The Pennsylvania State University
The Graduate School
College of Medicine

JOINT MODELING OF A LONGITUDINAL BIOMARKER,
RECURRENT EVENTS AND A TERMINAL EVENT IN A
MATCHED STUDY

A Dissertation in
Biostatistics
by
Cong Xu

© 2017 Cong Xu

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

August 2017
The dissertation of Cong Xu was reviewed and approved* by the following:

Vernon M. Chinchilli  
Distinguished Professor of Public Health Sciences  
Chair of Department of Public Health Sciences  
Dissertation Co-Adviser  
Co-Chair of Committee

Ming Wang  
Assistant Professor of Public Health Sciences  
Dissertation Co-Adviser  
Co-Chair of Committee

Lan Kong  
Professor of Public Health Sciences

W. Brian Reeves  
Chair of Department of Medicine  
University of Texas Health Science Center at San Antonio

Arthur Berg  
Associate Professor of Public Health Sciences  
Head of Biostatistics Program

*Signatures are on file in the Graduate School.
Abstract

In longitudinal studies, matched designs are often employed to control the potential confounding effects in the field of biomedical research and public health. It is common to collect both repeated measures of risk factors (e.g., biomarkers) and time-to-event data for each subject, such as recurrent events and death. There are existing standard approaches to model the data separately. Mixed effects models are commonly used to model the association between repeated measures and covariates, which can incorporate the correlation among repeated measures. The Cox PH models or accelerated failure time (AFT) model are often used to estimate the covariate effects on the risk of the event. However, separate modeling may lead to biased results and are less efficient when the two processes are related through some unobserved variables. In many instances, the terminal event of death may prevent the observations and even the occurrence of any further recurrent events, but not vice versa. Thus, the common assumption of independent censoring for recurrent events is violated due to the competing risk of death because these two event processes are often correlated. For instance, if recurrent events (e.g., heart attacks) have a substantially negative effect on health condition, then the hazard for death could be increased. In addition, longitudinal biomarkers are often measured repeatedly over time for investigating their association with the event recurrence or death, thus identifying the candidate biomarker with enhance predictive accuracy is crucial for clinical practice.

Motivated by the the Assessment, Serial Evaluation, and Subsequent Sequelae in Acute Kidney Injury (ASSESS-AKI) study, several challenges are recognized for joint modeling: 1) A certain large portion of subjects may not have any recurrent events during the study period due to non-susceptibility to events or censoring; 2) there exists left-censoring issue for some longitudinal biomarkers due to inherent
limit of detection; 3) The correlation within matched cohorts need to be incorporated; 4) the informative censoring due to competing risk of death need to be adjusted. In this dissertation, first, we propose a joint frailty model with zero-inflated recurrent events and death in a matched study, where a matched logistic model is adopted to adjust for structural zero recurrent events. We incorporated two frailties to measure the dependency between subjects within a match pair and that among recurrent events within each individual. By sharing the random effects, two event processes of recurrent events and death are dependent with each other. Furthermore, because of left-censoring of the assay used to quantify the marker, longitudinal data could be complicated by left-censoring of some measures. Next, we propose a joint model of longitudinal biomarkers, recurrent events and death which can accommodate left-censoring biomarkers. The maximum likelihood based approach is applied for parameter estimation, where the Monte Carlo Expectation-Maximization (MCEM) algorithm is adopted and implemented in R. In addition, alternative estimation methods such as Gaussian quadrature (PROC NLMIXED) and a Bayesian approach (PROC MCMC) are also considered for comparison to show our method’s superiority. Extensive simulations are conducted and a real data application on acute ischemic studies is provided.
# Table of Contents

List of Figures viii  
List of Tables ix  
Acknowledgments xii  

## Chapter 1 Motivation Study 1  
1.1 Introduction ........................................ 1  
1.1.1 The Assessment, Serial Evaluation, and Subsequent Sequela in Acute Kidney Injury (ASSESS-AKI) Consortium ........................................ 2  
1.1.1.1 Recurrent events .................................... 3  
1.1.1.2 Deaths ............................................ 3  
1.1.1.3 Dropouts .......................................... 4  
1.1.1.4 Longitudinal Biomarkers .......................... 5  
1.2 Research hypotheses ................................... 5  
1.2.1 Joint frailty model of zero-inflated recurrent events and death 6  
1.2.2 Joint modeling of left-censored longitudinal biomarkers, recurrent events with death as informative censoring ........... 7  

## Chapter 2 Literature Review 8  
2.1 Recurrent events ....................................... 8  
2.1.1 Frailty model ......................................... 10  
2.1.1.1 Counting process .................................. 10  
2.1.2 Mixture cure model .................................. 13  
2.2 Recurrent events with a terminal event ................... 15  
2.3 Joint models for longitudinal data and survival data ............. 17
A.1 Conditional expectation ........................................... 81
A.2 Metropolis-Hastings algorithm ................................. 82
A.3 Newton-Raphson algorithm ................................. 83
A.4 Data generation ...................................................... 86
A.5 Supplementary simulation results ................................. 87

Appendix B Chapter 4 .................................................. 92
B.1 Baseline intensities .................................................. 92
  B.1.1 Recurrent events .............................................. 92
  B.1.2 Death event .................................................. 94
B.2 Variance components .................................................. 96
B.3 Newton-Raphson algorithm ........................................... 96

Appendix C Chapter 5 .................................................. 106
C.1 Baseline intensities .................................................. 106
  C.1.1 Recurrent events .............................................. 106
  C.1.2 Death event .................................................. 108
C.2 Variance components .................................................. 110
C.3 Newton-Raphson algorithm ........................................... 111

Bibliography ..................................................................... 112
List of Figures

3.1 Sample data on recurrent acute ischemic stroke 39

4.1 Sample simulated data of longitudinal biomarker, recurrent events and death 62
## List of Tables

1.1 Matched AKI and non-AKI individuals by baseline CKD status . . . 2
1.2 Matched AKI and non-AKI individuals by clinical research center . 3
1.3 Subjects with recurrent episodes of AKI by baseline status . . . . . 3
1.4 Descriptive statistics of subjects with recurrent episodes of AKI . . 3
1.5 The deaths by baseline status of CKD/AKI and clinical research center . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 4
1.6 Distribution of recurrent episodes of AKI in deaths . . . . . . . . . 4
1.7 The dropouts by baseline status of CKD/AKI and clinical research center . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 5

3.1 Simulation results of Setting I: $\beta_1 = 1, \lambda_0^R = 0.5, \lambda_0^D = 0.1$: 48% subjects are observed with at least one recurrent event. Another 22% subjects are susceptible but have not yet experienced any events during the follow-up period. The censoring rate for death is 40%. Full model denotes the proposed joint model of zero-inflated recurrent events and death event; Model I denotes the model without considering the matched pair correlation, where matched pair random effect $\mu$ is removed; Model II denotes the model without considering zero-inflation for recurrent events. . . . . . . . . . . . . 37
3.2 Simulation results of Setting II: $\beta_1 = 0.5, \lambda_0^R = 0.25, \lambda_0^D = 0.1$: 33% subjects are observed with at least one recurrent event. Another 27% subjects are susceptible but have not yet experienced any events during the follow-up period. The censoring rate for death is 40%. Full model denotes the proposed joint model of zero-inflated recurrent events and death event; Model I denotes the model without considering the matched pair correlation, where matched pair random effect $\mu$ is removed; Model II denotes the model without considering zero-inflation for recurrent events. . . . . . . . . . . . . 38
3.3 Analysis of recurrent ischemic stroke data: Full model denotes the proposed joint model of zero-inflated recurrent events and death event; Model I denotes the model without considering the matched pair correlation, where matched pair random effect $\mu$ is removed; Model II denotes the model without considering zero-inflation for recurrent events.

4.1 Simulation results of two stage model Setting I: $\eta_R = 0.5$, $\eta_D = 1$ and $\lambda_0^R = 0.1$: 60% subjects are observed with at least one recurrent event and death percentage is 40%. Full model denotes the proposed joint model of biomarker, recurrent events and death event; Model I denote the model without considering the joint modeling of biomarker. The piecewise constant approach is adopted to estimate the baseline intensities for both recurrent and death event processes, where 5 intervals are considered.

4.2 Simulation results of two stage model Setting II: $\eta_R = 0.4$, $\eta_D = 0.7$ and $\lambda_0^R = 0.065$: 50% subjects are observed with at least one recurrent event and death percentage is 25%. Full model denotes the proposed joint model of biomarker, recurrent events and death event; Model I denote the model without considering the joint modeling of biomarker. The piecewise constant approach is adopted to estimate the baseline intensities for both recurrent and death event processes, where 5 intervals are considered.

4.3 Simulation results of two stage model Setting III: the percentage of left-censoring longitudinal observations is 30%. 60% subjects are observed with at least one recurrent event and death percentage is 40%. Full model denotes the proposed joint model of left-censored biomarker, recurrent events and death event; Model I denote the model without considering the left-censoring biomarker.

5.1 Simulation results of joint model Setting I: $\sigma_c^2 = 0.01$ and $n_{ij} = 10$: 60% subjects are observed with at least one recurrent event and death percentage is 40%. Full model denotes the proposed joint model of biomarker, recurrent events and death event; Model I denote the two-stage joint model.
A.1 Simulation results of Setting I for Model II with Breslow baseline estimator: \( N = 250, \lambda_0^R = 0.5, \lambda_0^D = 0.1 \): 65% subjects are observed with at least one recurrent event. The percentage of death is 40%. Model II denotes the model without considering zero-inflation for recurrent events, which is the full model here; Model III denotes the model without considering the matched pair correlation based on Model II, where matched pair random effect \( \mu \) is removed; . . .

A.2 Simulation results of Setting II for Model II with Breslow baseline estimator: \( N = 500, \lambda_0^R = 0.5, \lambda_0^D = 0.1 \): 65% subjects are observed with at least one recurrent event. The percentage of death is 40%. Model II denotes the model without considering zero-inflation for recurrent events, which is the full model here; Model III denotes the model without considering the matched pair correlation based on Model II, where matched pair random effect \( \mu \) is removed; . . .
Acknowledgments

I would gratefully and sincerely like to thank my advisors, Dr. Vernon M. Chin-chilli and Dr. Ming Wang for their guidance, encouragement, support and deep consideration throughout my dissertation work. I have been greatly impressed by their inspiration and immense knowledge, which help me move forward. I have learned from them not only in academic research, but also in future career development. I extremely appreciate their time and patience with me. They always read and responded to the drafts of each page of my work more quickly than I could have hoped.

I am also gratefully to thank my committee members, Dr. W. Brian Reeves and Dr. Lan Kong for sharing their insightful comments and valuable suggestions on this dissertation. I shall thank Dr. Douglas Leslie, Dr. Guodong Liu and M.S. Djibril Ba for data preparation and permission to use the data from the Marketscan research database.

Moreover, I would like to thank Dr. Arthur Berg, Dr. Rongling Wu and Dr. Jay Zhu who as my supervisors guided my RA consulting work and cared deeply about me. Also, I wish to thank the department of public health sciences, every professor I took courses from. And, thank Mardi for all help and convenience she provided during my Ph.D. study.

Special thanks go to my friends and classmates for their support and encouragement. Last, I can never find words to express my gratitude to my family who are always standing right next to me.
Dedication

To my family
Motivation Study

1.1 Introduction

Acute kidney injury (AKI) is a sudden decrease in kidney function and is associated with a high risk of adverse outcomes including mortality, development of chronic kidney disease (CKD) and other organ dysfunction (Shusterman et al., 1987; Liangos et al., 2006). This under-recognized disorder is an imperative clinical and public health problem with a rising incidence and mortality in the U.S. even with modern medical technology (Hou et al., 1983). It has been known that AKI is associated with a high risk of death in severe cases (> 30%) (Go et al., 2010). However, the effect of AKI on “long-term” outcome has been listed as a “critically important knowledge gap” by American Society of Nephrology Renal Research Report (Berl, 2005). It is clinically important to investigate the recurrence of AKI, and explore its relationship with mortality and other clinical outcomes.
1.1.1 The Assessment, Serial Evaluation, and Subsequent Sequelae in Acute Kidney Injury (ASSESS-AKI) Consortium

Our research is motivated by the Assessment, Serial Evaluation, and Subsequent Sequelae of Acute Kidney Injury (ASSESS-AKI) Consortium (Go et al., 2010). This ASSESS-AKI study is funded by the National Institute of Diabetes, Digestive and Kidney Disease (NIDDK) with a multi-center collaboration network including four clinical research centers (CRCs). These four CRCs follow the same inclusion and exclusion criteria recruited a diverse prospective parallel, matched cohort involving children and adults (aged 18 to 89 years old) with and without AKI based on serum creatinine-based diagnostic criteria. AKI is defined as 50% relative increase and/or absolute increase of 0.3 mg/dL (26 mol/L) in peak inpatient serum creatinine compared with baseline outpatient serum creatinine.

In this multi-center matched cohort study design, the AKI and non-AKI individuals are matched with the ratio of a minimum 1:1 on major baseline confounders based on a prioritized set of criteria (i.e., Clinical cardiovascular disease, diabetes mellitus, category of baseline eGFR, adult age category and hospital location where AKI episode occurred and so on) within each stratum according to baseline CKD status (using CKD-EPI estimated glomerular filtration rate (GFR) threshold of $< 60$ or $\geq 60 \text{ml/min/1.73m}^2$) and CRCs (Go et al., 2010). The recruitment has been completed with the total of 1,603 adults with up to 84-months maximal follow-up, where 1,538 subjects are 1:1 matched on AKI status (Tables 1.1 and 1.2), and also 618 (38.55%) have CKD and 985 (61.45%) don’t have CKD. Of note, currently the follow-up is still ongoing.

Table 1.1. Matched AKI and non-AKI individuals by baseline CKD status

<table>
<thead>
<tr>
<th></th>
<th>CKD</th>
<th>Non-CKD</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>AKI</td>
<td>306</td>
<td>463</td>
<td>769</td>
</tr>
<tr>
<td>Non-AKI</td>
<td>306</td>
<td>463</td>
<td>769</td>
</tr>
<tr>
<td>Total</td>
<td>612</td>
<td>926</td>
<td>1538</td>
</tr>
</tbody>
</table>
Table 1.2. Matched AKI and non-AKI individuals by clinical research center

<table>
<thead>
<tr>
<th></th>
<th>Center 1</th>
<th>Center 2</th>
<th>Center 3</th>
<th>Center 4</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>AKI</td>
<td>156</td>
<td>208</td>
<td>251</td>
<td>154</td>
<td>769</td>
</tr>
<tr>
<td>Non-AKI</td>
<td>156</td>
<td>208</td>
<td>251</td>
<td>154</td>
<td>769</td>
</tr>
<tr>
<td>Total</td>
<td>312</td>
<td>416</td>
<td>502</td>
<td>308</td>
<td>1538</td>
</tr>
</tbody>
</table>

1.1.1.1 Recurrent events

Based on updated database by Oct 2015, 1,192 participants (77.50%) don’t experience any recurrent episodes of AKI. From Table 1.3, we can see that among subjects who have recurrent episodes of AKI, participants who have both AKI and CKD account for the largest percentage (32.95%) and individuals who don’t suffer from both AKI and CKD at baseline have the least percentage (16.18%). It turns out that the stratum of baseline AKI and CKD has the highest risk for recurrent episodes of AKI. From Table 1.4, participants who are in the AKI an CKD stratum at baseline have the largest average number of recurrent episodes of AKI which is 1.54. The maximum number of recurrent episodes of AKI is 10, which is much higher than the number in other strata. Accordingly, the non-AKI and non-CKD group has the smallest average number of recurrent episodes of AKI is 1.25.

Table 1.3. Subjects with recurrent episodes of AKI by baseline status

<table>
<thead>
<tr>
<th></th>
<th>Non-AKI/Non-CKD</th>
<th>Non-AKI/CKD</th>
<th>AKI/Non-CKD</th>
<th>AKI/CKD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recurrent AKI</td>
<td>56 (16.18%)</td>
<td>78 (22.54%)</td>
<td>98 (28.32%)</td>
<td>114 (32.95%)</td>
</tr>
<tr>
<td>No Recurrent AKI</td>
<td>407 (34.14%)</td>
<td>228 (19.13%)</td>
<td>365 (30.62%)</td>
<td>192 (16.11%)</td>
</tr>
<tr>
<td>Total</td>
<td>463</td>
<td>306</td>
<td>463</td>
<td>306</td>
</tr>
</tbody>
</table>

Table 1.4. Descriptive statistics of subjects with recurrent episodes of AKI

<table>
<thead>
<tr>
<th></th>
<th>Non-AKI/Non-CKD</th>
<th>Non-AKI/CKD</th>
<th>AKI/Non-CKD</th>
<th>AKI/CKD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average episodes</td>
<td>1.25</td>
<td>1.38</td>
<td>1.28</td>
<td>1.54</td>
</tr>
<tr>
<td>Range</td>
<td>1-3</td>
<td>1-3</td>
<td>1-3</td>
<td>1-10</td>
</tr>
</tbody>
</table>

1.1.1.2 Deaths

Deaths are identified primarily through surveys of subjects or their proxy contacts and review of medical records or death certificates, if available (Go et al., 2010). Until Oct 2015, there are 159 (9.92%) participants who died. From Table 1.5, we can see that participants who have both AKI and CKD at baseline have a higher
risk for death. Among all the dead subjects, 78 (49.06%) individuals experienced recurrent episodes of AKI, which is much higher than the percentage based on all the participants 22.5%. For dead participants who have both CKD and AKI at baseline, 50% (Table 1.6) of them experienced recurrent episodes of AKI and the percentage is much higher than that of other strata. There seems to be an obvious association between recurrent episodes of AKI and death.

Table 1.5. The deaths by baseline status of CKD/AKI and clinical research center

<table>
<thead>
<tr>
<th></th>
<th>Center 1</th>
<th>Center 2</th>
<th>Center 3</th>
<th>Center 4</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>AKI/CKD</td>
<td>19</td>
<td>12</td>
<td>25</td>
<td>11</td>
<td>67 (42.13%)</td>
</tr>
<tr>
<td>AKI/Non-CKD</td>
<td>6</td>
<td>17</td>
<td>12</td>
<td>7</td>
<td>42 (26.42%)</td>
</tr>
<tr>
<td>Non-AKI/CKD</td>
<td>7</td>
<td>4</td>
<td>7</td>
<td>5</td>
<td>23 (14.47%)</td>
</tr>
<tr>
<td>Non-AKI/Non-CKD</td>
<td>8</td>
<td>9</td>
<td>5</td>
<td>5</td>
<td>27 (16.98%)</td>
</tr>
<tr>
<td>Total</td>
<td>40</td>
<td>42</td>
<td>49</td>
<td>28</td>
<td>159</td>
</tr>
</tbody>
</table>

Table 1.6. Distribution of recurrent episodes of AKI in deaths

<table>
<thead>
<tr>
<th></th>
<th>Recurrent AKI</th>
<th>No recurrent AKI</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>AKI/CKD</td>
<td>39 (50%)</td>
<td>28 (34.57%)</td>
<td>67</td>
</tr>
<tr>
<td>AKI/Non-CKD</td>
<td>16 (20.51%)</td>
<td>26 (32.10%)</td>
<td>42</td>
</tr>
<tr>
<td>Non-AKI/CKD</td>
<td>16 (20.51%)</td>
<td>7 (9.64%)</td>
<td>23</td>
</tr>
<tr>
<td>Non-AKI/Non-CKD</td>
<td>7 (8.97%)</td>
<td>20 (24.69%)</td>
<td>27</td>
</tr>
<tr>
<td>Total</td>
<td>78 (49.06%)</td>
<td>81 (50.94%)</td>
<td>159</td>
</tr>
</tbody>
</table>

1.1.1.3 Dropouts

According to the protocol of ASSESS-AKI, a participant is considered as a dropout if the subject is not dead and the date of censoring time is more than 15 months from today. By the end of Oct 2015, there are 179 (11.17%) individuals who dropped out from the study. Among these dropouts, 145 individuals (85.29%) did not experience any recurrent AKI events before they dropped out. For the remaining 28 subjects, the range of recurrent episodes of AKI is 1-3 and the average number of recurrent episodes of AKI is 1.24. From Table 1.7, we can see that there is no specific pattern among CKD/AKI strata and clinical research center. Thus, it is reasonable to assume the dropouts arise from non-informative censoring.
Table 1.7. The dropouts by baseline status of CKD/AKI and clinical research center

<table>
<thead>
<tr>
<th></th>
<th>Center 1</th>
<th>Center 2</th>
<th>Center 3</th>
<th>Center 4</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>AKI and CKD</td>
<td>6</td>
<td>3</td>
<td>17</td>
<td>4</td>
<td>30 (16.76%)</td>
</tr>
<tr>
<td>AKI and Non-CKD</td>
<td>11</td>
<td>18</td>
<td>21</td>
<td>8</td>
<td>58 (32.40%)</td>
</tr>
<tr>
<td>Non-AKI and CKD</td>
<td>3</td>
<td>2</td>
<td>16</td>
<td>8</td>
<td>29 (16.20%)</td>
</tr>
<tr>
<td>Non-AKI and Non-CKD</td>
<td>11</td>
<td>12</td>
<td>24</td>
<td>15</td>
<td>62 (34.64%)</td>
</tr>
<tr>
<td>Sum</td>
<td>31</td>
<td>35</td>
<td>78</td>
<td>35</td>
<td>179</td>
</tr>
</tbody>
</table>

1.1.1.4 Longitudinal Biomarkers

One of the major goals of ASSESS-AKI is to evaluate the utility of urine and blood biomarkers for improving the diagnosis and risk stratification after a hospitalized episode of AKI (Go et al., 2010). A set of urine biomarkers (IL-18, NGAL, KIM-1, cystatin C, L-FABP and NAG) and blood biomarkers (serum cystatin C, serum NGAL and plasma IL-6) are collected at each visit. We can see that some biomarkers are left censored (i.e., urine albumin with 0.02 (mg/L) as the lower limit and urine protein with 0.06 (g/L) as the lower limit). Accordingly, the urine ratios (i.e., urine albumin-to-creatinine ratio (ACR) and urine protein-creatinine ratio (PCR) are calculated based on the lower limits if the biomarkers are left censored. Of note is that the ratios are not regularly censored data rigorously. There are missing values in longitudinal biomarkers during the follow-up, and there exist different patterns. Given the current data supporting the use of these markers for the detection of AKI, it will be important to know whether these markers predict short or long term outcomes.

1.2 Research hypotheses

Based on the preliminary analysis of the ASSESS-AKI data, several research hypotheses are framed. First, we observe that it seems that a higher intensity of recurrent episodes of AKI is associated with a higher death rate. Second, several biomarkers have positive or negative effects on the recurrence of AKI and death. Therefore, we would like to investigate the dependency among recurrence of AKI, death and further explore the prediction of longitudinal biomarkers on the recur-
rence of AKI and death. Also, investigate the effects of important covariates, such as baseline AKI, demographic, clinical variables etc., on recurrence of AKI with death as a competing risk. Therefore, joint modeling of biomarkers, recurrent AKI and death is considered.

1.2.1 Joint frailty model of zero-inflated recurrent events and death

We first consider joint modeling of recurrent events along with death that discontinues further observations. Recurrent events occur frequently during the follow-up period in longitudinal clinical studies. For example, patients may experience relapsed tumors, recurrent strokes or hypoglycemia events on multiple occasions. For recurrent events analysis, different timescales can be used (Kelly and Lim, 2000). Here, we consider the total timescale which is one of the most often used scales. With total time, the time to any recurrent event is measured from time-origin which could be a fixed calendar time, onset of treatment, or a biological event. In many instances, the terminal event of death is commonly encountered during follow-up, which prevents the observations and even the occurrence of any further recurrent events, but not vice versa (Fine et al., 2001). Thus, the common assumption of independent censoring for recurrent events is violated due to the competing risk of death because these two event processes are often correlated. For instance, if recurrent events (e.g., heart attacks) have a substantially negative effect on health condition, then the hazard for death could be diminished. The association between recurrent events and death has attracted increasing interest recently, and a joint analysis taking their correlation into account is needed for valid inference (Lancaster and Intrator, 1998). In some scenarios, a certain large portion of subjects may not have any recurrent events during the study period. These patients are believed to being cured by treatment or non-susceptibility to events or censoring. Thus, the zero-inflated nature of data should be considered in analysis. To model such a complicated system, we propose a joint frailty model with recurrent events and death is proposed to adjust for zero-inflation and matched designs. We incorporate two frailties to measure the dependency between subjects within a matched
pair and that among recurrent events within each individual. By sharing the random effects, two event processes of recurrent events and death are dependent with each other.

1.2.2 Joint modeling of left-censored longitudinal biomarkers, recurrent events with death as informative censoring

Next, we focus on the joint analysis of left-censored longitudinal biomarker, recurrent events along with death as informative censoring. In many longitudinal clinical studies, longitudinal biomarkers are often repeated measured over time for investigating their association with the outcome, where there exists left-censoring issue due to the inherent limit of detection (e.g., HIV-1 RNA level, tumor size). Due to clinical interest, the outcome of recurrent time-to-event data are captured during the follow-up. Meanwhile, the terminal event of death is always encountered, which should be taken into account for valid inference because of informative censoring. Therefore, longitudinal data, recurrent events and death are often associated in some ways. For example, the time to event may be associated with the longitudinal trajectories. Separate analyses of the three processes may lead to inefficient or biased results (Tsiatis and Davidian, 2004; Liu et al., 2008; Wu et al., 2011). Thus, a joint model for a left-censored longitudinal biomarker, recurrent events and death is proposed for a matched study. In particular, a linear mixed effects model is used to model longitudinal data and a joint frailty model is used for recurrent events and death. We adopted two frameworks for the joint model: two-stage and joint likelihood frameworks. We incorporate random effect to account for the correlation within longitudinal biomarker measures as well as two frailties to measure the dependency between subjects within a matched pair and that among recurrent events within each individual. By sharing the random effects, death may be dependent on repeated biomarkers and recurrent event history.
Chapter 2

Literature Review

In this chapter, we review literature on statistical methods for recurrent events in the presence of a terminal event in Section 2.1 and 2.2, for joint modeling of longitudinal outcome, recurrent events and a terminal event in Section 2.3.

2.1 Recurrent events

Cox’s proportional hazard (Cox PH) models are commonly used as linear models with uncensored data in survival analysis (Cox, 1992). However, in some situations, Cox PH model may not be applied directly. For example, for recurrent events, some unobserved factors will cause the heterogeneity among subject which violates the assumption of Cox PH model that univariate failure times are independent of each other given all covariates. Recurrent events or repeated events occur frequently during the follow-up in longitudinal clinical studies (Liu et al., 2004; Kim et al., 2012; Belot et al., 2014). For example, patients may experience relapsed tumors or recurrent hospitalizations. More generally, recurrent events can be treated as a special case of clustered survival data, which are called multivariate survival data. Clustered survival data are quite common in clinical studies. For another example, in randomized clinical trial, when subjects are recruited in at multiple clinical centers, factors that may vary by center, including patient characteristics
and medical practice patterns, may exert a powerful influence on study outcomes (Glidden and Vittinghoff, 2004). In both examples, such clustered effects (e.g. recurrent events, center effects) results in dependencies of failure times among subjects that violates the independence assumption of the Cox PH model. If clustered effects are sufficiently powerful, the inference will be invalid if the effects are ignored.

Many researchers have extended the Cox PH model to handle the dependence in clustered survival data, mainly including marginal (or population-averaged) or conditional (or cluster-specific) models. Marginal models have been studied in different settings by Lee et al. (1992), Liang et al. (1993), Lin (1994), Spiekerman and Lin (1998), Huang and Chen (2003) and Cong et al. (2007). Marginal models ignore the cluster effects by treating the survival times as independent when estimating the regression model. Moreover, copula models could be considered due to their convenience by specifying the joint survival function in terms of the marginal function for each cluster (Nelsen, 1997; Roy and Mukherjee, 1998; Pipper and Martinussen, 2003; Andersen et al., 2005). But copula models are not appropriate for data with large or imbalanced cluster size. For conditional model, the popular approach are stratified models and frailty models. Stratified Cox model assume different and unspecified baseline hazard for each cluster, which makes the stratified models the most general of the conditional models (Holt and Prentice, 1974; Kalbfleisch and Prentice, 2011). When the clustering is of no intrinsic interest, the stratified Cox model is an appealing tool by applying the partial likelihood approach. However, this model requires discarding a considerable amount of information and the loss of information increases as the number of clusters increases, which leads to less efficient estimation. Frailty models, which are random effects models, have been popularly employed as a way of accommodating the heterogeneity among the clusters and the dependence among subjects within a cluster. Frailty models provides a flexible, efficient frame working in a variety of practical settings, which is our focus in this dissertation. Next, in section 2.1.1, we review frailty models in the analysis of clustered survival data.
2.1.1 Frailty model

A frailty model is a multiplicative hazard model consisting of three components: a frailty (random effect), a baseline hazard function (parametric or nonparametric), and a term modeling the influence of observed covariates (fixed effects) (Wienke, 2010).

The term “frailty” was first introduced by (Vaupel et al., 1979) for the analysis of time to death. Nielsen et al. (1992) proposed a frailty model for recurrent events by incorporating the counting process (Andersen and Gill, 1982). Pénichoux et al. (2015) compared the performance of conditional and mixture likelihood approaches estimating models with these frailty effects in censored bivariate survival data. They found that mixture methods were robust to misspecification of the frailty distribution, which can be a log-gamma or normal distribution. There are two fundamental ways to model recurrent events, one is using event times (gap time, calendar time or cumulative time), the other one is using event counts based on the counting process. If the trend of the intensity function is of sole interest, then the latter one would be preferred. Kelly and Lim (2000) first systematically identified the components of model PWP-GT (Prentice, Williams and Peterson, gap time) and TT-R (“total time”-restricted) that are appropriate for recurrent event data, among the other extended survival models. In this paper, we adopted the Poisson process which is one of the most widely-used counting processes to model recurrent events in the frailty model settings. Here, we give a brief introduction of the counting process, which is the fundamental modeling background in our research.

2.1.1.1 Counting process

This approach was first developed by Aalen (1978) who combined the elements of stochastic integration, continuous time martingale theory and counting process theory into a methodology allowing for development of inference techniques for survival quantities based on censored or truncated data. Andersen and Gill (1982) extended Cox PH models to a multivariate counting process for analyzing the time to recurrent events.
Suppose $0 = t_0 < t_1 < t_2 < \cdots < t_n = \tau \leq \infty$, during study period $[0, \tau]$, define a counting process $\{N(t) : 0 \leq t \leq \tau\}$. As a stochastic process with the properties that $N(0) = 0$; $N(t) < \infty$ with probability one. For example, let a right-censored sample $T = \min(X, C)$, where $X$ is the death time and $C$ is the right-censoring time. The process $N_i(t) = I[T_i \leq t, \delta_i = 1]$ is zero until individual $i$ dies and then jumps to one is counting process, where $\delta_i$ indicates whether the lifetime $X$ is observed ($\delta_i = 1$), if $X \leq C$, or not ($\delta_i = 0$), if $X > C$.

The process $N(t) = \sum_{i=1}^{n} N_i(t) = \sum_{i=1}^{n} I[T_i \leq t, \delta_i = 1]$ counts the number of deaths in the samples at or prior to time $t$ is also a counting process. The process $Y_i(t) = I(T_i \leq t)$ is the referred to the at risk process, indicating whether the subject is at risk at time $t$. Define the history or filtration $F_t$ of the counting process at time $t$ as the accumulated knowledge about what has happened to patients up to time $t$.

For the counting process, define $dN(t)$ to be the change in the process $N(t)$ over a short time interval $[t, t + dt)$, which is $dN(t) = N[(t + dt)^-] - N[(t^-)]$ (Here $t^-$ is the time just prior to $t$). Next, if we define $Y(t) = \sum_{i=1}^{n} Y_i(t)$ as the number of subjects with a study time $T_i \geq t$, then

$$E[dN(t) | F_t^-] = E[\# \text{ of observations with } t \leq X_i < t + dt, C_i \geq t + dt | F_t^-] = Y(t) \lambda_0(t) dt$$

The process $\lambda(t) = Y(t) \lambda_0(t)$ is the intensity process of the counting process, where $\lambda_0(t)$ is a baseline hazard function for death. $\lambda(t)$ depends on the information contained in the history process $F_t$ though $Y(t)$. $\Lambda(t) = \int_0^t \lambda(s) ds$ is the cumulative intensity process.

Thus, the probability density of an outcome of $n$ independent events, $(t_1, t_2, \cdots, t_n)$ is

$$\prod_{i=1}^{n} \lambda(t_i | F_t^-) \exp[-\int_0^{t} \lambda(t | F_t^-) dt]$$

A counting process $\{N(t) : 0 \leq t < \infty\}$ is a Poisson counting process with the rate $\lambda(t) > 0$ if it has the following three properties:
1. \( N(0) = 0; \)

2. \( N(t) \) has independent increments;

3. The number of events in any interval of length \( t > 0 \) has \( \text{Poisson}(\lambda(t)t) \) distribution.

Let

\[
\lambda(t \mid H(t)) = \lim_{\Delta t \to 0} \frac{Pr(\Delta N(t) = 1 \mid H(t))}{\Delta t} = \lim_{\Delta t \to 0} \frac{Pr(\Delta N(t) = 1)}{\Delta t} = p(t),
\]

which is independent of event history. The mean cumulative function: \( \mu(t) = E(N(t)) = \int_0^t p(s)ds. \)

The probability density of \( n \) independent events \((t_1, t_2, \ldots, t_n)\) is

\[
\prod_{j=1}^n p(t_j) \exp(-\int_0^t p(s)ds) = \prod_{j=1}^n p(t_j) \exp(-\mu(t))
\]

Andersen and Gill (1982) extended Cox PH models by introducing the Poisson process with covariates to model recurrent events:

\[
p(t \mid x) = p_0(t) \exp(x^T \beta)
\]

Nielsen et al. (1992) and Andersen et al. (1997) presented shared frailty modes in a counting process framework with statistical inference. \( x_i \) is the covariate, and \( \gamma_i \) is a frailty term to incorporate the correlation among the recurrent events and within subjects:

\[
p(t \mid x_i) = p_0(t) \gamma_i \exp(x_i^T \beta)
\]

Let \( \mu_i = \log(\gamma_i) \), then the model can be rewritten as

\[
p(t \mid x_i) = p_0(t) \exp(x_i^T \beta + \mu_i)
\]

The process is the Poisson process, but the process is not marginal by a Poisson
process. Thus, the point density function for the $i^{th}$ process is,

$$f(t_1, t_2, \cdots, t_n, \gamma_i | x_i) = \prod_{i=1}^{n} p_0(t) \gamma_i \exp(x_i^T \beta) \exp[-\gamma_i \mu_i(\tau_i)] g(\gamma_i | \theta)$$

The gamma distribution is most commonly used since the frailty terms in the conditional likelihood can be integrated out and therefore the full likelihood has a closed-form expression. Hou et al. (1983) first introduced the positive stable distribution in multivariate survival analysis, which has a nice property that the proportionality can be inherited from the conditional hazard to the marginal or population hazard, such as Normal and Weibull distributions. We study the normal distributed frailties model in this dissertation.

### 2.1.2 Mixture cure model

In the ASSESS-AKI study, there are a large proportion of subjects who are not susceptible to recurrent AKI events which manifest the “zero-inflated” nature of the data. Zero inflated models have been proposed for many types of data, such as continuous data (Olsen and Schafer, 2001; Tooze et al., 2002; Liu et al., 2016) and count data (Lambert, 1992; Min and Agresti, 2005; Garay et al., 2011; Lee et al., 2011). Lambert (1992) first proposed the zero-inflated Poisson model, which concerned a random event containing excess zero-count data in unit time. In these models, a proportional of the data have an outcome $Y = 0$ with probability $p$. Therefore, zero-inflated models employ two components that correspond to two zero generating processes. The first process is a binary distribution that generates structural zeros with probability $p$. Another process is governed by a specific distribution, e.g., Gaussian, Poisson, or negative binomial that generates the outcomes. When the outcomes are count data, such as the number of heart attacks, excessive zeros can be categorized as structural zeros and random zeros. Structural zeros are usually produced because a certain population is not at risk. Zeros are called random (or sampling) zeros when those subjects who are at risk may still produce a zero-outcome due to sampling variability. Poisson or negative binomial models
have the capability to accommodate the random zeros, which can be classified as a special cases of latent class models.

Similar to zero-inflated Poisson or negative binomial model for count data, we may observe zero events (e.g. death, recurrent events) for some subjects in some longitudinal clinical trials. The subjects who are “cured” could not experience any episode of the disease by the end of follow-up. In this situation, if a significant number of patients are “cured” and thus risk free of recurrences, the population is then a mixture of susceptible and non-susceptible subjects (Rondeau et al., 2013). Therefore, estimation from standard Cox PH models model would not always be appropriate, because they assume that all subjects will experience the disease over a sufficient period of follow-up. The estimation of patients who are cured and the failure time distribution of uncured patients are very important. Farewell (1982) used a logistic regression for the mixture proportion and Weibull regression for the latency in a toxicological animal experiment. Kuk and Chen (1992) introduced a semi-parametric generalization of a mixture cure model using a Cox PH model in the susceptible group. Sy and Taylor (2000) developed a maximum likelihood technique for the joint estimation of the incidence and latency regression parameters in this model using the nonparametric form of the likelihood and an expectation-maximization (EM) algorithm. Rondeau et al. (2013) extended the PH cure model to recurrent events in the frailty model framework. Liu et al. (2015) proposed two joint frailty models for zero-inflated recurrent events in the presence of a terminal event. Peng and Taylor (2011) derived a Monte Carlo Expectation-Maximization (MCEM) algorithm to apply a cure PH model for clustered survival data with random effects.

Next, we briefly introduced the framework of the mixture cure model. Given the covariate $x_i$, a mixture cure model is the mixture of a certain proportion $\pi(x_i)$ belonging to the cured subpopulation and the remaining fraction $1 - \pi(x_i)$ for not cured subpopulation,

$$S_{pop}(t \mid x_i) = \pi(x_i) + [1 - \pi(x_i)]S(t)$$

$S(t)$ is the conditional survival function for the uncured population (susceptible). It
is assumed that all patients in the non-cured population will eventually experience the event. The nonsusceptible subpopulation who are cured will never experience the event. For the estimation of the cure fraction $\pi(x_i)$, the logistic regression has been popular adopted:

$$\pi(x_i) = \frac{\exp(\beta x_i)}{1 + \exp(\beta x_i)}$$

2.2 Recurrent events with a terminal event

In the previous section, we have reviewed the modeling approaches for recurrent events. In many instances, the terminal event of death is commonly encountered during follow-up, which prevents the observations and even the occurrence of any further recurrent events, but not vice versa (Fine et al., 2001). Thus, the common assumption of independent censoring for recurrent events is violated due to the competing risk of death because these two event processes are often correlated. For instance, if recurrent events (e.g., heart attacks) have a substantially negative effect on health condition, then the hazard for death could be increased. The association between recurrent events and death has attracted increasing interest recently, and a joint analysis taking their dependency into account is needed for valid inference (Lancaster and Intrator, 1998).

With regards to the joint analysis of recurrent events and death, two general approaches can be adopted to accommodate the dependent censoring, namely, marginal models and frailty models. Marginal models regard the terminating event of death as a censoring event for each recurrent event, or estimate the marginal mean of the cumulative number of recurrent events over time (Cook and Lawless, 1997; Ghosh and Lin, 2002; Chen and Cook, 2004). However, the dependence between recurrent events and death cannot be specified by a marginal model (Liu et al., 2004). Moreover, copula models could be considered due to their convenience, but undesirable issues still exist, such as strong parametric assumptions on the association structure, and mis-specification of the copula may lead to unreliable estimates (Li and Cheng, 2016). Frailty models or shared random effects models, not only capture the correlation among recurrent events but also incorporate the
dependence between two event processes of recurrent events and death. Lancaster and Intrator (1998) initially considered a correlation between recurrent events and death via a person-specific frailty term by proposing a joint parametric modeling of the repeated inpatient episodes (via Poisson process) and survival time in a HIV study. Wang et al. (2001) assumed a subject-specific nonstationary Poisson process via a latent variable. However, the proposed model is not directly applicable to situations where inferences for both the recurrent and terminal events are of interest. They modeled the occurrence rate function for recurrent events with informative censoring in semiparametric and nonparametric ways. Huang and Wolfe (2002) proposed a joint frailty model for clustered data with informative censoring, where the risk to be censored was affected by the risk of failure by sharing the common log normal frailty. Liu et al. (2004) proposed a joint model by incorporating a shared gamma frailty that was included in both intensity functions of recurrent events and death to account for their dependency. Without parametric assumptions on the latent variables and censoring time, Box-Steffensmeier and De Boef (2006) developed a conditional frailty model accounting for three common conditions: heterogeneity across individuals, dependence across the number of events, and both heterogeneity and event dependence by using a frailty term. In these approaches, the estimation was carried out through a MCEM algorithm with Metropolis-Hastings sampler in the E-step. Zeng and Lin (2009) generalized joint frailty models of recurrent events and death using a broad class of transformation models, rather than Cox Proportional Hazards (PH) models. Recently, Zeng et al. (2014) proposed a joint frailty model that accommodated multiple recurrent events and multivariate informative censoring. They also developed a nonparametric EM algorithm without increasing the computational burden. Other than the nonparametric EM algorithm, Rondeau et al. (2007) and Belot et al. (2014) adopted penalized likelihood estimation to jointly estimate recurrence and death processes within a parametric framework.
2.3 Joint models for longitudinal data and survival data

In many biomedical studies, it is common to collect both repeated measures of risk factors (e.g., biomarkers) and time-to-event data for each subject. There are existing standard approaches to model the data separately. Mixed effects model (Laird and Ware, 1982) are commonly used to model the association between repeated measures and covariates, which can incorporate the correlation among repeated measures. The Cox PH models or accelerated failure time (AFT) model are often used to estimate the covariate effects on the risk of the event. However, separate modeling may lead to biased results and are less efficient when the two processes are related through some unobserved variables (Wu et al., 2011). Usually, the objective is to estimate the hazard of the events (e.g., death, cardiovascular disease) and the impact of prognostic biomarkers on the hazard of the event. Therefore, joint modeling of longitudinal and survival data is of interest to many researchers to study the dependent dropout in the repeated measure settings.

Commonly, joint models include survival models with longitudinal measurement errors and missing data as time-dependent covariates which is called two-stage (or two-step) method, or jointly model longitudinal models with informative dropouts, which is called joint likelihood method. Self and Pawitan (1992) proposed a two-stage model by imputing the fitted values of the longitudinal trajectories as a time-dependent covariate in the survival model and inference was based on usual partial likelihood. Conditional on the covariate history up to current time $t$, Tsiatis et al. (1995) replaced the true longitudinal trajectories by an empirical Bayes estimate of the conditional expectation of the covariate. However, the major concerns existed in the two-stage model is that the uncertainty of the estimation in the first stage is not incorporated in the second stage of the survival estimation, which result in the underestimation of the standard error. Wu et al. (2011) indicated that the bias resulted from the naive two-stage method was caused by the fact that the covariate trajectory was related to the length of follow-up. Thus, Albert and Shih (2010) proposed to recapture the missing measurements for those dropouts that were caused by death. They simulated missing data from the conditional
distribution of the covariate given the event time, which was similar to multiple imputation method with non-ignorable missing data.

Joint likelihood methods that simultaneously model the longitudinal data and survival data are the most widely used approach in the literature. Wulfsohn and Tsiatis (1997) jointly model longitudinal and survival data by sharing the random effects. Ratcliffe et al. (2004) developed a joint model for longitudinal and survival data which incorporated both subject and cluster-level random effects (frailties) with subjects nested within clusters. Sometimes, the longitudinal data, recurrent events and death are correlated. To jointly model longitudinal data, recurrent events and death, Liu et al. (2008) proposed a joint random effects model that includes three submodels: (a) a frailty model for the intensity of recurrent hospital admission times, (b) a random effects model for markers taken at recurrent visits, and (c) a proportional hazard model for death. They jointly modeled the cost accrual process, described by the time to recurrent hospital visits and the medical cost for each visit, in the presence of potential death. Motivated by a HIV study, Liu and Huang (2009) joint modeled the CD4 cell counts, the intensity of opportunistic disease and the hazard of death. Gaussian quadrature technique is adopted for the estimation with piecewise baseline hazard intensities. Kim et al. (2012) analyzed longitudinal data with recurrent events and a terminal event by a semiparametric joint model using general transformation models for both the recurrent events and the terminal event. Król et al. (2016) propose a joint model for a simultaneous analysis of three types of data: a left-censored longitudinal marker, recurrent events, and a terminal event. This model allowed to determine in a randomized clinical trial on which particular component treatment acts mostly.
Chapter 3

Joint model of zero-inflated recurrent events and death

3.1 Introduction

In this chapter, our motivation example explores recurrent acute ischemic stroke. Based on a 2016 report on heart disease and stroke statistics from the American Heart Association the estimated direct and indirect cost of stroke was $33 billion in the United States from 2011 to 2012 (Mozaffarian et al., 2016). Also stroke is the third leading cause of death in the United States. The literature has shown that recurrent stroke can be one of the major causes of morbidity and mortality among stroke victims, and provided valuable information about the risk of stroke recurrence in stroke survivors and the associated effects of comorbidities (Dhamoon et al., 2006; Azarpazhooh et al., 2008). For instance, based on a study with 10,399 patients with a primary diagnosis of stroke, discharged from the hospital during the year 2002 in South Carolina, the risk of recurrent stroke or death is much higher in the first year with an 8.0% risk of stroke recurrence and 24.5% risk of all-cause death at the end of 1 year; then the cumulative risk increases at a relatively
steady rate each year. By the end of four years after a stroke, the risk of another stroke was 18% and the risk of death was 41% (Feng et al., 2010). However, most studies adopted the basic Cox PH model to analyze the stroke recurrence without considering the competing risk of death or a matched design because the majority are randomized trials or epidemiological studies. Our motivation data is a matched cohort of patients with and without acute ischemic stroke at baseline from MarketScan research database year 2011-2014 (TruvenHealth, 1990). The ICD-9 code is utilized to identify the episode of acute ischemic stroke, and the 1:1 exact matching strategy is performed based on several important confounding variables, such as age, gender, admission date and the follow-up window. Our objective is to investigate the effect of initial stroke detected at baseline and comorbidity (e.g., diabetes mellitus, hypertension and heart disease) on the risk of recurrent stroke occurrence and death during index hospitalization, and also explore the dependency between recurrent stroke events and death.

With regards to this joint analysis, two main issues require attention, namely, zero-inflated recurrent events and the matched design. In our study, recurrent stroke events are observed in 597 out of 2,122 patients, thus a substantial portion of patients (71.8%) have no recurrent stroke events during the follow-up period due to non-susceptibility to events or censoring; thus, the zero-inflated nature of data should be considered in analysis. Zero-inflated Poisson or negative binomial models have been proposed and popularly used for count data to handle the over-dispersion (DeSantis et al., 2014; Liu et al., 2015). With regards to zero-inflated recurrent events data, the models have been extended from the cure models for single-event data, where the population consists of a mixture of non-cured (susceptible to the event) and cured subjects (non-susceptible to the event). For estimation of the cure rate and survival distribution of the non-cured subjects, a logistic regression for mixture proportions has been widely considered (Farewell, 1982; Kuk and Chen, 1992; Sy and Taylor, 2000; Peng and Taylor, 2011). Recently, Rondeau et al. (2013) and Liu et al. (2015) extended the Cox PH cure models to recurrent events by incorporating the frailties. The other issue is the matched design, where the correlation within each matched pair should not be ignored; otherwise, the inference will be invalid. The common strategy is to incorporate
random effects (i.e., frailties) to capture it (DeSantis et al., 2014). Therefore, a hierarchical or nested structure of correlation will be employed, where the frailties that take the correlation among recurrent stroke events are nested to the ones that capture matched-pair correlation. In recent years, such kind of clinical studies are commonly encountered indicating the needs of advanced model development, for instance, a child survival study in Northeast Brazil with data being collected according to families and communities (Sastry, 1997), and the Assessment, Serial Evaluation, and Subsequent Sequelae of Acute Kidney Injury (ASSESS-AKI) study in which patients are matched by baseline AKI status within each stratum according to baseline chronic kidney disease and clinical research center (Go et al., 2010). Several studies using nested frailty models to account for the hierarchical clustering of the data have been discussed, but limited work exist for recurrent time-to-event data (Sastry, 1997; Yau, 2001; Rondeau et al., 2006).

In this chapter, due to specific issues on zero-inflated nature of data and matched designs, we propose a novel joint frailty modeling for recurrent events and a terminal event of death by accommodating a conditional logistic model for subjects with zero recurrent events and nested frailties accounting for the hierarchical correlation structure. In particular, two nested frailties are to measure the dependency (1) between stroke and non-stroke subjects within a matched pair and (2) among recurrent stroke events within one individual. By sharing these frailties, death is dependent on recurrent stroke event process. The remainder of this chapter is organized as follows. In Section 3.2, we introduce the joint modeling of recurrent events and death in a matched cohort study. We provide the theoretical work on parameter estimation via the Monte Carlo Expectation-Maximization (MCEM) algorithm in Section 3.3. Results from simulation and a real-data analysis are presented in Sections 3.4 and 3.5, respectively. We summarize the conclusions and a discussion of future work in Section 3.6.
3.2 Joint Modeling

3.2.1 Notations

Let \( i \) denote the index of matched pairs and \( j \) denote the baseline exposure variable, \( i = 1, 2, \ldots, I; j = 1, 2, \ldots, J \). For the \( ij^{th} \) subject, let \( C_{ij}, M_{ij} \) and \( D_{ij} \) be the follow-up time, drop-out time and the death time respectively. Let \( T_{ij} = \min(D_{ij}, C_{ij}, M_{ij}) \) be the observed follow-up time with \( \Delta_{ij} = I(D_{ij} \leq \min(C_{ij}, M_{ij})) \) as the death indicator, where \( I(\cdot) \) is the indicator function. Denote \( \Psi_{ij}(t) = I(T_{ij} \geq t) \) as the “at-risk” indicator to show whether the subject is still under observation at time \( t \) or not. \( X_{ij} \) denotes the vector of the observed covariates including baseline risk factors, such as gender, race and smoking status and so on. Currently, we consider time-independent covariates, but the time-varying covariates can be incorporated in a straightforward manner.

Define \( N_{ij}^D(t) = I(D_{ij} \leq t, \Delta_{ij} = 1) \) as the observed death process, and \( N_{ij}^{D*}(t) = I(D_{ij} \leq t) \) as the actual death process by time point \( t \). In addition, denote \( N_{ij}^{R*}(t) = N\{\min(D_{ij}, t)\} \) and \( N_{ij}^R(t) = N\{\min(T_{ij}, t)\} = N_{ij}^{R*}(\min(T_{ij}, t)) \) as the actual and observed number of recurrent events by time point \( t \) separately. \( N_{ij}^D \) and \( N_{ij}^R \) are both the observed parts of the counting processes \( N_{ij}^{D*} \) and \( N_{ij}^{R*} \). The number of recurrent events that occur for the \( ij^{th} \) subject over the small interval \([t, t + dt)\) is \( dN_{ij}^R(t) = N_{ij}^{R*}((t + dt)^-) - N_{ij}^{R*}(t^-) \) as \( dt \to 0 \) and \( dN_{ij}^R = \Psi_{ij}(t) dN_{ij}^{R*} \). \( R_{ij} = I(N_{ij}^R > 0) \) denotes the observed indicator for whether the \( ij^{th} \) subject has at least one recurrent stroke during the follow-up. Let \( t_{ijk} \) be the \( k^{th} \) recurrent event time and \( \delta_{ijk} \) denote the indicator of recurrent events at time \( t_{ijk}, k = 1, 2, \ldots, N_{ij}^R \). Therefore, the data of the \( ij^{th} \) individual at time \( t \) is then \( O_{ij}(t) = \{\Psi_{ij}(u), N_{ij}^D(u), N_{ij}^R(u), 0 < u \leq t\} \). Then the entire set of observed data for individual \( ij \) is \( O_{ij} = \{O_{ij}(u), 0 < u \leq T_{ij}\} \).

3.2.2 Model

As we mentioned previously, there may be a high proportion patients (i.e., > 50%) who didn’t experience any recurrent events by the end of follow-up period. Let \( Y_{ij} \)
denote the indicator that the $ij^{th}$ subject will eventually ($Y_{ij} = 1$) or never ($Y_{ij} = 0$) experience the recurrent events, with probability $p_{ij} = Pr(Y_{ij} = 1)$. Define $y_{ij}$ as the value taken by the random variable $Y_{ij}$. It follows that, if $R_{ij} = 1$, $y_{ij} = 1$ and if $R_{ij} = 0$, $y_{ij}$ is unobserved, then this individual could fall in either of the two groups. A logistic regression model is used to model the probability of susceptible or not cured $p_{ij}$:

$$
p_{ij} = \frac{\exp(\beta_1^T \tilde{X}_{ij} + \mu_i)}{1 + \exp(\beta_1^T \tilde{X}_{ij} + \mu_i)} \quad (3.1)
$$

where $\beta_1$ is the vector of regression coefficients, indicating the effect of potential covariates $\tilde{X}_{ij}$. Note that the intercept term is absorbed into $\tilde{X}_{ij}$ which could be overlapped with $X_{ij}$.

Let $\mu_i$ denote the frailty measuring the dependency between subjects within a matched pair. $Y_{ij}'s$ from individuals in the same matched pair tend to be correlated and which also can be captured by $\mu_i$. Denote $\omega_{ij}$ as the frailty measuring the dependency among the recurrent events within the $ij^{th}$ individual. $\mu_i$ and $\omega_{ij}$ are shared by the recurrent and terminal events to induce their dependency. Assume that $(\mu_i, \omega_{ij})^T \sim MVN(0, \Sigma)$. For simplicity, we can assume that $\mu_i \perp \omega_{ij}$, where $\mu_i \sim N(0, \sigma_\mu^2)$ and $\omega_{ij} \sim N(0, \sigma_\omega^2)$. If $\sigma_\mu^2$ and $\sigma_\omega^2$ equal 0, then it implies there is no dependency between recurrent events and death, and the heterogeneity in both event processes is solely explained by covariate $X_{ij}$. Conditional on frailties $\mu_i$, $\omega_{ij}$ and covariates $X_{ij}$, the zero-inflated model for recurrent events is defined as follows which is a special case of latent models with two classes:

$$
\lambda_{ij}^R(t \mid X_{ij}, \mu_i, \omega_{ij}) = \begin{cases} 
\lambda_0^R(t) \exp(\beta_2^T X_{ij} + \mu_i + \omega_{ij}) & \text{with } p_{ij} (Y_{ij} = 1) \\
0 & \text{with } 1 - p_{ij} (Y_{ij} = 0) 
\end{cases} \quad (3.2)
$$

where $\beta_2$ is the vector of regression coefficients and $\lambda_0^R(t)$ is the baseline intensity function.

Conditional on $\mu_i$ and $\omega_{ij}$, the recurrent event process and death are independent of each other. The association between death and recurrent events is quantified by the shared frailties $\mu_i$ and $\omega_{ij}$ through $\phi_\mu$ and $\phi_\omega$ in equation (3.3), and a higher intensity of recurrent events is associated with a higher hazard rate for mortality.
Thus, the hazard function for death is given by

\[ \lambda_{ij}^D(t \mid X_{ij}, \mu_i, \omega_{ij}) = \lambda_0^D(t) \exp(\beta_3^T X_{ij} + \phi_\mu \mu_i + \phi_\omega \omega_{ij}) \] (3.3)

where \( \beta_3 \) is the vector of regression coefficients and \( \lambda_0^D(t) \) is the baseline hazard function of death. Of note, \( X_{ij} \) is assumed to be the same for \( \lambda_{ij}^R \) and \( \lambda_{ij}^D \) for simplicity, but can be different due to clinical perspectives in data applications.

Combining equations (3.2) and (3.3), we have a joint model of recurrent events and death adjusted for zero-inflation and a matched design.

Given \( \theta = (\lambda_0^R(\cdot), \lambda_0^D(\cdot), \beta_1, \beta_2, \beta_3, \phi_\mu, \phi_\omega, \sigma_\omega^2, \sigma_\mu^2)^T \), the marginal likelihood is

\[
L(\theta) = \int \int \prod_{i=1}^I \prod_{j=1}^J L(\theta \mid \mu_i, \omega_{ij}) f(\mu_i) f(\omega_{ij}) d\mu_i d\omega_{ij} \\
= \prod_{i=1}^I \prod_{j=1}^J \left\{ \left( L_{ij}^0 \right)^{1-R_{ij}} \left( L_{ij}^1 \right)^{R_{ij}} L_{ij}^D \right\} f(\mu_i) \right\} d\mu_i \] (3.4)

with

\[
L_{ij}^0 = 1 - p_{ij} + p_{ij} \exp[- \int_0^\infty \Psi_{ij}(t) \lambda_0^R(t) dt \exp(\beta_2^T X_{ij} + \mu_i + \omega_{ij})]
\]

\[
L_{ij}^1 = p_{ij} \prod_{k} \left[ \lambda_0^R(t) \exp(\beta_2^T X_{ij} + \mu_i + \omega_{ij}) \right]^{\delta_{ijk}} \cdot \exp[- \int_0^\infty \Psi_{ij}(t) \lambda_0^R(t) dt \exp(\beta_2^T X_{ij} + \mu_i + \omega_{ij})]
\]

\[
L_{ij}^D = \left[ (\lambda_0^D(T_{ij}) \exp(\beta_3^T X_{ij} + \phi_\mu \mu_i + \phi_\omega \omega_{ij}) \right]^{\Delta_{ij}} \cdot \exp(- \int_0^\infty \Psi_{ij}(t) \lambda_0^D(t) dt \exp(\beta_3^T X_{ij} + \phi_\mu \mu_i + \phi_\omega \omega_{ij}))
\]

where, \( L_{ij}^0 \) is the likelihood of observing zero recurrent events \( (R_{ij} = 0) \), which includes those participants who are at risk for recurrent events but censored or died before the first event. \( L_{ij}^1 \) is the likelihood of observing at least 1 recurrent event \( (R_{ij} = 1) \). \( L_{ij}^D \) is the likelihood for the terminal event of death. \( f(\mu_i) \) and \( f(\omega_{ij}) \) are normal densities. In addition, the cumulative baseline intensities of recurrent events and terminal event denoted by \( \Lambda_{0}^R(t) \) and \( \Lambda_{0}^D(t) \) are given by
\[ \Lambda^R_0(t) = \int_0^\infty \Psi_{ij}(t) \lambda^R_0(t) \, dt \] and \[ \Lambda^D_0(t) = \int_0^\infty \Psi_{ij}(t) \lambda^D_0(t) \, dt. \]

### 3.3 Estimation

The parameter estimation can be achieved by maximizing \( L(\theta) \). There are several commonly approaches used for estimation in the literature. The first approach is based on numerical integration techniques, such as Gaussian quadrature, to integrate out the frailties and then maximize the integrated likelihood. This approach can be easily implemented in SAS PROC NLMIXED and has been adopted by many researchers (Liu et al., 2015; Rondeau et al., 2013; Liu and Huang, 2008). Gaussian quadrature uses a weighted average of the integrand assessed at predetermined quadrature points over the random effects. Another approach is a Bayesian framework with Markov Chain Monte Carlo (MCMC) technique, which has emerged as a popular tool of joint frailty analysis (Ibrahim et al., 2004; Ouyang et al., 2013; Wen et al., 2016). The MCMC algorithm is able to draw inferences from a complex posterior distribution on a high-dimensional parameter space. By simulating a Markov chain that has the posterior distribution as its stationary distribution, MCMC algorithm draw inferences from a complex posterior distribution on a high-dimensional parameter space. Another alternative approach, the most popular for estimation in joint random effects models, is the EM algorithm (Liu et al., 2004; Zeng and Lin, 2009; Peng and Taylor, 2011; Klein, 1992; Vaida et al., 2000; Kim et al., 2012; Zeng et al., 2014). The EM algorithm iteratively computes maximum likelihood estimates from an incomplete data set by treating the frailties as missing data. In some cases, the joint likelihood functions involving multiple integrals do not yield closed-form expressions or analytic solutions for parameter estimation. Therefore, several numerical integration techniques, such as Monte Carlo (Liu et al., 2004; Peng and Taylor, 2011; Vaida et al., 2000) or Laplace approximation (Rondeau et al., 2006; Król et al., 2016), can be utilized.

We adopt the MCEM algorithm for estimation. In the E step, we find the expectation of the conditional log-likelihood, and integrate out the frailties by numerical integration with the Metropolis-Hasting algorithm.
We define $q_{1ij} = \exp(\beta_1^T \tilde{X}_{ij} + \mu_i)$, $q_{2ij} = \exp(\beta_2^T X_{ij} + \mu_i + \omega_{ij})$, and $q_{3ij} = \exp(\beta_3^T X_{ij} + \phi \mu_i + \phi \omega_{ij})$. In addition, the cumulative baseline intensities of recurrent events and the terminal event, denoted by $\Lambda^R(t)$ and $\Lambda^D(t)$, are given by $\Lambda^R(t) = \int_0^\infty \Psi_{ij}(t) \lambda^R(t) dt$ and $\Lambda^D(t) = \int_0^\infty \Psi_{ij}(t) \lambda^D(t) dt$. Thus given the values of latent variable $y$, frailties $\mu$ and $\omega$, the complete log likelihood function denoted by $l(\theta)$ can be split into four parts:

\[
l(\theta) = \sum_{i=1}^I \sum_{j=1}^J \log \{ p_{ij}^{y_{ij}} (1 - p_{ij})^{1-y_{ij}} \} \\
+ \sum_{i=1}^I \sum_{j=1}^J \log \{ (\prod_k (\lambda^R(t_{ijk}) \cdot q_{2ij})^{y_{ij}} \exp(-\Lambda^R(t) \cdot q_{2ij}))^{y_{ij}} \} \\
+ \sum_{i=1}^I \sum_{j=1}^J \log \{ (\lambda^D(T_{ij}) \cdot q_{3ij})^{\Delta_{ij}} \exp(-\Lambda^D(t) \cdot q_{3ij})\} \\
+ \sum_{i=1}^I \sum_{j=1}^J \log \{ -\frac{1}{\sqrt{2\pi}\sigma^2_\mu} \exp(-\frac{\mu^2_i}{2\sigma^2_\mu}) - \frac{1}{\sqrt{2\pi}\sigma^2_\omega} \exp(-\frac{\omega^2_{ij}}{2\sigma^2_\omega}) \} = l_1(\beta_1) + l_2(\beta_2, \lambda^R) + l_3(\beta_3, \lambda^D, \phi, \phi) + l_4(\sigma^2_\mu, \sigma^2_\omega)\]

The expectation of the log-likelihood $l(\theta)$ conditional on the observed data and the current parameter estimate $\hat{\theta}^{(k)}$ is:

\[
Q(\theta \mid \hat{\theta}^{(k)}) = E[l(\theta) \mid \hat{\theta}^{(k)}]
\]

\[
= E[l_1(\beta_1) \mid \hat{\theta}^{(k)}] + E[l_2(\beta_2, \lambda^R) \mid \hat{\theta}^{(k)}]
+ E[l_3(\beta_3, \lambda^D, \phi, \phi) \mid \hat{\theta}^{(k)}] + E[l_4(\sigma^2_\mu, \sigma^2_\omega) \mid \hat{\theta}^{(k)}]
\]

\[
= Q_1(\beta_1 \mid \hat{\theta}^{(k)}) + Q_2(\beta_2, \lambda^R \mid \hat{\theta}^{(k)})
+ Q_3(\beta_3, \lambda^D, \phi, \phi \mid \hat{\theta}^{(k)}) + Q_4(\sigma^2_\mu, \sigma^2_\omega \mid \hat{\theta}^{(k)})
\]

where

\[
Q_1(\beta_1 \mid \hat{\theta}^{(k)}) = \sum_{i=1}^I \sum_{j=1}^J \left\{ E[y_{ij} \mid \hat{\theta}^{(k)}] + E[y_{ij}] \beta_1^T \tilde{X}_{ij} - E[\log(1 + q_{1ij}) \mid \hat{\theta}^{(k)}] \right\}
\]
\[ Q_2(\beta_2, \lambda_0^R, \lambda_0^D | \hat{\theta}^{(k)}) \]
\[
= \sum_{i=1}^{I} \sum_{j=1}^{J} \{ R_{ij} \left( \sum_{m=1}^{N_{ij}} (\log(\lambda_0^R(t_{ijk})) + \beta_2^T X_{ij} + E[\mu_i | \hat{\theta}^{(k)}] + E[\omega_{ij} | \hat{\theta}^{(k)}]) \right) \\
- \Lambda_0^R(T_{ij}) \exp(\beta_2^T X_{ij} + \log E[y_{ij} \exp(\mu_i + \omega_{ij}) | \hat{\theta}^{(k)}]) \} 
\]

\[ Q_3(\beta_3, \lambda_0^D, \phi, \phi \omega | \hat{\theta}^{(k)}) \]
\[
= \sum_{i=1}^{I} \sum_{j=1}^{J} \{ \Delta_{ij} \left( \log(\lambda_0^D(T_{ij})) + \beta_3^T X_{ij} + \phi E[\mu_i | \hat{\theta}^{(k)}] + \phi \omega E[\omega_{ij} | \hat{\theta}^{(k)}] \right) \\
- \Lambda_0^D(T_{ij}) \exp(\beta_3^T X_{ij}) E[\exp(\phi \mu_i + \phi \omega_{ij}) | \hat{\theta}^{(k)}] \} 
\]

\[ Q_4(\sigma_\mu^2, \sigma_\omega^2 | \hat{\theta}^{(k)}) = \sum_{i=1}^{I} \sum_{j=1}^{J} \{ -\frac{1}{2} \left( \log(2\pi) + \log \sigma_\mu^2 + \frac{E[\omega_{ij}^2 | \hat{\theta}^{(k)}]}{\sigma_\mu^2} \right) \\
- \frac{1}{2} \left( \log(2\pi) + \log \sigma_\omega^2 + \frac{E[\mu_i^2 | \hat{\theta}^{(k)}]}{\sigma_\omega^2} \right) \} 
\]

More details on the computation of the conditional expectation above are given in Appendix A.1. Expectation terms that involve \( y_{ij} \) in equation (4.28) in E-step are: \( E[y_{ij} | \hat{\theta}^{(k)}] \), \( E[y_{ij} \mu_i | \hat{\theta}^{(k)}] \), \( E[y_{ij} \exp(\mu_i + \omega_{ij}) | \hat{\theta}^{(k)}] \). \( E[y_{ij} \mu_i | \hat{\theta}^{(k)}] \) is not involved in the M-step, and thus not considered. The expectation terms are evaluated with respect to \( p(y, \mu, \omega) \):

\[
p(y, \mu, \omega | \hat{\theta}^{(k)}) \propto \\
\prod_{i=1}^{I} \prod_{j=1}^{J} \left\{ \left[ N_{ij}^R \exp(\mu_i + \omega_{ij}) \right]^{R_{ij}} \exp(-y_{ij} \Lambda_0^R(T_{ij}) q_{1ij}^{(k)}) \right\} \\
\times \exp(y_{ij} (\beta_1^{T(k)} \tilde{X}_{ij} + \mu_i)) / \left( 1 + \exp(\beta_1^{T(k)} \tilde{X}_{ij} + \mu_i) \right) \\
\frac{1}{\sqrt{2\pi \sigma_\mu^{2(k)}}} \frac{1}{\sqrt{2\pi \sigma_\omega^{2(k)}}} \exp(-\frac{\mu_i^2}{2\sigma_\mu^{2(k)}}) \exp(-\frac{\omega_{ij}^2}{2\sigma_\omega^{2(k)}}) \} 
\]

(3.7)

where, \( q_{1ij}^{(k)} = \exp(\beta_1^{T(k)} \tilde{X}_{ij} + \mu_i) \), \( q_{2ij}^{(k)} = \exp(\beta_2^{T(k)} X_{ij} + \mu_i + \omega_{ij}) \),
\(q_{3ij}^{(k)} = \exp(\beta_3^{(k)} X_{ij} + \phi_3^{(k)} \mu_i + \phi_3^{(k)} \omega_{ij})\), and \(p_{ij}^{(k)} = \frac{\exp(\beta_1^{(k)} x_{ij} + \mu_i)}{1 + \exp(\beta_1^{(k)} x_{ij} + \mu_i)}\).

We know that if \(R_{ij} = 1\), then \(y_{ij} = 1\). If \(R_{ij} = 0\), then \(y_{ij}\) follows the binomial distribution below:

\[
y_{ij} \mid \mu, \omega \sim \text{bin}\left\{1, \pi = \frac{p_{ij}^{(k)} \exp(-\Lambda_0^{(k)}(T_{ij})q_{2ij}^{(k)})}{1 - p_{ij}^{(k)} + p_{ij}^{(k)} \exp(-\Lambda_0^{(k)}(T_{ij})q_{2ij}^{(k)})}\right\}
\]

Therefore, conditional on the observed data and the current parameter estimation \(\hat{\theta}^{(k)}\), we can evaluate the following expectations as:

\[
E[y_{ij} \mid \hat{\theta}^{(k)}] = \begin{cases} 
E \left[ \frac{p_{ij}^{(k)} \exp(-\Lambda_0^{(k)}(T_{ij})q_{2ij}^{(k)})}{1 - p_{ij}^{(k)} + p_{ij}^{(k)} \exp(-\Lambda_0^{(k)}(T_{ij})q_{2ij}^{(k)})} \right] \hat{\theta}^{(k)} & \text{if } R_{ij} = 0 \\
1 & \text{otherwise}
\end{cases}
\]

\[
E[y_{ij} \exp(\mu_i + \omega_{ij}) \mid \hat{\theta}^{(k)}] = \begin{cases} 
E \left[ \frac{\exp(\mu_i + \omega_{ij}) p_{ij}^{(k)} \exp(-\Lambda_0^{(k)}(T_{ij})q_{2ij}^{(k)})}{1 - p_{ij}^{(k)} + p_{ij}^{(k)} \exp(-\Lambda_0^{(k)}(T_{ij})q_{2ij}^{(k)})} \right] \hat{\theta}^{(k)} & \text{if } R_{ij} = 0 \\
E[\exp(\mu_i + \omega_{ij}) \mid \hat{\theta}^{(k)}] & \text{otherwise}
\end{cases}
\]

where the expectation are taken with respect to the \(p(\mu, \omega \mid \hat{\theta}^{(k)})\).

\[
p(\mu, \omega \mid \hat{\theta}^{(k)}) \propto \prod_{i=1}^{I} \prod_{j=1}^{J} \left\{ 1 - p_{ij}^{(k)} + p_{ij}^{(k)} \exp(-\int_0^\infty \Psi_{ij}(t) \lambda_0^{(k)}(t) dt) q_{2ij}^{(k)} \right\}

\prod_{k} (q_{2ij}^{(k)})^{\Delta_{ij,k}} \exp(-\int_0^\infty \Psi_{ij}(t) \lambda_0^{(k)}(t) dt) q_{2ij}^{(k)}

(q_{3ij}^{(k)})^{\Delta_{ij}} \exp(-\int_0^\infty \Psi_{ij}(t) \lambda_0^{(k)}(t) dt) q_{3ij}^{(k)}

\frac{1}{\sqrt{2\pi\sigma^2_\mu^{(k)}}} \exp\left(-\frac{\mu_i^2}{2\sigma^2_\mu^{(k)}}\right)

\frac{1}{\sqrt{2\pi\sigma^2_\omega^{(k)}}} \exp\left(-\frac{\omega_i^2}{2\sigma^2_\omega^{(k)}}\right)
\]

Since there is no closed-form expression of \(p(\mu, \omega \mid \hat{\theta}^{(k)})\), Monte Carlo methods in combination with the Metropolis-Hastings algorithm are used to approximate the posterior distributions of \(u_i's\) and \(\omega_i's\). Given \(\hat{\theta}^{(k)}\), \(N\) random samples are generated for \(\mu_i^{(m)} (m = 1, \ldots M)\) and \(\omega_i^{(m)} (m = 1, \ldots M)\) to estimate the expectation of the sufficient statistics involving frailties. Thus, \(E[f(\mu_i) \mid \hat{\theta}^{(k)}] = \frac{1}{N} \sum_{n=1}^{N} f(\mu_i^{(n)})\)
and \( E[f(\omega_{ij}) \mid \hat{\theta}^{(k)}] = \frac{1}{N} \sum_{n=1}^{N} f(\omega_{ij})^{(n)} \), where \( f(\cdot) \) could be any smooth and monotone function. A brief introduction to the Metropoli-Hastings algorithm is given in A.2.

In the M-step, we use a Newton-Raphson procedure to maximize \( Q_1(\beta_1 \mid \hat{\theta}^{(k)}) \), \( Q_2(\beta_2, \lambda_0^R \mid \hat{\theta}^{(k)}) \), \( Q_3(\beta_3, \lambda_0^D, \phi_\mu, \phi_\omega \mid \hat{\theta}^{(k)}) \), \( Q_4(\sigma^2_\mu, \sigma^2_\omega \mid \hat{\theta}^{(k)}) \) to estimate \( \theta \). For estimation of baseline intensity, Sy and Taylor (2000) suggested that the Breslow estimator does not work well because of the inability to approach zero even when the data indicate a leveling off of the marginal survival curve when there exists zero-inflation. We use the piecewise constant baseline intensity function which retains enough model structure, while it provides more flexibility over the a priori choices of baseline intensity distribution (e.g., Exponential, Weibull). Partition of \( M \) intervals with cutpoints \( 0 \leq c_0 < c_1 < \ldots < c_M = \infty \) on the time duration can be based on the quantiles of event times, where \( c_0 = 0 \) or the smallest event time.

The baseline intensity function is assumed to be constant within each of the \( M \) intervals, so that

\[
\tilde{\lambda}_0(t) = \sum_{m=1}^{M} \lambda_m I(c_{m-1} < t \leq c_m) \quad (3.12)
\]

Therefore we model the baseline intensity \( \tilde{\lambda}_0(t) \) using \( M \) parameters \( \lambda_1, \lambda_2, \ldots, \lambda_m \). The cumulative baseline intensity is

\[
\tilde{\Lambda}_0(t) = \sum_{m=1}^{M} \lambda_m \max(0, \min(c_m - c_{m-1}, t - c_{m-1})) \quad (3.13)
\]

Of note, we found that when there are enough recurrent events, the estimates of Breslow estimator for baseline intensities performs well. In A.5, we provide results of a simulation study of the proposed full model without zero-inflation for recurrent events with Breslow estimator for baseline intensities.

The variance of the MLE \( \hat{\theta} = (\hat{\lambda}_0^R(t), \hat{\lambda}_0^D(t), \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \hat{\phi}_\mu, \hat{\phi}_\omega, \hat{\sigma}_\mu^2, \hat{\sigma}_\omega^2)^T \) in the MCEM algorithm cannot be obtained directly from the algorithm. Louis’s formula was used to obtain it Louis (1982). The variance-covariance matrix can be estimated using the inverse of an observed information matrix. The observed
information matrix $I(\hat{\theta})$ is given by

$$ I(\hat{\theta}) = -E\left[ \frac{\partial^2 l(\theta)}{\partial \theta \partial \theta^T} \right | \hat{\theta} ] - \text{Var}\left[ \frac{\partial l(\theta)}{\partial \theta} \right | \hat{\theta} ] $$

(3.14)

where \( \text{Var}\left[ \frac{\partial l(\theta)}{\partial \theta} \right | \hat{\theta} ] = E\left[ \frac{\partial l(\theta)}{\partial \theta} \frac{\partial l(\theta)}{\partial \theta^T} \right | \hat{\theta} ] \). The last term becomes zero due to the MLE \( \hat{\theta} \) and all of these terms are evaluated at the last iteration of the EM algorithm. The details of applying the Newton-Raphson algorithm are given in Appendix A.3.

### 3.4 Simulation

In this section, we conduct simulation studies to evaluate our proposed joint frailty model. For simplicity, we only consider one covariate for $X$, which takes values of 0 or 1 each with probability 0.5, and also only the intercept term for the logistic regression model. Considering exponential distributions for both recurrent event times and death, the following models are shown below

$$ \logit P(Y_{ij} = 1 | \mu_i) = \beta_1 + \mu_i $$

$$ \lambda_{ij}^R(t | \mu_i, \omega_{ij}, Y_{ij} = 1) = \lambda_0^R(t) \exp(\beta_2 X_{ij} + \mu_i + \omega_{ij}) $$

$$ \lambda_{ij}^D(t | \mu_i, \omega_{ij}) = \lambda_0^D(t) \exp(\beta_3 X_{ij} + \phi_\mu \mu_i + \phi_\omega \omega_{ij}) $$

(3.15)

The censoring time is taken as $C_{ij} = 6 \ast \text{Unif}(0,1)$, where \( \text{Unif}(0,1) \) is a random number generated from a uniform distribution in $[0,1]$. We assume $\mu_i \sim N(0, \sigma^2_\mu)$, $\omega_{ij} \sim N(0, \sigma^2_\omega)$ and set $\sigma^2_\mu = 1$, $\sigma^2_\omega = 0.25$. We set the parameters as $\beta_2 = 1.2$, $\beta_3 = 1$, $\phi_\mu = 1$, $\phi_\omega = 0.5$, but invoke different values for $\beta_1$ and recurrent events baseline intensity $\lambda_0^R$.

There are two set-ups: 1) $\beta_1 = 1$, $\lambda_0^R = 0.5$: 48% subjects are observed with at least one recurrent event. 22% subjects are susceptible but have not yet experienced any events during the follow-up period; 2) $\beta_1 = 0.5$, $\lambda_0^R = 0.25$. 33% subjects are observed with at least one recurrent event. 27% subjects are susceptible but have not yet experienced any events during the follow-up period. In both set-ups, $\lambda_0^D = 0.1$ and the censoring rate for death is 40%. The major difference
between the two set-ups is the percentage of subjects who have at least one recurrent event (48% and 33%), we recognize these two percentages are reasonable to consider the zero-inflation, and obviously if the percentage is too low, the convergence issue could be challenging.

A total of 500 Monte Carlo replicates are generated, each with a sample size of 500 subjects. A non-homogeneous Poisson process is adopted to simulate recurrent event times and the detailed procedures are provided in Appendix A.4. Of note, for comparison, we also consider the estimation approaches based on Gaussian quadrature and a Bayesian approach through SAS PROC MCMC to further evaluate our algorithm.

The piecewise constant approach is adopted to estimate the baseline intensity for recurrent events and baseline hazard for death, where five intervals are chosen. The proposed MCEM algorithm is implemented in R software, and the functions are available upon request. With regards to the Gaussian quadrature method, we invoke the adaptive Gaussian quadrature option with five quadrature points in SAS PROC NLMIXED. The Bayesian method in SAS PROC MCMC is implemented by the random walk Metropolis algorithm to obtain posterior samples with random walk step size 2.38. A Normal (0, 1000) prior is adopted for $\beta_1, \beta_2, \beta_3, \phi_\mu, \phi_\omega$. An Inverse gamma (0.01, 0.01) prior is adopted for $\sigma_\mu^2$ and $\sigma_\omega^2$. A Gamma (0.01,0.01) prior is adopted for five piecewise baseline intensities for both $\tilde{\lambda}_R^0(t)$ and $\tilde{\lambda}_D^0(t)$. The posterior sample size is 50,000. Also, due to the poor mixing and slow convergence, we thin a chain by keeping every 50th simulated draw from each sequence.

The resulting parameter estimates are shown in Table 3.1 and Table 3.2. It can be seen that in both set-ups for the full model, MCEM estimation method yields satisfactory results and all parameter estimates have very small empirical biases. MCEM has better performance than the Gaussian quadrature and the Bayesian in SAS, especially on the estimation of parameter $\phi_\omega$ which has the smallest bias. Also, Monte Carlo standard error (MCSE) and MSE estimates from the EM algorithm are smaller than those from PROC NLMIXED and PROC MCMC. Comparing the parameter estimates between settings I and II, the empirical biases tend to increase when the percentage of recurrent events decreases from 48% to 33%. We
also compute the mean of the standard error estimate for the full model based on our derivation, which are very close to the MCSE. Therefore, here we only report MCSE in summary results for simulation.

For comparisons with miss-specified models, if we don’t consider the correlation $\mu$ in the matched pair (denoted by Model I), the estimation of $\sigma^2_\omega$ and $\phi_\omega$ are overestimated. The correlation that was induced by matching is accounted for by recurrent events. This indicates that the hierarchical structure of the data needs to be taken into account to obtain accurate inference. If we ignore the zero-inflation from the full model (denoted by Model II), this will yield poor estimates in variance and coefficients of both frailties $\omega$ and $\mu$. For example, in Setting I, the estimate of $\sigma^2_\omega$ is about two times larger than the true value, suggesting much more heterogeneity in recurrent events if not accounting for zero-inflation. Therefore, our model is recommended in practice for valid inference, and insufficient models could lead to large biased estimates.

3.5 Application- Stroke study

We apply the proposed model to a real data application on recurrent acute ischemic stroke. Our real data example is obtained from the MarketScan database between January, 2011 and December 2014, including a matched cohort of patients with and without acute ischemic stroke at baseline, who are aged 45-54 with surgical and medical admission for inpatient acute care hospitalization. The episodes of acute ischemic stroke are diagnosed by the ICD9-CM codes with 434.x and 436.x Goldstein (1998). A recurrent stroke is defined as any recurrent stroke occurring more than 28 days after the incident stroke Coull and Rothwell (2004). Figure 3.1 shows the sample data from random six matched pairs with and without stroke at baseline. Both cohorts are tracked for in-hospitalized mortality and pre-existing comorbidity conditions including diabetes mellitus, hypertension and heart disease during the past twelve-month which could potentially influence outcomes.

The total sample size in this study is 2,122, and the number of recurrent stroke ranges from 0 to 43, whereas 71.8% of the patients are not observed with recurrent
strokes. The wide range of the number of recurrent stroke suggests a large variation across subjects. 139 patients (6.55%) died in the follow-up period and 405 (20%) withdrew from the study. The preliminary analysis potentially shows that patients with baseline stroke have a higher risk of mortality (57.55%). Thus, censoring due to death is very likely to be informative. Joint frailty models of recurrent events in the presence of death can circumvent this problem by modeling recurrent events and death with shared frailties to capture their dependency. Baseline covariates included in the analysis are stroke status: Stroke (1=yes, 2=no) and Comorbidity (1=yes, 2=no). Comorbidity is defined as 1 if the patient has any one of diabetes mellitus, hypertension and heart disease the past twelve-month. Our final model is

$$\text{logit} P(Y = 1) = \beta_0 + \beta_1 \text{Stroke} + \mu$$

$$\lambda^R(t \mid Y = 1) = \lambda_0^R(t) \exp(\beta_2 \text{Stroke} + \beta_3 \text{Comorbidity} + \mu + \omega)$$

$$\lambda^D(t) = \lambda_0^D(t) \exp(\beta_4 \text{Stroke} + \beta_5 \text{Comorbidity} + \phi_\mu \mu + \phi_\omega \omega)$$

We summarize the results in Table 3.3. We can see that both the baseline stroke and comorbidity covariates are significant in the proposed model. Patients with stroke at baseline had an odds ratio of 1.493 (= exp(0.401)) to be "susceptible". Baseline stroke and comorbidity also have significant effects on both intensity of recurrent stroke among those "susceptible" and death hazard among all subjects. The intensity ratio of having recurrent stroke for patients with stroke at baseline is 2.942 (= exp(1.079)) compared to those without stroke at baseline. Compared to patients who don’t have comorbidity at baseline, the intensity ratio of having recurrent events is 1.104 (= exp(0.099)) for patients who have comorbidity. Finally, the frailty variance estimate of $\mu$ is 0.669, suggesting the existence of heterogeneity in the matched stroke and non-stroke pair. The frailty variance estimate for $\omega$ is 0.907, which may due to the wide range of the number of recurrent stroke events among patients. The estimates of $\phi_\mu$ (4.001) and $\phi_\omega$ (0.430) are significant greater than 0, which implies that the recurrent strokes events and death rates are positively associated. Thus, higher frailty ($\omega$) will lead to a higher risk of recurrence and a higher risk of death. Also, higher frailty ($\mu$) will result in a higher risk of recurrence, a higher risk of death and a higher probability to develop
a new stroke event.

We also fit two reduced joint frailty models for recurrent stroke, i.e. without zero-inflation and $\mu$. The results from these two models are shown in Table 3.3. We notice that some parameter estimates are quite different from those in the full model, indicating the necessity to adjust for zero-inflation and matched correlation, i.e., without considering zero-inflation or matched pair correlation, the effect of comorbidity at baseline on having recurrent events is enlarged (The rate of having recurrent stroke increases from 1.104 to 1.359 (=$\exp(0.307)$) and 1.328 (=$\exp(0.284)$) compared to patients who don’t have comorbidity). We also note that the frailty variance estimates are much larger than that in the full model, i.e., not accounting for the zero-inflation would leads to more heterogeneity for recurrent events in the reduced model (variance estimate of $\mu$ increased from 0.669 to 1.551). Ignoring the correlation induced by the matched design result in severely overlook the importance of the correlation in recurrent events (variance estimate of $\omega$ increased from 0.907 to 2.923). Two information criteria for model assessment, the Akaike information criterion (AIC) and the Bayesian information criterion (BIC) are also calculated, and both AIC and BIC statistics indicate that the proposed full model has the best fit with the smallest AIC=11923 and BIC=12003.

### 3.6 Discussion

In this Chapter, we proposed a joint frailty model of recurrent events and a terminal event adjusted for zero-inflation and the matched design. Compared to the latest research on similar topics, the major advantage of our model is the ability to quantify the dependency between recurrent events and the terminal event that are correlated in a hierarchical structure (e.g. matched case-control study, meta-analysis). We have shown by simulation and a real data application that using a reduced model (ignoring matched-pair correlation) instead of the proposed full model when there are two levels of clustering can lead to biased estimates, with an overestimation of the correlation in recurrent events. This calls into question the validity of traditional statistical techniques such as the shared frailty model.
in studies with matched-design. Another major advantage is the consideration of zero-inflation of the recurrent event, which results in more accurate parameter estimation. In general, the price of omitting the feature of zero-inflation of the events in the data from the models is biased estimates and overlooking the importance of certain cluster effects. Also, in medical research, investigators can evaluate the treatment effect by estimating the fraction of cured subjects, which is a substantially important question.

We also implemented a MCEM algorithm with a piecewise baseline intensity in R software to compute the maximum likelihood estimates of the model parameters, and the R functions for estimation are available upon request, which are flexibly adjusted for model extension and other research purposes. Most of the latest research on similar topics adopt the Gaussian quadrature in SAS PROC NLMIXED for estimation. Although SAS PROC NLMIXED is easy to implement for parameter estimation, there exist several limitations in practice. In our situation, we have two levels of nested random effects, which leads to dramatically computationally intense and convergence issues with about 3% failed optimization due to the fact that Hessian matrix is not positive-definite. The number of failure optimizations may highly increase with a more sophisticated joint frailty model (e.g. more levels of random effects). Due to the large volume of parameters for estimation and complexity of the posterior distribution, the adaptive MCMC algorithm in SAS also requires high computation burden because a large number of posterior samples are required. Sometimes, poor mixing or slow convergence issues emerge because the model parameters may be correlated with each other. Also, it is a little bit unstable because it yields relatively large standard errors on some parameter estimates, especially the parameters related with frailties.

There exist some limitations in our data. In the Marketscan research database, it’s not possible to obtain the death information when patients are discharged from the hospital. The study cohort was tracked for in-hospitalized mortality only. Therefore, the death rate in our data under-represents the true rate. The dependency between recurrent acute ischemic stroke and death may be under-estimated. For the future work, our model can be extended in several directions. First, other functional forms of the frailties can be incorporated in the proposed joint model,
such as gamma frailties, which have been applied by many researchers (Liu et al., 2004; Rondeau et al., 2006). With only a minor modification, the corresponding likelihood and the estimation algorithm are readily available. Second, time-varying covariates also can be incorporated (Wang et al., 2001). Third, we can consider more complex joint model settings. For example, we can jointly model longitudinal biomarkers with the current model (Kim et al., 2012; Król et al., 2016). Last, our proposed model will be applied to the ASSESS-AKI (Go et al., 2010) study to investigate the association between recurrent acute kidney injury (AKI) and death.
Table 3.1. Simulation results of Setting I: $\beta_1 = 1$, $\lambda_0^R = 0.5$, $\lambda_0^D = 0.1$: 48% subjects are observed with at least one recurrent event. Another 22% subjects are susceptible but have not yet experienced any events during the follow-up period. The censoring rate for death is 40%. Full model denotes the proposed joint model of zero-inflated recurrent events and death event; Model I denotes the model without considering the matched pair correlation, where matched pair random effect $\mu$ is removed; Model II denotes the model without considering zero-inflation for recurrent events.

<table>
<thead>
<tr>
<th>Setting I</th>
<th>Model I</th>
<th>Model II</th>
<th>Full model</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N = 500$</td>
<td>Parameters</td>
<td>True AB</td>
<td>MCSE MSE</td>
</tr>
<tr>
<td>MCEM</td>
<td>$\beta_1$</td>
<td>1 0.203</td>
<td>0.140</td>
</tr>
<tr>
<td></td>
<td>$\beta_2$</td>
<td>1.2 0.020</td>
<td>0.088</td>
</tr>
<tr>
<td></td>
<td>$\beta_3$</td>
<td>1 0.014</td>
<td>0.147</td>
</tr>
<tr>
<td></td>
<td>$\phi_\mu$</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\phi_\omega$</td>
<td>0.5 0.538</td>
<td>0.141</td>
</tr>
<tr>
<td></td>
<td>$\sigma^2_\mu$</td>
<td>0.25</td>
<td>0.926</td>
</tr>
<tr>
<td>GQ</td>
<td>$\beta_1$</td>
<td>1 0.192</td>
<td>0.205</td>
</tr>
<tr>
<td></td>
<td>$\beta_2$</td>
<td>1.2 0.013</td>
<td>0.124</td>
</tr>
<tr>
<td></td>
<td>$\beta_3$</td>
<td>1 0.056</td>
<td>0.180</td>
</tr>
<tr>
<td></td>
<td>$\phi_\mu$</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\phi_\omega$</td>
<td>0.5 0.541</td>
<td>0.157</td>
</tr>
<tr>
<td></td>
<td>$\sigma^2_\mu$</td>
<td>0.25</td>
<td>0.986</td>
</tr>
<tr>
<td>MCMC</td>
<td>$\beta_1$</td>
<td>1 0.178</td>
<td>0.218</td>
</tr>
<tr>
<td></td>
<td>$\beta_2$</td>
<td>1.2 0.008</td>
<td>0.154</td>
</tr>
<tr>
<td></td>
<td>$\beta_3$</td>
<td>1 0.076</td>
<td>0.204</td>
</tr>
<tr>
<td></td>
<td>$\phi_\mu$</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\phi_\omega$</td>
<td>0.5 0.565</td>
<td>0.175</td>
</tr>
<tr>
<td></td>
<td>$\sigma^2_\mu$</td>
<td>0.25</td>
<td>1.05</td>
</tr>
</tbody>
</table>

MCEM, Monte Carlo EM algorithm; GQ, Gaussian quadrature approach; MCMC, Markov chain Monte Carlo method
AB is the absolute value of the difference between Monte Carlo mean of the parameter estimates (bases on 500 replicates) and the true value; MCSE is the Monte Carlo standard error; MSE is the mean square error.
Table 3.2. Simulation results of Setting II: $\beta_1 = 0.5$, $\lambda^R_0 = 0.25$, $\lambda^D_0 = 0.1$: 33% subjects are observed with at least one recurrent event. Another 27% subjects are susceptible but have not yet experienced any events during the follow-up period. The censoring rate for death is 40%. Full model denotes the proposed joint model of zero-inflated recurrent events and death event; Model I denotes the model without considering the matched pair correlation, where matched pair random effect $\mu$ is removed; Model II denotes the model without considering zero-inflation for recurrent events.

<table>
<thead>
<tr>
<th>Set-up II</th>
<th>N = 500</th>
<th>Parameters</th>
<th>True</th>
<th>AB</th>
<th>MCSE</th>
<th>MSE</th>
<th>AB</th>
<th>MCSE</th>
<th>MSE</th>
<th>AB</th>
<th>MCSE</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>MCEM</td>
<td>$\beta_1$</td>
<td>0.5</td>
<td>0.102</td>
<td>0.128</td>
<td>0.027</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\beta_2$</td>
<td>1.2</td>
<td>0.040</td>
<td>0.116</td>
<td>0.015</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\beta_3$</td>
<td>1</td>
<td>0.042</td>
<td>0.163</td>
<td>0.028</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\phi_\mu$</td>
<td>1</td>
<td>0.282</td>
<td>0.111</td>
<td>0.092</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\phi_\omega$</td>
<td>0.5</td>
<td>0.643</td>
<td>0.182</td>
<td>0.446</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma^2_{\mu}$</td>
<td>0.25</td>
<td>0.935</td>
<td>0.289</td>
<td>0.958</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma^2_{\omega}$</td>
<td>0.5</td>
<td>0.080</td>
<td>0.237</td>
<td>0.062</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GQ</td>
<td>$\beta_1$</td>
<td>0.5</td>
<td>0.080</td>
<td>0.237</td>
<td>0.062</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\beta_2$</td>
<td>1.2</td>
<td>0.014</td>
<td>0.152</td>
<td>0.023</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\beta_3$</td>
<td>1</td>
<td>0.084</td>
<td>0.195</td>
<td>0.045</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\phi_\mu$</td>
<td>1</td>
<td>0.258</td>
<td>0.129</td>
<td>0.083</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\phi_\omega$</td>
<td>0.5</td>
<td>0.597</td>
<td>0.187</td>
<td>0.391</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma^2_{\mu}$</td>
<td>0.25</td>
<td>1.114</td>
<td>0.305</td>
<td>1.334</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma^2_{\omega}$</td>
<td>0.5</td>
<td>0.070</td>
<td>0.292</td>
<td>0.090</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCMC</td>
<td>$\beta_1$</td>
<td>0.5</td>
<td>0.080</td>
<td>0.237</td>
<td>0.062</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\beta_2$</td>
<td>1.2</td>
<td>0.003</td>
<td>0.177</td>
<td>0.031</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\beta_3$</td>
<td>1</td>
<td>0.114</td>
<td>0.226</td>
<td>0.064</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\phi_\mu$</td>
<td>1</td>
<td>0.210</td>
<td>1.281</td>
<td>1.685</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\phi_\omega$</td>
<td>0.5</td>
<td>0.652</td>
<td>0.234</td>
<td>0.479</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma^2_{\mu}$</td>
<td>0.25</td>
<td>1.169</td>
<td>0.318</td>
<td>1.468</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

MCEM, Monte Carlo EM algorithm; GQ, Gaussian quadrature approach; MCMC, Markov chain Monte Carlo method.

AB is the absolute value of the difference between Monte Carlo mean of the parameter estimates (bases on 500 replicates) and the true value; MCSE is the Monte Carlo standard error; MSE is the mean square error.
Figure 3.1. Sample data on recurrent acute ischemic stroke
Table 3.3. Analysis of recurrent ischemic stroke data: Full model denotes the proposed joint model of zero-inflated recurrent events and death event; Model I denotes the model without considering the matched pair correlation, where matched pair random effect $\mu$ is removed; Model II denotes the model without considering zero-inflation for recurrent events.

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Model I</th>
<th></th>
<th>Model II</th>
<th></th>
<th>Full model</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>SE</td>
<td>p-value</td>
<td>Estimate</td>
<td>SE</td>
<td>p-value</td>
</tr>
<tr>
<td>Incidence</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>-0.725</td>
<td>0.218</td>
<td>&lt;0.001</td>
<td>-0.300</td>
<td>0.114</td>
<td>0.009</td>
</tr>
<tr>
<td>Stroke</td>
<td>2.055</td>
<td>0.244</td>
<td>&lt;0.001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reurrent stroke</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stroke</td>
<td>1.152</td>
<td>0.242</td>
<td>&lt;0.001</td>
<td>1.045</td>
<td>0.038</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Comorbidity</td>
<td>0.284</td>
<td>0.127</td>
<td>0.025</td>
<td>0.307</td>
<td>0.119</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Death</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stroke</td>
<td>1.812</td>
<td>0.356</td>
<td>&lt;0.001</td>
<td>1.315</td>
<td>0.147</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Comorbidity</td>
<td>1.148</td>
<td>0.322</td>
<td>&lt;0.001</td>
<td>1.025</td>
<td>0.230</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>$\phi_\mu$</td>
<td>3.041</td>
<td>0.234</td>
<td>0.042</td>
<td>4.001</td>
<td>0.275</td>
<td>0.009</td>
</tr>
<tr>
<td>$\phi_\omega$</td>
<td>2.634</td>
<td>0.216</td>
<td>&lt;0.001</td>
<td>0.275</td>
<td>0.191</td>
<td>0.009</td>
</tr>
<tr>
<td>$\sigma^2_{\mu}$</td>
<td>1.551</td>
<td>0.091</td>
<td>&lt;0.001</td>
<td>0.669</td>
<td>0.074</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>$\sigma^2_{\omega}$</td>
<td>2.923</td>
<td>0.349</td>
<td>&lt;0.001</td>
<td>2.502</td>
<td>0.374</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>AIC</td>
<td>14630</td>
<td></td>
<td></td>
<td>14966</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BIC</td>
<td>14709</td>
<td></td>
<td></td>
<td>15036</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

SE is the standard error of the parameter estimate;
AIC is the Akaike information criterion; BIC is the Bayesian information criterion.
The two-stage joint model of a longitudinal biomarker, recurrent events and death

4.1 Introduction

In this chapter, we propose a two-stage joint model of longitudinal biomarkers, recurrent events and death which is an extending of the joint model in Chapter 3. The longitudinal biomarker is incorporated into the joint model of recurrent events and death as time-dependent covariates. There are two different categories of time-dependent covariates in survival analysis, namely, external or exogenous covariate and internal or endogenous covariate (Rizopoulos, 2012). We should distinguish between the different types, because that an internal covariate requires special treatment compared to an external one. Measurements taken on the subjects under study, such as biomarkers and clinical parameters are endogenous covariates. There some important features of endogenous covariates. First, endogenous covariates typically require the survival of the subject for their existence. Second, the measurements are typically measured with errors. Thus, for such covariates, it would be more reasonable to assume that the observed marker levels are biased
from the true marker levels. Moreover, in our model, the implication with endoge-
nous covariates is that the levels of a biomarker for a patient are only known for
the specific occasions that this patient visited the study center, and not in between
these visit times. Therefore, their complete path to any time \( t \) is not fully observed.

However, because of left-censoring of the assay used to quantify the marker, lon-
gitudinal data could be complicated by left-censoring of some measures. For ex-
ample, HIV-1 RNA level or viral load has been recognized as the best prognostic
marker with CD4 + cell counts. The major problem with HIV-1 RNA level is
due to measurement limitations, generally between 500 and 20 copies/ml, so the
measures are left-censored. In the literature, several approaches have been pro-
posed for the analysis of left censored longitudinal data. One is the naive method
which is carried out by simply replacing the left censored value by the fixed values,
which can be zero, the detection limit or half of the detection limits (Keet et al.,
presented model for the analysis of some antibody assay data sets, which is a mix-
ture of a censored lognormal distribution and a point distribution located at some
point below the detection level. This approach avoids clumping of the replaced
values, when there are several non-detectable values share a common detection
limit. Paxton et al. (1997) employed a two-stage iterative imputation procedure
to replace the censored value by half of their lower detection limits in the first
stage and then by imputing the new estimates, the model is fitted again at the
second stage. Many researcher have demonstrated that these naive approaches did
not adjust standard errors of parameter estimation for the loss of censoring due
to censoring. Therefore, approaches that handle the left censoring longitudinal
data directly have been proposed. Hughes (1999); Jacqmin-Gadda et al. (2000);
Lyles et al. (2000) considered the likelihood based approach to analyze the cen-
sored data. The results indicated that consideration of a detection limit reduces
the bias of estimated parameters in comparison to the traditional naive method.
Hughes (1999) adopted a Monte Carlo EM algorithm to solve the intractable in-
tegrals involved in the modified likelihood that can accommodate left censoring.
The estimation approach proposed by Jacqmin-Gadda et al. (2000) is based on the
Marquardt algorithm. Lyles et al. (2000) implemented the estimation approach
by directly maximizing the likelihood in SAS. Król et al. (2016) take into account left censoring to include the quantification limit of determining the size of tumors when joint modeling the longitudinal data, recurrent events and a terminal event.

Despite known results in the literature for the two-stage joint model methods, our work is still necessary for the following reasons: (1) although in the literature, the two-stage method is known to be less desirable than the joint likelihood method, the results are mostly based on simple results or basic joint models, such as joint linear mixed effects model and a Cox PH model. Therefore, it is less clear if these results still hold for more complex joint models or to what extent these results still hold (e.g., whether the biases become more/less severe for joint longitudinal model and frailty models) (Ye and Wu, 2017); (2) Under our complicated modeling system, e.g., nested cluster effects, left-censored biomarker, joint modeling recurrent events and death, currently available software for joint models is still limited to simple or basic joint models. When the number of random effects is not small, potential computational difficulties such as slow or nonconvergences of EM algorithms for joint likelihood method are very likely to increase. With intractable integral that caused by multiple random effects, the computation for the full likelihood method can be highly demanding. Thus, extends our proposed joint model in 3 to two-stage model by incorporating longitudinal data is still needed for such complex joint models.

In this chapter, due to the matched designs, we propose a two-stage joint frailty modeling for left-censored longitudinal biomarkers, recurrent events and a terminal event of death by accommodating nested frailties accounting for the hierarchical correlation structure. The remainder of this chapter is organized as follows. In Section 4.2, we introduce the joint modeling of left-censored longitudinal biomarker recurrent events and death in a matched cohort study. We provide the theoretical work on parameter estimation via the Monte Carlo Expectation-Maximization (MCEM) algorithm in Section 4.3. Results from simulation studies are presented in Sections 4.4.
4.2 Model

4.2.1 Notations

Let $i$ denote the index of matched pairs and $j$ denote the baseline exposure variable, $i = 1, 2, \ldots, I; j = 1, 2$ with the value of $j = 1$ (non-stroke at baseline) and $j = 2$ (stroke at baseline) in our stroke study. For the $ij^{th}$ subject, let $C_{ij}$, $M_{ij}$, and $D_{ij}$ be the follow-up time, drop-out time and the death time respectively. Let $T_{ij} = \min(D_{ij}, C_{ij}, M_{ij})$ be the observed follow-up time with $\Delta_{ij} = I(D_{ij} \leq \min(C_{ij}, M_{ij}))$ as the death indicator, where $I(\cdot)$ is the indicator function. Denote $\Psi_{ij}(t) = I(T_{ij} \geq t)$ as the “at-risk” indicator to show whether the subject is still under observation at time $t$ or not. Define $N_{ij}^D(t) = I(D_{ij} \leq t)$ as the observed death process, and $N_{ij}^R(t) = N\{min(T_{ij}, t)\}$ as the actual death process by time point $t$. In addition, denote $N_{ij}^{R*}(t) = N\{\min(D_{ij}, t)\}$ and $N_{ij}^{R*}(\min(T_{ij}, t))$ as the actual and observed number of recurrent events by time point $t$ separately. $N_{ij}^D$ and $N_{ij}^R$ are both the observed parts of the counting processes $N_{ij}^{R*}$ and $N_{ij}^{R*}$. The number of recurrent events that occur for the $ij^{th}$ subject over the small interval $[t, t + dt]$ is $dN_{ij}^{R*}(t) = N_{ij}^{R*}((t + dt)^-) - N_{ij}^{R*}(t^-)$ as $dt \to 0$ and $dN_{ij}^{R} = \Psi_{ij}(t) dN_{ij}^{R*}$. $R_{ij} = I(N_{ij}^R > 0)$ denotes the observed indicator for whether the $ij^{th}$ subject has at least one recurrent stroke during the follow-up. Let $t_{ijk}$ be the $k^{th}$ recurrent event time and $\delta_{ijk}$ denote the indicator of recurrent events at time $t_{ijk}$, $k = 1, 2, \ldots, N_{ij}^R$.

For subject $ij$, the number of biomarker measurements are not equal due to the missing values. The repeated measurements $y_{ij} = \{y_{ij}(t_{ij1}), y_{ij}(t_{ij2}), \ldots, y_{ij}(t_{ijm}), \ldots, y_{ij}(t_{ijn})\}^T$, $m = 1, \ldots, n_{ij}$. Measurements $y_{ij}(t_{ijm})$ and recurrent events $t_{ijk}$ can be possibly observed at the same moments, for example, patient visits. Neither $y_{ij}(t_{ijm})$ and $t_{ijk}$ can be observed after $T_{ij}$. The right-censoring does not interrupt the processes and they are simply no longer observed.
4.2.2 Longitudinal measurements

The longitudinal response $y_{ij}$ is usually modeled using a linear mixed effects model:

$$y_{ij} = m_{ij} + \epsilon_{ij} = X_{ij}^{(L)} \beta_L + Z_{ij} b_{ij} + \epsilon_{ij}$$

where $X_{ij}^{(L)} = (X_{ij1}, \ldots, X_{ijp})$ are covariates and $\beta_L$ is the $p \times 1$ vector for fixed-effects parameter. The covariates for random effects are $Z_{ij} = (Z_{ij1}, \ldots, Z_{ijp})$, which could be a subset of $X_{ij}^{(L)}$ including unit component. $b_{ij}$ is the vector of random effects for subject $ij$ with $b_{ij} \sim N(0, \Sigma_b)$. The error vector $\epsilon_{ij} = (\epsilon_{ij}(t_{ij1}), \ldots, \epsilon_{ij}(t_{ijn}))^T$ is usually specified as $\epsilon_{ij} \sim N_{n_{ij}}(0, \sigma^2_{\epsilon} I_{n_{ij}})$, where $I_{n_{ij}}$ represents the identity matrix of dimension $n_{ij}$, $\text{cov}(\epsilon_{ij}(t), \epsilon_{ij}(t')) = 0$, for $t \neq t'$. Note that $b_{ij}$ is independent with $\epsilon_{ij}$. Therefore, $\text{var}(y_{ij}) = V_{ij} = Z_{ij} \Sigma_b Z_{ij}^T + \sigma^2_{\epsilon} I_{n_{ij}}$. The observed longitudinal outcome $y_{ij}$ is represented by its true value $m_{ij}$ and measurement error $\epsilon_{ij}$.

In many joint models, the longitudinal component uses random effects with functions of time only (Tsiatis and Davidian, 2004). The covariate $Z_{ij}$ usually includes only random intercept and random slope effects, or at most a random quadratic effect of time. Here, we specify $b_{ij} = (b_{ij0}, b_{ij1})$, where $b_{ij0}$ is the random intercept and $b_{ij1}$ is the random slope. Then for subject $ij$, the mixed effects model is:

$$y_{ij} = m_{ij} + \epsilon_{ij} = X_{ij}^{(L)} \beta_L + 1 b_{ij0} + t_{ij} b_{ij1} + \epsilon_{ij} \quad (4.1)$$

where $1 = (1, \ldots, 1)^T$ is a $n_{ij}$ unit vector, $t_{ij} = (t_1, t_2, \ldots, t_{n_{ij}})^T$.

4.2.2.1 Left censored longitudinal measurements

As we mentioned in the Section 4.1, longitudinal or grouped data could be complicated by left-censoring of some measures because of a detection limit of the assay used to quantify the biomarker.

Assuming there are $n^c_{ij}$-vector of censored observations, let $c_{ij}$ be the censoring threshold for subject $ij$. Let $y^o_{ij}$ be the $n^c_{ij}$-vector of completely observed outcomes
for subject $ij$, and $y^c_{ij}$ the $n^c_{ij}$-vector of censored observations ($n_{ij} = n^o_{ij} + n^c_{ij}$).

Specifically, we define the observed and censored data:

$$y^o_{ij} = \begin{cases} y_{ij} & \text{if } y_{ij} > c_{ij} \\ c_{ij} & \text{if } y_{ij} \leq c_{ij} \end{cases}$$

and the censored data $y^c_{ij} = (y_{ij} | y_{ij} \leq c_{ij})$.

In Section 4.1, we have reviewed the estimation approaches to handle left-censoring longitudinal data. We adopted likelihood based approach for inference, which provides more accurate parameter estimates and standard errors. First we partitioned the following matrices:

$$X^{(L)}_{ij} \begin{bmatrix} X^{(L) o}_{ij} \\ X^{(L) c}_{ij} \end{bmatrix}, \ y_{ij} \begin{bmatrix} y^o_{ij} \\ y^c_{ij} \end{bmatrix}, \ V_{ij} = \begin{bmatrix} V^o_{ij} & V^{coT}_{ij} \\ V^{co}_{ij} & V^c_{ij} \end{bmatrix}$$

Using the above partition models, according to model 4.1, $y^o_{ij}$ has a multivariate Gaussian probability density function $f^o(y^o_{ij})$. Let $\Omega$ denote the parameter space. Then, the likelihood function of the censored data takes the form:

$$L(\Omega) = \prod_{i=1}^I \prod_{j=1}^J f^o(y^o_{ij} | \Omega)P(y^c_{ij} \leq c_{ij} | Y^o_{ij}, \Omega)$$

Following the properties of multivariate normal distribution, the conditional distribution of $Y^c_{ij}$ given $Y^o_{ij}$ is normally distributed with expectation $\mu^{c|o}_{ij}$ and variance $V^{c|o}_{ij}$ given by the following expressions respectively:

$$\mu^{c|o}_{ij} = X^{(L) o}_{ij} \beta + V^{co}_{ij} V^{-1}^o_{ij} \left[ y^o_{ij} - \mu^o_{ij} \right]$$

$$V^{c|o}_{ij} = V^c_{ij} - V^{co}_{ij} V^{-1}^o_{ij} V^{coT}_{ij}$$

Then the likelihood can be rewritten as:

$$L(\Omega) = \prod_{i=1}^I \prod_{j=1}^J f^o(y^o_{ij} | \Omega)F^{c|o}_{ij}(c_{ij} | \Omega)$$
where \( F_{ij}^{c|o} \) denote the multivariate normal distribution function for \( Y_{ij}^c \) given \( Y_{ij}^o \). The likelihood can be further simplified by the density and cumulative distribution of multivariate normal distribution:

\[
L(\Omega) = \prod_{i=1}^{I} \prod_{j=1}^{J} \frac{1}{\sqrt{(2\pi)^{n_i} |V_{ij}^o|^{1/2}}} \exp \left[ -\frac{1}{2} (y_{ij}^o - \mu_{ij}^o)^T V_{ij}^o^{-1} (y_{ij}^o - \mu_{ij}^o) \right] - \\
\int_{-\infty}^{c_{i1}} \int_{-\infty}^{c_{i2}} \cdots \int_{-\infty}^{c_{in_i}} \frac{1}{\sqrt{(2\pi)^{n_i} |V_{ij}^c|^{1/2}}} \exp \left[ -\frac{1}{2} (\mu - \mu_{ij}^c)^T V_{ij}^c|o^{-1} (\mu - \mu_{ij}^c) \right] d\mu
\]

(4.4)

\[4.2.3\] The joint model of recurrent events and death

Let \( \mu_i \) denote the frailty measuring the dependency between stroke and non-stroke subjects within a matched pair. Denote \( \omega_{ij} \) as the frailty measuring the dependency among the recurrent acute ischemic stroke events within the \( ij \)th individual. \( \mu_i \) and \( \omega_{ij} \) are shared by the recurrent and terminal events to induce their dependency. Assume that \((\mu_i, \omega_{ij})^T \sim MVN(0, \Sigma)\). For simplicity, we can assume that \( \mu_i \perp \omega_{ij} \), where \( \mu_i \sim N(0, \sigma^2_\mu) \) and \( \omega_{ij} \sim N(0, \sigma^2_\omega) \). If \( \sigma^2_\mu \) and \( \sigma^2_\omega \) equal 0, then it implies there is no dependency between recurrent events and death, and the heterogeneity in both event processes is solely explained by covariates \( X_{ij}^{(R)} \) and \( X_{ij}^{(D)} \).

The model for recurrent events is defined as follows:

\[
\lambda_{ij}^{R}(t \mid X_{ij}^{(R)}, \mu_i, \omega_{ij}) = \lambda_0^R(t) \exp(\beta_{ij}^{T}X_{ij}^{(R)} + \mu_i + \omega_{ij})
\]

(4.5)

where \( \beta_{ij} \) is the vector of regression coefficients, \( \lambda_0^R(t) \) is the baseline intensity function. \( X_{ij}^{(R)} \) denotes the vector of the observed covariates including baseline risk factors, such as gender, race and smoking status and so on. Currently, we focus on time-independent covariates.

Conditional on \( \mu_i \) and \( \omega_{ij} \), the recurrent event process and death are independent of each other. Thus, the association between death and recurrent events is quantified by the shared frailties \( \mu_i \) and \( \omega_{ij} \) through \( \phi_{\mu} \) and \( \phi_{\omega} \) in equation (5.7).
Consequently, a higher intensity of recurrent events is associated with a higher mortality rate. Note that by extending the work from Liu et al. (2015), the correlation within matched pairs also is incorporated through \( \mu_i \). Thus, the frailty proportional hazard model for death is

\[
\lambda_{ij}^D(t \mid X_{ij}^{(D)}, \mu_i, \omega_{ij}) = \lambda_0^D(t) \exp(\beta_T^D X_{ij}^{(D)} + \phi_\mu\mu_i + \phi_\omega\omega_{ij}) \tag{4.6}
\]

where \( \beta_D \) is vector of regression coefficients, and \( \lambda_0^D(t) \) is the baseline hazard function of death. \( X_{ij}^{(D)} \) denotes the vector of the observed covariates in the survival model for death. Of note, \( X_{ij}^{(R)} \) and \( X_{ij}^{(D)} \) are assumed to be the same for \( \lambda_{ij}^R \) and \( \lambda_{ij}^D \) for simplicity, but can be different in terms of clinical perspectives for real data applications.

### 4.2.4 Two-stage joint model

We jointly model the following process:

\[
y_{ij}(t) = m_{ij}(t) + \epsilon_{ij}(t) = \beta_T^L X_{ij}^{(L)}(t) + b_{ij0} + b_{ij1}t + \epsilon_{ij}(t) \tag{4.7}
\]

\[
\lambda_{ij}^R(t \mid X_{ij}^{(R)}, X_{ij}^{(L)}, \mu_i, \omega_{ij}, b_{ij}) = \lambda_0^R(t) \exp(\beta_R^T X_{ij}^{(R)} + \mu_i + \omega_{ij} + \eta_R^T h_R^L(X_{ij}^{(L)}, b_{ij}, t)) \tag{4.8}
\]

\[
\lambda_{ij}^D(t \mid X_{ij}^{(D)}, X_{ij}^{(L)}, \mu_i, \omega_{ij}, b_{ij}) = \lambda_0^D(t) \exp(\beta_D^T X_{ij}^{(D)} + \phi_\mu\mu_i + \phi_\omega\omega_{ij} + 
\eta_D^T h_D^L(X_{ij}^{(L)}, b_{ij}, t)) \tag{4.9}
\]

We explore the nature of the dependency between the longitudinal repeated measurements, the risk of recurrence and death by considering the link functions \( h_R^L(X_{ij}^{(L)}, b_{ij}, t) \) and \( h_D^L(X_{ij}^{(L)}, b_{ij}, t) \). The form that \( h_R^L(X_{ij}^{(L)}, b_{ij}, t) \) and \( h_D^L(X_{ij}^{(L)}, b_{ij}, t) \) take determine the type of joint model that are fit. \( \eta_R \) and \( \eta_D \) are the vector of parameters that links the longitudinal process and survival process and determine the association’s strength. Different approaches can be adopted to construct the trajectory function, e.g. include mean response which composed by the fixed effects model only, or use the full longitudinal trajectory with both fixed effects and random effect components (Chen et al., 2014). \( X_{ij}^{(L)}(t), X_{ij}^{(R)} \) and \( X_{ij}^{(D)} \) can be the same or different in the three processes.
We adopt the full longitudinal trajectory as the link function, which indicates the current true level of the repeated measurements is predictive of the risk of recurrent events and death. This joint model is a kind of two stage model. A simple naive two-stage method is as follows (Wu et al., 2011).

1. Fit the mixed effects longitudinal model, and obtain an estimated trajectory of the biomarker for each subject during the follow-up.

2. Fit the survival model separately, with the missing or unobserved true covariate values substituted by their estimates from the first stage as if they were observed values and then proceed with the usual survival analysis.

In this model, we assume

\[ h^R(X^{(L)}_{ij}, b_{ij}, t) = \hat{m}^R_{ij}(t) \]

\[ h^D(X^{(L)}_{ij}, b_{ij}, t) = \hat{m}^D_{ij}(t) \]

\( \eta_R \) and \( \eta_D \) are the scalar parameters. Here the link function \( \hat{m}^R_{ij}(t) \) and \( \hat{m}^D_{ij}(t) \) are the fitted values from 5.5 which can be the same or different.

First, we fit the longitudinal model with marginal likelihood:

\[
\ell(\beta_l, \Sigma_b, \sigma^2_e) = \prod_{i=1}^{I} \prod_{j=1}^{J} \int_{b_{ij}} \left\{ \prod_{m=1}^{n_{ij}} f(y_{ij}(t_{ijm}) \mid b_{ij}, \sigma^2_e) \right\} f(b_{ij} \mid \Sigma_b) db_{ij} \tag{4.10} \]

where

\[
f(y_{ij}(t_{ijm}) \mid b_{ij}, \sigma^2_e) = \frac{1}{(2\pi \sigma^2_e)^{1/2}} \exp\left\{ -\frac{1}{2} \left( (y_{ij}(t_{ijm}) - \beta^T L X^{(L)}_{ij}(t_{ijm}) - b_{ij0} - b_{ij1} t_{ijm})^2 / \sigma^2_e \right) \right\}
\]

\[
f(b_{ij} \mid \Sigma_b) = \frac{1}{(2\pi |\Sigma_b|)^{1/2}} \exp\left\{ -\frac{1}{2} b_{ij}^T \Sigma_b^{-1} b_{ij} / 2 \right\}
\]

Second, we incorporate the fitted \( \hat{m}_{t_{ij}} \) in the joint model of recurrent events and death event. Given \( \theta = (\lambda^R_0(\cdot), \lambda^D_0(\cdot), \beta_R, \beta_D, \phi_\mu, \phi_\omega, \eta_R, \eta_D, \sigma^2_\omega, \sigma^2_\mu)^T \), the marginal
likelihood is

\[
L(\theta) = \int \int \prod_{i=1}^{l} \prod_{j=1}^{J} \int L(\theta \mid \mu_i, \omega_{ij}) f(\mu_i) f(\omega_{ij}) d\mu_i d\omega_{ij} \\
= \prod_{i=1}^{l} \int \left\{ \prod_{j=1}^{J} L_R^{i,j} f(\omega_{ij}) d\omega_{ij} \right\} f(\mu_i) d\mu_i
\]

(4.11)

with

\[
L_R^{i,j} = \prod_k \left[ \lambda_R^0(t) \exp(\beta_R^T X_{ij}^{(R)}) + \mu_i + \omega_{ij} + \eta_R \hat{m}_{ij}(t) \right]^{\delta_{ijk}} \cdot \exp\left[ - \int_0^\infty \psi_{ij}(t) \lambda_R^0(t) \exp(\beta_R^T X_{ij}^{(R)}) + \mu_i + \omega_{ij} + \eta_R \hat{m}_{ij}(t) dt \right]
\]

\[
L_D^{i,j} = \left[ (\lambda_D^0(T_{ij}) \exp(\beta_D^T X_{ij}^{(D)}) + \phi_\mu \mu_i + \phi_\omega \omega_{ij} + \eta_D \hat{m}_{ij}(t)) \right]^{\Delta_{ij}} \cdot \exp\left[ - \int_0^\infty \psi_{ij}(t) \lambda_D^0(t) \exp(\beta_D^T X_{ij}^{(D)}) + \phi_\mu \mu_i + \phi_\omega \omega_{ij} + \eta_D \hat{m}_{ij}(t) dt \right]
\]

\(L_R^{i,j}\) is the likelihood of observing at least 1 recurrent event \((R_{ij} = 1)\). \(L_D^{i,j}\) is the likelihood for terminal event death. \(f(\mu_i)\) and \(f(\omega_{ij})\) are normal densities. In addition, the cumulative baseline intensities of recurrent events and terminal event denoted by \(\Lambda_R^0(t)\) and \(\Lambda_D^0(t)\) are given by \(\Lambda_R^0(t) = \int_0^\infty \psi_{ij}(t) \lambda_R^0(t) dt\) and \(\Lambda_D^0(t) = \int_0^\infty \psi_{ij}(t) \lambda_D^0(t) dt\).

### 4.3 Estimation

Similar to Chapter 3, in this Chapter, we adopt the Monte Carlo EM (MCEM) algorithm for estimation, which is described next. In the E step, we find the expectation of the conditional log-likelihood, and integrate out the frailties by numerical integration with the Metropolis-Hasting algorithm.
4.3.1 Estimation of longitudinal model

First we fit the longitudinal model 4.1. Usually, the estimation of the parameters of linear mixed effects model is often based on maximum likelihood (ML) or restricted maximum likelihood (REML) principles. Since our sample size is large, we adopt the maximum likelihood principles for estimation. The marginal density of the observed biomarker data for the \( ij \)th subject is:

\[
f(y_{ij} | b_{ij}, \sigma_e^2) = \int_{b_{ij}} \{ \prod_{m=1}^{n_{ij}} f(y_{ij}(t_{ijm}) | b_{ij}, \sigma_e^2) \} f(b_{ij} | \Sigma_b) \, db_{ij} \quad (4.12)
\]

Since a convolution of normal is normal, take advantage of the fact that above integral has a closed-form solution, and lead to an \( n_i \)-dimensional normal distribution

\[
Y | \Sigma_b, \sigma_e^2 \sim \prod_{i=1}^{I} \prod_{j=1}^{J} N(X_{ij}^{(L)} \beta_L, V_{ij})
\]

where \( \text{var}(y_{ij}) = V_{ij} = Z_{ij} \Sigma_b Z_{ij}^T + \sigma_e^2 I_{n_{ij}} \). Given \( \theta_L = (\beta_l, \Sigma_b, \sigma_e^2) \), then the log likelihood is

\[
l(\theta_L) = \frac{1}{2} \sum_{i=1}^{I} \sum_{j=1}^{J} n_{ij} \log(2\pi) - \frac{1}{2} \sum_{i=1}^{I} \sum_{j=1}^{J} \log|V_{ij}| - \frac{1}{2} \sum_{i=1}^{I} \sum_{j=1}^{J} (y_{ij} - X_{ij}^{(L)} \beta_L)^T V_{ij}^{-1} (y_{ij} - X_{ij}^{(L)} \beta_L) \quad (4.13)
\]

The score function for \( \beta_L \) is

\[
\frac{\partial l(\theta_L)}{\beta_L} = \sum_{i=1}^{I} \sum_{j=1}^{J} X_{ij}^{(L)^T} V_{ij}^{-1} y_{ij} - \sum_{i=1}^{I} \sum_{j=1}^{J} X_{ij}^{(L)^T} V_{ij}^{-1} X_{ij}^{(L)} \beta_L
\]

and yields the MLE for \( \beta_L \) as

\[
\hat{\beta}_L = \left( \sum_{i=1}^{I} \sum_{j=1}^{J} X_{ij}^{(L)^T} V_{ij}^{-1} X_{ij}^{(L)} \right)^{-1} \left( \sum_{i=1}^{I} \sum_{j=1}^{J} X_{ij}^{(L)^T} V_{ij}^{-1} y_{ij} \right) \quad (4.14)
\]

The variance of \( \hat{\beta}_L \) can be obtained either directly from 4.14 or from the second
derivative of the log likelihood. Since
\[
\frac{\partial^2 l(\theta_L)}{\beta_L \beta_L^T} = -\sum_{i=1}^{I} \sum_{j=1}^{J} X_{ij}^{(L)T} V_{ij}^{-1} X_{ij}^{(L)}
\]

The variance of \( \hat{\beta}_L \) is
\[
\text{var}(\hat{\beta}_L) = -\text{E}\left[ \frac{\partial^2 l(\theta_L)}{\beta_L \beta_L^T} \right] = \sum_{i=1}^{I} \sum_{j=1}^{J} X_{ij}^{(L)T} V_{ij}^{-1} X_{ij}^{(L)} \quad (4.15)
\]

The MLE of \( \Sigma_b, \sigma_e^2 \) from 4.13 in general, has no closed-form solution. Therefore, the EM algorithm and Newton-Raphson algorithms are frequently used, whose implementation for linear mixed effects models can be found in Laird and Ware (1982) and Lindstrom and Bates (1988). We only provide the outline of the estimation. The “missing data” in EM algorithm are the random effects \( b_{ij} \) and the errors \( \epsilon_{ij} \).

Given \( k \)th estimate \( \hat{\theta}_L^{(k)} \), in E step, estimate the expectation of sufficient statistics. In M step, given \( b_{ij} \) and \( \epsilon_{ij} \), obtain the estimates:
\[
\hat{\Sigma}_b = \frac{\text{E}(\sum_{i=1}^{I} \sum_{j=1}^{J} b_{ij}^T b_{ij} | \hat{\theta}_L^{(k)})}{IJ} \quad (4.16)
\]
\[
\hat{\sigma}_e^2 = \frac{\text{E}(\sum_{i=1}^{I} \sum_{j=1}^{J} \epsilon_{ij}^T \epsilon_{ij} | \hat{\theta}_L^{(k)})}{\sum_{i=1}^{I} \sum_{j=1}^{J} n_{ij}} \quad (4.17)
\]

Then based on the MLE \( \hat{\theta}_L \), the predictor of random effect \( b \) is
\[
\hat{b}_{ij} = \hat{\Sigma}_b Z_{ij}^T V_{ij}^{-1}(y_{ij} - X_{ij}^{(L)} \hat{\beta}_L) \quad (4.18)
\]

Therefore, the fitted value for \( y_{ij} \) is
\[
\hat{y}_{ij} = \hat{m}_{ij} = X_{ij}^{(L)} \hat{\beta}_L + Z_{ij} \hat{b}_{ij} \quad (4.19)
\]

To considering the left-censoring biomarker, we can simplify the expression of 4.2. Given the left censoring indicator \( d_{ij}(t) \) (1=Yes; 0=No), the complete likelihood
function is: If the biomarker is left-censored, then let \( L \) denote the lower detection limit, and we can have the cumulative distribution function for the left-censored measurement at time \( t \):

\[
F(L | b_{ij}) = \int_{-\infty}^{L} f(y_{ij}(t) | b_{ij}) \, dy_{ij}(t) \tag{4.20}
\]

For each subject, given the left censoring indicator \( d_{ij}(t) \) (1=Yes; 0=No), the complete likelihood function is:

\[
L_{ij}(\theta) = \int \prod_{m=1}^{n_{ij}} f(y_{ij}(t_{ijm}) | b_{ij})^{I(d_{ij}(t_{ijm})=0)} F(L | b_{ij})^{I(d_{ij}(t_{ijm})=1)} f(b_{ij}) \, db_{ij} \tag{4.21}
\]

Adaptive Gaussian quadrature (Thiébaut and Jacqmin-Gadda, 2004), MCEM algorithm (Hughes, 1999; Vaida et al., 2007) can be adopted to solve the integrals and obtain the marginal likelihood for maximization. Here, we adopt the MCEM algorithm proposed by Vaida et al. (2007).

4.3.2 Estimation of the joint model of recurrent events and death

Next, given the fitted value \( \tilde{m}_{ij} \), we estimate 4.11. We adopted similar MCEM algorithm as in Chapter 3 with a few modifications. Since we don’t consider the zero-inflation, the MCEM is adapted with \( p_{ij} = 1 \) in 3.2.

Similar to Chapter 3, we adopt the piecewise baseline intensity function for recurrent events and death. Partition the time duration into \( M \) intervals with cutpoints \( 0 \leq c_0 < c_1 < \ldots < c_M = \infty \), which can be quantiles of event times and where \( c_0 = 0 \) or the smallest event time. The baseline intensity function is assumed to be piecewise constant within each of the \( M \) intervals, so that

\[
\tilde{\lambda}_0(t) = \sum_{m=1}^{M} \lambda_m I(c_{m-1} < t \leq c_m) \tag{4.22}
\]

Denote the baseline intensity of recurrent events is \( \tilde{\lambda}_0^R(t) \), using \( M \) parameters
\[ \lambda_1^R, \lambda_2^R, \ldots, \lambda_m^R. \] The baseline hazard for death is \( \tilde{\lambda}_0^D(t) \), using \( M \) parameters \( \lambda_1^D, \lambda_2^D, \ldots, \lambda_m^D. \) Note that the selection of cutpoint \( c_m \) and \( M \) for recurrent events and death can be the same or different. The detailed estimation of \( \tilde{\lambda}_0^R(t) \) and \( \tilde{\lambda}_0^D(t) \) in provided in Appendix B.1.

For simplicity, we only consider the linear relationship and denote \( \beta_t \) is the fixed coefficient for \( t \) in longitudinal model, which is included in \( \beta_L \). Now, the cumulative baseline intensities of recurrent events is

\[
\tilde{\Lambda}_0^R(t) = \sum_{m=1}^{M} \int_{t_{m-1}^R}^{t_m^R} \lambda_m^R \exp(\eta_R(\hat{\beta}_t + \hat{b}_{1ij})t) dt
\]

\[
= \sum_{m=1}^{M} \frac{\lambda_m^R}{\eta_R(\hat{\beta}_t + \hat{b}_{1ij})} \left[ \exp(\eta_R(\hat{\beta}_t + \hat{b}_{1ij})t_m^R) - \exp(\eta_R(\hat{\beta}_t + \hat{b}_{1ij})t_{m-1}^R) \right]
\]  

(4.23)

Here, \( t_m^R \) and \( t_{m-1}^R \) are upper and lower bound at \( m \)th interval.

\[ t_m^R = \max(0, \min(c_m - c_{m-1}, t - c_{m-1})) \]. And, the cumulative baseline intensities of death event is

\[
\tilde{\Lambda}_0^D(t) = \sum_{m=1}^{M} \int_{t_{m-1}^D}^{t_m^D} \lambda_m^D \exp(\eta_D(\hat{\beta}_t + \hat{b}_{1ij})t) dt
\]

\[
= \sum_{m=1}^{M} \frac{\lambda_m^D}{\eta_D(\hat{\beta}_t + \hat{b}_{1ij})} \left[ \exp(\eta_D(\hat{\beta}_t + \hat{b}_{1ij})t_m^D) - \exp(\eta_D(\hat{\beta}_t + \hat{b}_{1ij})t_{m-1}^D) \right]
\]  

(4.24)

Here, \( t_m^D \) and \( t_{m-1}^D \) are upper and lower bound at \( m \)th interval.

\[ t_m^D = \max(0, \min(c_m - c_{m-1}, t - c_{m-1})) \].

In the E step, we find the expectation of the conditional log-likelihood, and integrate out the frailties by numerical integration with the Metropolis-Hasting algorithm. We define that \( X_{1ij}^{(L)} \) is the covariate vector \( X_{ij}^{(L)} \) that don’t contain the time variable \( t_{ij} \). \( \beta_{L_i} \) is the coefficient vector that exclude \( \beta_t \). Let \( q_{1ij} = \beta_T^{R}X_{ij}^{(R)} + \eta_R(X_{1ij}^{(L)}(t)\hat{\beta}_L + \hat{b}_{0ij}), \) and \( q_{2ij} = \beta_T^{D}X_{ij}^{(D)} + \eta_D(X_{1ij}^{(L)}(t)\hat{\beta}_L + \hat{b}_{0ij}). \) Thus given the values of frailties \( \mu \) and \( \omega \), the complete log likelihood function
denoted by \( l(\theta) \) can be split into three parts:

\[
l_1(\beta_R, \eta_R, \lambda^R_0(t)) = \sum_{i=1}^{I} \sum_{j=1}^{J} \log \left\{ \left( \prod_{k} \left( \lambda^R_0(t_{ijk}) \cdot \exp(\eta_R(\hat{\beta}_t + \hat{b}_{1ij})t_{ijk} + q_{1ij} + \mu_i + \omega_{ij}) \right) \right) \delta_{ijk} \right\} \exp \left( -\lambda^R_0(t) \cdot \exp(q_{1ij} + \mu_i + \omega_{ij}) \right)
\]

\[
(4.25)
\]

\[
l_2(\beta_D, \phi_\mu, \phi_\omega, \eta_D, \lambda^D_0(t)) = \sum_{i=1}^{I} \sum_{j=1}^{J} \log \left\{ \left( \lambda^D_0(T_{ij}) \cdot \exp(\eta_D(\hat{\beta}_t + \hat{b}_{1ij})T_{ij} + q_{2ij} + \phi_\mu \mu_i + \phi_\omega \omega_{ij}) \right) \Delta_{ij} \right\} \exp \left( -\lambda^D_0(t) \exp(q_{2ij} + \phi_\mu \mu_i + \phi_\omega \omega_{ij}) \right)
\]

\[
(4.26)
\]

\[
l_3(\sigma^2_\mu, \sigma^2_\omega) = \sum_{i=1}^{I} \sum_{j=1}^{J} \log \left\{ \frac{1}{\sqrt{2\pi \sigma^2_\mu}} \exp(-\frac{\mu^2_i}{2\sigma^2_\mu}) \frac{1}{\sqrt{2\pi \sigma^2_\omega}} \exp(-\frac{\omega^2_{ij}}{2\sigma^2_\omega}) \right\}
\]

\[
(4.27)
\]

The expectation of the log-likelihood \( l(\theta) \) conditional on the observed data and the current parameter estimate \( \hat{\theta}^{(k)} \) is:

\[
Q(\theta \mid \hat{\theta}^{(k)}) = E[l(\theta) \mid \hat{\theta}^{(k)}] = E[l_1(\beta_R, \eta_R, \lambda^R_0(t)) \mid \hat{\theta}^{(k)}] + E[l_2(\beta_D, \phi_\mu, \phi_\omega, \eta_D, \lambda^D_0(t)) \mid \hat{\theta}^{(k)}] +
E[l_3(\sigma^2_\mu, \sigma^2_\omega) \mid \hat{\theta}^{(k)}]
\]

\[
= Q_1(\beta_R, \eta_R, \lambda^R_0(t) \mid \hat{\theta}^{(k)}) + Q_2(\beta_D, \phi_\mu, \phi_\omega, \eta_D, \lambda^D_0(t) \mid \hat{\theta}^{(k)}) + Q_3(\sigma^2_\mu, \sigma^2_\omega \mid \hat{\theta}^{(k)})
\]

\[
(4.28)
\]

where

\[
Q_1(\beta_R, \eta_R, \lambda^R_0(t) \mid \hat{\theta}^{(k)}) = \sum_{i=1}^{I} \sum_{j=1}^{J} \left\{ \sum_{k=1}^{N_{ij}^R} \left( \log(\lambda^R_0(t_{ijk})) + \eta_R(\hat{\beta}_t + \hat{b}_{1ij})t_{ijk} + q_{1ij} + E[\mu_i \mid \hat{\theta}^{(k)}] \right)
\right\} + E[\omega_{ij} \mid \hat{\theta}^{(k)}] - \lambda^R_0(T_{ij}) \exp(q_{1ij})E[\exp(\mu_i + \omega_{ij}) \mid \hat{\theta}^{(k)}]
\]
\[ Q_2(\beta_D, \phi, \phi, \lambda_0^D (t) \mid \hat{\theta}^{(k)}) \]
\[ = \sum_{i=1}^{I} \sum_{j=1}^{J} \left\{ \Delta_{ij} \left( \log(\tilde{\lambda}_0^D (T_{ij})) + \eta_D (\tilde{\beta}_t + \tilde{b}_{1ij})T_{ij} + q_{2ij} + \phi_E [\mu_i \mid \hat{\theta}^{(k)}] \right. \\
+ \phi_E E[\omega_{ij} \mid \hat{\theta}^{(k)}] \right\} - \tilde{\lambda}_0^D (T_{ij}) \exp(q_{2ij})E[\exp(\phi_E \mu_i + \phi_E \omega_{ij}) \mid \hat{\theta}^{(k)}] \right\} \]
\[ Q_3(\sigma_\mu^2, \sigma_\omega^2 \mid \hat{\theta}^{(k)}) = \sum_{i=1}^{I} \sum_{j=1}^{J} \left\{ - \frac{1}{2} \left( \log(2\pi) + \log \sigma_\omega^2 \right) + \frac{E[\omega_{ij}^2 \mid \hat{\theta}^{(k)}]}{\sigma_\omega^2} \right\} \]
\[ - \frac{1}{2} \left( \log(2\pi) + \log \sigma_\mu^2 \right) + \frac{E[\mu_i^2 \mid \hat{\theta}^{(k)}]}{\sigma_\mu^2} \right\} \}

The computation of the conditional expectation above are given in Appendix A.1.

Similar to the algorithm in Chapter 3, \[ p(\mu, \omega \mid \hat{\theta}^{(k)}) \propto l(\theta \mid \hat{\theta}^{(k)}) \]. Since there is no closed-form expression of \[ p(\mu, \omega \mid \hat{\theta}^{(k)}) \], we adopted the Monte Carlo methods in combination with the Metropolis-Hasting algorithm are used to approximate the posterior distributions of \[ \mu_i \]’s and \[ \omega_{ij} \]’s. Given \[ \hat{\theta}^{(k)} \], \[ N \] random samples are generated for \[ \mu_i^{(m)} \] (\[ m = 1, \ldots, M \]) and \[ \omega_{ij}^{(m)} \] (\[ m = 1, \ldots, M \]) to estimate the expectation of the sufficient statistics involving frailties. Thus, \[ E[f(\mu_i) \mid \hat{\theta}^{(k)}] = \frac{1}{N} \sum_{n=1}^{N} f(\mu_i^{(n)}) \] and \[ E[f(\omega_{ij}) \mid \hat{\theta}^{(k)}] = \frac{1}{N} \sum_{n=1}^{N} f(\omega_{ij}^{(n)}) \], where \[ f(\cdot) \] could be any smooth and monotone function. A brief introduction to the Metropolis-Hasting algorithm is given in B.3.

In the M-step, we use a Newton-Raphson procedure to maximize
\[ Q_1(\beta_R, \eta_R, \lambda_0^R (t) \mid \hat{\theta}^{(k)}), Q_2(\beta_D, \phi, \phi, \eta_D, \lambda_0^D (t) \mid \hat{\theta}^{(k)}), Q_3(\sigma_\mu^2, \sigma_\omega^2 \mid \hat{\theta}^{(k)}) \] to estimate \[ \theta \]. For the parameter estimation, all the score components and second partial derivatives are given in ??.

The variance of the MLE \[ \hat{\theta} = (\hat{\lambda}_0^R (t), \hat{\lambda}_0^D (t), \hat{\beta}_R, \hat{\beta}_D, \hat{\phi}, \hat{\phi}_R, \hat{\eta}_R, \hat{\eta}_D, \hat{\sigma}_\mu^2, \hat{\sigma}_\omega^2)^T \] in the MCEM algorithm cannot be obtained directly from the algorithm. Louis’s formula was used to obtain it Louis (1982). The variance-covariance matrix can be estimated using the inverse of an observed information matrix. The observed information matrix \[ I(\hat{\theta}) \] is given by
\[
I(\hat{\theta}) = -E \left[ \frac{\partial^2 l(\theta)}{\partial \theta \partial \theta^T} \mid \hat{\theta} \right] - \text{Var} \left[ \frac{\partial l(\theta)}{\partial \theta} \mid \hat{\theta} \right]
\]
(4.29)
where \( \text{Var} \left[ \frac{\partial l(\theta)}{\partial \theta} \mid \hat{\theta} \right] = E \left[ \frac{\partial l(\theta)}{\partial \theta} \frac{\partial l(\theta)}{\partial \theta^T} \mid \hat{\theta} \right] \). The last term becomes zero due to the MLE \( \hat{\theta} \) and all of these terms are evaluated at the last iteration of the EM algorithm.

### 4.4 Simulation studies

In this section, we conduct simulation studies to evaluate the performance of the proposed two stage joint frailty model considering longitudinal data with left-censoring. We compare the performances of different models based on the biases of the estimates, and the coverage rates of the confidence intervals under several scenarios. First, we give a brief description about the data simulation procedure. Then, we compare the results under different simulation scenarios and draw conclusions.

#### 4.4.1 Data generation

The most common parametric choices for time-to-event simulation are Exponential, Weibull and Gompertz distributions. As noted by (Bender et al., 2005), only these three distributions share the assumption of proportional hazards. Austin (2012) described data generating processes for the Cox proportional hazards model with time-varying covariates. Denote the time-varying covariate by \( z(t) \), while other covariates, \( x \) are time-invariant. Assume the logarithmic link function is used to relate the hazard function to the linear predictor: \( h(t \mid x(t)) = h_0(t) \exp(\beta_t z(t) + \beta^Tx) \). Then, the cumulative hazard function is \( H(t, x, z(t)) = \int_0^t \exp(\beta_t z(u) + \beta^Tx)h_0(u)du \).

In our model, the time-varying covariate \( z(t) \) is a biomarker which is a continuous variable. Then we assume that \( z(t) \) is proportional to \( t \): \( z(t) = kt \), with \( k > 0 \). If survival times follow an exponential distribution, \( h_0(t) = \lambda \), an event time can be generated as

\[
T = \frac{1}{\beta_t k} \log(1 + \frac{\beta_t k(- \log(u))}{\lambda \exp(\beta^Tx)})
\]
If the survival times follow a Weibull distribution $h_0(t) = \lambda \nu t^{\nu - 1}$,

$$T = \left[ \frac{1}{k} \log(1 + \frac{(1 + \nu)(-\log(u))}{\beta t \exp(\beta^T x) \lambda \nu}) \right]^{1/(1+\nu)}$$

where $u \sim \text{Unif}[0, 1]$

The recurrent event time in our time is specified in a calendar timescale. For a vector of external covariates $X$, the common multiplicative proportional intensity $\lambda(t \mid X = x) = \lambda_0(t) \exp(\beta^T x)$. Non-homogeneous Poisson process is adopted to generate the recurrent events time. Inversion and thinning methods are common approaches for recurrent event times simulation, and here we adopt the inversion method.

Referring to the inversion method by Pénichoux et al. (2015), for a given subject, let $T_j$ denote the time elapsed from the origin to the $j^{th}$ event, $T_0 = 0$, and let $W_j = T_j - T_{j-1}$ denote the $j^{th}$ waiting time between two consecutive events. The $W_j$ are referred to as gap times. The number of events observed for a non-homogeneous Poisson process with intensity $\lambda(\cdot)$ between two times, s and $t$, follows a Poisson distribution with parameter $\int_s^t \lambda(u) du$. Then

$$N(T_{j-1} + w) - N(T_{j-1}) \sim P\left( \int_{T_{j-1}}^{T_{j-1} + w} \lambda(u) du \right)$$

, which leads to

$$F_j(w) = 1 - P(W_j > w) = 1 - \exp\left( - \int_{T_{j-1}}^{T_{j-1} + w} \lambda(u) du \right)$$

Therefore, for $j^{th}$ event, $W_j = F_j^{-1}(u_j)$ and the event time $T_j = T_{j-1} + W_j$.

Exponential and Weibull intensity are common choices. For example, exponential distribution with constant baseline constant $\lambda_0(t) = \lambda$, the recurrent event process can be considered equivalently as an homogeneous Poisson process or a sequence of exponentially distributed gaps. The event times are generated by $T = -\log(u) / \lambda \exp(-\beta^T x)$,

$$T_j = T_{j-1} + W_j = -\log(u_j) / \lambda \exp(-\beta^T x) + T_{j-1}$$

where $u_j \sim \text{Unif}[0, 1]$.

With time-varying covariate $z(t)$, $\lambda(t \mid x(t)) = \lambda_0(t) \exp(\beta_t z(t) + \beta^T x)$. We adopted
the exponential intensity $\lambda_0(t) = \lambda$, in which $F_j(\cdot)$ is invertible. The cumulative intensity function

$$\Lambda_j(w) = \frac{\lambda \exp(\beta^T x)}{\beta_t k} \left( \exp(\beta_t k(T_{j-1} + w)) - \exp(\beta_t k T_{j-1}) \right)$$

, and $F_j(w) = 1 - \Lambda_j(w)$. Then the $j^{th}$ recurrent event times are generated by

$$T_j = T_{j-1} + W_j = \frac{1}{\beta_t k} \log \left( \frac{-\log(u_j) \beta_t k}{\lambda \exp(\beta^T x)} + \exp(\beta_t k T_{j-1}) \right)$$

where $u_j \sim \text{Unif}[0, 1]$.

### 4.4.2 Simulation results

For simplicity, we consider a continuous covariate of $X_1$, which are randomly sampled from the uniform distribution on $[0, 1]$ and binary variable $X_2$, which takes values of 0 or 1 with probability 0.5. And also, only the intercept term for the logistic regression model. Considering the exponential distributions for both recurrent event times and death, the following models are shown below

$$Y_{ij}(t) = m_{ij}(t) + \epsilon_{ij}(t) = \beta_{t0} + \beta_{11} t + \beta_{12} X_{1ij} + b_{ij0} + b_{ij1} t + \epsilon_{ij}(t)$$

$$\lambda^R_{ij}(t \mid \mu_i, \omega_{ij}, Y_{ij} = 1) = \lambda_0^R(t) \exp(\beta_R X_{2ij} + \mu_i + \omega_{ij} + \eta_R \ast \hat{m}_{ij}(t))$$

$$\lambda^D_{ij}(t \mid \mu_i, \omega_{ij}) = \lambda_0^D(t) \exp(\beta_D X_{2ij} + \phi_{i} \mu_i + \phi_{\omega} \omega_{ij} + \eta_D \ast \hat{m}_{ij}(t))$$

(4.30)

The censoring time is taken as $C_{ij} = 6 \ast \text{Unif}(0, 1)$, where Unif(0,1) is a random number generated from uniform distribution in $[0, 1]$. The initial time points for each subject is $n_{ij} = 10$, which is truncated by death and censoring. We assume $\mu_i \sim N(0, \sigma_{\mu}^2)$, $\omega_{ij} \sim N(0, \sigma_{\omega}^2)$ and set $\sigma_{\mu}^2 = 1$, $\sigma_{\omega}^2 = 0.25$. $b_{ij0}$ and $b_{ij1}$ follows multivariate normal distribution with $\sigma_{b0}^2 = 0.1$, $\sigma_{b1}^2 = 0.04$ and $\text{cov}(b_0, b_1) = 0.01$. The error term $\epsilon_{ij}(t) \sim N(0, \sigma_{\epsilon}^2)$ and $\sigma_{\epsilon}^2 = 0.01$. We set the common parameters as $\beta_{t0} = 0.5$, $\beta_{t1} = 1$, $\beta_{t2} = 1$, $\beta_R = 1.2$, $\beta_D = 1, \phi_{i} = 1$, $\phi_{\omega} = 0.5$, $\lambda_0^D = 0.005$, but different values for $\eta_R$ and $\eta_D$, and recurrent event baseline intensity $\lambda_0^R$. 
There are three set-ups: Setting I: \( \eta_R = 0.5, \eta_D = 1 \) and \( \lambda_0^R = 0.1 \): 60% subjects are observed with at least one recurrent event and death percentage is 40%; Setting II: \( \eta_R = 0.4, \eta_D = 0.7 \) and \( \lambda_0^R = 0.065 \): 50% subjects are observed with at least one recurrent event and death percentage is 25%. The major difference between the Setting I and Setting II is the percentage of death (40% and 25%), we recognize these two percentages are reasonable to represent a normal and a low percentage of death to assess the performance of the proposed two stage joint model, and obviously if the percentage is too low, the convergence issue could be challenging. For Setting III, we consider the biomarker with left-censoring. Setting III: \( \eta_R = 0.5, \eta_D = 1 \) and \( \lambda_0^R = 0.1 \): 60% subjects are observed with at least one recurrent event and death percentage is 40%; The percentage of left-censoring longitudinal observations is 30%. A total of 500 Monte Carlo replicates are generated, each with a sample size of 500 subjects. Figure 4.1 is the sample simulated data of Setting I.

We adopted the piecewise constant approach is adopted to estimate the baseline intensities for both recurrent and death event processes, where 5 intervals are considered. The resulting parameter estimates are shown in Table 4.1 and Table 4.2. It can be seen that in both set-ups for the full model, MCEM estimation method yield satisfactory results with relatively small biases. But on the estimation of parameter \( \phi_\omega \), the proposed two-stage model result in quite large bias. Comparing the parameter estimates between settings I and II, the empirical biases tend to increase when the percentage of recurrent events and death decreases from 60% to 50% and 40% to 25%. For comparisons with miss-specified models, if we don’t consider the correlation between longitudinal biomarker and survival data (remove \( \eta_R \) and \( \eta_D \)), the estimation of \( \sigma_\omega^2 \) and \( \phi_\omega \) are overestimated. The correlation effect that was induced by longitudinal biomarkers is accounted for by recurrent events.

Next, we estimate the data from Setting III with considering the left-censored longitudinal biomarker to compare the naive approach and maximum likelihood approach with below-detection-limits data. The resulting parameter estimates are shown in Table 4.3. It is clear to see that with quite high percentage of left-censored longitudinal outcomes (30%), the naive approach (Model I) produce estimates with significant biases. Hughes (1999) have shown that the bias increases
as the proportion of the data is censored increases. Compared to the full model, the estimates of the joint model of recurrent events and death are also biased, especially on the estimation of parameters that account for the correlation among the three processes, such as $\phi_\omega$, $\eta_R$ and $\eta_D$. 
Figure 4.1. Sample simulated data of longitudinal biomarker, recurrent events and death
Table 4.1. Simulation results of two stage model Setting I: $\eta_R = 0.5$, $\eta_D = 1$ and $\lambda_0^R = 0.1$: 60% subjects are observed with at least one recurrent event and death percentage is 40%. Full model denotes the proposed joint model of biomarker, recurrent events and death event; Model I denote the model without considering the joint modeling of biomarker. The piecewise constant approach is adopted to estimate the baseline intensities for both recurrent and death event processes, where 5 intervals are considered.

<table>
<thead>
<tr>
<th>Setting II Parameters</th>
<th>True</th>
<th>Bias</th>
<th>MCSE</th>
<th>MSE</th>
<th>Bias</th>
<th>MCSE</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_R$</td>
<td>1.2</td>
<td>-0.035</td>
<td>0.100</td>
<td>0.011</td>
<td>0.010</td>
<td>0.101</td>
<td>0.011</td>
</tr>
<tr>
<td>$\beta_D$</td>
<td>1</td>
<td>-0.149</td>
<td>0.158</td>
<td>0.047</td>
<td>0.013</td>
<td>0.160</td>
<td>0.026</td>
</tr>
<tr>
<td>$\phi_{\mu}$</td>
<td>1</td>
<td>-0.010</td>
<td>0.120</td>
<td>0.026</td>
<td>0.011</td>
<td>0.118</td>
<td>0.014</td>
</tr>
<tr>
<td>$\phi_{\omega}$</td>
<td>0.5</td>
<td>0.167</td>
<td>0.297</td>
<td>0.121</td>
<td>-0.064</td>
<td>0.228</td>
<td>0.056</td>
</tr>
<tr>
<td>$\eta_R$</td>
<td>0.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\eta_D$</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma^2_{\mu}$</td>
<td>1</td>
<td>-0.073</td>
<td>0.144</td>
<td>0.028</td>
<td>0.015</td>
<td>0.133</td>
<td>0.018</td>
</tr>
<tr>
<td>$\sigma^2_{\omega}$</td>
<td>0.25</td>
<td>0.048</td>
<td>0.091</td>
<td>0.010</td>
<td>0.008</td>
<td>0.065</td>
<td>0.004</td>
</tr>
</tbody>
</table>

Bias is the value of Monte Carlo mean of the parameter estimates (bases on 500 replicates) minus the true value; MCSE is the empirical standard error of the parameter estimate; MSE is the mean square error (MSE).
Table 4.2. Simulation results of two stage model Setting II: $\eta_R = 0.4$, $\eta_D = 0.7$ and $\lambda_0^R = 0.065$: 50% subjects are observed with at least one recurrent event and death percentage is 25%. Full model denotes the proposed joint model of biomarker, recurrent events and death event; Model I denote the model without considering the joint modeling of biomarker. The piecewise constant approach is adopted to estimate the baseline intensities for both recurrent and death event processes, where 5 intervals are considered.

<table>
<thead>
<tr>
<th>Setting II</th>
<th>Model I</th>
<th>Full model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameters</td>
<td>True</td>
<td>Bias</td>
</tr>
<tr>
<td>$\beta_R$</td>
<td>1.2</td>
<td>-0.026</td>
</tr>
<tr>
<td>$\beta_D$</td>
<td>1</td>
<td>-0.052</td>
</tr>
<tr>
<td>$\phi_\mu$</td>
<td>1</td>
<td>-0.042</td>
</tr>
<tr>
<td>$\phi_\omega$</td>
<td>0.5</td>
<td>0.035</td>
</tr>
<tr>
<td>$\eta_R$</td>
<td>0.4</td>
<td></td>
</tr>
<tr>
<td>$\eta_D$</td>
<td>0.7</td>
<td></td>
</tr>
<tr>
<td>$\sigma_\mu^2$</td>
<td>1</td>
<td>-0.032</td>
</tr>
<tr>
<td>$\sigma_\omega^2$</td>
<td>0.25</td>
<td>0.046</td>
</tr>
</tbody>
</table>

Bias is the value of Monte Carlo mean of the parameter estimates (bases on 500 replicates) minus the true value; MCSE is the empirical standard error of the parameter estimate; MSE is the mean square error (MSE).
Table 4.3. Simulation results of two stage model Setting III: the percentage of left-censoring longitudinal observations is 30%. 60% subjects are observed with at least one recurrent event and death percentage is 40%. Full model denotes the proposed joint model of left-censored biomarker, recurrent events and death event; Model I denote the model without considering the left-censoring biomarker.

<table>
<thead>
<tr>
<th>Setting II</th>
<th>Model I</th>
<th>Full model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameters</td>
<td>True Bias RB MCSE MSE</td>
<td>Bias RB MCSE MSE</td>
</tr>
<tr>
<td>$\beta_R$</td>
<td>1.2 0.029 2.412 0.119 0.015</td>
<td>0.011 0.092 0.110 0.010</td>
</tr>
<tr>
<td>$\beta_D$</td>
<td>1 0.032 3.191 0.181 0.034</td>
<td>0.021 1.126 0.160 0.026</td>
</tr>
<tr>
<td>$\phi_\mu$</td>
<td>1 0.023 2.296 0.101 0.010</td>
<td>0.014 1.239 0.094 0.009</td>
</tr>
<tr>
<td>$\phi_\omega$</td>
<td>0.5 -0.084 16.821 0.356 0.134</td>
<td>-0.065 12.213 0.242 0.062</td>
</tr>
<tr>
<td>$\eta_R$</td>
<td>0.5 -0.027 5.444 0.049 0.003</td>
<td>-0.010 1.212 0.034 0.001</td>
</tr>
<tr>
<td>$\eta_D$</td>
<td>1 -0.075 4.516 0.181 0.038</td>
<td>-0.031 3.012 0.119 0.015</td>
</tr>
<tr>
<td>$\sigma^2_\mu$</td>
<td>1 0.024 2.123 0.148 0.022</td>
<td>0.019 1.517 0.146 0.022</td>
</tr>
<tr>
<td>$\sigma^2_\omega$</td>
<td>0.25 0.009 3.611 0.066 0.004</td>
<td>0.010 3.211 0.069 0.004</td>
</tr>
<tr>
<td>$\beta_{l0}$</td>
<td>0.5 0.638 127.541 0.016 0.407</td>
<td>0.001 0.214 0.033 0.001</td>
</tr>
<tr>
<td>$\beta_{l1}$</td>
<td>1 0.635 63.492 0.032 0.404</td>
<td>0.001 0.066 0.055 0.003</td>
</tr>
<tr>
<td>$\beta_{l2}$</td>
<td>1 0.179 17.872 0.015 0.032</td>
<td>0.004 0.446 0.012 0.000</td>
</tr>
<tr>
<td>$\sigma^2_{\beta_0}$</td>
<td>0.1 0.092 91.573 0.002 0.008</td>
<td>0.000 0.062 0.008 0.000</td>
</tr>
<tr>
<td>$\sigma^2_{\beta_1}$</td>
<td>0.04 0.024 58.560 0.014 0.001</td>
<td>0.000 0.446 0.004 0.000</td>
</tr>
<tr>
<td>$\text{cov}_{\beta_0\beta_1}$</td>
<td>0.01 0.003 29.196 0.003 0.000</td>
<td>0.000 4.394 0.004 0.000</td>
</tr>
<tr>
<td>$\sigma^2_{\beta_1}$</td>
<td>0.01 0.027 265.572 0.002 0.001</td>
<td>0.000 0.221 0.001 0.000</td>
</tr>
</tbody>
</table>

Bias is the value of Monte Carlo mean of the parameter estimates (bases on 500 replicates) minus the true value; RB is the relative bias which is the absolute bias divided by the true value times 100; MCSE is the empirical standard error of the parameter estimate; MSE is the mean square error (MSE).
Joint model of a longitudinal biomarker and recurrent events with the informative death event

Due to the existing drawbacks, such as the underestimates of standard error to the proposed two-stage in Chapter 4, we simultaneously jointly analyze the longitudinal biomarker, recurrent events and death.

5.1 Model

5.1.1 Notations

Let $i$ denote the index of matched pairs and $j$ denote the baseline exposure variable, $i = 1, 2, \ldots, I; j = 1, 2$ with the value of $j = 1$ (non-stroke at baseline) and $j = 2$ (stroke at baseline) in our stroke study. For the $ij^{th}$ subject, let $C_{ij}$, $M_{ij}$ and $D_{ij}$ be the follow-up time, drop-out time and the death time respectively. Let $T_{ij} = \min(D_{ij}, C_{ij}, M_{ij})$ be the observed follow-up time with $\Delta_{ij} = I(D_{ij} \leq \min(C_{ij}, M_{ij}))$ as the death indicator, where $I(\cdot)$ is the indicator function. Denote $\Psi_{ij}(t) = I(T_{ij} \geq t)$ as the “at-risk” indicator to show whether the subject is still
under observation at time t or not. Define \( N_{ij}^{D}(t) = I(D_{ij} \leq t, \Delta_{ij} = 1) \) as the observed death process, and \( N_{ij}^{R*}(t) = I(D_{ij} \leq t) \) as the actual death process by time point \( t \). In addition, denote \( N_{ij}^{R}(t) = N\{\min(D_{ij}, t)\} \) and \( N_{ij}^{R}(t) = N\{\min(T_{ij}, t)\} = N_{ij}^{R*}(\min(T_{ij}, t)) \) as the actual and observed number of recurrent events by time point \( t \) separately. \( N_{ij}^{D} \) and \( N_{ij}^{R} \) are both the observed parts of the counting processes \( N_{ij}^{D*} \) and \( N_{ij}^{R*} \). The number of recurrent events that occur for the \( ij^{th} \) subject over the small interval \([t, t + dt)\) is \( dN_{ij}^{R*}(t) = N_{ij}^{R*}((t + dt)^-) - N_{ij}^{R*}(t^-) \) as \( dt \to 0 \) and \( dN_{ij}^{R*} = \Psi_{ij}(t) dN_{ij}^{R*} \). \( R_{ij} = I(N_{ij}^{R} > 0) \) denotes the observed indicator for whether the \( ij^{th} \) subject has at least one recurrent stroke during the follow-up. Let \( t_{ijk} \) be the \( k^{th} \) recurrent event time and \( \delta_{ijk} \) denote the indicator of recurrent events at time \( t_{ijk}, k = 1, 2, \ldots, N_{ij}^{R} \).

For subject \( ij \), the number of biomarker measurements are not equal due to the missing values. The repeated measurements \( y_{ij} = \{y_{ij}(t_{ij1}), y_{ij}(t_{ij2}), \ldots, y_{ij}(t_{ijm}), \ldots, y_{ij}(t_{ijn_{ij}})\}^T, m = 1, \ldots, n_{ij} \). Measurements \( y_{ij}(t_{ijm}) \) and recurrent events \( t_{ijk} \) can be possibly observed at the same moments, for example, patient visits. Neither \( y_{ij}(t_{ijm}) \) and \( t_{ijk} \) can be observed after \( T_{ij} \). The right-censoring does not interrupt the processes and they are simply no longer observed.

### 5.1.2 Longitudinal measurements

The longitudinal response \( y_{ij} \) is modeled using a linear mixed effects model:

\[
y_{ij} = m_{ij} + \epsilon_{ij} = X_{ij}^{(L)} \beta_{L} + Z_{ij} \beta_{ij} + \epsilon_{ij}
\]

(5.1)

where \( X_{ij}^{(L)} = (X_{ij1}, \ldots, X_{ijp}) \) are covariates and \( \beta_{L} \) is the \( p \times 1 \) vector for fixed-effects parameter. The covariates for random effects are \( Z_{ij} = (Z_{ij1}, \ldots, Z_{ijp}) \), which could be a subset of \( X_{ij}^{(L)} \) including unit component. \( b_{ij} \) is the vector of random effects for subject \( ij \) with \( b_{ij} \sim N(0, \Sigma_{b}) \). The error vector \( \epsilon_{ij} = (\epsilon_{ij}(t_{ij1}), \ldots, \epsilon_{ij}(t_{ijn_{ij}}))^T \) is usually specified as \( \epsilon_{ij} \sim N_{n_{ij}}(0, \sigma_{e}^2 I_{n_{ij}}) \), where \( I_{n_{ij}} \) represents the identity matrix of dimension \( n_{ij} \), \( \text{cov}(\epsilon_{ij}(t), \epsilon_{ij}(t')) = 0, \) for \( t \neq t' \). Note that \( b_{ij} \) is independent with \( \epsilon_{ij} \). Therefore, \( \text{var}(y_{ij}) = V_{ij} = Z_{ij} \Sigma_{b} Z_{ij}^T + \sigma_{e}^2 I_{n_{ij}} \). The observed longitudinal outcome \( y_{ij} \) is represented by its
true value \( m_{ij} \) and measurement error \( \epsilon_{ij} \).

In many joint models, the longitudinal component uses random effects with functions of time only Tsiatis and Davidian (2004). The covariate \( Z_{ij} \) usually includes only random intercept and random slope effects, or at most a random quadratic effect of time. Here, we specify \( b_{ij} = (b_{ij0}, b_{ij1}) \), where \( b_{ij0} \) is the random intercept and \( b_{ij1} \) is the random slope. Then for subject \( ij \), the mixed effects model is:

\[
y_{ij} = m_{ij} + \epsilon_{ij} = X_{ij}^{(L)} \beta_{L} + 1b_{ij0} + t_{ij}b_{ij1} + \epsilon_{ij} \tag{5.2}
\]

where \( 1 = (1, \ldots, 1)^T \) is a \( n_{ij} \) unit vector, \( t_{ij} = (t_1, t_2, \ldots, t_{n_{ij}})^T \).

### 5.1.3 The joint model of recurrent events and death event

Let \( \mu_i \) denote the frailty measuring the dependency between stroke and non-stroke subjects within a matched pair. Denote \( \omega_{ij} \) as the frailty measuring the dependency among the recurrent acute ischemic stroke events within the \( ij^{th} \) individual. \( \mu_i \) and \( \omega_{ij} \) are shared by the recurrent and terminal events to induce their dependency. Assume that \((\mu_i, \omega_{ij})^T \sim MVN(0, \Sigma)\). For simplicity, we can assume that \( \mu_i \perp \omega_{ij} \), where \( \mu_i \sim N(0, \sigma_{\mu}^2) \) and \( \omega_{ij} \sim N(0, \sigma_{\omega}^2) \). If \( \sigma_{\mu}^2 \) and \( \sigma_{\omega}^2 \) equal 0, then it implies there is no dependency between recurrent events and death, and the heterogeneity in both event processes is solely explained by covariates \( X_{ij}^{(R)} \) and \( X_{ij}^{(D)} \).

The model for recurrent events is defined as follows:

\[
\lambda_{ij}^R(t \mid X_{ij}^{(R)}, \mu_i, \omega_{ij}) = \lambda_0^R(t) \exp(\beta_R^TX_{ij}^{(R)} + \mu_i + \omega_{ij}) \tag{5.3}
\]

where \( \beta_R \) is the vector of regression coefficients, \( \lambda_0^R(t) \) is the baseline intensity function. \( X_{ij}^{(R)} \) denotes the vector of the observed covariates including baseline risk factors, such as gender, race and smoking status and so on. Currently, we focus on time-independent covariates.

Conditional on \( \mu_i \) and \( \omega_{ij} \), the recurrent event process and death are independent of each other. Thus, the association between death and recurrent events is quantified by the shared frailties \( \mu_i \) and \( \omega_{ij} \) through \( \phi_\mu \) and \( \phi_\omega \) in equation (5.7).
Consequently, a higher intensity of recurrent events is associated with a higher mortality rate. Note that by extending the work from Liu et al. (2015), the correlation within matched pairs also is incorporated through $\mu_i$. Thus, the frailty proportional hazard model for death is

$$
\lambda^D_{ij}(t \mid X^{(D)}_{ij}, \mu_i, \omega_{ij}) = \lambda^D_0(t) \exp(\beta^T_D X^{(D)}_{ij} + \phi \mu_i + \phi \omega \omega_{ij})
$$

(5.4)

where $\beta_D$ is vector of regression coefficients, and $\lambda^D_0(t)$ is the baseline hazard function of death. $X^{(D)}_{ij}$ denotes the vector of the observed covariates in the survival model for death. Of note, $X^{(R)}_{ij}$ and $X^{(D)}_{ij}$ are assumed to be the same for $\lambda^R_{ij}$ and $\lambda^D_{ij}$ for simplicity, but can be different in terms of clinical perspectives for real data applications.

### 5.1.4 Joint model

In this chapter, we simultaneously perform a joint analysis the following process:

$$
y_{ij}(t) = m_{ij}(t) + \epsilon_{ij}(t) = \beta^T_L X^{(L)}_{ij}(t) + b_{ij0} + b_{ij1}t + \epsilon_{ij}(t)
$$

(5.5)

$$
\lambda^R_{ij}(t \mid X^{(R)}_{ij}, X^{(L)}_{ij}, \mu_i, \omega_{ij}, b_{ij}) = \lambda^R_0(t) \exp(\beta^T_R X^{(R)}_{ij} + \mu_i + \omega_{ij} + \eta^T_R h^R(X^{(L)}_{ij}, b_{ij}, t))
$$

(5.6)

$$
\lambda^D_{ij}(t \mid X^{(D)}_{ij}, X^{(L)}_{ij}, \mu_i, \omega_{ij}, b_{ij}) = \lambda^D_0(t) \exp(\beta^T_D X^{(D)}_{ij} + \phi \mu_i + \phi \omega \omega_{ij} + \eta^T_D h^D(X^{(L)}_{ij}, b_{ij}, t))
$$

(5.7)

We explore the nature of the dependency between the longitudinal repeated measurements, the risk of recurrence and death by considering the link functions $h^R(X^{(L)}_{ij}, b_{ij}, t)$ and $h^D(X^{(L)}_{ij}, b_{ij}, t)$. The form that $h^R(X^{(L)}_{ij}, b_{ij}, t)$ and $h^D(X^{(L)}_{ij}, b_{ij}, t)$ take determine the type of joint model that are fit. $\eta_R$ and $\eta_D$ are the vector of parameters that links the longitudinal process and survival process and determine the association’s strength. Different approaches can be adopted to construct the trajectory function, e.g. include mean response which composed by the fixed effects model only, or use the full longitudinal trajectory with both fixed effects and random effect components (Chen et al., 2014). $X^{(L)}_{ij}(t)$, $X^{(R)}_{ij}$ and $X^{(D)}_{ij}$ can be the same or different in the three processes.
The individual deviations can be directly included in the survival model to explain the impact of biomarker dynamics. The function linking both processes is time-independent. The link between the longitudinal biomarker and the recurrent events, and the biomarker and death is explained partly by the true mean profile that incorporated \( b_{ij} \) Wulfsohn and Tsiatis (1997); Kim et al. (2012).

\[
\begin{align*}
    h^R(X_{ij}^{(L)}, b_{ij}, t) &= m_{ij}^{R}(t) = \beta_L^T X_{ij}^{(L)}(t) + b_{ij0} + b_{ij1}t \\
    h^D(X_{ij}^{(L)}, b_{ij}, t) &= m_{ij}^{D}(t) = \beta_L^T X_{ij}^{(L)}(t) + b_{ij0} + b_{ij1}t
\end{align*}
\]

Given \( \theta = (\lambda^R_0(\cdot), \lambda^D_0(\cdot), \beta_L, \beta_R, \beta_D, \phi_\mu, \phi_\omega, \eta_\mu, \eta_\omega, \eta_D, \sigma^2_\mu, \sigma^2_\omega, \sigma^2_\omega, \Sigma_b)^T \), the marginal likelihood is

\[
L(\theta) = \int \int \int \int \cdots \int \prod_{i=1}^{I} \prod_{j=1}^{J} L(\theta \mid \mu_i, \omega_{ij}, b_{ij}) f(\mu_i) f(\omega_{ij}) f(b_{ij} \mid \Sigma_b) \, db_{ij} \, d\omega_{ij} \, d\mu_i
\]

\[
= \prod_{i=1}^{I} \int \prod_{j=1}^{J} \left\{ \prod_{m=1}^{n_{ij}} L^L_{ijm}(t_{ijm}) \right\} L^R_{ij} L^D_{ij} f(\omega_{ij}) f(b_{ij} \mid \mu_i) \, d\mu_i
\]

(5.8)

with

\[
L^L_{ijm}(t_{ijm}) = (2\pi \sigma^2)^{-1/2} \exp\{-y_{ij}(t_{ijm}) - \beta_L^T X_{ij}^{(L)}(t_{ijm}) - b_{ij0} - b_{ij1}t_{ijm})^2/2\sigma^2\}
\]

\[
L^R_{ij} = \prod_k [\lambda^R_0(t) \exp(\beta_R^T X_{ij}^{(R)} + \mu_i + \omega_{ij} + \eta_R m_{ij}(t))]^{\delta_{ijk}} \cdot \exp[-\int_{0}^{\infty} \Psi_{ij}(t) \lambda^R_0(t) \exp(\beta_R^T X_{ij}^{(R)} + \mu_i + \omega_{ij} + \eta_R m_{ij}(t)) \, dt]
\]

\[
L^D_{ij} = [\lambda^D_0(T_{ij}) \exp(\beta_D^T X_{ij}^{(D)} + \phi_\mu \mu_i + \phi_\omega \omega_{ij} + \eta_D m_{ij}(t))]^{\Delta_{ij}} \cdot \exp[-\int_{0}^{\infty} \Psi_{ij}(t) \lambda^D_0(t) \exp(\beta_D^T X_{ij}^{(D)} + \phi_\mu \mu_i + \phi_\omega \omega_{ij} + \eta_D m_{ij}(t)) \, dt]
\]

\[
f(b_{ij} \mid \Sigma_b) = (2\pi |\Sigma_b|)^{-1/2} \exp\{-(b_{ij} \Sigma_b^{-1} b_{ij})^2/2\}
\]

\( L^L_{ijm}(t_{ijm}) \) is the likelihood of the longitudinal measurements at time \( t_{ijm} \) for subject \( ij \). \( L^R_{ij} \) is the likelihood of observing at least 1 recurrent event \( (R_{ij} = 1) \). \( L^D_{ij} \) is
the likelihood for terminal event death. $f(\mu_i)$ and $f(\omega_{ij})$ are normal densities. In addition, the cumulative baseline intensities of recurrent events and terminal event denoted by $\Lambda^R_0(t)$ and $\Lambda^D_0(t)$ are given by $\Lambda^R_0(t) = \int_0^\infty \Psi_{ij}(t)\lambda^R_0(t)dt$ and $\Lambda^D_0(t) = \int_0^\infty \Psi_{ij}(t)\lambda^D_0(t)dt$.

5.2 Estimation

We adopted similar MCEM algorithm as in Chapter 4 with a few modifications.

In the E step, we find the expectation of the conditional log-likelihood, and integrate out the frailties by numerical integration with the Metropolis-Hasting algorithm. We define that $X_{1ij}^{(L)}$ is the covariate vector $X_{ij}^{(L)}$ that doesn’t contain the time variable $t_{ij}$. Thus, $\beta_{L_1}$ is the coefficient vector that exclude $\beta_t$ and $\theta = (\lambda^R_0(\cdot), \lambda^D_0(\cdot), \beta_{L_1}, \beta_t, \beta_R, \beta_D, \phi_\mu, \phi_\omega, \eta_R, \eta_D, \sigma^2_\mu, \sigma^2_\omega, \sigma^2_e, \Sigma_b)^T$. Let $q_{1ij} = \beta^T_R X_{ij}^{(R)} + \eta_R X_{1ij}^{(L)}(t)\beta_{L_1}$, and $q_{2ij} = \beta^T_D X_{ij}^{(D)} + \eta_D X_{1ij}^{(L)}(t)\beta_{L_1}$. Thus, given the
values of \( b, \mu \) and \( \omega \), the complete log likelihood function denoted by \( l(\theta) \) is:

\[
l(\theta) = \sum_{i=1}^{I} \sum_{j=1}^{J} \left\{ \log \left\{ \prod_{m} (2\pi \sigma_e^2)^{-1/2} \exp\left\{ -\left( y_{ij} (t_{ijm}) - \beta_L^T X_{1ij}^{(L)} (t_{ijm}) - \beta_t t_{ijm} - b_{ij0} - b_{ij1} t_{ijm} \right)^2 / 2\sigma_e^2 \right\} \right\} + \log \left\{ \prod_{k} \left( \lambda_0^R(t_{ijk}) \cdot \exp(\eta_R (\beta_t + b_{1ij}) t_{ijk} + q_{1ij} + \mu_i + \omega_{ij} + \eta_R b_{0ij}) \right)^{\delta_{ijk}} \right\} \exp \left\{ -\Lambda_0^R(t) \cdot \exp(q_{1ij} + \mu_i + \omega_{ij} + \eta_R b_{0ij}) \right\} \right\} + \log \left\{ \left( \lambda_0^D(T_{ij}) \cdot \exp(\eta_D (\beta_t + b_{1ij}) T_{ij} + q_{2ij} + \phi_\mu \mu_i + \phi_\omega \omega_{ij} + \eta_R b_{0ij}) \right)^{\Delta_{ij}} \exp \left\{ -\Lambda_0^D(t) \exp(q_{2ij} + \phi_\mu \mu_i + \phi_\omega \omega_{ij} + \eta_R b_{0ij}) \right\} \right\} + \log \left\{ \frac{1}{\sqrt{2\pi \sigma_\mu^2}} \exp\left( -\frac{\mu_i^2}{2\sigma_\mu^2} \right) \frac{1}{\sqrt{2\pi \sigma_\omega^2}} \exp\left( -\frac{\omega_{ij}^2}{2\sigma_\omega^2} \right) \right\} + \log \left\{ (2\pi |\Sigma_b|)^{-1/2} \exp \left( -b_{ij}^T \Sigma_b^{-1} b_{ij} / 2 \right) \right\} \}
\]

(5.9)

The expectation of the log-likelihood \( l(\theta) \) conditional on the observed data and
the current parameter estimate $\hat{\theta}^{(k)}$ is:

$$Q(\theta | \hat{\theta}^{(k)}) = E[l(\theta) | \hat{\theta}^{(k)}]$$

$$= \sum_{i=1}^{I} \sum_{j=1}^{J} \left\{ \frac{1}{2} \sum_{m} \left( \log(2\pi) + \log(\sigma^2_{\omega}) - (y_{ij} t_{ijm}) - \beta L_{1i}^{(L)} (t_{ijm}) - \beta t_{ijm} - b_{ij0} - b_{ij1} t_{ijm} \right)^2 / 2 \sigma^2_{\omega} \right\} +$$

$$\left\{ \sum_{k=1}^{K} \left( \log(\Lambda_0^{R}(t_{ijk})) + q_{1ij} + \eta_R \beta t_{ijk} + \eta_R E[b_{1ij} t_{ijk} | \hat{\theta}^{(k)}] + \eta_R E[b_{0ij} | \hat{\theta}^{(k)}] + \right.$$  

$$E[\mu_i | \hat{\theta}^{(k)}] + E[\omega_{ij} | \hat{\theta}^{(k)}] - \exp(q_{1ij}) E[\Lambda_0^{R}(t) \exp(\mu_i + \omega_{ij} + \eta_R b_{0ij}) | \hat{\theta}^{(k)}] \} +$$

$$\left\{ \Delta_{ij}( \log(\Lambda_0^{D}(T_{ij})) + q_{2ij} + \eta_D \beta T_{ij} + \eta_D E[b_{1ij} T_{ij} | \hat{\theta}^{(k)}] + \eta_D E[b_{0ij} | \hat{\theta}^{(k)}] + \right.$$  

$$\phi_{\omega} E[\mu_i | \hat{\theta}^{(k)}] + \phi_{\omega} E[\omega_{ij} | \hat{\theta}^{(k)}] \} -$$

$$\exp(q_{2ij}) E[\Lambda_0^{D}(T_{ij}) \exp(\phi_{\omega} \mu_i + \phi_{\omega} \omega_{ij} + \eta_D b_{0ij}) | \hat{\theta}^{(k)}] \} +$$

$$\left\{ - \frac{1}{2} \left( \log(2\pi) + \log(\sigma^2_{\omega}) + \frac{E[\omega_{ij}^2 | \hat{\theta}^{(k)}]}{\sigma^2_{\omega}} \right) - \frac{1}{2} \left( \log(2\pi) + \log(\sigma^2_{\mu}) + \frac{E[\mu_i^2 | \hat{\theta}^{(k)}]}{\sigma^2_{\mu}} \right) \} +$$

$$\left\{ - \frac{1}{2} \left( \log(2\pi) + \log(|\Sigma_b|) + b_{ij}^{T} \Sigma^{-1}_{b} b_{ij} \right) \} \right\} \right\}$$

$$= \sum_{i=1}^{I} \sum_{j=1}^{J} \left\{ \frac{1}{2} \sum_{m} \left( \log(2\pi) + \log(\sigma^2_{\omega}) - (y_{ij} t_{ijm}) - \beta L_{1i}^{(L)} (t_{ijm}) - \beta t_{ijm} - b_{ij0} - b_{ij1} t_{ijm} \right)^2 / 2 \sigma^2_{\omega} \right\}$$

(5.10)

More details on the computation of the conditional expectation above are given in Appendix A.1.

Since there is no closed-form expression of $p(b, \mu, \omega | \hat{\theta}^{(k)})$, Monte Carlo methods, specifically the Metropolis-Hasting algorithm can be used to approximate the posterior distributions of $b_{0i}'s$, $b_{1i}'s$, $u_i'$s and $\omega_{ij}'s$.

In the M-step, we use a Newton-Raphson procedure to maximize $Q(\theta | \hat{\theta}^{(k)})$ to estimate $\theta$. For estimation of baseline intensity, similar to Chapter 4, we adopt the piecewise baseline intensity function for recurrent events and death. Partition the time duration into $M$ intervals with cutpoints $0 \leq c_0 < c_1 < \ldots < c_M = \infty$,...
which can be quantiles of event times and where \(c_0 = 0\) or the smallest event time. The baseline intensity function is assumed to be piecewise constant within each of the \(M\) intervals, so that

\[
\tilde{\lambda}_0(t) = \sum_{m=1}^{M} \lambda_m I(c_{m-1} < t \leq c_m)
\]  

(5.11)

Denote the baseline intensity of recurrent events is \(\tilde{\lambda}_R^0(t)\), using \(M\) parameters \(\lambda_1^R, \lambda_2^R, \ldots, \lambda_m^R\). And the baseline hazard for death is \(\tilde{\lambda}_D^0(t)\), using \(M\) parameters \(\lambda_1^D, \lambda_2^D, \ldots, \lambda_m^D\). Note that the selection of cutpoint \(c_m\) and \(M\) for recurrent events and death can be the same or different. The detailed estimation of \(\tilde{\lambda}_R^0(t)\) and \(\tilde{\lambda}_D^0(t)\) in provided in Appendix C.1.

Now, the cumulative baseline intensities of recurrent events is

\[
\tilde{\Lambda}_R^0(t) = \sum_{m=1}^{M} \int_{t_{m-1}}^{t_m} \lambda_m^R \exp(\eta_R(\beta_t + b_{1ij})t)dt
\]

(5.12)

\[
= \sum_{m=1}^{M} \frac{\lambda_m^R}{\eta_R(\beta_t + b_{1ij})} \left[ \exp(\eta_R(\beta_t + b_{1ij})t_m^R) - \exp(\eta_R(\beta_t + b_{1ij})t_{m-1}^R) \right]
\]

Here, \(t_m^R\) and \(t_{m-1}^R\) are upper and lower bound at \(m\)th interval. \(t_m^R = \max(0, \min(c_m - c_{m-1}, t - c_{m-1}))\). And, the cumulative baseline intensities of death event is

\[
\tilde{\Lambda}_D^0(t) = \sum_{m=1}^{M} \int_{t_{m-1}}^{t_m} \lambda_m^D \exp(\eta_D(\beta_t + b_{1ij})t)dt
\]

(5.13)

\[
= \sum_{m=1}^{M} \frac{\lambda_m^D}{\eta_D(\beta_t + b_{1ij})} \left[ \exp(\eta_D(\beta_t + b_{1ij})t_m^D) - \exp(\eta_D(\beta_t + b_{1ij})t_{m-1}^D) \right]
\]

Here, \(t_m^D\) and \(t_{m-1}^D\) are upper and lower bound at \(m\)th interval. \(t_m^D = \max(0, \min(c_m - c_{m-1}, t - c_{m-1}))\).

For the parameter estimation, all the score components and second partial derivatives are given in Appendix C.3.

The variance of the MLE \(\hat{\theta} = (\hat{\lambda}_0^R(t), \hat{\lambda}_0^D(t), \hat{\beta}_R, \hat{\beta}_D, \hat{\phi}_\mu, \hat{\phi}_\omega, \hat{\eta}_R, \hat{\eta}_D, \hat{\sigma}_\mu^2, \hat{\sigma}_\omega^2)^T\) in the
MCEM algorithm cannot be obtained directly from the algorithm. Louis’s formula was used to obtain it Louis (1982). The variance-covariance matrix can be estimated using the inverse of an observed information matrix. The observed information matrix $I(\hat{\theta})$ is given by

$$I(\hat{\theta}) = -E\left[\frac{\partial^2l(\theta)}{\partial \theta \partial \theta^T} \mid \hat{\theta}\right] - \text{Var}\left[\frac{\partial l(\theta)}{\partial \theta} \mid \hat{\theta}\right]$$  \hspace{1cm} (5.14)$$

where $\text{Var}\left[\frac{\partial l(\theta)}{\partial \theta} \mid \hat{\theta}\right] = E\left[\frac{\partial^2 l(\theta)}{\partial \theta \partial \theta^T} \mid \hat{\theta}\right]$. The last term becomes zero due to the MLE $\hat{\theta}$ and all of these terms are evaluated at the last iteration of the EM algorithm.

### 5.3 Simulation studies

In this section, we conduct simulation studies to evaluate the performance of the proposed two stage joint frailty model considering longitudinal data with left-censoring. We compare the performances of different models based on the biases of the estimates, and the coverage rates of the confidence intervals under several scenarios. Then, we compare the results under different simulation scenarios and draw conclusions.

For simplicity, we consider a continuous covariate of $X_1$, which are randomly sampled from the uniform distribution on $[0, 1]$ and binary variable $X_2$, which takes values of 0 or 1 with probability 0.5. And also, only the intercept term for the logistic regression model. Considering the exponential distributions for both recurrent event times and death, the following models are shown below

$$Y_{ij}(t) = m_{ij}(t) + \epsilon_{ij}(t) = \beta_{01} + \beta_{11}t + \beta_{12}X_{1ij} + b_{ij0} + b_{ij1}t + \epsilon_{ij}(t)$$

$$\lambda_{ij}^R(t \mid \mu_i, \omega_{ij}, Y_{ij} = 1) = \lambda_0^R(t) \exp(\beta_R X_{2ij} + \mu_i + \omega_{ij} + \eta_R \ast \hat{m}_{ij}(t))$$

$$\lambda_{ij}^D(t \mid \mu_i, \omega_{ij}) = \lambda_0^D(t) \exp(\beta_D X_{2ij} + \phi_d \mu_i + \phi_{dw} \omega_{ij} + \eta_D \ast \hat{m}_{ij}(t))$$

(5.15)

The censoring time is taken as $C_{ij} = 6 \ast \text{Unif}(0,1)$, where Unif$(0,1)$ is a random number generated from uniform distribution in $[0, 1]$. We assume $\mu_i \sim N(0, \sigma^2_\mu)$,
\( \omega_{ij} \sim N(0, \sigma^2_\omega) \) and set \( \sigma^2_\mu = 1, \sigma^2_\omega = 0.25 \). \( b_{ij0} \) and \( b_{ij1} \) follows multivariate normal distribution with \( \sigma^2_{b0} = 0.1, \sigma^2_{b1} = 0.04 \) and \( \text{cov}(b_0, b_1) = 0.01 \). The error term \( \epsilon_{ij}(t) \sim N(0, \sigma^2_\epsilon) \) and \( \sigma^2_\epsilon = 0.01 \). The coefficients \( \beta_0 = 0.5, \beta_1 = 1, \beta_2 = 1, \beta_R = 1.2, \beta_D = 1, \phi_\mu = 1, \phi_\omega = 0.5, \eta_R = 0.5, \eta_D = 1 \). The baseline intensities \( \lambda^R_0 = 0.1 \) and \( \lambda^D_0 = 0.005 \). As a result, 60% subjects are observed with at least one recurrent event and death percentage is 40%. A total of 500 Monte Carlo replicates are generated, each with a sample size of 500 subjects.

First, we evaluate the estimation of the standard error and coverage probabilities of the proposed the joint models and compared, we adopt the constant baseline to estimate the baseline intensities for both recurrent and death event processes. The resulting parameter estimates are shown in Table 5.1. It can be seen that the two-stage joint model (Model I) produce relatively larger biases and lower coverage rates compared to the estimates from the full model. First, the longitudinal data is truncated by death and censoring in the data generation step, which strengthens the association between the longitudinal and survival processes. The longitudinal covariate trajectory is related to the length of follow-up; the truncation causes the loss of information. On the other hand, the proposed joint model incorporates the association among longitudinal, recurrent and death processes, which recapture the missing measurements due to death or censoring.
### Table 5.1. Simulation results of joint model Setting I: $\sigma^2_e = 0.01$ and $n_{ij} = 10$: 60% subjects are observed with at least one recurrent event and death percentage is 40%. Full model denotes the proposed joint model of biomarker, recurrent events and death event; Model I denote the two-stage joint model.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>True</th>
<th>Setting I</th>
<th>Model I</th>
<th>Full model</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_R$</td>
<td>1.2</td>
<td>0.833</td>
<td>0.089</td>
<td>0.110</td>
</tr>
<tr>
<td>$\beta_D$</td>
<td>1</td>
<td>1.311</td>
<td>0.140</td>
<td>0.161</td>
</tr>
<tr>
<td>$\phi_\mu$</td>
<td>1</td>
<td>1.038</td>
<td>0.095</td>
<td>0.118</td>
</tr>
<tr>
<td>$\phi_\omega$</td>
<td>0.5</td>
<td>12.833</td>
<td>0.189</td>
<td>0.227</td>
</tr>
<tr>
<td>$\eta_R$</td>
<td>0.5</td>
<td>1.151</td>
<td>0.029</td>
<td>0.035</td>
</tr>
<tr>
<td>$\eta_D$</td>
<td>1</td>
<td>3.089</td>
<td>0.093</td>
<td>0.121</td>
</tr>
<tr>
<td>$\sigma^2_R$</td>
<td>1</td>
<td>1.196</td>
<td>0.095</td>
<td>0.143</td>
</tr>
<tr>
<td>$\sigma^2_\omega$</td>
<td>0.25</td>
<td>3.191</td>
<td>0.032</td>
<td>0.065</td>
</tr>
<tr>
<td>$\beta_{l0}$</td>
<td>0.5</td>
<td>0.276</td>
<td>0.029</td>
<td>0.028</td>
</tr>
<tr>
<td>$\beta_{l1}$</td>
<td>1</td>
<td>0.088</td>
<td>0.029</td>
<td>0.028</td>
</tr>
<tr>
<td>$\beta_{l2}$</td>
<td>1</td>
<td>0.028</td>
<td>0.011</td>
<td>0.011</td>
</tr>
<tr>
<td>$\sigma^2_{b_l}$</td>
<td>0.1</td>
<td>0.367</td>
<td>0.007</td>
<td>0.000</td>
</tr>
<tr>
<td>$\sigma^2_{b_l}$</td>
<td>0.04</td>
<td>0.552</td>
<td>0.003</td>
<td>0.000</td>
</tr>
<tr>
<td>$\text{cov}_{b_l}$</td>
<td>0.01</td>
<td>0.636</td>
<td>0.003</td>
<td>0.000</td>
</tr>
<tr>
<td>$\sigma^2_{b_l}$</td>
<td>0.01</td>
<td>0.164</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>

RB is the relative bias which is the absolute differences between the value of Monte Carlo mean of the parameter estimates (bases on 500 replicates) and the true value divided by the true value times 100; SE is the mean of the standard error estimates; MCSE is the empirical standard deviation of the parameter estimate; MSE is the mean square error (MSE); CP is the 95% coverage probability.
Summary and future work

In this dissertation, we have proposed two models to jointly study the longitudinal biomarker, recurrent events and death processes in matched studies, where the models of each component were connected through the shared random effects. Particularly, in Chapter 3, we proposed a joint frailty model of recurrent events and a terminal event adjusted for zero-inflation, which results in more accurate parameter estimation. In general, the price of omitting the feature of zero-inflation of the events in the data from the models is biased estimates and overlooking the importance of certain cluster effects. Also, in medical research, investigators can evaluate the treatment effect by estimating the fraction of cured subjects, which is a substantially important question. In Chapter 4, we studied the two-stage joint model of longitudinal data via the linear mixed effects model, recurrent events and a terminal event. In Chapter 5, we simultaneously model the three processes and compare the performance of the two-stage method and the jointly likelihood method based on the biases of the estimates, and the coverage rates of the confidence intervals under several scenarios. Meanwhile, the maximum likelihood approach is adopted to accommodate the longitudinal biomarkers subject to left censoring.

In all of the methods, the maximum likelihood approach was adopted for parameter estimation. We assumed the baseline cumulative intensities and hazard functions to be the piecewise step functions. By treating the frailties or random effects as
missing data, we adopted MCEM algorithms to iteratively compute maximum likelihood estimates. Based on various scenarios, we also investigated the properties of the proposed methods via extensive simulation studies. We also compare the performance between the MCEM algorithm and adaptive Gaussian quadrature and Bayesian approach based on SAS procedures. Due to the large volume of parameters for estimation and intractable integrals, our adopted MCEM algorithm has better performance. The proposed methods were also applied to real data examples for illustration. In particular, we apply the proposed model in Chapter 3 to a real data application on recurrent acute ischemic stroke, which are obtained from Marketscan research database. We jointly modeling the recurrent ischemic stroke events in the presence of death with shared frailties to capture their dependency. We investigate the effect of initial stroke detected at baseline and comorbidity (e.g., diabetes mellitus, hypertension and heart disease) on the risk of recurrent stroke occurrence and death during index hospitalization, and also explored the dependency between recurrent stroke events and death.

For the future work, the proposed methods in this dissertation research can be extended in several directions. First, other functional forms of the frailties can be incorporated in the proposed joint model, such as gamma frailties, which have been applied by many researchers (Liu et al., 2004; Rondeau et al., 2006). With only a minor modification, the corresponding likelihood and the estimation algorithm are readily available. Also, we may consider the conjugate random effects to relax the often-restrictive mean-variance prescription in the non-Gaussian outcome (Njagi et al., 2016). Second, we can consider more flexible joint model settings. For example, we may consider the accelerated failure time (AFT) models for recurrent events and death. When modeling the longitudinal data, nonlinear and generalized linear models may considered for different types of longitudinal outcomes (Ye and Wu, 2017). Third, to adjust for the zero-inflation of recurrent events, we can consider alternative mixture cure models. The latency can be modeled by negative binomial (NB) distribution and the time for the corresponding cause to produce the event of interest (death or recurrence) can be modeled by parametric mixture cure models, such as log normal and generalized gamma distributions (Rodrigues et al., 2009; Cordeiro et al., 2016). Last, our proposed model can be applied to
the ASSESS-AKI (Go et al., 2010) study to investigate the association among longitudinal biomarkers, recurrent acute kidney injury (AKI) and death.
Chapter 3

A.1 Conditional expectation

The computation of the conditional expectation equation (4.28) is not trivial. Denote the function of frailties as \( h(\gamma_{ij}) \) and \( \gamma_{ij} = (y_{ij}, \mu_i, \omega_{ij})^T \). Thus, the expectation of complete data likelihood is \( \text{E}[h(\gamma_{ij} | \hat{\theta}^{(k)})] \). The term \( \hat{\theta}^{(k)} \) denotes all current parameters estimated in the M-step. The conditional density of \( \gamma_{ij} \) given the observed data and current estimate of the parameters is:

\[
    f(\gamma_{ij} | \hat{\theta}^{(k)}) = \frac{L(\hat{\theta}^{(k)} | \gamma_{ij}) f(\gamma_{ij})}{\int_{-\infty}^{\infty} L(\hat{\theta}^{(k)} | \gamma_{ij}) f(\gamma_{ij}) d\gamma_{ij}} \tag{A.1}
\]

Thus, the conditional expectation for any function \( h \) of the random effects is:

\[
    \text{E}[h(\gamma_{ij} | \hat{\theta}^{(k)})] = \int_{-\infty}^{\infty} h(\gamma_{ij}) f(\gamma_{ij} | \hat{\theta}^{(k)}) d\gamma_{ij} = \frac{\int_{-\infty}^{\infty} h(\gamma_{ij}) k(\gamma_{ij} | \hat{\theta}^{(k)}) d\gamma_{ij}}{\int_{-\infty}^{\infty} k(\gamma_{ij} | \hat{\theta}^{(k)}) d\gamma_{ij}} \tag{A.2}
\]

where \( k(\gamma_{ij} | \hat{\theta}^{(k)}) \) is the kernel density after removing the parts that do not depend on \( \gamma_{ij} \).
A.2 Metropolis-Hastings algorithm

It is difficult to sample directly from

\[ p(\mu, \omega \mid \hat{\theta}^{(k)}) \propto \prod_{i=1}^{l} \prod_{j=1}^{j} \left\{ 1 - p_{ij}^{(k)} + p_{ij}^{(k)} \exp(-\int_{0}^{\infty} \Psi_{ij}(t)\lambda_{0}^{R(k)}(t)dt q_{2ij}^{(k)}) \right\} \]

\[ p_{ij}^{(k)} \prod_{k} (q_{2ij}^{(k)})^{\delta_{ik}} \exp(-\int_{0}^{\infty} \Psi_{ij}(t)\lambda_{0}^{R(k)}(t)dt q_{2ij}^{(k)}) \]

\[ (q_{3ij}^{(k)})^{\Delta_{ij}} \exp(-\int_{0}^{\infty} \Psi_{ij}(t)\lambda_{0}^{D(k)}(t)dt q_{3ij}^{(k)}) \]

\[ \frac{1}{\sqrt{2\pi\sigma_{\mu}^{2(k)}}} \exp(-\frac{\mu_{i}^{2(k)}}{2\sigma_{\mu}^{2(k)}}) \frac{1}{\sqrt{2\pi\sigma_{\omega}^{2(k)}}} \exp(-\frac{\omega_{ij}^{2}}{2\sigma_{\omega}^{2(k)}}) \]  

\[ (A.3) \]

Metropolis-Hastings algorithm is a Markov chain Monte Carlo (MCMC) method for generating a sequence of random samples from a probability distribution for which direct sampling is difficult. Hastings (1970). Suppose we are at the kth E-step with current estimates superscripted with (k) and we want to generate the random number chain \( \mu_{i}^{(m)} \) and \( \omega_{ij}^{(m)} \) \((m = 1, \ldots M)\). Let \( v_{ij} = (\mu_{i}^{(m)}, \omega_{ij}^{(m)}) \). The Metropolis-Hastings chain starts with an initial value \( v_{ij}^{1} \). We use the normal densities as the proposal function. After we obtain \( v_{ij}^{m} \), new values \( \tilde{v}_{ij} \) are sampled from normal density with variance \( \sigma_{\mu}^{2(k)} \) and \( \sigma_{\omega}^{2(k)} \).

Given

\[ p(v_{ij} \mid \hat{\theta}^{(k)}) = \frac{p(\hat{\theta}^{(k)} \mid v_{ij})f(v_{ij})}{p(\hat{\theta}^{(k)})} = \frac{p(\hat{\theta}^{(k)} \mid v_{ij})f(v_{ij})}{\int v_{ij} p(\hat{\theta}^{(k)} \mid v_{ij})f(v_{ij})} \]

\( v_{ij}^{(m+1)} \) is obtained as:

\[ v_{ij}^{(m+1)} = \begin{cases} \tilde{v}_{ij} & \text{if } a \leq \min \left(1, \frac{p(\hat{\theta}^{(k)} \mid v_{ij})}{p(\hat{\theta}^{(k)} \mid v_{ij}^{(m)})} \right), \\ v_{ij}^{(m)} & \text{otherwise.} \end{cases} \]  

\[ (A.4) \]
A.3 Newton-Raphson algorithm

In the M-step, first we have closed-form estimators for \( \hat{\lambda}_0^R \), \( \hat{\lambda}_0^D \), \( \sigma_\mu^2 \) and \( \sigma_\omega^2 \). For the estimation of \( \hat{\lambda}_0^R \), first we introduce several notations. Assume there are \( M \) intervals with cutpoints \( 0 \leq c_0^R < c_1^R < \ldots < c_M^R = \infty \), which can be quantiles of recurrent event times, where \( c_0^R = 0 \) or the smallest event time. \( I_m^R(t) = I(c_m^R - 1 < t \leq c_m^R) \). Denote \( E_{ij,k,m}^R = I(R(T_{ij} \geq c_m^R - 1)(c_m^R \wedge T_{ij} - c_m^R - 1) \) the total time individual \( ij \) is at risk in the \( m^{th} \) recurrent events time interval \( (c_m^R - 1, c_m^R) \).

\[
E_{ij,m}^R = \sum_{i=1}^{M} \sum_{j=1}^{J} E_{ij,m}^R \exp(X_{ij}\beta_k^2)E(y_{ij} \exp(\mu_i + \omega_{ij}) | \hat{\theta}(k))
\]

To estimate \( \hat{\lambda}_0^D \), similarly, we have the closed form:

\[
\hat{\lambda}_0^{D(k+1)} = \frac{\sum_{m}^{M} \sum_{i=1}^{I} \sum_{j=1}^{J} O_{ij,m}^D}{\sum_{m}^{M} \sum_{i=1}^{I} \sum_{j=1}^{J} E_{ij,m}^D \exp(X_{ij}\beta_k^3)E(\exp(\phi_{\mu}\mu_i + \phi_\omega \omega_{ij}) | \hat{\theta}(k))}
\]

where \( 0 \leq c_0^D < c_1^D < \ldots < c_M^D = \infty \) is the quantiles of death event times. \( I_m^D(t) = I(c_m^D - 1 < t \leq c_m^D) \). \( E_{ij,k,m}^D = I(D(T_{ij} \geq c_m^D - 1)(c_m^D \wedge T_{ij} - c_m^D - 1) \) the total time individual \( ij \) is at risk in the \( m^{th} \) death events time interval \( (c_m^D - 1, c_m^D) \).

\[
O_{ij,m}^D = I_m^D(T_{ij}) \Delta_{ij}, \text{ the number of death event for individual } ij \text{ in the } m^{th} \text{ interval.}
\]

The score components of \( \sigma_\mu^2 \) and \( \sigma_\omega^2 \) are:

\[
S(\sigma_\mu^2) = \frac{\partial Q_4}{\partial \sigma_\mu^2} = \sum_{i=1}^{I} \left( \frac{1}{\sigma_\mu^2} - \frac{E(\mu_i^2 | \hat{\theta}(k))}{\sigma_\mu^2} \right)
\]

\[
S(\sigma_\omega^2) = \frac{\partial Q_5}{\partial \sigma_\omega^2} = \sum_{i=1}^{I} \sum_{j=1}^{J} \left( \frac{1}{\sigma_\omega^2} - \frac{E(\omega_{ij}^2 | \hat{\theta}(k))}{\sigma_\omega^2} \right)
\]
The estimates for $\sigma^2_\mu$ and $\sigma^2_\omega$ are:

$$\hat{\sigma}^2_\mu = \frac{1}{I} \sum_{i=1}^{I} \mathbb{E}(\mu_i^2 \mid \hat{\theta}^{(k)})$$  \hspace{1cm} (A.7)

$$\hat{\sigma}^2_\omega = \frac{1}{IJ} \sum_{i=1}^{I} \sum_{j=1}^{J} \mathbb{E}(\omega_{ij}^2 \mid \hat{\theta}^{(k)})$$  \hspace{1cm} (A.8)

For the other parameters, which don’t have the closed-form estimator, the Newton-Raphson algorithm is used to solve the expected log-likelihood iteratively. Given the $k^{th}$ estimate $\hat{\theta}^{(k)}$, the $(k+1)^{th}$ estimate is obtained by

$$\hat{\beta}^{(k+1)}_2 = \hat{\beta}^{(k)}_2 + I^{-1}_2 (\hat{\beta}^{(k)}_2) S_\beta_2 (\hat{\beta}^{(k)}_2)$$  \hspace{1cm} (A.9)

Denote the parameters in $Q_3(\beta^T_3, \lambda^D_0, \phi_\mu, \phi_\omega)$ as $\tau = (\beta_3, \phi_\mu, \phi_\omega)$. The gradient function $g(\tau) = \left(\frac{\partial Q_3}{\partial \beta_3}, \frac{\partial Q_3}{\partial \phi_\mu}, \frac{\partial Q_3}{\partial \phi_\omega}\right)$, and the Hessian matrix can be obtained and denoted as $H$. Then, given current estimate $\hat{\theta}^{(k)}$, the $(k+1)^{th}$ estimate is obtained by

$$\hat{\tau}^{(k+1)} = \hat{\tau}^{(k)} - H^{-1} \hat{\tau}^{(k)} g(\hat{\tau}^{(k)})$$  \hspace{1cm} (A.10)

Similarly, the MLE of $\beta_1$ can be updated by:

$$\hat{\beta}^{(k+1)}_1 = \hat{\beta}^{(k)}_1 + I^{-1}_1 (\hat{\beta}^{(k)}_1) S_\beta_1 (\hat{\beta}^{(k)}_1)$$  \hspace{1cm} (A.11)

where $S(\cdot)$ and $I(\cdot)$ are the score function and information matrix, respectively.

To simplify the formula, we assume here that $X_{ij}$ only contains one covariate, which is denoted as $X_{ij}$. Then the corresponding score functions and information matrices of the parameters are:

$$S(\beta_1) = \frac{\partial Q_1}{\partial \beta_1} = \sum_{i=1}^{I} \sum_{j=1}^{J} \{\mathbb{E}(y_{ij} \mid \hat{\beta}^{(k)})X_{ij} - \exp(\beta_1 X_{ij})X_{ij} E[\exp(\mu_i) \mid \hat{\theta}^{(k)}]\}$$

$$I(\beta_1) = -\frac{\partial^2 Q_1}{\partial \beta_1^2} = \sum_{i=1}^{I} \sum_{j=1}^{J} \{\exp(\beta_1 X_{ij})X_{ij}^2 E[\exp(\mu_i) \mid \hat{\theta}^{(k)}]/(1 + q_{ij})^2 \}$$
\[ S(\beta_2) = \frac{\partial Q_2}{\partial \beta_2} = \sum_{i=1}^{I} \sum_{j=1}^{J} \left\{ \sum_{k}^{N_{ij}} \delta_{ijk} X_{ij} - \tilde{\Lambda}_0^R(T_{ij}) X_{ij} \exp(X_{ij}\beta_2) \right\} \]

\[ I(\beta_2) = -\frac{\partial^2 Q_2}{\partial \beta_2^2} = \sum_{i=1}^{I} \sum_{j=1}^{J} \left\{ \tilde{\Lambda}_0^R(T_{ij}) X_{ij}^2 \exp(X_{ij}\beta_2) \exp(y_{ij} \exp(\mu_i + \omega_{ij}) \mid \theta^{(k)}) \right\} \]

\[ S(\beta_3) = \frac{\partial Q_3}{\partial \beta_3} = \sum_{i=1}^{I} \sum_{j=1}^{J} \left\{ \Delta_{ij} X_{ij} - \tilde{\Lambda}_0^P(T_{ij}) X_{ij} \exp(X_{ij}\beta_3) \right\} \]

\[ E(\exp(\phi_{\mu}\mu_i + \phi_{\omega}\omega_{ij}) \mid \theta^{(k)}) \]

\[ S(\phi_{\mu}) = \frac{\partial Q_3}{\partial \phi_{\mu}} = \sum_{i=1}^{I} \sum_{j=1}^{J} \left\{ \Delta_{ij} E(\mu_i \mid \theta^{(k)}) - \tilde{\Lambda}_0^P(T_{ij}) X_{ij} \exp(X_{ij}\beta_3) \right\} \]

\[ E(\mu_i \exp(\phi_{\mu}\mu_i + \phi_{\omega}\omega_{ij}) \mid \theta^{(k)}) \]

\[ S(\phi_{\omega}) = \frac{\partial Q_3}{\partial \phi_{\omega}} = \sum_{i=1}^{I} \sum_{j=1}^{J} \left\{ \Delta_{ij} E(\omega_{ij} \mid \theta^{(k)}) - \tilde{\Lambda}_0^P(T_{ij}) X_{ij} \exp(X_{ij}\beta_3) \right\} \]

\[ E(\omega_{ij} \exp(\phi_{\mu}\mu_i + \phi_{\omega}\omega_{ij}) \mid \theta^{(k)}) \]

\[ \frac{\partial^2 Q_3}{\partial \beta_3^2} = \sum_{i=1}^{I} \sum_{j=1}^{J} \left\{ - \tilde{\Lambda}_0^P(T_{ij}) X_{ij}^2 \exp(X_{ij}\beta_3) E(\exp(\phi_{\mu}\mu_i + \phi_{\omega}\omega_{ij}) \mid \theta^{(k)}) \right\} \]

\[ \frac{\partial^2 Q_3}{\partial \phi_{\mu}^2} = \sum_{i=1}^{I} \sum_{j=1}^{J} \left\{ - \tilde{\Lambda}_0^P(T_{ij}) \exp(X_{ij}\beta_3) E(\mu_i^2 \exp(\phi_{\mu}\mu_i + \phi_{\omega}\omega_{ij}) \mid \theta^{(k)}) \right\} \]

\[ \frac{\partial^2 Q_3}{\partial \phi_{\omega}^2} = \sum_{i=1}^{I} \sum_{j=1}^{J} \left\{ - \tilde{\Lambda}_0^P(T_{ij}) \exp(X_{ij}\beta_3) E(\omega_{ij}^2 \exp(\phi_{\mu}\mu_i + \phi_{\omega}\omega_{ij}) \mid \theta^{(k)}) \right\} \]

\[ \frac{\partial^2 Q_3}{\partial \beta_3 \phi_{\mu}} = \sum_{i=1}^{I} \sum_{j=1}^{J} \left\{ - \tilde{\Lambda}_0^P(T_{ij}) X_{ij} \exp(X_{ij}\beta_3) E(\mu_i \exp(\phi_{\mu}\mu_i + \phi_{\omega}\omega_{ij}) \mid \theta^{(k)}) \right\} \]

\[ \frac{\partial^2 Q_3}{\partial \beta_3 \phi_{\omega}} = \sum_{i=1}^{I} \sum_{j=1}^{J} \left\{ - \tilde{\Lambda}_0^P(T_{ij}) X_{ij} \exp(X_{ij}\beta_3) E(\omega_{ij} \exp(\phi_{\mu}\mu_i + \phi_{\omega}\omega_{ij}) \mid \theta^{(k)}) \right\} \]

\[ \frac{\partial^2 Q_3}{\partial \phi_{\mu} \phi_{\omega}} = \sum_{i=1}^{I} \sum_{j=1}^{J} \left\{ - \tilde{\Lambda}_0^P(T_{ij}) \exp(X_{ij}\beta_3) E(\mu_i \omega_{ij} \exp(\phi_{\mu}\mu_i + \phi_{\omega}\omega_{ij}) \mid \theta^{(k)}) \right\} \]
More components of the Information matrix are given below:

$$\frac{\partial^2 l(\theta)}{\partial \sigma^2_\mu} = -\frac{1}{2} \left\{ - \frac{I}{\sigma^4_\mu} + \frac{2 \sum_{i=1}^{I} \text{E}(\mu^2_i | \hat{\theta}^{(k)})}{\sigma^6_\mu} \right\}$$

$$\frac{\partial^2 l(\theta)}{\partial \sigma^2_\omega} = -\frac{1}{2} \left\{ - \frac{IJ}{\sigma^4_\omega} + \frac{2 \sum_{i=1}^{I} \sum_{j=1}^{J} \text{E}(\omega^2_{ij} | \hat{\theta}^{(k)})}{\sigma^6_\omega} \right\}$$

All other off-diagonal terms are zero. When $X_{ij}$ is a covariate vector, corresponding score functions and information matrices of the parameters can be easily adapted to matrix versions.

## A.4 Data generation

In this section, we give a brief description about the simulation procedure. Based on the definition of a Poisson process, for each $0 \leq s \leq t$, $N(t) - N(s)$ has a Poisson distribution with mean $m(t) - m(s) = \int_s^t \lambda(x)dx$. Therefore, we need to simulate recurrent event times from a Poisson process with any intensity function $\lambda(t)$. The most common parametric choices for time-to-event simulation are Exponential, Weibull and Gompertz distributions. As noted by Bender et al. (2005), among the commonly used distributions for survival times, only these three distributions share the assumption of proportional hazards, and the survival time, $T$, can be generated by $T = \Lambda^{-1}_0[- \log(u) \exp(-\beta^T x)]$, where $u \sim \text{Unif}[0, 1]$.

There are several available methods for recurrent event times simulation, and here we adopt the inversion method which is often used. Referring to the inversion method by Pénichoux et al. (2015), for a given subject, let $T_j$ denote the time elapsed from the origin to the $j^{th}$ event, $T_0 = 0$, and let $W_j = T_j - T_{j-1}$ denote the $j^{th}$ waiting time between two consecutive events. The $W_j$ are referred to as gap times. The number of events observed for a non-homogeneous Poisson process with intensity $\lambda(\cdot)$ between two times, $s$ and $t$ follows a Poisson distribution with $\int_s^t \lambda(v)dv$ which leads to

$$F_j(w) = 1 - \exp(- \int_{T_{j-1}}^{T_{j-1}+w} \lambda(v)dv)$$
\[ W_j = F_j^{-1}(u_j) \text{ and } T_j = T_{j-1} + W_j. \]

For simplicity, we adopt an exponential distribution with constant baseline constant \( \lambda_0(t) = \lambda \), the recurrent event process can be considered equivalently as an homogeneous Poisson process or a sequence of exponentially distributed gaps (Pénichoux et al., 2015). The survival times are generated by

\[
T = -\frac{\log(u)}{\lambda \exp(-\beta^T x)}
\]

\[
T_j = T_{j-1} + W_j = -\frac{\log(u_j)}{\lambda \exp(-\beta^T x)} + T_{j-1}
\]

For Weibull intensity with \( \lambda_0(t) = \lambda \nu t^{\nu-1} \) and \( \Lambda_0(t) = \lambda t^\nu \), the event rate increases (if \( \nu > 1 \)), decreases (if \( 0 < \nu < 1 \)), or stays constant (if \( \nu = 1 \)) over time. The inverse cumulative hazard function is \( \Lambda_0^{-1}(t) = (\lambda^{-1} t)^{\frac{1}{\nu}} \). The survival times are generated by

\[
T = \left(-\frac{\log(u)}{\lambda \exp(-\beta^T x)}\right)^\frac{1}{\nu}
\]

\[
T_j = T_{j-1} + W_j = \left(-\frac{\log(u_j)}{\lambda \exp(-\beta^T x)} + T_{j-1}^{\nu}\right)^\frac{1}{\nu}
\]

### A.5 Supplementary simulation results

In this section, we conduct simulation studies to evaluate the joint frailty model without inflation with Brewslow estimator baseline intensities (Denoted as Model II).

First, we introduce the estimation for Brewslow baseline intensities. Let \( \lambda_0^R(t) \) take values only at distinct recurrent events time \( t_{ijk} \) for which \( \delta_{ijk} = 1 \), and \( \lambda_0^D(t) \) take values only at distinct death event time \( D_{ij} \). For other time points, \( \lambda_0^R(t) \) and \( \lambda_0^D(t) \) are 0. To keep it simple, we assume that there are no ties.

As described before, \( \Psi_{ij}(t) = I(t \leq T_{ij}) \) is the “at-risk” indicator to show whether the subject is still under observation at time \( t \) or not. Then the cumulative baseline
intensities for recurrent events and the terminal event become:

\[
\Lambda_0^R(T_{ij}) = \sum_{t_{ijk} \leq T_{ij}} \lambda_0^R(t_{ijk}) = \sum_i \sum_j \sum_k \lambda_0^R(t_{ijk}) \Psi_{ijk}(t_{ijk})
\] (A.12)

\[
\Lambda_0^{D(k+1)}(T_{ij}) = \sum_{D_{ij} \leq T_{ij}} \lambda_0^D(D_{ij}) = \sum_i \sum_j \lambda_0^D(D_{ij}) \Psi_{ij}(D_{ij})
\] (A.13)

It can be shown that the value of \( \lambda_0^R \) which maximizes \( Q_2(\beta^R, \lambda_0^R) \) is \( \hat{\lambda}_0^R = (\hat{\lambda}_0^R(t_1), \ldots, \hat{\lambda}_0^R(t_{ijk}), \ldots)^T \).

\[
\hat{\lambda}_0^{R(k+1)}(t_{ijk}) = \frac{\delta_{ijk}}{\sum_i \sum_j \sum_k \exp(X_{ij} \beta_2) E(\exp(\mu_i + \omega_{ij}) | \hat{\theta}^{(k)}) \Psi_{ijk}(t_{ijk})}
\] (A.14)

Similarly, \( \hat{\lambda}_0^D = (\hat{\lambda}_0^D(D_1), \ldots, \hat{\lambda}_0^D(D_{ij}), \ldots)^T \), where

\[
\hat{\lambda}_0^{D(k+1)}(D_{ij}) = \frac{I(\Delta_{ij} = 1)}{\sum_i \sum_j \exp(X_{ij} \beta_3) E(\exp(\phi_\mu \mu_i + \phi_\omega \omega_{ij}) | \hat{\theta}^{(k)}) \Psi_{ij}(D_{ij})}
\] (A.15)

Other derivation is similar to the M step in A.3 with \( p_{ij} = 1 \) with Brewslow baseline intensity estimators.

For simplicity, we only consider one covariate for \( X \), which takes values of 0 or 1 each with probability 0.5. Considering exponential distributions for both recurrent event times and death, the following models are shown below

\[
\begin{align*}
\lambda_{ij}^R(t \mid \mu_i, \omega_{ij}) &= \lambda_0^R(t) \exp(\beta_2 X_{ij} + \mu_i + \omega_{ij}) \\
\lambda_{ij}^D(t \mid \mu_i, \omega_{ij}) &= \lambda_0^D(t) \exp(\beta_3 X_{ij} + \phi_\mu \mu_i + \phi_\omega \omega_{ij})
\end{align*}
\] (A.16)

The censoring time is taken as \( C_{ij} = 6 \ast \text{Unif}(0, 1) \), where Unif(0,1) is a random number generated from a uniform distribution in \([0, 1]\). We assume \( \mu_i \sim N(0, \sigma^2_\mu) \), \( \omega_{ij} \sim N(0, \sigma^2_\omega) \). We use exponential baseline for recurrent events and death: \( \lambda_0^R = 0.5 \) and \( \lambda_0^D = 0.1 \). We set the parameters as \( \beta_2 = 1.2, \beta_3 = 1, \phi_\mu = 1, \phi_\omega = 0.5, \sigma^2_\mu = 0.25 \), but invoke different values for \( \sigma^2_\omega \). We would like to see the effects of matched pair correlation in the model with and without considering \( \mu \) by setting
different values of $\sigma^2_\mu$. We let $\sigma^2_\mu = 1, 0.12, 0.01$. We consider two set-ups of sample size: 500 and 250.

A total of 500 Monte Carlo replicates are generated. A non-homogeneous Poisson process is adopted to simulate recurrent event times and the detailed procedures are provided in Appendix A.4. 65% subjects are observed with at least one recurrent event. The percentage of death is 40%.

The resulting parameter estimates are shown in Table A.1, where for each estimator Monte Carlo mean (denoted by MCM), Monte Carlo standard error (denoted by MCSE) and mean square error (MSE) are reported. In Table A.1, all parameter estimates have small biases in Model II for both sample size 250 and 500, except for parameter $\phi_\omega$, which has a little bit larger bias. To show the effects of matched pair correlation on the model, we also consider an alternative model (Model III in Table A.1), in which we removed $\mu$. It can be seen that ignoring the correlation in the matched pair will result in very biased parameter estimates in the variance and coefficient of random effects $\omega$ when the correlation is not negligible (e.g., $\sigma^2_\mu = 1$).
Table A.1. Simulation results of Setting I for Model II with Breslow baseline estimator: \( N = 250, \lambda_0^R = 0.5, \lambda_0^D = 0.1 \): 65% subjects are observed with at least one recurrent event. The percentage of death is 40%. Model II denotes the model without considering zero-inflation for recurrent events, which is the full model here; Model III denotes the model without considering the matched pair correlation based on Model II, where matched pair random effect \( \mu \) is removed:

<table>
<thead>
<tr>
<th>Setting I</th>
<th>Model II</th>
<th>Model III</th>
</tr>
</thead>
<tbody>
<tr>
<td>( N = 250 )</td>
<td>Parameters</td>
<td>True MCM MCSE MSE</td>
</tr>
<tr>
<td>( \sigma^2_\mu = 1 )</td>
<td>( \beta_2 )</td>
<td>1.2</td>
</tr>
<tr>
<td></td>
<td>( \beta_3 )</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>( \phi_\mu )</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>( \phi_\omega )</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>( \sigma^2_\beta )</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>( \sigma^2_\omega )</td>
<td>0.25</td>
</tr>
<tr>
<td>( \sigma^2_\mu = 0.12 )</td>
<td>( \beta_2 )</td>
<td>1.2</td>
</tr>
<tr>
<td></td>
<td>( \beta_3 )</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>( \phi_\mu )</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>( \phi_\omega )</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>( \sigma^2_\beta )</td>
<td>0.12</td>
</tr>
<tr>
<td></td>
<td>( \sigma^2_\omega )</td>
<td>0.25</td>
</tr>
<tr>
<td>( \sigma^2_\mu = 0.01 )</td>
<td>( \beta_2 )</td>
<td>1.2</td>
</tr>
<tr>
<td></td>
<td>( \beta_3 )</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>( \phi_\mu )</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>( \phi_\omega )</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>( \sigma^2_\beta )</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>( \sigma^2_\omega )</td>
<td>0.25</td>
</tr>
</tbody>
</table>

MCM is the Monte Carlo mean of the parameter estimates (bases on 500 replicates); MCSE is the Monte Carlo standard error; MSE is the mean square error.
Table A.2. Simulation results of Setting II for Model II with Breslow baseline estimator: \( N = 500, \lambda_0^R = 0.5, \lambda_0^D = 0.1 \): 65% subjects are observed with at least one recurrent event. The percentage of death is 40%. Model II denotes the model without considering zero-inflation for recurrent events, which is the full model here; Model III denotes the model without considering the matched pair correlation based on Model II, where matched pair random effect \( \mu \) is removed;

<table>
<thead>
<tr>
<th>Setting II</th>
<th>Model II</th>
<th>Model III</th>
</tr>
</thead>
<tbody>
<tr>
<td>( N = 500 )</td>
<td>Parameters</td>
<td>True</td>
</tr>
<tr>
<td>( \sigma^2_\mu = 1 )</td>
<td>( \beta_2 )</td>
<td>1.2</td>
</tr>
<tr>
<td></td>
<td>( \beta_3 )</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>( \phi_\mu )</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>( \phi_\omega )</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>( \sigma^2_\beta )</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>( \sigma^2_\omega )</td>
<td>0.25</td>
</tr>
<tr>
<td>( \sigma^2_\mu = 0.12 )</td>
<td>( \beta_2 )</td>
<td>1.2</td>
</tr>
<tr>
<td></td>
<td>( \beta_3 )</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>( \phi_\mu )</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>( \phi_\omega )</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>( \sigma^2_\beta )</td>
<td>0.12</td>
</tr>
<tr>
<td></td>
<td>( \sigma^2_\omega )</td>
<td>0.25</td>
</tr>
<tr>
<td>( \sigma^2_\mu = 0.01 )</td>
<td>( \beta_2 )</td>
<td>1.2</td>
</tr>
<tr>
<td></td>
<td>( \beta_3 )</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>( \phi_\mu )</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>( \phi_\omega )</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>( \sigma^2_\beta )</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>( \sigma^2_\omega )</td>
<td>0.25</td>
</tr>
</tbody>
</table>

MCM is the Monte Carlo mean of the parameter estimates (bases on 500 replicates); MCSE is the Monte Carlo standard error; MSE is the mean square error.
Chapter 4

In the M-step, first we have closed-form estimators for baseline intensities $\tilde{\lambda}_0^R(t)$, $\tilde{\lambda}_0^D(t)$ and variance components $\sigma^2_\mu, \sigma^2_\omega$.

B.1 Baseline intensities

B.1.1 Recurrent events

We introduce several notations for the estimation of $\tilde{\lambda}_0^R(t)$. Assume there are $M$ intervals with cutpoints $0 \leq c_0^R < c_1^R < \ldots < c_M^R = \infty$, which can be quantiles of recurrent event times, where $c_0^R = 0$ or the smallest event time. Denote the indicator function $I_{R_m}(t) = I(c_{m-1}^R < t \leq c_m^R)$. For $m = 1, 2, \ldots, M$,

$$\tilde{\lambda}_0^R(t) = \sum_{m=1}^{M} \lambda_m^R I(c_{m-1}^R < t \leq c_m^R) = \sum_{m=1}^{M} \lambda_m^R I_{R_m}(t) \quad (B.1)$$

Therefore, we denote by $\mathbf{\lambda}^R = (\lambda_1^R, \lambda_2^R, \ldots, \lambda_m^R, \ldots, \lambda_M^R)$ the parameter we aim to
estimate. The expected log likelihood part involving $\tilde{\lambda}_0(t)$ is

$$Q_1(\beta_R, \eta_R, \tilde{\lambda}_0^R(t) | \hat{\theta}^{(k)})$$

$$= \sum_{i=1}^{I} \sum_{j=1}^{J} \left\{ (\sum_{m=1}^{N_{ij}}) \left( \log(\tilde{\lambda}_0^R(t_{ijk})) + \eta_R(\tilde{\beta}_t + \hat{b}_{1ij})t_{ijk} + q_{1ij} + E[\mu_i | \hat{\theta}^{(k)}] 
+ E[\omega_{ij} | \hat{\theta}^{(k)}]) - \tilde{\Lambda}_0^R(T_{ij}) \exp(q_{1ij})E[\exp(\mu_i + \omega_{ij}) | \hat{\theta}^{(k)}] \right\}$$

For computational purpose, the log-likelihood can be written in a Poisson regression form:

$$Q_1(\lambda^R | \hat{\theta}^{(k)}) =$$

$$\sum_{m=1}^{M} \left\{ \sum_{i=1}^{I} \sum_{j=1}^{J} \left\{ \left( \sum_{k=1}^{N_{ij}} \left( \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{N_{ij}} \delta_{ijk} \frac{1}{\lambda_m^R} \right) - \sum_{i=1}^{I} \sum_{j=1}^{J} S_{ij,m}^{R} \right) \right\} \right\}$$

Denote $S(\cdot)$ is the score function component. Let $O_{ij,m}^R = \sum_{k=1}^{N_{ij}} I_{m}^R(t_{ijk}) \delta_{ijk}$, the number of recurrent events for individual $ij$ in the $m^{th}$ subinterval. Then

$$S(\lambda^R) = \frac{\partial Q_1(\lambda^R | \hat{\theta}^{(k)})}{\partial \lambda^R}$$

$$= \sum_{m=1}^{M} \left\{ \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{N_{ij}} \frac{1}{\lambda_m^R} - \sum_{i=1}^{I} \sum_{j=1}^{J} S_{ij,m}^{R} \right\}$$

$$= \sum_{m=1}^{M} \left\{ \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{O_{ij,m}^R}{\lambda_m^R} - \sum_{i=1}^{I} \sum_{j=1}^{J} S_{ij,m}^{R} \right\}$$
, where

\[ S_{ij,m}^{R} = \frac{\partial}{\partial \lambda^{R}} \left\{ \tilde{\lambda}^{R}(T_{ij}) \exp(q_{1ij}) \mathbb{E}[\exp(\mu_{i} + \omega_{ij}) \mid \hat{\theta}^{(k)}] \right\} \]

\[ = \frac{\partial}{\partial \lambda^{R}} \left\{ \int_{t_{m-1}^{R}}^{t_{m}^{R}} \lambda^{R}_{m} \exp(\eta_{R} p_{t} t) dT_{ij} \exp(q_{1ij}) \mathbb{E}[\exp(\mu_{i} + \omega_{ij}) \mid \hat{\theta}^{(k)}] \right\} \]

\[ = \frac{\partial}{\partial \lambda^{R}_{m}} \left\{ \frac{\lambda^{R}_{m}}{\eta_{R} p_{t}} \left[ \exp(\eta_{R} p_{t} t_{m}^{R}) - \exp(\eta_{R} p_{t} t_{m-1}^{R}) \right] \right. \]

\[ \times \exp(q_{1ij}) \mathbb{E}[\exp(\mu_{i} + \omega_{ij}) \mid \hat{\theta}^{(k)}] \right\} \]

Here, \( t_{m}^{R} \) and \( t_{m-1}^{R} \) are upper and lower bounds at \( m \)th interval.

\( t_{m}^{R} = \max(0, \min(c_{m}^{R} - c_{m-1}^{R}, T_{ij} - c_{m-1}^{R})) \). To simplify the expression, we denote \( p_{t} = \hat{\beta}_{t} + \hat{b}_{1ij} \).

Solve \( S(\lambda^{R}) = 0 \). The maximum likelihood estimator has an explicit solution:

\[ \hat{\lambda}_{m}^{R(k+1)} = \frac{\sum_{m=1}^{M} \left[ \sum_{i=1}^{I} \sum_{j=1}^{J} O_{ij,m}^{R} \right]}{\sum_{m=1}^{M} \left[ \sum_{i=1}^{I} \sum_{j=1}^{J} S_{ij,m}^{R} \right]} \] (B.2)

**B.1.2 Death event**

The estimation of \( \hat{\lambda}_{0}^{D} \) is similar to the estimation of \( \hat{\lambda}_{0}^{R} \). \( 0 \leq c_{0}^{D} < c_{1}^{D} < \ldots < c_{M}^{D} = \infty \) is the quantiles of death event times. Note that the selection of \( M \) for recurrent events and death can be the same or different. Denote the indicator function \( I_{m}^{D}(t) = I(c_{m-1}^{D} < t \leq c_{m}^{D}) \). For \( m = 1, 2, \ldots, M \),

\[ \hat{\lambda}_{0}^{D}(t) = \sum_{m=1}^{M} \lambda_{m}^{D} I(c_{m-1}^{D} < t \leq c_{m}^{D}) \] (B.3)
Therefore, we denote by $\mathbf{\lambda}^D = (\lambda_1^D, \lambda_2^D, \ldots, \lambda_m^D, \ldots, \lambda_M^D)$ the parameter we aim to estimate for baseline intensity function of death events.

For computational purpose, the log-likelihood can be written as:

$$Q_2(\mathbf{\lambda}^D \mid \hat{\theta}^{(k)}) =$$

$$\sum_{m=1}^{M} \left\{ \sum_{i=1}^{I} \sum_{j=1}^{J} \left\{ I_m^D(T_{ij}) \Delta_{ij} \left[ \log(\lambda_m^D) + \eta_D(\hat{\beta}_t + \hat{b}_{1ij})T_{ij} + q_{2ij} + \phi_\mu E[\mu_i \mid \hat{\theta}^{(k)}] + \phi_\omega E[\omega_{ij} \mid \hat{\theta}^{(k)}] \right] - \sum_{i=1}^{I} \sum_{j=1}^{J} \bar{\Lambda}_m^D(T_{ij}) \exp(q_{2ij})E[\exp(\phi_\mu \mu_i + \phi_\omega \omega_{ij}) \mid \hat{\theta}^{(k)}] \right\} \right\}$$

Let $O_{ij,m}^D = I_m^D(T_{ij})\Delta_{ij}$, which is the number of death event for individual $ij$ in the $m^{th}$ interval. The score component is:

$$S(\mathbf{\lambda}^D) = \frac{\partial Q_2(\mathbf{\lambda}^D \mid \hat{\theta}^{(k)})}{\partial \mathbf{\lambda}^D} = \sum_{m=1}^{M} \left\{ \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{O_{ij,m}^D}{\lambda_m^D} - \sum_{i=1}^{I} \sum_{j=1}^{J} S_{ij,m}^D \right\}$$

$$S_{ij,m}^D = \frac{\partial \left\{ \bar{\Lambda}_m^D(T_{ij}) \exp(q_{2ij})E[\exp(\phi_\mu \mu_i + \phi_\omega \omega_{ij}) \mid \hat{\theta}^{(k)}] \right\}}{\partial \lambda_m^D}$$

$$= \frac{\partial \left\{ \int_{t_{m-1}^D}^{t_m^D} \lambda_m^D \exp(\eta_D p_i T_{ij}) dt_{ij} \exp(q_{2ij})E[\exp(\phi_\mu \mu_i + \phi_\omega \omega_{ij}) \mid \hat{\theta}^{(k)}] \right\}}{\partial \lambda_m^D}$$

$$= \frac{\partial \left\{ \lambda_m^D \left[ \exp(\eta_D p_i t_m^D) - \exp(\eta_D p_i t_{m-1}^D) \right] \exp(q_{2ij})E[\exp(\phi_\mu \mu_i + \phi_\omega \omega_{ij}) \mid \hat{\theta}^{(k)}] \right\}}{\partial \lambda_m^D}$$

$$= \frac{1}{\eta_D p_i} \left[ \exp(\eta_D p_i t_m^D) - \exp(\eta_D p_i t_{m-1}^D) \right] \exp(q_{2ij})E[\exp(\phi_\mu \mu_i + \phi_\omega \omega_{ij}) \mid \hat{\theta}^{(k)}]$$

$t_m^D$ and $t_{m-1}^D$ are upper and lower bounds at $m^{th}$ interval. $t_m^D = \max(0, \min(c_m^D - c_{m-1}^D, T_{ij} - c_{m-1}^D))$. To simplify the expression, we denote $p_t = \hat{\beta}_t + \hat{b}_{1ij}$. 
Solve $S(\lambda^D) = 0$. The maximum likelihood estimator has an explicit solution:

$$
\hat{\lambda}_m^{D(k+1)} = \frac{\sum_{m=1}^M \left[ \sum_{i=1}^I \sum_{j=1}^J O^D_{ij,m} \right]}{\sum_{m=1}^M \left[ \sum_{i=1}^I \sum_{j=1}^J S^D_{ij,m} \right]}
$$

(B.4)

**B.2 Variance components**

Then the corresponding score functions and second partial derivatives for $\sigma^2_\mu$ and $\sigma^2_\omega$ are:

$$
S(\sigma^2_\mu) = \frac{\partial Q_3}{\partial \sigma^2_\mu} = \sum_{i=1}^I \left\{ - \frac{1}{2\sigma^2_\mu} + \frac{E(\mu_i^2 \ | \ \hat{\theta}^{(k)})}{2\sigma^4_\mu} \right\}
$$

$$
S(\sigma^2_\omega) = \frac{\partial Q_3}{\partial \sigma^2_\omega} = \sum_{i=1}^I \sum_{j=1}^J \left\{ - \frac{1}{2\sigma^2_\omega} + \frac{E(\omega^2_{ij} \ | \ \hat{\theta}^{(k)})}{2\sigma^4_\omega} \right\}
$$

The maximum likelihood estimators for $\sigma^2_\mu$ and $\sigma^2_\omega$ are:

$$
\hat{\sigma}^2_\mu^{(k+1)} = \frac{1}{I} \sum_{i=1}^I E(\mu_i^2 \ | \ \hat{\theta}^{(k)})
$$

(B.5)

$$
\hat{\sigma}^2_\omega^{(k+1)} = \frac{1}{IJ} \sum_{i=1}^I \sum_{j=1}^J E(\omega^2_{ij} \ | \ \hat{\theta}^{(k)})
$$

(B.6)

of note that $IJ$ is the total number of subjects.

**B.3 Newton-Raphson algorithm**

For the other parameters, which don’t have the closed-form estimator, the Newton-Raphson algorithm is used to solve the expected log-likelihood iteratively. Denote $\tau = (\beta_R, \beta_D, \phi_\mu, \phi_\omega, \eta_R, \eta_D)$. The gradient function $G(\tau)$ is a score vector, and $H_{\tau}(\cdot)$ is the Hessian matrix. Then a second Taylor series expansion of the log
liklihood \( l(\tau) \), th estimates at iteration \( k \) gives:

\[
g(\tau)^{(k)} = l(\tau) + G^{(k)T}(\tau - \tau^{(k)}) + \frac{1}{2}(\tau - \tau^{(k)})^T H_\tau(\tau - \tau^{(k)})
\]

differentiating and setting equal to zero gives the next estimate

\[
\hat{\tau}^{(k+1)} = \hat{\tau}^{(k)} - H_\tau^{-1}(\hat{\tau}^{(k)}) G_\tau(\hat{\tau}^{(k)})
\] (B.7)

Recall that we denote \( q_{1ij} = \beta^T R X^{(R)}_{ij} + \eta_R(X^{(L)}_{1ij}(t) \hat{\beta}_{L1} + \hat{b}_{0ij}) \), and \( q_{2ij} = \beta^T D X^{(D)}_{ij} + \eta_D(X^{(L)}_{1ij}(t) \hat{\beta}_{L1} + \hat{b}_{0ij}) \). To simplify the formula, we assume here that \( X^{R}_{ij} \) and \( X^{D}_{ij} \) only share the same covariate (i.e. baseline event status), which is denoted as \( X_{ij} \).

Then the corresponding score functions and second partial derivatives the parameters are:

\[
S(\beta_R) = \frac{\partial Q_1}{\partial \beta_R} = \sum_{i=1}^{I} \sum_{j=1}^{J} \{ \sum_{k=1}^{K_R} \delta_{ijk} X_{ij} - X_{ij} \tilde{\Lambda}_0^R(T_{ij}) \exp(q_{1ij}) E[\exp(\mu_i + \omega_{ij}) | \hat{\theta}^{(k)}] \}
\]

\[
S(\beta_D) = \frac{\partial Q_2}{\partial \beta_D} = \sum_{i=1}^{I} \sum_{j=1}^{J} \{ \Delta_{ij} X_{ij} - X_{ij} \tilde{\Lambda}_0^D(T_{ij}) \exp(q_{2ij}) E[\exp(\phi_\mu \mu_i + \phi_\omega \omega_{ij}) | \hat{\theta}^{(k)}] \}
\]

\[
S(\phi_\mu) = \frac{\partial Q_2}{\partial \phi_\mu} = \sum_{i=1}^{I} \sum_{j=1}^{J} \{ \Delta_{ij} E(\mu_i | \hat{\theta}^{(k)}) - \tilde{\Lambda}_0^D(T_{ij}) \exp(q_{2ij}) E[\mu_i \exp(\phi_\mu \mu_i + \phi_\omega \omega_{ij}) | \hat{\theta}^{(k)}] \}
\]

\[
S(\phi_\omega) = \frac{\partial Q_2}{\partial \phi_\omega} = \sum_{i=1}^{I} \sum_{j=1}^{J} \{ \Delta_{ij} E(\omega_{ij} | \hat{\theta}^{(k)}) - \tilde{\Lambda}_0^D(T_{ij}) \exp(q_{2ij}) E[\omega_{ij} \exp(\phi_\mu \mu_i + \phi_\omega \omega_{ij}) | \hat{\theta}^{(k)}] \}
\]

\[
S(\eta_R) = \frac{\partial Q_1}{\partial \eta_R} = \sum_{i=1}^{I} \sum_{j=1}^{J} \{ \sum_{k=1}^{K_R} \delta_{ijk} ((\hat{\beta}_i + \hat{b}_{1ij})t_{ijk} + X^{(L)}_{1ij}(t) \hat{\beta}_{L1} + \hat{b}_{0ij}) \} - P_{ \eta_R} \]
\[ S(\eta_D) = \frac{\partial Q_2}{\partial \eta_D} = \sum_{i=1}^{I} \sum_{j=1}^{J} \{ \Delta_{ij} (\hat{\beta}_t + \hat{b}_1 t)_{ij k} + X_{ij}^{(L)}(t) \hat{\beta}_{L1} + \hat{b}_0 i j \} - P^{\eta_D} \]

To simplify the expression, we denote \( p_t = \hat{\beta}_t + \hat{b}_1 i j \) and \( f_t = X_{ij}^{(L)}(t) \hat{\beta}_{L1} + \hat{b}_0 i j \).

\[ P^{\eta_R} = \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{\partial}{\partial \eta_R} \left\{ \frac{\tilde{\Lambda}_R^R(T_{ij}) \exp(q_{1ij})E[\exp(\mu_i + \omega_{ij}) | \hat{\theta}^{(k)}]}{\partial \eta_R} \right\} \]

\[ = \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{\partial}{\partial \eta_R} \left\{ \sum_{m=1}^{M} \lambda_m^R \exp(\eta_R p_t t^R_m) dt \exp(q_{1ij})E[\exp(\mu_i + \omega_{ij}) | \hat{\theta}^{(k)}] \right\} \]

\[ = \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{1}{f_i \eta_R^2} \exp(q_{1ij})E[\exp(\mu_i + \omega_{ij}) | \hat{\theta}^{(k)}] \]

\[ \sum_{m=1}^{M} \lambda_m^R \left[ \exp((f_i + p_t t^R_m) \eta_R - 1) \exp(\eta_R p_t t^R_m) - \exp((f_i + p_t t^R_{m-1}) \eta_R - 1) \exp(\eta_R p_t t^R_{m-1}) \right] \]
The second partial derivative for the parameters are:

\[
\frac{\partial^2 Q_1}{\partial \beta_R^2} = - \sum_{i=1}^I \sum_{j=1}^J X_{ij}^2 \tilde{\Lambda}_0^R (T_{ij}) \exp(q_{1ij}) E[\exp(\mu_i + \omega_{ij}) | \hat{\theta}^{(k)}]
\]

\[
\frac{\partial^2 Q_2}{\partial \beta_D^2} = - \sum_{i=1}^I \sum_{j=1}^J X_{ij}^2 \tilde{\Lambda}_0^D (T_{ij}) \exp(q_{2ij}) E[\exp(\phi_{\mu} + \phi_{\omega}) | \hat{\theta}^{(k)}]
\]

\[
\frac{\partial^2 Q_2}{\partial \phi_D^2} = - \sum_{i=1}^I \sum_{j=1}^J \left\{ \tilde{\Lambda}_0^D (T_{ij}) \exp(q_{2ij}) E(\mu_i^2 \exp(\phi_{\mu} + \phi_{\omega}) | \hat{\theta}^{(k)}) \right\}
\]

\[
\frac{\partial^2 Q_2}{\partial \phi_D^2} = - \sum_{i=1}^I \sum_{j=1}^J \left\{ \tilde{\Lambda}_0^D (T_{ij}) \exp(q_{2ij}) E(\omega_i^2 \exp(\phi_{\mu} + \phi_{\omega}) | \hat{\theta}^{(k)}) \right\}
\]

\[
\frac{\partial^2 Q_1}{\partial \eta_R^2} = - \frac{\partial P_{nr}}{\partial \eta_R}
\]

\[
\frac{\partial^2 Q_2}{\partial \eta_D^2} = - \frac{\partial P_{nd}}{\partial \eta_D}
\]
where

\[
\frac{\partial P^{\eta R}}{\partial \eta_R} = \partial \left\{ \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{1}{f_i \eta_R^2} \exp(q_{ij}) E[\exp(\mu_i + \omega_{ij}) | \hat{\theta}^{(k)}] \right. \\
+ \sum_{m=1}^{M} \lambda_m^{R} \left[ ((f_i + p_t R) \eta_R - 1) \exp(\eta_R p_t R) - \\
((f_i + p_t R_{m-1}) \eta_R - 1) \exp(\eta_R p_t R_{m-1}) \right] / \partial \eta_R \\
\left. \right\} \\
= \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{1}{f_i \eta_R^3} \exp(q_{ij}) E[\exp(\mu_i + \omega_{ij}) | \hat{\theta}^{(k)}] \\
+ \sum_{m=1}^{M} \lambda_m^{R} \left[ ((f_i^2 + p_t R) \eta_R)^2 - 2\left[ (f_i + p_t R) \eta_R + 1 \right] \exp(\eta_R p_t R) - \\
\left[ (f_i^2 + p_t R_{m-1}) \eta_R \right]^2 - 2\left[ (f_i + p_t R_{m-1}) \eta_R + 1 \right] \exp(\eta_R p_t R_{m-1}) \right] \\
\]

\[
\frac{\partial P^{\eta D}}{\partial \eta_D} = \partial \left\{ \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{1}{f_i \eta_D^2} \exp(q_{2ij}) E[\exp(\phi_i \mu_i + \phi_i \omega_{ij}) | \hat{\theta}^{(k)}] \right. \\
+ \sum_{m=1}^{M} \lambda_m^{D} \left[ ((f_i + p_t D) \eta_D - 1) \exp(\eta_D p_t D) - \\
((f_i + p_t D_{m-1}) \eta_D - 1) \exp(\eta_D p_t D_{m-1}) \right] / \partial \eta_D \\
\left. \right\} \\
= \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{1}{f_i \eta_D^3} \exp(q_{2ij}) E[\exp(\phi_i \mu_i + \phi_i \omega_{ij}) | \hat{\theta}^{(k)}] \\
+ \sum_{m=1}^{M} \lambda_m^{D} \left[ ((f_i^2 + p_t D) \eta_D)^2 - 2\left[ (f_i + p_t D) \eta_D + 1 \right] \exp(\eta_D p_t D) - \\
\left[ (f_i^2 + p_t D_{m-1}) \eta_D \right]^2 - 2\left[ (f_i + p_t D_{m-1}) \eta_D + 1 \right] \exp(\eta_D p_t D_{m-1}) \right] \\
\]

\[
\frac{\partial^2 Q_3}{\partial \sigma^2 \mu} = - \sum_{i=1}^{I} \frac{1}{2} \left\{ - \frac{1}{\sigma^4} + \frac{2E(\mu_i^2 | \hat{\theta}^{(k)})}{\sigma^6} \right\} \\
\frac{\partial^2 Q_3}{\partial \sigma^2 \omega} = - \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{1}{2} \left\{ - \frac{1}{\sigma^4} + \frac{2E(\omega_{ij}^2 | \hat{\theta}^{(k)})}{\sigma^6} \right\} \\
\]
\begin{align*}
\frac{\partial^2 Q_1(\lambda^R \mid \hat{\theta}^{(k)})}{\partial \lambda^R \partial^2} &= - \sum_{m=1}^{M} \left\{ \frac{\sum_{i=1}^{I} \sum_{j=1}^{J} O_{ij,m}^R}{\lambda^R_m} \right\} \\
\frac{\partial^2 Q_1(\lambda^D \mid \hat{\theta}^{(k)})}{\partial \lambda^D \partial^2} &= - \sum_{m=1}^{M} \left\{ \frac{\sum_{i=1}^{I} \sum_{j=1}^{J} O_{ij,m}^D}{\lambda^D_m} \right\} \\
\frac{\partial^2 Q_1}{\partial \beta R \eta_R} &= - \frac{\partial \{ \sum_{i=1}^{I} \sum_{j=1}^{J} X_{ij} \tilde{\Lambda}^0 R (T_{ij}) \exp(q_{1ij}) \mathbb{E}[\exp(\mu_i + \omega_{ij}) \mid \hat{\theta}^{(k)}] \} }{\partial \eta_D} \\
&= \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{P_{ij}^{\eta_R}}{\lambda^R} X_{ij} \\
&= \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{1}{f_i^{\eta_R}} X_{ij} \exp(q_{1ij}) \mathbb{E}[\exp(\mu_i + \omega_{ij}) \mid \hat{\theta}^{(k)}] \\
&= \sum_{m=1}^{M} \lambda^R_m \left[ (f_i + p_i t_m^R)(\eta_R - 1) \exp(\eta_R p_i t_m^R) - \\
&[ (f_i + p_i t_{m-1}^R)(\eta_R - 1) \exp(\eta_R p_i t_{m-1}^R) \right] \\
\frac{\partial^2 Q_1}{\partial \beta \lambda^R} &= - \frac{\partial \{ \sum_{i=1}^{I} \sum_{j=1}^{J} X_{ij} \tilde{\Lambda}^0 R (T_{ij}) \exp(q_{1ij}) \mathbb{E}[\exp(\mu_i + \omega_{ij}) \mid \hat{\theta}^{(k)}] \} }{\partial \lambda^R} \\
&= - \sum_{m=1}^{M} \left\{ \frac{1}{\eta_D p_i} X_{ij} \left[ \exp(\eta_R p_i t_m^R) - \exp(\eta_R p_i t_{m-1}^R) \right] \\
\exp(q_{1ij}) \mathbb{E}[\exp(\mu_i + \omega_{ij}) \mid \hat{\theta}^{(k)}] \right\} 
\end{align*}
\[
\frac{\partial^2 Q_1}{\partial \eta_R \lambda^R} = - \frac{\partial P^R}{\partial \lambda^R} \\
= - \partial \left\{ \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{1}{f_i R} \exp(q_{1ij}) \mathbb{E}[\exp(\mu_i + \omega_{ij}) | \hat{\theta}^{(k)}] \right\} \\
= \sum_{m=1}^{M} \lambda^R_m \left[ ([f_i + pt^R_m] \eta_R - 1] \exp(\eta_R pt^R_m) - ([f_i + pt^R_{m-1}] \eta_R - 1] \exp(\eta_R pt^R_{m-1}) \right] / \partial \lambda^R \\
= - \sum_{m=1}^{M} \left\{ \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{1}{f_i R} \exp(q_{1ij}) \mathbb{E}[\exp(\mu_i + \omega_{ij}) | \hat{\theta}^{(k)}] \right\} \\
\frac{\partial^2 Q_2}{\partial \beta_D \phi_\mu} = - \sum_{i=1}^{I} \sum_{j=1}^{J} \left\{ X_{ij} \tilde{\Lambda}_0^D (T_{ij}) \exp(q_{2ij}) \mathbb{E}[\exp(\phi_\mu \mu_i + \phi_\omega \omega_{ij}) | \hat{\theta}^{(k)}] \right\} \\
\frac{\partial^2 Q_2}{\partial \beta_D \phi_\omega} = - \sum_{i=1}^{I} \sum_{j=1}^{J} \left\{ X_{ij} \tilde{\Lambda}_0^D (T_{ij}) \exp(q_{2ij}) \mathbb{E}[\exp(\phi_\mu \mu_i + \phi_\omega \omega_{ij}) | \hat{\theta}^{(k)}] \right\} \\
\frac{\partial^2 Q_2}{\partial \beta_D \eta_D} = \partial \left\{ \sum_{i=1}^{I} \sum_{j=1}^{J} X_{ij} \tilde{\Lambda}_0^D (T_{ij}) \exp(q_{2ij}) \mathbb{E}[\exp(\phi_\mu \mu_i + \phi_\omega \omega_{ij}) | \hat{\theta}^{(k)}] \right\} / \partial \eta_D \\
= \sum_{i=1}^{I} \sum_{j=1}^{J} P_{ij}^D \eta_D X_{ij} \\
= \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{1}{f_i R} X_{ij} \exp(q_{2ij}) \mathbb{E}[\exp(\phi_\mu \mu_i + \phi_\omega \omega_{ij}) | \hat{\theta}^{(k)}] \\
\sum_{m=1}^{M} \lambda^D_m \left[ ([f_i + pt^D_m] \eta_D - 1] \exp(\eta_D pt^D_m) - ([f_i + pt^D_{m-1}] \eta_D - 1] \exp(\eta_D pt^D_{m-1}) \right] 
\]
\[
\frac{\partial^2 Q_2}{\partial D \lambda^D} = - \frac{\partial}{\partial \lambda^D} \left\{ \sum_{i=1}^{I} \sum_{j=1}^{J} X_{ij} \tilde{\Lambda}_0^D (T_{ij}) \exp(q_{ij}) E[\mu_i \exp(\phi_{\mu_i} + \phi_{\omega} \omega_{ij}) | \hat{\theta}^{(k)}] \right\}
\]

\[
= - \sum_{m=1}^{M} \left\{ \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{1}{\eta_{DPi}^D} X_{ij} \exp(\eta_{DPi}^D t_m^D) - \exp(\eta_{DPi}^D t_{m-1}^D) \right\} \exp(q_{ij}) E[\exp(\phi_{\mu_i} + \phi_{\omega} \omega_{ij}) | \hat{\theta}^{(k)}] \}
\]

\[
\frac{\partial^2 Q_2}{\partial \phi_{\mu} \phi_{\omega}} = \sum_{i=1}^{I} \sum_{j=1}^{J} \{ - \tilde{\Lambda}_0^D (T_{ij}) \exp(q_{ij}) E[\mu_i \exp(\phi_{\mu_i} + \phi_{\omega} \omega_{ij}) | \hat{\theta}^{(k)}] \}
\]

\[
\frac{\partial^2 Q_2}{\partial \phi_{\mu} \eta_D} = - \frac{\partial}{\partial \eta_D} \left\{ \sum_{i=1}^{I} \sum_{j=1}^{J} \tilde{\Lambda}_0^D (T_{ij}) \exp(q_{ij}) E[\mu_i \exp(\phi_{\mu_i} + \phi_{\omega} \omega_{ij}) | \hat{\theta}^{(k)}] \right\}
\]

\[
= \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{1}{f_i \eta_D} \exp(q_{ij}) E[\mu_i \exp(\phi_{\mu_i} + \phi_{\omega} \omega_{ij}) | \hat{\theta}^{(k)}] \]

\[
- \sum_{m=1}^{M} \lambda_m^D \left[ ((f_i + p_i t_m^D) \eta_D - 1) \exp(\eta_{DPi}^D t_m^D) - \right]

\[
- ((f_i + p_i t_{m-1}^D) \eta_D - 1) \exp(\eta_{DPi}^D t_{m-1}^D) \}
\]

\[
\frac{\partial^2 Q_2}{\partial \phi_{\mu} \lambda^D} = - \frac{\partial}{\partial \lambda^D} \left\{ \sum_{i=1}^{I} \sum_{j=1}^{J} \tilde{\Lambda}_0^D (T_{ij}) \exp(q_{ij}) E[\mu_i \exp(\phi_{\mu_i} + \phi_{\omega} \omega_{ij}) | \hat{\theta}^{(k)}] \right\}
\]

\[
= - \sum_{m=1}^{M} \left\{ \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{1}{\eta_{DPi}^D} \left[ \exp(\eta_{DPi}^D t_m^D) - \exp(\eta_{DPi}^D t_{m-1}^D) \right] \right\} \exp(q_{ij}) E[\mu_i \exp(\phi_{\mu_i} + \phi_{\omega} \omega_{ij}) | \hat{\theta}^{(k)}] \}
\]
\[
\frac{\partial^2 Q_2}{\partial \phi \partial \eta_D} = - \partial \left\{ \sum_{i=1}^I \sum_{j=1}^J \lambda^D_0 (T_{ij}) \exp(q_{2ij}) E [\omega_{ij} \exp(\phi_{\mu \mu_i} + \phi_{\omega \omega_{ij}}) | \dot{\theta}^{(k)}] \right\} \\
\frac{\partial}{\partial \eta_D} = \sum_{i=1}^I \sum_{j=1}^J \frac{1}{f_D \eta_D^2} \exp(q_{2ij}) E [\omega_{ij} \exp(\phi_{\mu \mu_i} + \phi_{\omega \omega_{ij}}) | \dot{\theta}^{(k)}] \\
\sum_{m=1}^M \lambda^D_m \left[ [(f_i + p_{it_m^D}) \eta_D - 1] \exp(\eta_D p_{it_m^D}) - \\
[(f_i + p_{it_{m-1}^D}) \eta_D - 1] \exp(\eta_D p_{it_{m-1}^D}) \right] \right\} / \partial \lambda^D \\
\frac{\partial^2 Q_2}{\partial \phi \partial \lambda^D} = - \partial \left\{ \sum_{i=1}^I \sum_{j=1}^J \frac{1}{f_D \eta_D} \exp(q_{2ij}) E [\exp(\phi_{\mu \mu_i} + \phi_{\omega \omega_{ij}}) | \dot{\theta}^{(k)}] \right\} \\
\sum_{m=1}^M \lambda^D_m \left[ [(f_i + p_{it_m^D}) \eta_D - 1] \exp(\eta_D p_{it_m^D}) - \\
[(f_i + p_{it_{m-1}^D}) \eta_D - 1] \exp(\eta_D p_{it_{m-1}^D}) \right] \right\} / \partial \lambda^D \\
\frac{\partial^2 Q_2}{\partial \eta_D \partial \lambda^D} = - \frac{\partial^2 P^{nd}}{\partial \lambda^D} \\
= - \partial \left\{ \sum_{i=1}^I \sum_{j=1}^J \frac{1}{f_D \eta_D^2} \exp(q_{2ij}) E [\exp(\phi_{\mu \mu_i} + \phi_{\omega \omega_{ij}}) | \dot{\theta}^{(k)}] \right\} \\
\sum_{m=1}^M \lambda^D_m \left[ [(f_i + p_{it_m^D}) \eta_D - 1] \exp(\eta_D p_{it_m^D}) - \\
[(f_i + p_{it_{m-1}^D}) \eta_D - 1] \exp(\eta_D p_{it_{m-1}^D}) \right] \right\} / \partial \lambda^D \\
= - \sum_{m=1}^M \left\{ \sum_{i=1}^I \sum_{j=1}^J \frac{1}{f_D \eta_D^2} \exp(q_{2ij}) E [\exp(\phi_{\mu \mu_i} + \phi_{\omega \omega_{ij}}) | \dot{\theta}^{(k)}] \right\} \\
\left[ [(f_i + p_{it_m^D}) \eta_D - 1] \exp(\eta_D p_{it_m^D}) - \\
[(f_i + p_{it_{m-1}^D}) \eta_D - 1] \exp(\eta_D p_{it_{m-1}^D}) \right] \right\}
All other off-diagonal terms are zero, such as

\[ \frac{\partial^2 Q}{\partial \beta_R \beta_D} = 0 \]

\[ \frac{\partial^2 Q}{\partial \beta_R \phi_\mu} = 0 \]

\[ \frac{\partial^2 Q}{\partial \beta_R \phi_\omega} = 0 \]
Chapter 5

In the M-step, first we have closed-form estimators for baseline intensities $\tilde{\lambda}_0^R(t)$, $\tilde{\lambda}_0^D(t)$ and variance components: $\sigma^2_\mu$, $\sigma^2_\omega$, $\Sigma_b$ and $\sigma^2_e$.

C.1 Baseline intensities

C.1.1 Recurrent events

We introduce several notations for the estimation of $\tilde{\lambda}_0^R(t)$. Assume there are $M$ intervals with cutpoints $0 \leq c^R_0 < c^R_1 < \ldots < c^R_M = \infty$, which can be quantiles of recurrent event times, where $c^R_0 = 0$ or the smallest event time.

Denote the indicator function $I_R^m(t) = I(c^R_{m-1} < t \leq c^R_m)$. For $m = 1, 2, \ldots, M$,

$$\tilde{\lambda}_0^R(t) = \sum_{m=1}^M \lambda^R_m I(c^R_{m-1} < t \leq c^R_m)$$

(C.1)

Then, we denote by $\lambda^R = (\lambda^R_1, \lambda^R_2, \ldots, \lambda^R_m, \ldots, \lambda^R_M)$ the parameter we aim to
estimate. We denote the expected log likelihood part which involves \( \hat{\lambda}_0^R(t) \) as

\[
Q_1(\hat{\lambda}_0^R(t) | \hat{\theta}^{(k)}) = \sum_{i=1}^{I} \sum_{j=1}^{J} \left\{ \sum_{m=1}^{N^R_{ij}} \left( \log(\hat{\lambda}_0^R(t_{ijk})) + q_{1ij} + \eta_R \beta_{t_{ijk}} + \eta_R \mathbb{E}[b_{1ij} | \hat{\theta}^{(k)}] + \eta_R \mathbb{E}[b_{0ij} | \hat{\theta}^{(k)}] \right) \right. \\
- \exp(q_{1ij}) \mathbb{E}[\hat{\Lambda}_0^R(t) \exp(\mu_i + \omega_{ij} + \eta_R b_{0ij}) | \hat{\theta}^{(k)}] \right\}
\]

For computational purpose, the log-likelihood can be written in a Poisson regression form:

\[
Q_1(\lambda^R | \hat{\theta}^{(k)}) = \sum_{m=1}^{M} \left\{ \sum_{i=1}^{I} \sum_{j=1}^{J} \left( \left( \sum_{k=1}^{N^R_{ij}} \log(\lambda^R_{m} + q_{1ij} + \eta_R \beta_{t_{ijk}} + \eta_R \mathbb{E}[b_{1ij} | \hat{\theta}^{(k)}] + \eta_R \mathbb{E}[b_{0ij} | \hat{\theta}^{(k)}]) \right) - \sum_{i=1}^{I} \sum_{j=1}^{J} \exp(q_{1ij}) \mathbb{E}[\bar{\lambda}_0^R(T_{ij}) \exp(\mu_i + \omega_{ij} + \eta_R b_{0ij}) | \hat{\theta}^{(k)}] \right) \right\}
\]

Denote \( S(\cdot) \) is the score function component. Let \( O^R_{ij,m} = \sum_{k=1}^{N^R_{ij}} I^R_{m}(t_{ijk}) \delta_{ijk} \), the number of recurrent events for individual \( ij \) in the \( m^{th} \) subinterval. To simplify the expression, we denote \( p_t = \beta_t + b_{1ij} \). Recall that \( q_{1ij} = \beta_T^{(R)} X_{ij}^{(R)} + \eta_R X_{1ij}^{(L)}(t) \beta_{L} \), and \( q_{2ij} = \beta_T^{(D)} X_{ij}^{(D)} + \eta_D X_{1ij}^{(L)}(t) \beta_{L} \). Then

\[
S(\lambda^R) = \frac{\partial Q_1(\lambda^R | \hat{\theta}^{(k)})}{\partial \lambda^R} = \sum_{m=1}^{M} \left\{ \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{N^R_{ij}} \frac{1}{\lambda^R_m} S^R_{ij,m} - \sum_{i=1}^{I} \sum_{j=1}^{J} S^R_{ij,m} \right\} = \sum_{m=1}^{M} \left\{ \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{O^R_{ij,m}}{\lambda^R_m} - \sum_{i=1}^{I} \sum_{j=1}^{J} S^R_{ij,m} \right\}
\]
where

\[ S_{ij,m}^R = \frac{\partial}{\partial \lambda^R_m} \left\{ \exp(q_{1ij}) E[\Lambda^R_m(T_{ij}) \exp(\mu_i + \omega_{ij} + \eta_{Rb_{0ij}}) | \hat{\theta}^{(k)}] \right\} \]

\[ = \frac{\partial}{\partial \lambda^R_m} \left\{ \exp(q_{1ij}) E[\int_{t^R_{m-1}}^{t^R_m} \lambda^R_m \exp(\eta_{Rp_t} T_{ij} dT_{ij}) \exp(\mu_i + \omega_{ij} + \eta_{Rb_{0ij}}) | \hat{\theta}^{(k)}] \right\} \]

\[ = \frac{\partial}{\partial \lambda^R_m} \left\{ \exp(q_{1ij}) E\left[ \frac{1}{\eta_{Rp_t}} \left( \exp(\eta_{Rp_t} t^R_m) - \exp(\eta_{Rp_t} t^R_{m-1}) \right) \exp(\mu_i + \omega_{ij} + \eta_{Rb_{0ij}}) | \hat{\theta}^{(k)} \right] \right\} \]

Here, \( t^R_m \) and \( t^R_{m-1} \) are upper and lower bounds at \( m \)th interval. \( t^R_m = \max(0, \min(c^R_m - c^R_{m-1}, T_{ij} - c^R_{m-1})) \).

Solve \( S(\lambda^R) = 0 \). The maximum likelihood estimator has an explicit solution:

\[ \hat{\lambda}^R_{m(k+1)} = \frac{\sum_{m=1}^{M} \left[ \sum_{i=1}^{I} \sum_{j=1}^{J} O_{ij,m}^R \right]}{\sum_{m=1}^{M} \left[ \sum_{i=1}^{I} \sum_{j=1}^{J} S_{ij,m}^R \right]} \]  \hspace{1cm} (C.2)

### C.1.2 Death event

The estimation of \( \tilde{\lambda}^D_0 \) is similar to the estimation of \( \tilde{\lambda}^R_0 \). \( 0 \leq c^D_0 < c^D_1 < \ldots < c^D_M = \infty \) is the quantiles of death event times. Note that the selection of \( M \) for recurrent events and death can be the same or different. Denote the indicator function \( I^D_m(t) = I(c^D_{m-1} < t \leq c^D_m) \). For \( m = 1, 2, \ldots, M \),

\[ \tilde{\lambda}^D_0(t) = \sum_{m=1}^{M} \lambda^D_m I(c^D_{m-1} < t \leq c^D_m) \]

\[ = \sum_{m=1}^{M} \lambda^D_m I^D_m(t) \]  \hspace{1cm} (C.3)
Therefore, we denote by \( \boldsymbol{\lambda}^D = (\lambda_1^D, \lambda_2^D, \ldots, \lambda_m^D, \ldots, \lambda_M^D) \) the parameter we aim to estimate for baseline intensity function of death events.

For computational purpose, the log-likelihood can be written as:

\[
Q_2(\boldsymbol{\lambda}^D \mid \hat{\theta}^{(k)}) = \\
\sum_{m=1}^{M} \left\{ \sum_{i=1}^{I} \sum_{j=1}^{J} \left\{ I_m^D(T_{ij}) \Delta_{ij} \left[ \log(\lambda_m^D) + q_{2ij} + \eta_D \beta_i T_{ij} + \eta_D E[b_{0ij} T_{ij} \mid \hat{\theta}^{(k)}] + \eta_D E[b_{0ij} \mid \hat{\theta}^{(k)}] + \phi_\mu E[\mu_i \mid \hat{\theta}^{(k)}] + \phi_\omega E[\omega_{ij} \mid \hat{\theta}^{(k)}] \right] \\
- \sum_{i=1}^{I} \sum_{j=1}^{J} \exp(q_{2ij}) E[\tilde{\Lambda}_m^D(T_{ij}) \exp(\phi_\mu \mu_i + \phi_\omega \omega_{ij} + \eta_D b_{0ij} \mid \hat{\theta}^{(k)})] \right\} \right\}
\]

Let \( O_{ij,m}^D = I_m^D(T_{ij}) \Delta_{ij} \), which is the number of death event for individual \( ij \) in the \( m \)th interval. The score component is:

\[
S(D) = \frac{\partial Q_2(\boldsymbol{\lambda}^D \mid \hat{\theta}^{(k)})}{\partial \lambda_m^D} = \\
\sum_{m=1}^{M} \left\{ \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{O_{ij,m}^D}{\lambda_m^D} - \sum_{i=1}^{I} \sum_{j=1}^{J} S_{ij,m}^D \right\}
\]

\[
S_{ij,m}^D = \frac{\partial \left\{ \exp(q_{2ij}) E[\tilde{\Lambda}_m^D(T_{ij}) \exp(\phi_\mu \mu_i + \phi_\omega \omega_{ij} + \eta_D b_{0ij} \mid \hat{\theta}^{(k)})] \right\}}{\partial \lambda_m^D} = \\
\frac{\partial \left\{ \exp(q_{2ij}) E[\int_{t_{m-1}^D}^{t_m^D} \lambda_m^D \exp(\eta_D p_t T_{ij} dT_{ij} \exp(\phi_\mu \mu_i + \phi_\omega \omega_{ij} + \eta_D b_{0ij} \mid \hat{\theta}^{(k)})] \right\}}{\partial \lambda_m^D} = \\
\frac{\partial \left\{ \exp(q_{2ij}) E\left[ \frac{\lambda_m^D}{\eta_D p_t} \left[ \exp(\eta_D p_t t_{m}^D) - \exp(\eta_D p_t t_{m-1}^D) \right] \right] \right\}}{\partial \lambda_m^D} = \\
\exp(\phi_\mu \mu_i + \phi_\omega \omega_{ij} + \eta_D b_{0ij} \mid \hat{\theta}^{(k)}) \}
\]

\( t_m^D \) and \( t_{m-1}^D \) are upper and lower bounds at \( m \)th interval. \( t_m^D = \max(0, \min(c_m^D - \ldots) \ldots) \).
\[ c_{m-1}^D, T_{ij} - c_{m-1}^D \].

Solve \( S(\lambda^D) = 0 \). The maximum likelihood estimator has an explicit solution:

\[
\hat{\lambda}_{m}^{D(k+1)} = \frac{\sum_{m=1}^{M} \left[ \sum_{i=1}^{I} \sum_{j=1}^{J} O_{ij,m}^D \right]}{\sum_{m=1}^{M} \left[ \sum_{i=1}^{I} \sum_{j=1}^{J} S_{ij,m}^D \right]}
\]  \hspace{1cm} (C.4)

### C.2 Variance components

Then the corresponding score functions and second partial derivatives for \( \sigma^2_\mu \), \( \sigma^2_\omega \), \( \Sigma_b \) and \( \sigma^2_e \) are:

\[
S(\sigma^2_\mu) = \frac{\partial Q_3}{\partial \sigma^2_\mu} = \sum_{i=1}^{I} \{ -\frac{1}{2\sigma^2_\mu} + \frac{E(\mu_i^2 \mid \hat{\theta}^{(k)})}{2\sigma^4_\mu} \}
\]

\[
S(\sigma^2_\omega) = \frac{\partial Q_3}{\partial \sigma^2_\omega} = \sum_{i=1}^{I} \sum_{j=1}^{J} \{ -\frac{1}{2\sigma^2_\omega} + \frac{E(\omega_{ij}^2 \mid \hat{\theta}^{(k)})}{2\sigma^4_\omega} \}
\]

The maximum likelihood estimators for \( \sigma^2_\mu \), \( \sigma^2_\omega \), \( \Sigma_b \) and \( \sigma^2_e \) are:

\[
\hat{\sigma}^2(\mu)^{(k+1)} = \frac{1}{I} \sum_{i=1}^{I} \left\{ E(\mu_i^2 \mid \hat{\theta}^{(k)}) \right\}
\]  \hspace{1cm} (C.5)

\[
\hat{\sigma}^2(\omega)^{(k+1)} = \frac{1}{IJ} \sum_{i=1}^{I} \sum_{j=1}^{J} E(\omega_{ij}^2 \hat{\theta}^{(k)})
\]  \hspace{1cm} (C.6)

\[
\hat{\Sigma}_b^{(k+1)} = \frac{\sum_{i=1}^{I} \sum_{j=1}^{J} E[b_{ij}^T b_{ij} \mid \hat{\theta}^{(k)}]}{IJ}
\]  \hspace{1cm} (C.7)

\[
\hat{\sigma}^2(e)^{(k+1)} = \frac{\sum_{i=1}^{I} \sum_{j=1}^{J} E[\epsilon_{ij}^T \epsilon_{ij} \mid \hat{\theta}^{(k)}]}{\sum_{i=1}^{I} \sum_{j=1}^{J} n_{ij}}
\]  \hspace{1cm} (C.8)

of note that \( IJ \) is the total number of subjects.
C.3 Newton-Raphson algorithm

For the other parameters, which don’t have the closed-form estimator, the Newton-Raphson algorithm is used to solve the expected log-likelihood iteratively. Denote \( \tau = (\beta_R, \beta_D, \phi_\mu, \phi_\omega, \eta_R, \eta_D) \). The gradient function \( G(\tau) \) is a score vector, and \( H_\tau(\cdot) \) is the Hessian matrix. Then a second Taylor series expansion of the log likelihood \( l(\tau) \), th estimates at iteration \( k \) gives:

\[
g((\tau)^{(k)} = l(\tau) + G(\tau)^T(\tau - \tau^{(k)}) + \frac{1}{2}(\tau - \tau^{(k)})^T H_\tau(\tau - \tau^{(k)})
\]

differentiating and setting equal to zero gives the next estimate

\[
\hat{\tau}^{(k+1)} = \hat{\tau}^{(k)} - H_\tau^{-1}(\hat{\tau}^{(k)})G_\tau(\hat{\tau}^{(k)})
\]  
(C.9)
Bibliography


Bender, R., Augustin, T., and Blettner, M. (2005). Generating survival times to


analysis for large numbers of small groups of correlated failure time observations. In *Survival analysis: state of the art*, pages 237–247. Springer.


Vita

Cong Xu

Ph.D in Biostatistics August 2013 - present
Department of Public Health Sciences
Pennsylvania State University, Hershey, PA, USA
Overall GPA: 3.98/4.00, Major GPA: 4.00/4.00

M.S. in Statistics August 2011 - May 2013
Department of Statistics
University of Illinois at Urbana-Champaign, Urbana, IL, USA
Overall GPA: 3.76/4.00

B.Ec. in Statistics September 2007 - July 2011
School of Statistics
Shandong Institute of Business and Technology, Yantai, China Overall
GPA: 3.77/4.00, Major: 3.97/4.00