively flat trajectory in the next 4 days.

### 2.3 GROWTH MIXTURE MODEL FOR INR MODELING

The aim of using GMM was to examine whether variability in change of INR up to 7 days can be captured by identifying patterns of trajectories from their distinctive natural progress. The time course of change in INR during the first 7 days into PALF was modeled in the 884 participants with at least one INR measured. As it was our aim to classify PALF patients based on INR trajectory patterns only, no covariate was included in the growth mixture models. To establish the best classification and most interpretable solutions based on the longitudinal pattern of INR trajectory, we fit a series of growth mixture models allowing both linear and quadratic trajectories. However, these growth models did not provide adequate fits due to a more rapid change rate from day 1 to day 3 than from day 4 to day 7 in a portion of the PALF patients as shown in Figure 1. Thus, piecewise models were applied to separate time periods, with distinct growth curves by transition points. Piecewise models allow flexible shapes with straight or curved trajectories between the transition points to improve the local fit (10, 12). Two-piece growth mixture models were discovered to better fit the observed data than the models with simple linear and quadratic trends. The two periods were the first 3 days and the next 4 days with linear trend over the first period and the quadratic curve in the second period. This was determined by the observed data.

To explicitly express model (2), we have

$$\begin{bmatrix} y_{i1} \\ y_{i2} \\ y_{i3} \\ y_{i4} \\ y_{i5} \\ y_{i6} \\ y_{i7} \end{bmatrix} = \begin{bmatrix} 1 & t_{i1} & 0 & 0 \\ 1 & t_{i2} & 0 & 0 \\ 1 & t_{i3} & 0 & 0 \\ 1 & t_{i3} & t_{i4} - t_{i3} & (t_{i4} - t_{i3})^2 \\ 1 & t_{i3} & t_{i5} - t_{i3} & (t_{i5} - t_{i3})^2 \\ 1 & t_{i3} & t_{i6} - t_{i3} & (t_{i6} - t_{i3})^2 \\ 1 & t_{i3} & t_{i7} - t_{i3} & (t_{i7} - t_{i3})^2 \end{bmatrix} \begin{bmatrix} \eta_{i0} \\ \eta_{i1} \\ \eta_{i2} \\ \eta_{i3} \end{bmatrix} + \begin{bmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \varepsilon_{i3} \\ \varepsilon_{i4} \\ \varepsilon_{i5} \\ \varepsilon_{i6} \\ \varepsilon_{i7} \end{bmatrix}$$

where  $\eta_{i0}$  represents INR status for individual  $i$  at the entry of study;  $\eta_{i1}$  represents the linear slope of growth trajectory for individual  $i$  in the first time-period;  $\eta_{i2}$  denotes the linear slope of growth trajectory for individual  $i$  in the second time-period;  $\eta_{i3}$  denotes the quadratic slope of growth trajectory for individual  $i$  in the second time-period.

The design matrix is given by

$$A_y = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 2 & 0 & 0 \\ 1 & 2 & 1 & 1 \\ 1 & 2 & 2 & 4 \\ 1 & 2 & 3 & 9 \\ 1 & 2 & 4 & 16 \end{bmatrix}$$

To incorporate the latent class by letting the individual intercepts and slopes vary as functions of the membership of latent class, with a 3-class model, model (3) was written as

$$\begin{bmatrix} \eta_{i0} \\ \eta_{i1} \\ \eta_{i2} \\ \eta_{i3} \end{bmatrix} = \begin{bmatrix} \omega_{10} & \omega_{20} & \omega_{30} \\ \omega_{11} & \omega_{21} & \omega_{31} \\ \omega_{12} & \omega_{22} & \omega_{32} \\ \omega_{13} & \omega_{23} & \omega_{33} \end{bmatrix} \begin{bmatrix} c_{i1} \\ c_{i2} \\ c_{i3} \end{bmatrix} + \begin{bmatrix} \zeta_{i0} \\ \zeta_{i1} \\ \zeta_{i2} \\ \zeta_{i3} \end{bmatrix}$$

If a participant was in class 1, the first column  $\omega_1 = (\omega_{10}, \omega_{11}, \omega_{12}, \omega_{13})$  is the class-specific parameters to capture the features of trajectory for INR developmental pattern for the participants in class 1. Similarly,  $\omega_2 = (\omega_{20}, \omega_{21}, \omega_{22}, \omega_{23})$  and  $\omega_3 = (\omega_{30}, \omega_{31}, \omega_{32}, \omega_{33})$  are class-specific parameters for the participants in class 2 and class 3 respectively.

To keep the model identifiable, the variance-covariance matrix was assumed to be class invariant. i.e., the class-specific features of the residuals are assumed to be similar. Furthermore, we assumed that the correlations only existed between the intercept and the linear slope over the first time-period, as well as between the linear slope over the first time-period and the linear slope over the second time-period;  $\Omega$  was class-invariant across the 3 classes with normal distribution of mean 0 and variance-covariance matrix

$$\begin{bmatrix} \text{var}(\eta_{i0}) & \text{cov}(\eta_{i0}, \eta_{i1}) & 0 & 0 \\ \text{cov}(\eta_{i0}, \eta_{i1}) & \text{var}(\eta_{i1}) & \text{cov}(\eta_{i1}, \eta_{i2}) & 0 \\ 0 & \text{cov}(\eta_{i1}, \eta_{i2}) & \text{var}(\eta_{i2}) & 0 \\ 0 & 0 & 0 & \text{var}(\eta_{i3}) \end{bmatrix}$$

The proposed 3-class model described above is given graphically in Figure 2.

## 2.4 GROWTH MIXTURE MODEL SELECTION

There are several criteria to evaluate the fit of growth mixture models. The Bayesian Information Criterion (BIC) measures how adequately the model describes the data without too many parameters. BIC calculates the maximized likelihood with a penalty for overfitting due to adding parameters in the model. A smaller value indicates a better fit (12). An alternative to BIC, the Akaike Information Criterion (AIC) has been shown to overestimate the number of components

from finite mixture models such that the BIC consistently outperforms AIC by picking the correct model (13). The simulations to decide the number of classes in growth mixture modeling proved superiority of BIC to AIC by correctly indicating the correct numbers across different kinds of models and different sample sizes (14). The Lo-Mendell-Rubin likelihood ratio test is specifically used for mixture models to compare the likelihood between the model with K classes and the model with K-1 classes. A p-value less than 0.05 is considered to favor the K-class model significantly over the K-1 class model (14, 15). In addition, entropy values are calculated for models with more than one class, to evaluate the accuracy and uncertainty of classification of individuals into latent classes. Entropy values range from 0 to 1, with 0 indicating complete randomness and 1 is perfect classification (16). To prevent overfitting in mixture modeling, a latent class including a very small fraction of individuals is not recommended (17).

## **2.5 MISSING DATA**

Missing data data are a challenge in the PALF study due to early occurrence of outcome (prior to 7 days in the study), or lack of a blood draw to obtain INR which may be due to investigator's judgment not to measure INR on a certain day. Of the 377 PALF patients who died, underwent transplantation or were discharged prior to 7 days, 28.3% had at least one INR measure missing during their hospitalization; whereas among the 507 patients who stayed in the study for at least 7 days, 35.5% had at least one missing INR. M-plus (9) conducts maximum likelihood estimation for datasets containing missing data without the need to impute missing values and

provides unbiased estimates under the relatively unrestrictive missing at random (MAR) assumption (18). We assumed missing at random for INR among PALF patients in the analyses.

## **3.0 RESULTS**

### **3.1 STUDY SAMPLE**

Among the 884 participants, 389 (44.0%) had an unknown etiology, diagnosis as indeterminate, while the remaining 495 participants had a specified final diagnosis with 116 (13.1%) as acetaminophen toxicity, 82 (9.3%) as metabolic disease, 61 (6.9%) as autoimmune hepatitis, 66 (7.5%) as viral infection hepatitis, 170 (19.2%) as other miscellaneous specified diagnoses including hemophagocytic syndrome, shock/ischemia, drug-induced hepatitis. Compared to the participants with a specified final diagnosis, the participants with an indeterminate final diagnosis were on average younger at enrollment, more likely to be male, and more likely to have hepatic encephalopathy at enrollment. The clinical lab tests at enrollment reflected the indeterminate participants having worse liver function with significantly higher total bilirubin and INR than the participants with a specified diagnosis. A participant with an indeterminate final diagnosis had a higher probability of undergoing liver transplantation than those with a specified diagnosis within the first 3 weeks from enrollment (Table 1).

### **3.2 NUMBER OF LATENT CLASSES**

Exploring models with different numbers of classes suggested that PALF patients were not homogeneous in terms of change in INR (Table 2). The 3-class piecewise GMM was favored over both the single-class and the 2-class piecewise models using the Lo-Mendell-Rubin likelihood ratio test. BIC results consistently favored the 3-class piecewise model. The 4-class model provided a marginal improvement in the Lo-Mendell-Rubin likelihood ratio test ( $p=0.04$ ) over the 3-class model, but indicated a worse quality of classification (entropy=0.89 for the 3-class model vs. entropy=0.81 for the 4-class model). The 5-class model provided no advantage over the 4-class model based on the Lo-Mendell-Rubin likelihood ratio test ( $p=0.10$ ). Finally, the 3-class model was chosen as the final model as each subgroup contained a reasonable proportion of participants whereas the 4-class model might cause overfitting as it contained a subgroup with less than 10% of participants.

### **3.3 CLASSIFICATIONS FROM A 3-CLASS MODEL**

The time course of change in INR during the first 7 days after enrolling in the PALF study was modeled in the 884 PALF participants with at least one INR measured. The 3-class model classified the participants into subgroups which we labeled as high-INR with 91 (10.3%) participants, middle-INR with 307 (34.7%) participants and low-INR with 486 (55.0%) participants. The observed and the estimated means of INR in the three groups from the 3-class model are given in Figure 3. The mean INR in the high-INR group was 7.26 at enrollment



followed by a rapid decline to 6.38 on day 3, then rebounding back to 6.70 on day 4 and day 5. It then decreased to 6.28 and 5.50 on day 6 and day 7. The mean INR in the middle-INR group started at 3.62, and decreased to 3.05 on day 3, then stayed at a stable level at 3.02 until the end of 7 days. The mean INR in the low-INR group was 2.32 at enrollment with moderate decrease to 1.81 on day 3, and then declined during the last 4 days to 1.57 on day 7.

Table 3 provides some information regarding the associations of the class membership based on the fitted 3-class model with diagnoses of different comorbidities. The participants with indeterminate diagnosis were most likely to belong to the high-INR group and least likely to belong to the low-INR group as compared to the participants with a specified diagnosis ( $p < 0.0001$ ). The association of 3 latent class groups was significant ( $p < 0.0001$ ) with the study endpoint of the earliest event from death without transplantation, transplantation and survival without transplantation (Table 4). In the high-INR group, 19% died without transplantation, 61% received transplantation and 20% survived with native liver without transplantation. In the middle-INR group, 17% died, 46% received transplantation and 37% survived with native liver. In the low-INR group, 7% of the participants died, 13% received liver transplantation and 79% survived with native liver without transplantation. The mortality rates without liver transplantation within 21 days were similar ( $p = 0.70$ ) in the high-INR group (19%) and in the middle-INR group (17%), whereas the rate of transplantation was significantly higher ( $p = 0.01$ ) in the high-INR group (61%) than those in the middle-INR group (46%), which remained significant after adjustment for multiple comparisons with Bonferroni correction.

## 4.0 DISCUSSION

The exploration of the INR trajectory showed important heterogeneity in the temporal pattern of change over first 7 days following enrollment in PALF participants. PALF participants have rapid onset of disease, severe progress and uncertain prognosis, with substantial mortality and high probability of liver transplantation. The aim in this study was to discover methods and identify participants who have a high likelihood of surviving without transplantation from those who are unlikely to survive without transplantation based on INR. We found that modeling with GMM can provide potential useful information. Using a GMM model to identify distinctive growth trajectories, we classified the PALF participants into 3 subgroups (latent classes) according to INR trajectory. The configuration including the starting INR level with INR's change in the first week for each latent class provided an insight of how INR progressed quantitatively. There was similar mortality without live transplantation between the model-identified high-INR group and the middle-INR group, but the rate of transplantation was much higher in the high-INR group than in the middle-INR group which indicated that INR served as a factor in deciding whether the participant should undergo liver transplantation. This result provides the clinical investigators a second thought on whether INR level should be a pivotal indicator for determining a liver transplantation. The indeterminate patients were more likely to be classified into the high INR group than those with a specific diagnosis and more likely to

undergo liver transplantation. This may suggest that we need a more comprehensive evaluation to clearly delineate decision making regarding liver transplantation. Further work will focus on exploration of trajectories of other important clinical biomarkers, e.g. total bilirubin or hepatic encephalopathy, and their association with outcomes. We need to establish a comprehensive assessment to evaluate the severity of disease among indeterminate patients and to accurately predict the possibility of hepatic recovery with native liver to prevent from unnecessary liver transplantation.

## APPENDIX A- TABLES

**Table 1. Characteristics of PALF Participants by Diagnosis**

	Participants with a specified diagnosis N=495 N (%)	Participants with indeterminate diagnosis N=389 N (%)	P-value from either Wilcoxon or $\chi^2$ test
Age at enrollment (year)			
Median	7.0	4.2	0.002
Range	0.0, 18.0	0.0, 17.9	
Less than 1 yr	143 (28.9)	85 (21.9)	<0.0001
1-2 yrs old	48 (9.7)	71 (18.3)	
3-9 yrs old	89 (18.0)	139 (35.7)	
Greater than 10 yrs	215 (43.4)	94 (24.2)	
Sex			
Male	219 (44.2)	227 (58.4)	<0.0001
Encephalopathy at enrollment			
Missing	33	19	0.02
No	247 (53.4)	168 (45.3)	
Yes	216 (46.7)	203 (54.7)	
Total bilirubin at enrollment (mg/dl)			<0.0001
N	412	334	
Median	5.1	13.9	
Range	0.3, 59.8	0.1, 40.9	
AST at enrollment (Aspartate transaminase IU/L)			0.06
N	454	358	
Median	1666.5	2052.5	
Range	4.0, 46311.0	19.0, 32040.0	
Albumin at enrollment (mg/dl)			0.29
N	448	357	
Median	2.8	2.9	
Range	0.6, 4.9	1.1, 4.5	
INR at enrollment			0.002
N	459	368	
Median	2.5	2.8	
Range	1.0, 14.9	1.0, 26.4	
21-day outcome			
Died without transplantation	66 (13.3)	39 (10.0)	<0.0001
Transplantation	89 (18.0)	174 (44.9)	
Survival without transplantation	340 (68.7)	176 (45.1)	

**Table 2. Model Fit Criteria for Different Models**

Model	# of classes	BIC	Entropy	p-value from Lo-Mendell-Rubin tests	Proportion of subjects in class				
					1	2	3	4	5
Linear GMM									
1-class model	1	18407.26			.000				
2-class model	2	13332.36	0.91	<0.0001	.249	.751			
3-class model	3	12127.40	0.87	0.0003	.122	.407	.471		
4-class model	4	11804.78	0.83	0.006	.061	.197	.350	.392	
5-class model	5	11897.11	0.77	0.06	.050	.144	.261	.265	.281
Quadratic GMM									
1-class model	1	18193.25			.000				
2-class model	2	12927.70	0.91	<0.0001	.271	.729			
3-class model	3	11734.22	0.88	0.03	.118	.376	.506		
4-class model	4	11460.68	0.81	0.06	.080	.258	.317	.344	
5-class model	5	1162.53	0.78	0.16	.052	.143	.232	.267	.306
2-piecewise GMM									
1-class model	1	18285.88			.000				
2-class model	2	12905.89	0.91	<0.0001	.298	.702			
3-class model	3	11699.36	0.89	0.0003	.103	.347	.550		
4-class model	4	11479.74	0.81	0.04	.071	.272	.318	.338	
5-class model	5	11620.17	0.77	0.10	.051	.146	.237	.254	.312

**Table 3. Association of 3 Latent Class Groups from 3-class Model with Diagnosis**

Diagnosis	High INR group	Middle INR group	Low INR group	Total
Acetaminophen toxicity N (column percentage) (row percentage)	6 (6.6) (5.2)	27 (8.8) (23.3)	83 (17.1) (71.1)	116  (13.1)
Metabolic disease N (column percentage) (row percentage)	6 (6.6) (7.3)	8 (9.1) (34.2)	48 (9.9) (58.5)	82  (9.3)
Autoimmune hepatitis N (column percentage) (row percentage)	2 (2.2) (3.3)	10 (3.3) (16.4)	49 (10.1) (80.3)	61  (6.9)
Viral infection hepatitis N (column percentage) (row percentage)	9 (9.9) (13.6)	23 (7.5) (34.9)	34 (7.0) (51.5)	66  (7.5)
Other specified diagnosis N (column percentage) (row percentage)	8 (8.8) (4.7)	57 (18.6) (33.5)	105 (21.6) (61.8)	170  (19.2)
Indeterminate N (column percentage) (row percentage)	60 (65.9) (15.4)	162 (52.8) (41.7)	167 (34.4) (42.9)	389  (44.0)
Total N (row percentage)	91  (10.3)	307  (34.7)	486  (55.0)	884

**Table 4. Association of 3 Latent Class Groups from 3-class Model with 21-day Outcome**

21-day outcome	High INR group n=91 N (% <sup>#</sup> )	Middle INR group n=307 N (% <sup>#</sup> )	Low INR group n=486 N (% <sup>#</sup> )
Death without LT <sup>a</sup>	17 (19%)	52 (17%)	36 (7%)
LT <sup>b</sup>	56 (61%)	142 (46%)	65 (13%)
Survival without LT	18 (20%)	113 (37%)	384 (79%)

LT = Liver Transplantation

<sup>a</sup> P-value from chi-square test to compare death without LT between high INR group (19%) and middle INR group (17%) was 0.70.

<sup>b</sup> P-value from chi-square test to compare LT between high INR group (61%) and middle INR group (46%) was 0.01.

<sup>#</sup>% represents column percentage

## APPENDIX B- FIGURES

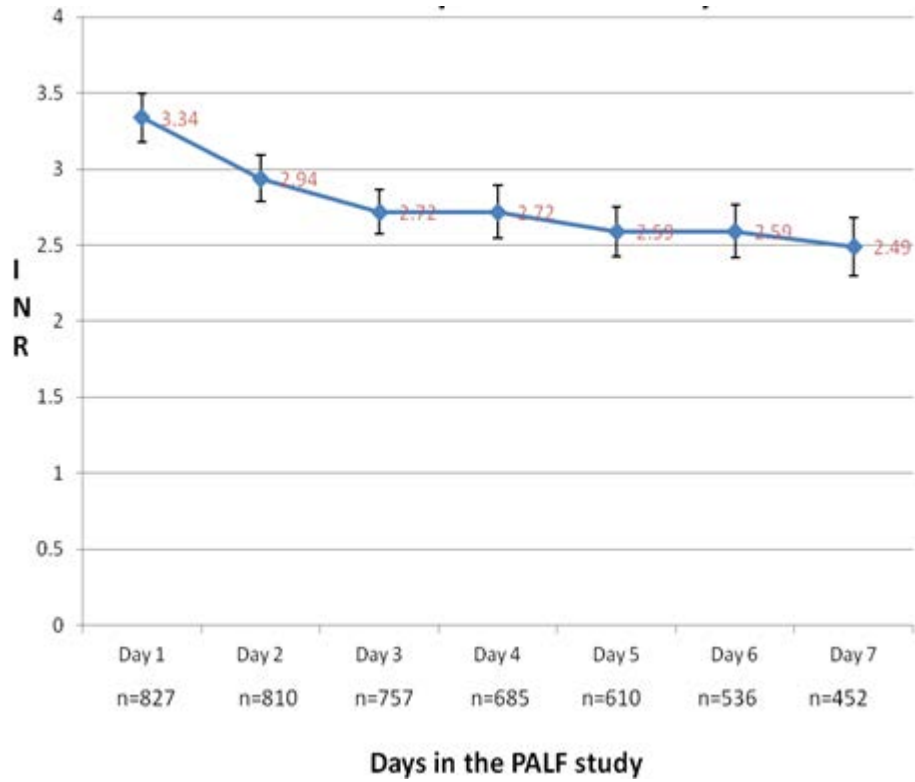


Figure 1. Means of INR and 95% Confidence Interval in the First 7 Days



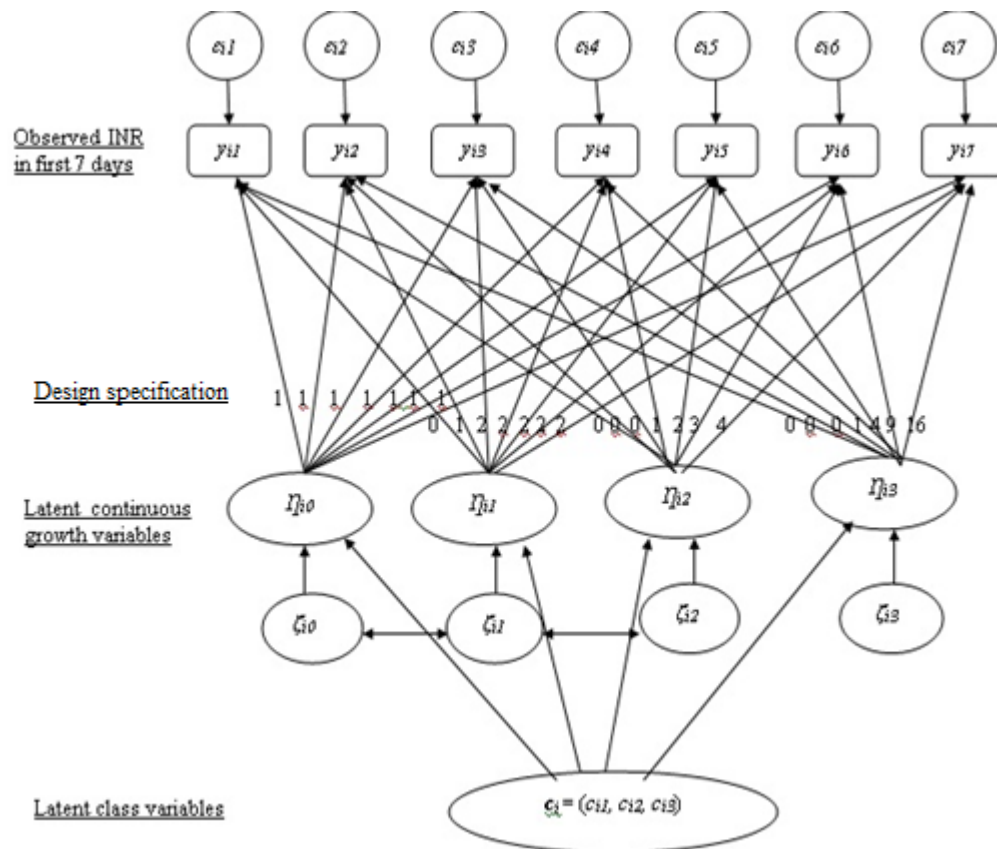


Figure 2. Diagram of 3-class, 2-piecewise GMM without Covariate for INR

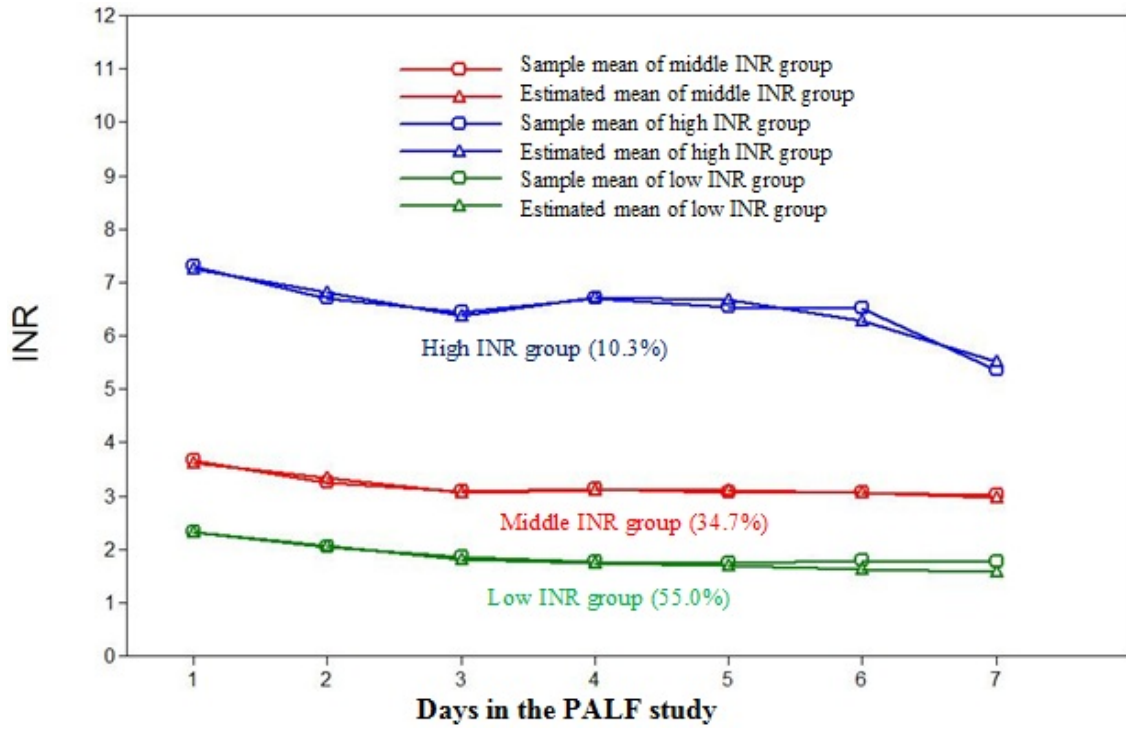


Figure 3. Sample Means And Estimated Means from the fitted 3-class Model

## BIBLIOGRAPHY

1. Squires RH, Schneider BL, Bucuvalas J, et al. (2006) *Acute liver failure in children: the first 348 patients in the pediatric acute liver failure study group*. *J Pediatr*, 148:652-658.
2. McDiarmid SV, Goodrich NP, Harper AM, Merion RM (2007). *Liver transplantation for status 1: the consequences of good intentions*. *Liver Transpl*;13:699-707.
3. Muthen B., Shedden B., (1999) *Finite mixture modeling with mixture outcomes using the EM algorithm*. *Biometrics*, 55, 463-469.
4. Muthen B., Asparouhov T., (2008) *Growth mixture modeling: Analysis with non-Gaussian random effects*. In *Longitudinal Data Analysis*, Chapman & Hall/CRC Press: Boca Raton, 143-165.
5. Muthen B., (2002) *Beyond SEM: General Latent Variable Modeling*. *Behaviormetrika*, 29:81-117.
6. Gheorghe L, Popescu I, Iacob R, Iacob S and Gheorghe C(2005) *Predictors of death on the waiting list for liver transplantation characterized by a long waiting time* *Transplant International* 18 572–576.
7. Laird, N. M. and Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics* **38**, 963–974.
8. McLachlan, G. J. and Peel, D. (2000). *Finite Mixture Models*. New York: Wiley.
9. Muthen B., & Muthen L., (2008) *Mplus user's guide (5th ed.)* Los Angeles: Methen & Muthen.
10. Li F, Duncan TE., Duncan SC., Hops H., (2001) *Piecewise growth mixture modeling of adolescent alcohol use data*. *Structural Equation Modeling: A Multidisciplinary Journal*, 8(2), 175-204.
11. Duncan TE., Duncan SC., Strycker LA., (2006) *Piecewise and pooled interrupted time series LGMs*. In *an introduction to latent variable growth curve modeling*, Lawrence Erlbaum Associates, Publishers: Mahwah, London. 151-164.

12. Schwarz, Gideon E. (1978). *Estimating the dimension of a model*. *Annals of Statistics* 6 (2): 461–464.
13. Roeder, K., & Wasserman, L. (1997). *Practical Bayesian density estimation using mixtures of normals*. *Journal of the American Statistical Association*, 92, 894–902.
14. Nylund K., Asparouhov T., Muthen B., (2008). *Deciding on the number of classes in latent class analysis and growth mixture modeling : A Monte Carlo simulation study*. *Structural Equation Modeling*, 14(4), 535-569.
15. Lo Y., Mendell NR., Rubin DB., (2001). *Testing the number of components in a normal mixture* *Biometrika* 88, 767-778.
16. Celeux G., Soromenho G., (1996) *An entropy criterion for assessing the number of clusters in a mixture model* *Journal of Classification* 13, 195-212.
17. Wang M., Bodner TE., (2007) *Growth mixture modeling: Identifying and predicting unobserved subpopulations with longitudinal data*. *Organizational Research Methods*, Volumn 10 Number 4, 635-656.
18. Little T., Rubin D., (1987) *Analysis with missing data*. John Wiley & Sons: New York.