An Introduction to the Joint Modeling of Longitudinal and Survival Data, with Applications in R

Dimitris Rizopoulos
Department of Biostatistics, Erasmus University Medical Center

d.rizopoulos@erasmusmc.nl

Data Science in Health
Jheronimus Academy of Data Science
April 3, 2020, Den Bosch
Contents

1 Introduction 1

1.1 Motivating Longitudinal Studies .................................................. 2
1.2 Research Questions ........................................................................... 10
1.3 Recent Developments ...................................................................... 13

2 Linear Mixed-Effects Models 15

2.1 Features of Longitudinal Data ......................................................... 16
2.2 The Linear Mixed Model .................................................................. 17
2.3 Mixed-Effects Models in R ................................................................. 25
2.4 Missing Data in Longitudinal Studies ............................................. 31
2.5 Missing Data Mechanisms ................................................................. 34

3 Relative Risk Models ........................................................................ 39

3.1 Features of Survival Data ................................................................. 40
3.2 Relative Risk Models ...................................................................... 43
3.3 Relative Risk Models in R ................................................................ 46
3.4 Time Dependent Covariates ............................................................ 48
3.5 Extended Cox Model ........................................................................ 53
4 The Basic Joint Model

4.1 Joint Modeling Framework

4.2 A Comparison with the TD Cox

4.3 Joint Models in R

4.4 Connection with Missing Data

5 Closing

5.1 Concluding Remarks

5.2 Additional References

5.3 Medical Papers with Joint Modeling

Practicals
What is This Course About

- Often in follow-up studies different types of outcomes are collected

  - **Explicit** outcomes
    - multiple longitudinal responses (e.g., markers, blood values)
    - time-to-event(s) of particular interest (e.g., death, relapse)

  - **Implicit** outcomes
    - missing data (e.g., dropout, intermittent missingness)
    - random visit times
• Methods for the separate analysis of such outcomes are well established in the literature

• Survival data:
  ▶ Cox model, accelerated failure time models, . . .

• Longitudinal data
  ▶ mixed effects models, GEE, marginal models, . . .
What is This Course About (cont’d)

Purpose of this course is to introduce the basics of

Joint Models for Longitudinal and Survival Data
Learning Objectives

- **Goals:** After this course participants will be able to
  - identify settings in which a joint modeling approach is required,
  - construct and fit an appropriate joint model, and
  - correctly interpret the obtained results

- The course will be explanatory rather than mathematically rigorous
  - emphasis is given on sufficient detail in order for participants to obtain a clear view on the different joint modeling approaches, and how they should be used in practice
Agenda

• **Part I:** Introduction
  ▶ Data sets that we will use throughout the course
  ▶ Categorization of possible research questions

• **Part II:** (brief) Review of Linear Mixed Models
  ▶ Features of repeated measurements data
  ▶ Linear mixed models
  ▶ Missing data in longitudinal studies
Agenda (cont’d)

• **Part III:** (brief) Review of Relative Risk Models
  - Features of survival data
  - Relative risk models
  - Time-dependent covariates

• **Part IV:** The Basic Joint Model
  - Definition
  - Estimation & Inference
  - Connection with the missing data framework
Structure of the Course & Material

- Lectures & short software practicals using R package JM and/or JMbayes

- Material (also available in http://www.drizopoulos.com/):
  - Course Notes
  - R code in soft format

- Within the course notes there are several examples of R code which are denoted by the symbol ‘R> ’
Joint modeling sources


* Extra references of papers using joint modeling available at pp. 86–93.
References (cont’d)

• Useful material for package **JM** can be found in the web sites:
  ▶ http://jmr.r-forge.r-project.org [R code used in the book]
  ▶ http://www.drizopoulos.com/
    [additional R script files]

• Useful material for package **JMbayes**
  ▶ a paper describing the current capabilities of the package is available on JSS
    http://dx.doi.org/10.18637/jss.v072.i07

• Blog about joint modeling http://iprogn.blogspot.nl/
• Other software packages capable of fitting joint models

▷ in R: joineR (by Philipson et al.), lcmm (by Proust-Lima et al.)
▷ in SAS: %JM macro (by Garcia-Hernandez and Rizopoulos – http://www.jm-macro.com/), %JMFIt macro (by Zhang et al.)
▷ in STATA: stjm (by Crowther)
Chapter 1
Introduction
1.1 Motivating Longitudinal Studies

• AIDS: 467 HIV infected patients who had failed or were intolerant to zidovudine therapy (AZT) (Abrams et al., NEJM, 1994)

• The aim of this study was to compare the efficacy and safety of two alternative antiretroviral drugs, didanosine (ddl) and zalcitabine (ddC)

• Outcomes of interest:
  ▶ time to death
  ▶ randomized treatment: 230 patients ddl and 237 ddC
  ▶ CD4 cell count measurements at baseline, 2, 6, 12 and 18 months
  ▶ prevOI: previous opportunistic infections
1.1 Motivating Longitudinal Studies (cont’d)
1.1 Motivating Longitudinal Studies (cont’d)
1.1 Motivating Longitudinal Studies (cont’d)

- **Research Questions:**
  - How strong is the association between CD4 cell count and the risk for death?
  - Is CD4 cell count a good biomarker?
    * if treatment improves CD4 cell count, does it also improve survival?
1.1 Motivating Longitudinal Studies (cont’d)

- **PBC**: Primary Biliary Cirrhosis:
  - a chronic, fatal but rare liver disease
  - characterized by inflammatory destruction of the small bile ducts within the liver

- Data collected by Mayo Clinic from 1974 to 1984 (Murtaugh et al., Hepatology, 1994)

- Outcomes of interest:
  - time to death and/or time to liver transplantation
  - randomized treatment: 158 patients received D-penicillamine and 154 placebo
  - longitudinal serum bilirubin levels
### 1.1 Motivating Longitudinal Studies (cont’d)

<table>
<thead>
<tr>
<th>Time (years)</th>
<th>log serum Bilirubin</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>38</td>
</tr>
<tr>
<td>5</td>
<td>39</td>
</tr>
<tr>
<td>10</td>
<td>51</td>
</tr>
<tr>
<td>15</td>
<td>68</td>
</tr>
<tr>
<td>20</td>
<td>70</td>
</tr>
<tr>
<td>25</td>
<td>82</td>
</tr>
<tr>
<td>30</td>
<td>90</td>
</tr>
<tr>
<td>35</td>
<td>93</td>
</tr>
<tr>
<td>40</td>
<td>134</td>
</tr>
<tr>
<td>45</td>
<td>148</td>
</tr>
<tr>
<td>50</td>
<td>173</td>
</tr>
<tr>
<td>55</td>
<td>200</td>
</tr>
<tr>
<td>60</td>
<td>216</td>
</tr>
<tr>
<td>65</td>
<td>242</td>
</tr>
<tr>
<td>70</td>
<td>269</td>
</tr>
<tr>
<td>75</td>
<td>290</td>
</tr>
</tbody>
</table>

Time (years) vs. log serum Bilirubin
1.1 Motivating Longitudinal Studies (cont’d)

Kaplan–Meier Estimate

Survival Probability

Time (years)

placebo
D–penicil
1.1 Motivating Longitudinal Studies (cont’d)

- **Research Questions:**
  - How strong is the association between bilirubin and the risk for death?
  - How the observed serum bilirubin levels could be utilized to provide predictions of survival probabilities?
  - Can bilirubin discriminate between patients of low and high risk?
1.2 Research Questions

- Depending on the questions of interest, different types of statistical analysis are required

- We will distinguish between two general types of analysis
  - separate analysis per outcome
  - joint analysis of outcomes

- Focus on each outcome separately
  - does treatment affect survival?
  - are the average longitudinal evolutions different between males and females?
  - . . .
1.2 Research Questions (cont’d)

- Focus on multiple outcomes

  ▶ Complex hypothesis testing: does treatment improve the average longitudinal profiles in all markers?

  ▶ Complex effect estimation: how strong is the association between the longitudinal evolution of CD4 cell counts and the hazard rate for death?

  ▶ Association structure among outcomes:
    * how the association between markers evolves over time (evolution of the association)
    * how marker-specific evolutions are related to each other (association of the evolutions)
1.2 Research Questions (cont’d)

▷ Prediction: can we improve prediction for the time to death by considering all markers simultaneously?

▷ Handling implicit outcomes: focus on a single longitudinal outcome but with dropout or random visit times
1.3 Recent Developments

- Up to now emphasis has been
  - restricted or coerced to separate analysis per outcome
  - or given to naive types of joint analysis (e.g., last observation carried forward)

- Main reasons
  - lack of appropriate statistical methodology
  - lack of efficient computational approaches & software
1.3 Recent Developments (cont’d)

- However, recently there has been an explosion in the statistics and biostatistics literature of joint modeling approaches

- Many different approaches have been proposed that
  - can handle different types of outcomes
  - can be utilized in pragmatic computing time
  - can be rather flexible
  - most importantly: can answer the questions of interest
Chapter 2

Linear Mixed-Effects Models
2.1 Features of Longitudinal Data

- Repeated evaluations of the same outcome in each subject in time
  - CD4 cell count in HIV-infected patients
  - Serum bilirubin in PBC patients

Measurements on the same subject are expected to be (positively) correlated

- This implies that standard statistical tools, such as the $t$-test and simple linear regression that assume independent observations, are not optimal for longitudinal data analysis.
2.2 The Linear Mixed Model

• **Basic idea:** Each subject in the population has her own subject-specific mean response profile over time
2.2 The Linear Mixed Model (cont’d)
2.2 The Linear Mixed Model (cont’d)

- The evolution of each subject in time can be described by a linear model

\[ y_{ij} = \tilde{\beta}_{i0} + \tilde{\beta}_{i1} t_{ij} + \epsilon_{ij}, \quad \epsilon_{ij} \sim \mathcal{N}(0, \sigma^2), \]

where

- \( y_{ij} \) the \( j \)th response of the \( i \)th subject
- \( \tilde{\beta}_{i0} \) is the intercept and \( \tilde{\beta}_{i1} \) the slope for subject \( i \)

- Assumption: Subjects are randomly sampled from a population \( \Rightarrow \) subject-specific regression coefficients are also sampled from a population of regression coefficients

\[ \tilde{\beta}_i \sim \mathcal{N}(\beta, D) \]
2.2 The Linear Mixed Model (cont’d)

- We can reformulate the model as

\[ y_{ij} = (\beta_0 + b_{i0}) + (\beta_1 + b_{i1})t_{ij} + \varepsilon_{ij}, \]

where

- \( \beta \)'s are known as the **fixed effects**
- \( b_i \)'s are known as the **random effects**

- In accordance for the random effects we assume

\[
\begin{bmatrix}
  b_{i0} \\
  b_{i1}
\end{bmatrix} \sim \mathcal{N}(0, D)
\]
2.2 The Linear Mixed Model (cont’d)

- Put in a general form

\[
\begin{align*}
  y_i &= X_i \beta + Z_i b_i + \varepsilon_i, \\
  b_i &\sim \mathcal{N}(0, D), \\
  \varepsilon_i &\sim \mathcal{N}(0, \sigma^2 I_{n_i}),
\end{align*}
\]

with

- $X$ design matrix for the fixed effects $\beta$
- $Z$ design matrix for the random effects $b_i$
- $b_i \perp \perp \varepsilon_i$
2.2 The Linear Mixed Model (cont’d)

- Interpretation:
  - $\beta_j$ denotes the change in the average $y_i$ when $x_j$ is increased by one unit.
  - $b_i$ are interpreted in terms of how a subset of the regression parameters for the $i$th subject deviates from those in the population.

- Advantageous feature: population + subject-specific predictions
  - $\beta$ describes mean response changes in the population.
  - $\beta + b_i$ describes individual response trajectories.
2.2 The Linear Mixed Model (cont’d)

- **Example:** We fit a linear mixed model for the AIDS dataset assuming
  - different average longitudinal evolutions per treatment group (**fixed part**)
  - random intercepts & random slopes (**random part**)

\[
\begin{cases}
  y_{ij} = \beta_0 + \beta_1 t_{ij} + \beta_2 \{ddI_i \times t_{ij}\} + b_{i0} + b_{i1}t_{ij} + \varepsilon_{ij}, \\
  b_i \sim \mathcal{N}(0, D), \\
  \varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2)
\end{cases}
\]

- **Note:** We did not include a main effect for treatment due to randomization
2.2 The Linear Mixed Model (cont’d)

<table>
<thead>
<tr>
<th>Value</th>
<th>Std.Err.</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_0$</td>
<td>7.189</td>
<td>0.222</td>
<td>32.359</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>-0.163</td>
<td>0.021</td>
<td>-7.855</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>0.028</td>
<td>0.030</td>
<td>0.952</td>
</tr>
</tbody>
</table>

- No evidence of differences in the average longitudinal evolutions between the two treatments
There are two primary packages in R for mixed models analysis:

- **Package nlme**
  * fits linear & nonlinear mixed effects models, and marginal models for normal data
  * allows for both random effects & correlated error terms
  * several options for covariances matrices and variance functions

- **Package lme4**
  * fits linear, nonlinear & generalized mixed effects models
  * uses only random effects
  * allows for nested and crossed random-effects designs
2.3 Mixed-Effects Models in R (cont’d)

R> We will only use package `nlme` because package `JM` accepts as an argument a linear mixed model fitted by `nlme`.

R> The basic function to fit linear mixed models is `lme()` and has three basic arguments:

▷ `fixed`: a formula specifying the response vector and the fixed-effects structure
▷ `random`: a formula specifying the random-effects structure
▷ `data`: a data frame containing all the variables.
2.3 Mixed-Effects Models in R (cont’d)

The data frame that contains all variables should be in the long format.

```
Subject y time gender age
1 5.1 0.0 male 45
1 6.3 1.1 male 45
2 5.9 0.1 female 38
2 6.9 0.9 female 38
2 7.1 1.2 female 38
2 7.3 1.5 female 38
... ... ... ... ...
```
2.3 Mixed-Effects Models in R (cont’d)

R> Using formulas in R

▷ CD4 = Time + Gender
⇒ \( \text{cd4} \sim \text{time} + \text{gender} \)

▷ CD4 = Time + Gender + Time*Gender
⇒ \( \text{cd4} \sim \text{time} + \text{gender} + \text{time:gender} \)
⇒ \( \text{cd4} \sim \text{time*gender} \) (the same)

▷ CD4 = Time + Time^2
⇒ \( \text{cd4} \sim \text{time} + \text{I(time}^2) \)

R> Note: the intercept term is included by default
The code used to fit the linear mixed model for the AIDS dataset (p. 23) is as follows

```R
lmeFit <- lme(CD4 ~ obstime + obstime:drug, data = aids, 
              random = ~ obstime | patient)

summary(lmeFit)
```
The same fixed-effects structure but only random intercepts

\[
\text{lme(CD4} \sim \text{obstime + obstime:drug, data = aids, random = } \sim \text{1 | patient)}
\]
2.4 Missing Data in Longitudinal Studies

- A major challenge for the analysis of longitudinal data is the problem of missing data
  - studies are designed to collect data on every subject at a set of prespecified follow-up times
  - often subjects miss some of their planned measurements for a variety of reasons
2.4 Missing Data in Longitudinal Studies (cont’d)

- Implications of missingness:
  - we collect less data than originally planned \(\Rightarrow\) \textit{loss of efficiency}
  - not all subjects have the same number of measurements \(\Rightarrow\) \textit{unbalanced datasets}
  - missingness may depend on outcome \(\Rightarrow\) \textit{potential bias}

- For the handling of missing data, we introduce the missing data indicator

\[
    r_{ij} = \begin{cases} 
    1 & \text{if } y_{ij} \text{ is observed} \\
    0 & \text{otherwise}
    \end{cases}
\]
2.4 Missing Data in Longitudinal Studies (cont’d)

- We obtain a partition of the complete response vector $y_i$
  - observed data $y^{o}_i$, containing those $y_{ij}$ for which $r_{ij} = 1$
  - missing data $y^{m}_i$, containing those $y_{ij}$ for which $r_{ij} = 0$

- For the remaining we will focus on dropout ⇒ notation can be simplified
  - Discrete dropout time: $r^{d}_i = 1 + \sum_{j=1}^{n_i} r_{ij}$ (ordinal variable)
  - Continuous time: $T^{*}_i$ denotes the time to dropout
2.5 Missing Data Mechanisms

- To describe the probabilistic relation between the measurement and missingness processes Rubin (1976, Biometrika) has introduced three mechanisms

- **Missing Completely At Random (MCAR):** The probability that responses are missing is unrelated to both $y_i^o$ and $y_i^m$

  $$p(r_i \mid y_i^o, y_i^m) = p(r_i)$$

- Examples
  - subjects go out of the study after providing a pre-determined number of measurements
  - laboratory measurements are lost due to equipment malfunction
• **Missing At Random (MAR):** The probability that responses are missing is related to $y_i^o$, but is unrelated to $y_i^m$

\[ p(r_i \mid y_i^o, y_i^m) = p(r_i \mid y_i^o) \]

• Examples
  - study protocol requires patients whose response value exceeds a threshold to be removed from the study
  - physicians give rescue medication to patients who do not respond to treatment
2.5 Missing Data Mechanisms (cont’d)

- **Missing Not At Random (MNAR):** The probability that responses are missing is related to $y_i^m$, and possibly also to $y_i^o$

  \[ p(r_i \mid y_i^m) \quad \text{or} \quad p(r_i \mid y_i^o, y_i^m) \]

- **Examples**
  - in studies on drug addicts, people who return to drugs are less likely than others to report their status
  - in longitudinal studies for quality-of-life, patients may fail to complete the questionnaire at occasions when their quality-of-life is compromised
2.5 Missing Data Mechanisms (cont’d)

- Features of MNAR
  - The observed data cannot be considered a random sample from the target population
  - Only procedures that explicitly model the joint distribution \( \{y^o_i, y^m_i, r_i\} \) provide valid inferences \( \Rightarrow \) analyses which are valid under MAR will not be valid under MNAR
2.5 Missing Data Mechanisms (cont’d)

We cannot tell from the data at hand whether the missing data mechanism is MAR or MNAR

Note: We can distinguish between MCAR and MAR
Chapter 3
Relative Risk Models
3.1 Features of Survival Data

- The most important characteristic that distinguishes the analysis of time-to-event outcomes from other areas in statistics is **Censoring**
  - the event time of interest is not fully observed for all subjects under study

- Implications of censoring:
  - standard tools, such as the sample average, the \( t \)-test, and linear regression **cannot** be used
  - inferences may be sensitive to misspecification of the distribution of the event times
3.1 Features of Survival Data (cont’d)

• Several types of censoring:
  ▶ Location of the true event time wrt the censoring time: right, left & interval
  ▶ Probabilistic relation between the true event time & the censoring time: informative & non-informative (similar to MNAR and MAR)

[Here we focus on non-informative right censoring]

• Note: Survival times may often be truncated; analysis of truncated samples requires similar calculations as censoring
3.1 Features of Survival Data (cont’d)

- Notation ($i$ denotes the subject)
  - $T^*_i$ ‘true’ time-to-event
  - $C_i$ the censoring time (e.g., the end of the study or a random censoring time)

- Available data for each subject
  - observed event time: $T_i = \min(T^*_i, C_i)$
  - event indicator: $\delta_i = 1$ if event; $\delta_i = 0$ if censored

Our aim is to make valid inferences for $T^*_i$ but using only $\{T_i, \delta_i\}$
3.2 Relative Risk Models

- **Relative Risk Models** assume a multiplicative effect of covariates on the hazard scale, i.e.,

\[
    h_i(t) = h_0(t) \exp(\gamma_1 w_{i1} + \gamma_2 w_{i2} + \ldots + \gamma_p w_{ip}) \Rightarrow
\]

\[
    \log h_i(t) = \log h_0(t) + \gamma_1 w_{i1} + \gamma_2 w_{i2} + \ldots + \gamma_p w_{ip},
\]

where

- \( h_i(t) \) denotes the hazard for an event for patient \( i \) at time \( t \)
- \( h_0(t) \) denotes the baseline hazard
- \( w_{i1}, \ldots, w_{ip} \) a set of covariates
• **Cox Model:** We make no assumptions for the baseline hazard function

• Parameter estimates and standard errors are based on the log partial likelihood function

\[
p\ell(\gamma) = \sum_{i=1}^{n} \delta_i \left[ \gamma^\top w_i - \log \left\{ \sum_{j: T_j \geq T_i} \exp(\gamma^\top w_j) \right\} \right],
\]

where only patients who had an event contribute
3.2 Relative Risk Models (cont’d)

- **Example:** For the PBC dataset we were interested in the treatment effect while correcting for sex and age effects

\[ h_i(t) = h_0(t) \exp(\gamma_1 D\text{penic}_i + \gamma_2 \text{Female}_i + \gamma_3 \text{Age}_i) \]

<table>
<thead>
<tr>
<th>Value</th>
<th>HR</th>
<th>Std.Err.</th>
<th>z-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma_1$</td>
<td>-0.138</td>
<td>0.871</td>
<td>0.156</td>
<td>-0.882</td>
</tr>
<tr>
<td>$\gamma_2$</td>
<td>-0.493</td>
<td>0.611</td>
<td>0.207</td>
<td>-2.379</td>
</tr>
<tr>
<td>$\gamma_3$</td>
<td>0.021</td>
<td>1.022</td>
<td>0.008</td>
<td>2.784</td>
</tr>
</tbody>
</table>
3.3 Relative Risk Models in R

R> The primary package in R for the analysis of survival data is the `survival` package.

R> A key function in this package that is used to specify the available event time information in a sample at hand is `Surv()`.

R> For right censored failure times (i.e., what we will see in this course) we need to provide the observed event times `time`, and the event indicator `status`, which equals 1 for true failure times and 0 for right censored times.

`Surv(time, status)`
3.3 Relative Risk Models in R (cont’d)

R> Cox models are fitted using function `coxph()`. For instance, for the PBC data the following code fits the Cox model that contains the main effects of ‘drug’, ‘sex’ and ‘age’:

```r
coxFit <- coxph(Surv(years, status2) ~ drug + sex + age,
                 data = pbc2.id)
summary(coxFit)
```

R> The two main arguments are a formula specifying the design matrix of the model and a data frame containing all the variables.
3.4 Time Dependent Covariates

- Often interest in the association between a time-dependent covariate and the risk for an event
  - treatment changes with time (e.g., dose)
  - time-dependent exposure (e.g., smoking, diet)
  - markers of disease or patient condition (e.g., blood pressure, PSA levels)
  - ...

- **Example:** In the PBC study, are the longitudinal bilirubin measurements associated with the hazard for death?
To answer our questions of interest we need to postulate a model that relates

- the serum bilirubin with
- the time-to-death

The association between baseline marker levels and the risk for death can be estimated with standard statistical tools (e.g., Cox regression)

When we move to the time-dependent setting, a more careful consideration is required
3.4 Time Dependent Covariates (cont’d)

- There are two types of time-dependent covariates
  (Kalbfleisch and Prentice, 2002, Section 6.3)
  - Exogenous (aka external): the future path of the covariate up to any time $t > s$ is not affected by the occurrence of an event at time point $s$, i.e.,
    \[
    \Pr\{\mathcal{Y}_i(t) \mid \mathcal{Y}_i(s), T^*_i \geq s\} = \Pr\{\mathcal{Y}_i(t) \mid \mathcal{Y}_i(s), T^*_i = s\},
    \]
    where $0 < s \leq t$ and $\mathcal{Y}_i(t) = \{y_i(s), 0 \leq s < t\}$
  - Endogenous (aka internal): not Exogenous
3.4 Time Dependent Covariates (cont’d)

• It is **very important** to distinguish between these two types of time-dependent covariates, because the type of covariate dictates the appropriate type of analysis.

• In our motivating examples all time-varying covariates are **Biomarkers** ⇒ These are always **endogenous** covariates:
  ▶ measured with error (i.e., biological variation)
  ▶ the complete history is not available
  ▶ existence directly related to failure status
3.4 Time Dependent Covariates (cont’d)

Subject 127

Follow-up Time (months)

0  5  10  15  20

6  8  10  12

CD4 cell count

Joint Models for Longitudinal and Survival Data: April 3, 2020, Den Bosch
3.5 Extended Cox Model

- The Cox model presented earlier can be extended to handle time-dependent covariates using the counting process formulation

\[ h_i(t \mid \mathcal{Y}_i(t), w_i) = h_0(t)R_i(t) \exp\{\gamma^\top w_i + \alpha y_i(t)\}, \]

where

- \( N_i(t) \) is a counting process which counts the number of events for subject \( i \) by time \( t \),
- \( h_i(t) \) denotes the intensity process for \( N_i(t) \),
- \( R_i(t) \) denotes the at risk process (‘1’ if subject \( i \) still at risk at \( t \)), and
- \( y_i(t) \) denotes the value of the time-varying covariate at \( t \)
3.5 Extended Cox Model (cont’d)

• Interpretation:

\[ h_i(t \mid \mathcal{Y}_i(t), w_i) = h_0(t) R_i(t) \exp\{\gamma^\top w_i + \alpha y_i(t)\} \]

\( \exp(\alpha) \) denotes the relative increase in the risk for an event at time \( t \) that results from one unit increase in \( y_i(t) \) at the same time point.

• Parameters are estimated based on the log-partial likelihood function

\[
\ell(\gamma, \alpha) = \sum_{i=1}^{n} \int_0^\infty \left\{ R_i(t) \exp\{\gamma^\top w_i + \alpha y_i(t)\} \right. \\
- \log \left[ \sum_j R_j(t) \exp\{\gamma^\top w_j + \alpha y_j(t)\} \right] \left. \right\} dN_i(t)
\]
3.5 Extended Cox Model (cont’d)

- How does the extended Cox model handle time-varying covariates?
  - assumes no measurement error
  - step-function path
  - existence of the covariate is not related to failure status
3.5 Extended Cox Model (cont’d)

![Diagram showing hazard function and longitudinal outcome over time.]
Therefore, the extended Cox model is only valid for exogenous time-dependent covariates.

**Treating endogenous covariates as exogenous may produce spurious results!**
Chapter 4

The Basic Joint Model
4.1 Joint Modeling Framework

• To account for the special features of endogenous covariates a new class of models has been developed

Joint Models for Longitudinal and Time-to-Event Data

• Intuitive idea behind these models
  1. use an appropriate model to describe the evolution of the marker in time for each patient
  2. the estimated evolutions are then used in a Cox model

• Feature: Marker level’s are not assumed constant between visits
4.1 Joint Modeling Framework (cont’d)
4.1 Joint Modeling Framework (cont’d)

- Some notation
  - $T^*_i$: True event time for patient $i$
  - $T_i$: Observed event time for patient $i$
  - $\delta_i$: Event indicator, i.e., equals 1 for true events
  - $y_i$: Longitudinal responses

- We will formulate the joint model in 3 steps – in particular, . . .
4.1 Joint Modeling Framework (cont’d)

- **Step 1:** Let’s assume that we know $m_i(t)$, i.e., the true & unobserved value of the marker at time $t$

- Then, we can define a standard relative risk model

$$h_i(t \mid M_i(t)) = h_0(t) \exp\{\gamma^\top w_i + \alpha m_i(t)\},$$

where

- $M_i(t) = \{m_i(s), 0 \leq s < t\}$ longitudinal history

- $\alpha$ quantifies the strength of the association between the marker and the risk for an event

- $w_i$ baseline covariates
4.1 Joint Modeling Framework (cont’d)

- **Step 2:** From the observed longitudinal response $y_i(t)$ reconstruct the covariate history for each subject

- Mixed effects model (we focus, for now, on continuous markers)

  $$y_i(t) = m_i(t) + \varepsilon_i(t)$$

  $$= x_i^T(t)\beta + z_i^T(t)b_i + \varepsilon_i(t), \quad \varepsilon_i(t) \sim \mathcal{N}(0, \sigma^2),$$

  where

  - $x_i(t)$ and $\beta$: Fixed-effects part
  - $z_i(t)$ and $b_i$: Random-effects part, $b_i \sim \mathcal{N}(0, D)$
4.1 Joint Modeling Framework (cont’d)

- **Step 3:** The two processes are associated \( \Rightarrow \) define a model for their joint distribution

- Joint Models for such joint distributions are of the following form

\[
p(y_i, T_i, \delta_i) = \int p(y_i \mid b_i) \left\{ h(T_i \mid b_i)^{\delta_i} S(T_i \mid b_i) \right\} p(b_i) \, db_i,
\]

where

- \( b_i \) a vector of random effects that explains the interdependencies
- \( p(\cdot) \) density function; \( S(\cdot) \) survival function
4.1 Joint Modeling Framework (cont’d)

- Key assumption: **Full Conditional Independence** ⇒ random effects explain all interdependencies

  ▶ the longitudinal outcome is independent of the time-to-event outcome
  ▶ the repeated measurements in the longitudinal outcome are independent of each other

\[
p(y_i, T_i, \delta_i \mid b_i) = p(y_i \mid b_i) \ p(T_i, \delta_i \mid b_i)
\]

\[
p(y_i \mid b_i) = \prod_j p(y_{ij} \mid b_i)
\]

**Caveat:** CI is difficult to be tested
4.1 Joint Modeling Framework (cont’d)

- The censoring and visiting processes are assumed non-informative:

- Decision to withdraw from the study or appear for the next visit
  - may depend on observed past history (baseline covariates + observed longitudinal responses)
  - no additional dependence on underlying, latent subject characteristics associated with prognosis

*The visiting process is defined as the mechanism (stochastic or deterministic) that generates the time points at which longitudinal measurements are collected.
4.2 A Comparison with the TD Cox

- **Example:** To illustrate the virtues of joint modeling, we compare it with the standard time-dependent Cox model for the AIDS data

\[
\begin{align*}
  y_i(t) &= m_i(t) + \varepsilon_i(t) \\
  &= \beta_0 + \beta_1 t + \beta_2 \{t \times d\text{d}I_i \} + b_0 + b_1 t + \varepsilon_i(t), \\
  \varepsilon_i(t) &\sim \mathcal{N}(0, \sigma^2), \\
  h_i(t) &= h_0(t) \exp\{\gamma d\text{d}I_i + \alpha m_i(t)\},
\end{align*}
\]

where

▷ \( h_0(t) \) is assumed piecewise-constant
4.2 A Comparison with the TD Cox (cont’d)

<table>
<thead>
<tr>
<th></th>
<th>JM</th>
<th>Cox</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>log HR (std.err)</td>
<td>log HR (std.err)</td>
</tr>
<tr>
<td>Treat</td>
<td>0.33 (0.16)</td>
<td>0.31 (0.15)</td>
</tr>
<tr>
<td>CD4^{1/2}</td>
<td>-0.29 (0.04)</td>
<td>-0.19 (0.02)</td>
</tr>
</tbody>
</table>

- Clearly, there is a considerable effect of ignoring the measurement error, especially for the CD4 cell counts
4.2 A Comparison with the TD Cox (cont’d)

- A unit decrease in CD4^{1/2}, results in a
  - **Joint Model**: 1.3-fold increase in risk (95% CI: 1.24; 1.43)
  - **Time-Dependent Cox**: 1.2-fold increase in risk (95% CI: 1.16; 1.27)

- Which one to believe?
  - a lot of theoretical and simulation work has shown that the Cox model underestimates the true association size of markers
Joint models are fitted using function `jointModel()` from package `JM`. This function accepts as main arguments a linear mixed model and a Cox PH model based on which it fits the corresponding joint model.

```r
lmeFit <- lme(CD4 ~ obstime + obstime:drug,
             random = ~ obstime | patient, data = aids)

coxFit <- coxph(Surv(Time, death) ~ drug, data = aids.id, x = TRUE)

jointFit <- jointModel(lmeFit, coxFit, timeVar = "obstime",
                       method = "piecewise-PH-aGH")

summary(jointFit)
```
4.3 Joint Models in R (cont’d)

R> The data frame given in \texttt{lme()} should be in the long format, while the data frame given to \texttt{coxph()} should have one line per subject*

▷ the ordering of the subjects needs to be the same

R> In the call to \texttt{coxph()} you need to set \texttt{x = TRUE} (or \texttt{model = TRUE}) such that the design matrix used in the Cox model is returned in the object \texttt{fit}

R> Argument \texttt{timeVar} specifies the time variable in the linear mixed model

* Unless you want to include exogenous time-varying covariates or handle competing risks
4.3 Joint Models in R (cont’d)

R> Argument method specifies the type of relative risk model and the type of numerical integration algorithm – the syntax is as follows:

<baseline hazard>-<parameterization>-<numerical integration>

Available options are:

- "piecewise-PH-GH": PH model with piecewise-constant baseline hazard
- "spline-PH-GH": PH model with B-spline-approximated log baseline hazard
- "weibull-PH-GH": PH model with Weibull baseline hazard
- "weibull-AFT-GH": AFT model with Weibull baseline hazard
- "Cox-PH-GH": PH model with unspecified baseline hazard

GH stands for standard Gauss-Hermite; using aGH invokes the pseudo-adaptive Gauss-Hermite rule
Joint models under the Bayesian approach are fitted using function `jointModelBayes()` from package `JMbayes`. This function works in a very similar manner as function `jointModel()`, e.g.,

```r
lmeFit <- lme(CD4 ~ obstime + obstime:drug,
              random = ~ obstime | patient, data = aids)

coxFit <- coxph(Surv(Time, death) ~ drug, data = aids.id, x = TRUE)

jointFitBayes <- jointModelBayes(lmeFit, coxFit, timeVar = "obstime")

summary(jointFitBayes)
```
Joint Models in R (cont’d)

R> **JMbayes** is more flexible (in some respects):
   ▶ directly implements the MCMC
   ▶ allows for categorical longitudinal data as well
   ▶ allows for general transformation functions
   ▶ penalized B-splines for the baseline hazard function
   ▶ . . .
4.3 Joint Models in R (cont’d)

R> In both packages methods are available for the majority of the standard generic functions + extras

▷ `summary()`, `anova()`, `vcov()`, `logLik()`
▷ `coef()`, `fixef()`, `ranef()`
▷ `fitted()`, `residuals()`
▷ `plot()`
▷ `xtable()` (you need to load package `xtable` first)
4.4 Connection with Missing Data

- So far we have attacked the problem from the survival point of view

- However, often, we may be also interested on the longitudinal outcome

- **Issue:** When patients experience the event, they dropout from the study
  - a direct connection with the missing data field
4.4 Connection with Missing Data (cont’d)

• To show this connection more clearly

  ▶ \( T_i^* \): true time-to-event

  ▶ \( y_i^o \): longitudinal measurements before \( T_i^* \)

  ▶ \( y_i^m \): longitudinal measurements after \( T_i^* \)

• **Important to realize** that the model we postulate for the longitudinal responses is for the complete vector \( \{ y_i^o, y_i^m \} \)

  ▶ implicit assumptions about missingness
4.4 Connection with Missing Data (cont’d)

• Missing data mechanism:

\[ p(T_i^* \mid y_i^o, y_i^m) = \int p(T_i^* \mid b_i) p(b_i \mid y_i^o, y_i^m) \, db_i \]

still depends on \( y_i^m \), which corresponds to nonrandom dropout

**Intuitive interpretation**: Patients who dropout show different longitudinal evolutions than patients who do not
4.4 Connection with Missing Data (cont’d)

- **Example:** In the AIDS data the association parameter $\alpha$ was highly significant, suggesting nonrandom dropout.

- A comparison between
  - linear mixed-effects model $\Rightarrow$ MAR
  - joint model $\Rightarrow$ MNAR

  is warranted

- MAR assumes that missingness depends only on the observed data

\[
p(T^*_i \mid y^o_i, y^m_i) = p(T^*_i \mid y^o_i)
\]
### 4.4 Connection with Missing Data (cont’d)

<table>
<thead>
<tr>
<th></th>
<th>LMM (MAR)</th>
<th>JM (MNAR)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>value (s.e.)</td>
<td>value (s.e)</td>
</tr>
<tr>
<td>Inter</td>
<td>7.19 (0.22)</td>
<td>7.22 (0.22)</td>
</tr>
<tr>
<td>Time</td>
<td>-0.16 (0.02)</td>
<td>-0.19 (0.02)</td>
</tr>
<tr>
<td>Treat:Time</td>
<td>0.03 (0.03)</td>
<td>0.01 (0.03)</td>
</tr>
</tbody>
</table>

- Minimal sensitivity in parameter estimates & standard errors
  ⇒ **Warning:** This does not mean that this is always the case!
Chapter 5

Closing
5.1 Concluding Remarks

• **When we need joint models for longitudinal and survival outcomes?**
  - to handle endogenous time-varying covariates in a survival analysis context
  - to account for nonrandom dropout in a longitudinal data analysis context

• **How joint models work?**
  - a mixed model for the longitudinal outcome
  - a relative risk model for the event process
  - explain interrelationships with shared random effects
5.1 Concluding Remarks (cont’d)

- **Where to pay attention when defining joint models?**
  - model flexibly the subject-specific evolutions for the longitudinal outcome
  - use parametric but flexible models for the baseline hazard function
  - consider how to model the association structure between the two processes
    ⇒ Parameterization

- **Extensions**
  - under the full conditional independence assumption we can easily extend the basic joint model
  - multiple longitudinal outcomes and/or multiple failure times
  - though more computationally intensive
5.1 Concluding Remarks (cont’d)

- **Individualized predictions**
  - joint models can provide subject-specific predictions for the longitudinal and survival outcomes
  - these are dynamically updated as extra information is recorded for the subjects
  - ⇒ joint models constitute an excellent tool for personalized medicine

- **What we did not cover**
  - diagnostics for joint models using residuals
  - . . .
The End!
5.2 Additional References


5.2 Additional References (cont’d)


5.2 Additional References (cont’d)


5.2 Additional References (cont’d)


5.2 Additional References (cont’d)


5.2 Additional References (cont’d)


5.3 Medical Papers with Joint Modeling


Practicals
Practical 1: A Simple Joint Model

- We will fit a simple joint model to the PBC dataset

- Start R and load package JM, using `library(JM)`

- The longitudinal (long format) and survival information for the PBC patients can be found in data frames `pbc2` and `pbc2.id`. The variables that we will need are:
Practical 1: A Simple Joint Model (cont’d)

▷ pbc2
  * id: patient id number
  * serBilir: serum bilirubin
  * year: follow-up times in years

▷ pbc2.id
  * years: observed event times in years
  * status: ‘alive’, ‘transplanted’, ‘dead’
  * drug: treatment indicator
• **T1:** Fit the linear mixed effects model for log serum bilirubin using function `lme()`, assuming simple linear evolutions in time for each subject, i.e., a simple random-intercepts and random-slopes structure and different average evolutions per treatment group (see pp. 26–30)

\[
y_i(t) = \beta_0 + \beta_1 t + \beta_2 \{D\text{-}penic_i \times t\} + b_i0 + b_i1 t + \varepsilon_i(t)
\]

• **T2:** Create the indicator for the composite event (i.e., ‘alive’ = 0, ‘transplanted’ or ‘dead’ = 1) using the code

\[
\text{pbc2.id}\$\text{status2} \leftarrow \text{as.numeric(pbc2.id}\$\text{status \!=\! "alive"})
\]
Practical 1: A Simple Joint Model (cont’d)

- **T3:** Fit the Cox PH model using \texttt{coxph()} that includes only treatment as baseline covariate, remember to set \( x = \text{TRUE} \) (see pp. 46–47)

- We want to fit the joint model

\[
\begin{align*}
  y_i(t) &= m_i(t) + \varepsilon_i(t) \\
  &= \beta_0 + \beta_1 t + \beta_2 \{D\text{-penic}_i \times t\} + b_{i0} + b_{i1} t + \varepsilon_i(t), \quad \varepsilon_i(t) \sim \mathcal{N}(0, \sigma^2), \\
  h_i(t) &= h_0(t) \exp\{\gamma D\text{-penic}_i + \alpha m_i(t)\},
\end{align*}
\]
Practical 1: A Simple Joint Model (cont’d)

- **T4:** Fit this joint model based on the fitted linear mixed and Cox models using function `jointModel()` (see pp. 70–72)
  - with piecewise-constant baseline hazard & the (pseudo) adaptive GH rule

- **T5:** Use the `summary()` method to obtain a detailed output of the fitted joint model – interpret the results

- **T6:** Produce 95% confidence intervals for the parameters in the longitudinal submodel, and for the hazard ratios in the survival submodel using function `confint()` (the `parm` argument of `confint()` can take as values "all" (default), "Longitudinal" and "Event")
Practical 1: A Simple Joint Model (cont’d)

- This model assumes that the strength of the association between the level of serum bilirubin and the risk for the composite event is the same in the the two treatment groups.

- To relax this additivity assumption we will add the interaction effect between serum bilirubin and treatment:

\[
\begin{align*}
    y_i(t) &= m_i(t) + \varepsilon_i(t) \\
        &= \beta_0 + \beta_1 t + \beta_2 \{D\text{-penic}_i \times t\} + b_{i0} + b_{i1} t + \varepsilon_i(t), \quad \varepsilon_i(t) \sim \mathcal{N}(0, \sigma^2), \\
    h_i(t) &= h_0(t) \exp\left[\gamma D\text{-penic}_i + \alpha_1 m_i(t) + \alpha_2 \{D\text{-penic}_i \times m_i(t)\}\right],
\end{align*}
\]
Practical 1: A Simple Joint Model (cont’d)

- To fit this model with package **JM** we need to define the `interFact` argument of `jointModel()`. This should be a named **list** with two elements:
  - **value**: a formula with the factors for which we wish to calculate the interaction terms
  - **data**: the data frame used to fit the Cox model

- **T7**: Define this list and fit the corresponding joint model. Use the `summary()` method to obtained a detailed output and interpret the results