

Tutorial I: Motivation for Joint Modeling & Joint Models for Longitudinal and Survival Data

Dimitris Rizopoulos

Department of Biostatistics, Erasmus University Medical Center

`d.rizopoulos@erasmusmc.nl`

Joint Modeling and Beyond

Meeting and Tutorials on Joint Modeling With Survival, Longitudinal, and Missing Data

April 14, 2016, Diepenbeek

Contents

1	Introduction	1
1.1	Motivating Longitudinal Studies	2
1.2	Research Questions	10
1.3	Recent Developments	13
1.4	Joint Models	15
2	Linear Mixed-Effects Models	18
2.1	Features of Longitudinal Data	19
2.2	The Linear Mixed Model	20

2.3	Missing Data in Longitudinal Studies	29
2.4	Missing Data Mechanisms	32
3	Relative Risk Models	37
3.1	Features of Survival Data	38
3.2	Relative Risk Models	41
3.3	Time Dependent Covariates	44
3.4	Extended Cox Model	49
4	The Basic Joint Model	54
4.1	Joint Modeling Framework	55

4.2 Estimation	66
4.3 A Comparison with the TD Cox	69
4.4 Joint Models in R	72
4.5 Connection with Missing Data	78

What are these Tutorials 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

What are these Tutorials About (cont'd)

- 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 are these Tutorials About (cont'd)

Purpose of these tutorials is to introduce the basics of popular

Joint Modelings Techniques

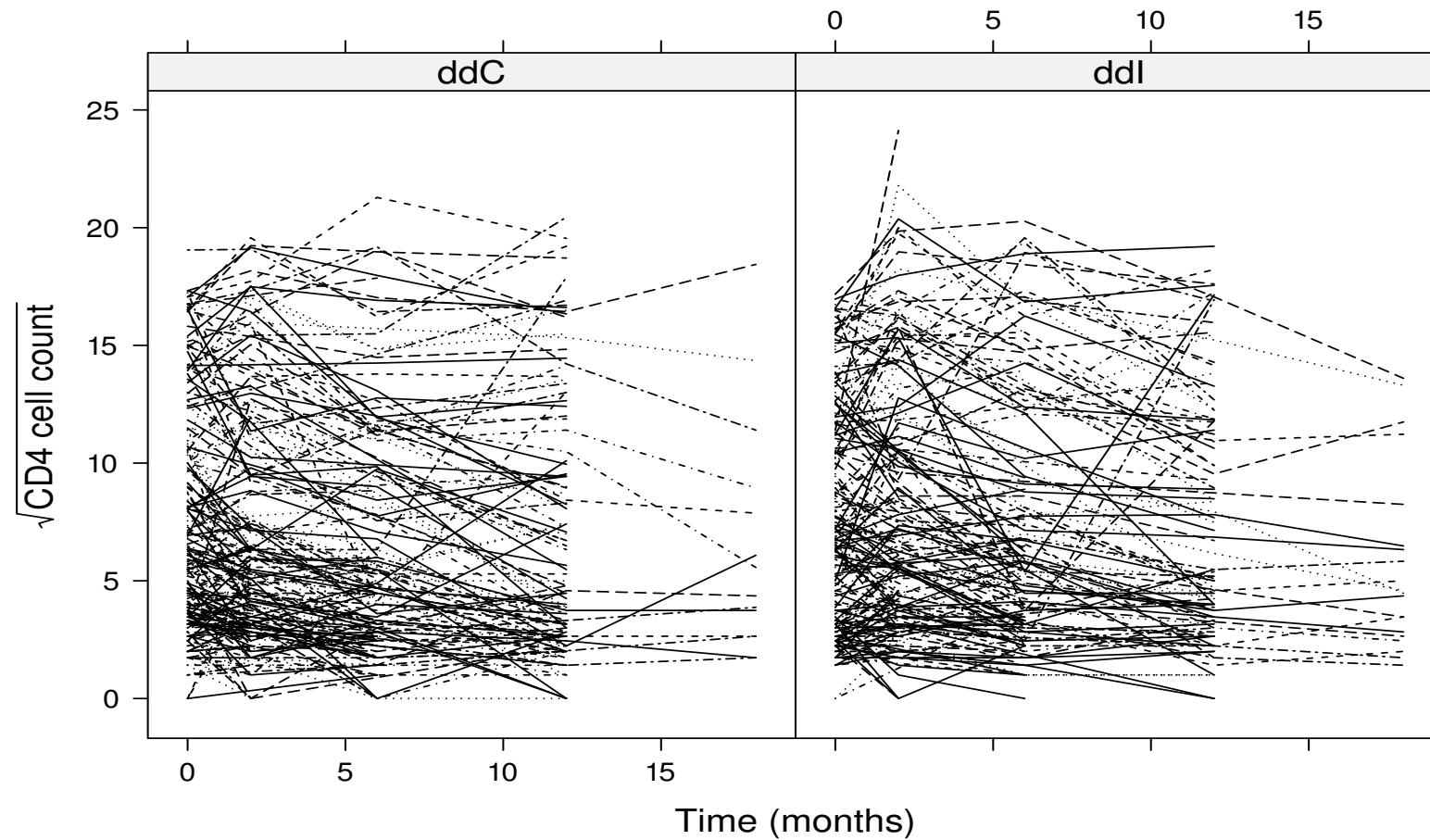
Chapter 1

Introduction

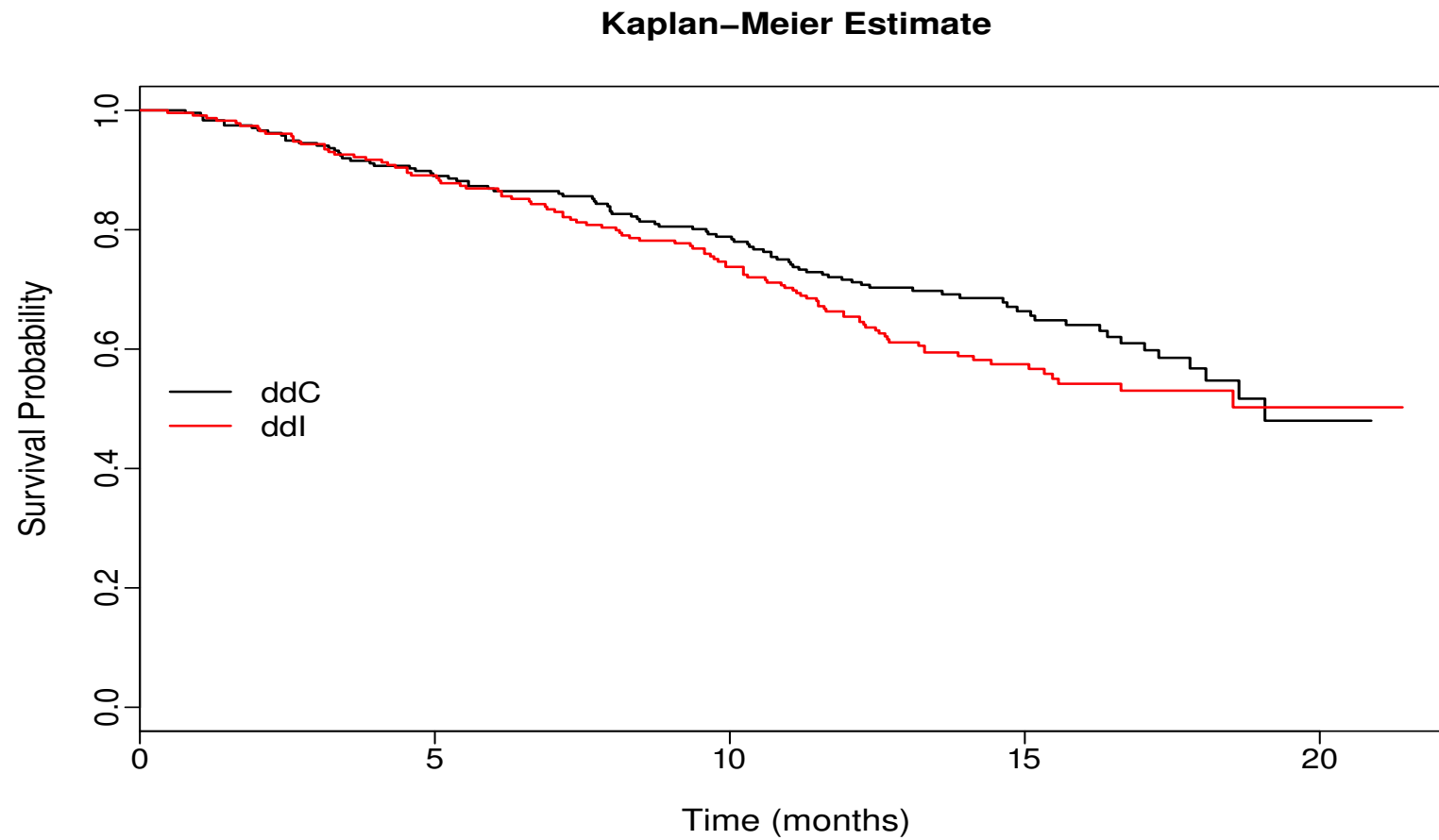
1.1 Motivating Longitudinal Studies

- **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)



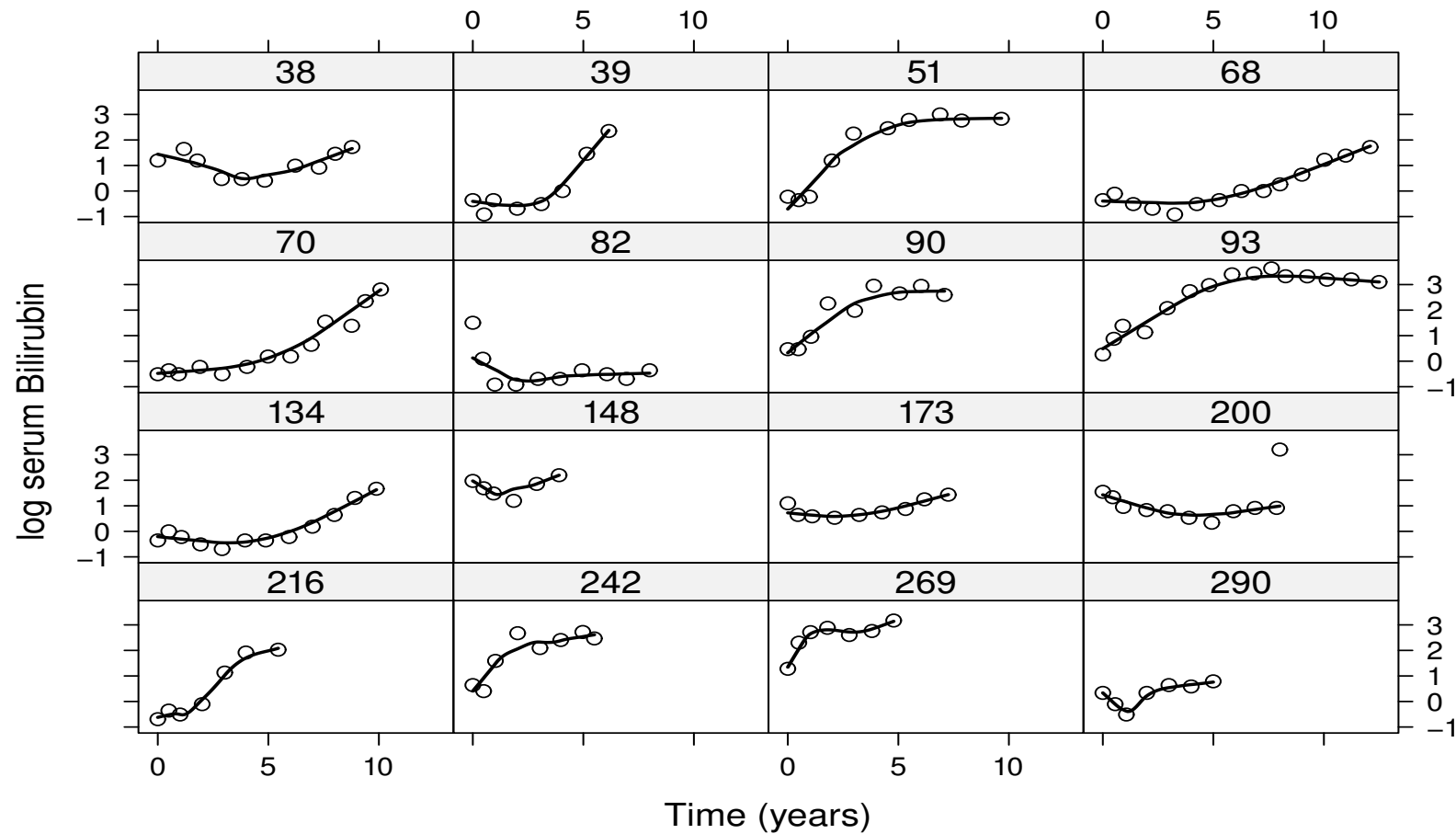
1.1 Motivating Longitudinal Studies (cont'd)

- 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?

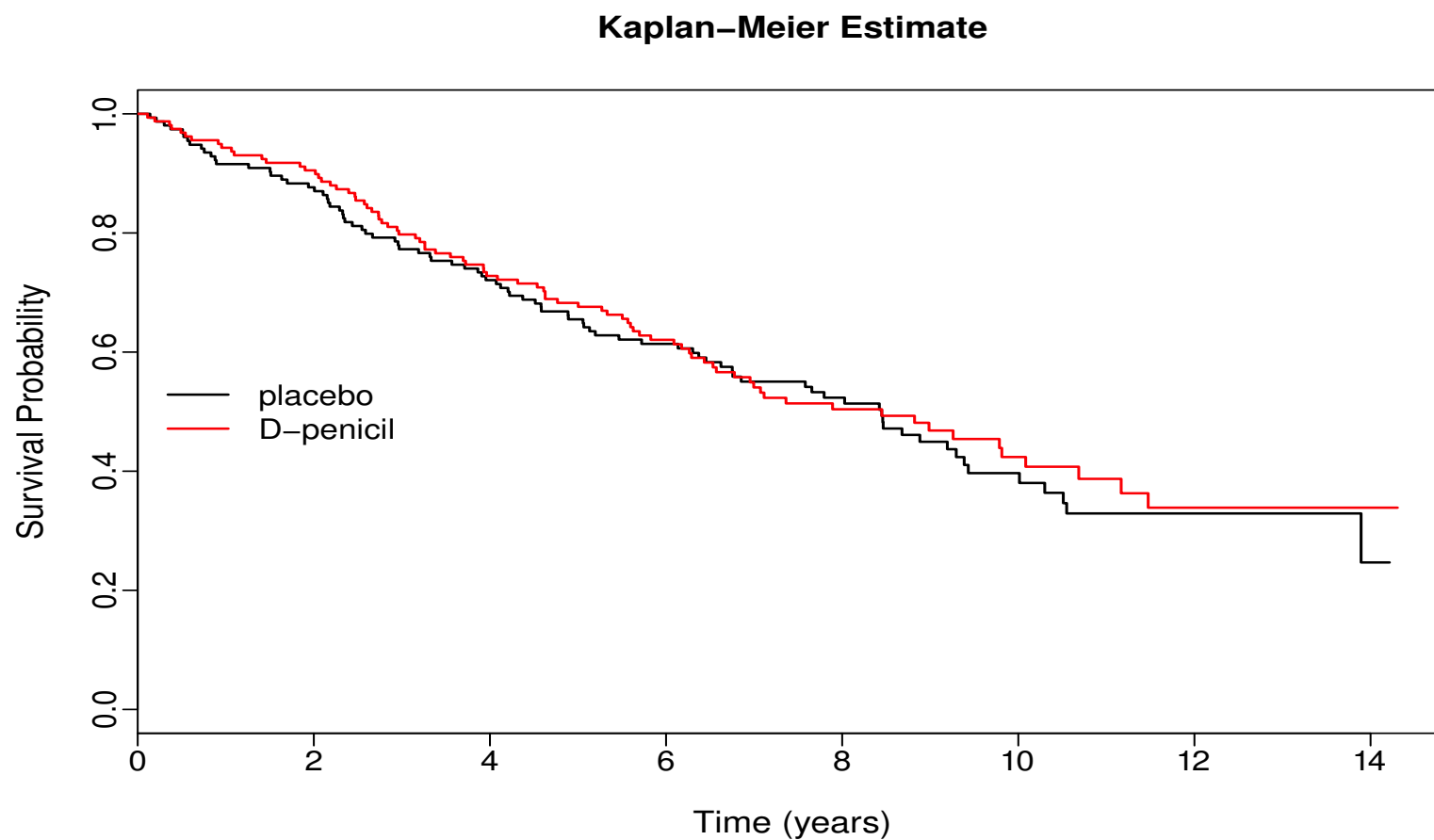
1.1 Motivating Longitudinal Studies (cont'd)

- **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)



1.1 Motivating Longitudinal Studies (cont'd)



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

1.4 Joint Models

- Let Y_1 and Y_2 two outcomes of interest measured on a number of subjects for which joint modeling is of scientific interest
 - ▷ both can be measured longitudinally
 - ▷ one longitudinal and one survival

- We have various possible approaches to construct a joint density $p(y_1, y_2)$ of $\{Y_1, Y_2\}$
 - ▷ Conditional models: $p(y_1, y_2) = p(y_1)p(y_2 | y_1)$
 - ▷ Copulas: $p(y_1, y_2) = c\{\mathcal{F}(y_1), \mathcal{F}(y_2)\}p(y_1)p(y_2)$

But **Random Effects Models** have (more or less) prevailed

1.4 Joint Models (cont'd)

- Random Effects Models specify

$$\begin{aligned}
 p(y_1, y_2) &= \int p(y_1, y_2 \mid b) p(b) db \\
 &= \int p(y_1 \mid b) p(y_2 \mid b) p(b) db
 \end{aligned}$$

- ▷ Unobserved random effects b explain the association between Y_1 and Y_2
- ▷ Conditional Independence assumption

$$Y_1 \perp\!\!\!\perp Y_2 \mid b$$

1.4 Joint Models (cont'd)

- Features:
 - ▷ Y_1 and Y_2 can be of different type
 - * one continuous and one categorical
 - * one continuous and one survival
 - * ...

 - ▷ Extensions to more than two outcomes straightforward

 - ▷ Specific association structure between Y_1 and Y_2 is assumed

 - ▷ Computationally intensive (especially in high dimensions)

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

- The direct approach to model correlated data \Rightarrow *multivariate regression*

$$y_i = X_i\beta + \varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}(0, V_i),$$

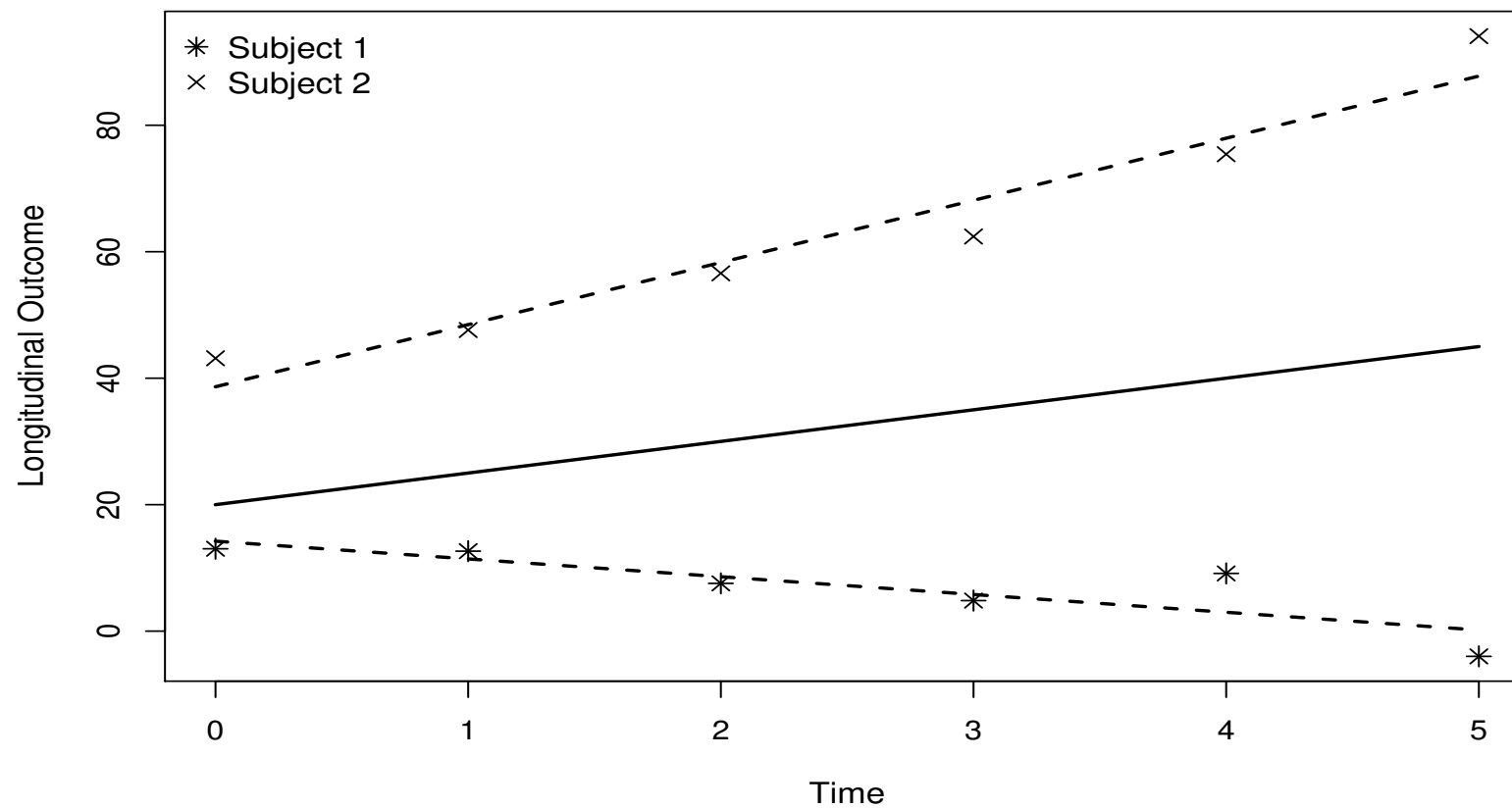
where

- ▷ y_i the vector of responses for the i th subject
 - ▷ X_i design matrix describing structural component
 - ▷ V_i covariance matrix describing the correlation structure
- There are several options for modeling V_i , e.g., compound symmetry, autoregressive process, exponential spatial correlation, Gaussian spatial correlation, ...

2.2 The Linear Mixed Model (cont'd)

- **Alternative intuitive approach:** 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

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

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

- ▷ β s are known as the *fixed effects*
- ▷ b_i s are known as the *random effects*

- In accordance for the random effects we assume

$$b_i = \begin{bmatrix} b_{i0} \\ b_{i1} \end{bmatrix} \sim \mathcal{N}(0, D)$$

2.2 The Linear Mixed Model (cont'd)

- Put in a general form

$$\begin{cases} y_i = X_i\beta + Z_ib_i + \varepsilon_i, \\ b_i \sim \mathcal{N}(0, D), \quad \varepsilon_i \sim \mathcal{N}(0, \sigma^2\mathbf{I}_{n_i}), \end{cases}$$

with

- ▷ X design matrix for the fixed effects β
- ▷ Z design matrix for the random effects b_i
- ▷ $b_i \perp\!\!\!\perp \varepsilon_i$

2.2 The Linear Mixed Model (cont'd)

- Interpretation:
 - ▷ β_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
 - ▷ β 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**)

$$\left\{ \begin{array}{l} y_{ij} = \beta_0 + \beta_1 t_{ij} + \beta_2 \{ \text{ddI}_i \times t_{ij} \} + b_{i0} + b_{i1} t_{ij} + \varepsilon_{ij}, \\ b_i \sim \mathcal{N}(0, D), \quad \varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2) \end{array} \right.$$

- Note: We did not include a main effect for treatment due to randomization

2.2 The Linear Mixed Model (cont'd)

	Value	Std.Err.	<i>t</i> -value	<i>p</i> -value
β_0	7.189	0.222	32.359	< 0.001
β_1	-0.163	0.021	-7.855	< 0.001
β_2	0.028	0.030	0.952	0.342

- No evidence of differences in the average longitudinal evolutions between the two treatments

2.3 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

2.3 Missing Data in Longitudinal Studies (cont'd)

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

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

$$r_{ij} = \begin{cases} 1 & \text{if } y_{ij} \text{ is observed} \\ 0 & \text{otherwise} \end{cases}$$

2.3 Missing Data in Longitudinal Studies (cont'd)

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

- **For the remaining we will focus on dropout** \Rightarrow notation can be simplified
 - ▷ Discrete dropout time: $r_i^d = 1 + \sum_{j=1}^{n_i} r_{ij}$ (ordinal variable)
 - ▷ **Continuous time**: T_i^* denotes the time to dropout

2.4 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

2.4 Missing Data Mechanisms (cont'd)

- *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

2.4 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

$$p(r_i | y_i^m) \quad \text{or} \quad p(r_i | y_i^o, y_i^m)$$

- 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

2.4 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 $\{y_i^o, y_i^m, r_i\}$ provide valid inferences \Rightarrow **analyses which are valid under MAR will not be valid under MNAR**

2.4 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.,

$$h_i(t) = h_0(t) \exp(\gamma_1 w_{i1} + \gamma_2 w_{i2} + \dots + \gamma_p w_{ip}) \Rightarrow$$

$$\log h_i(t) = \log h_0(t) + \gamma_1 w_{i1} + \gamma_2 w_{i2} + \dots + \gamma_p w_{ip},$$

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}, \dots, w_{ip} a set of covariates

3.2 Relative Risk Models (cont'd)

- **Cox Model:** We make no assumptions for the baseline hazard function
- Parameter estimates and standard errors are based on the log partial likelihood function

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

where only patients who had an event contribute

3.2 Relative Risk Models (cont'd)

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

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

	Value	HR	Std.Err.	z-value	p-value
γ_1	-0.138	0.871	0.156	-0.882	0.378
γ_2	-0.493	0.611	0.207	-2.379	0.017
γ_3	0.021	1.022	0.008	2.784	0.005

3.3 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?

3.3 Time Dependent Covariates (cont'd)

- 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

3.3 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 $\mathcal{Y}_i(t) = \{y_i(s), 0 \leq s < t\}$

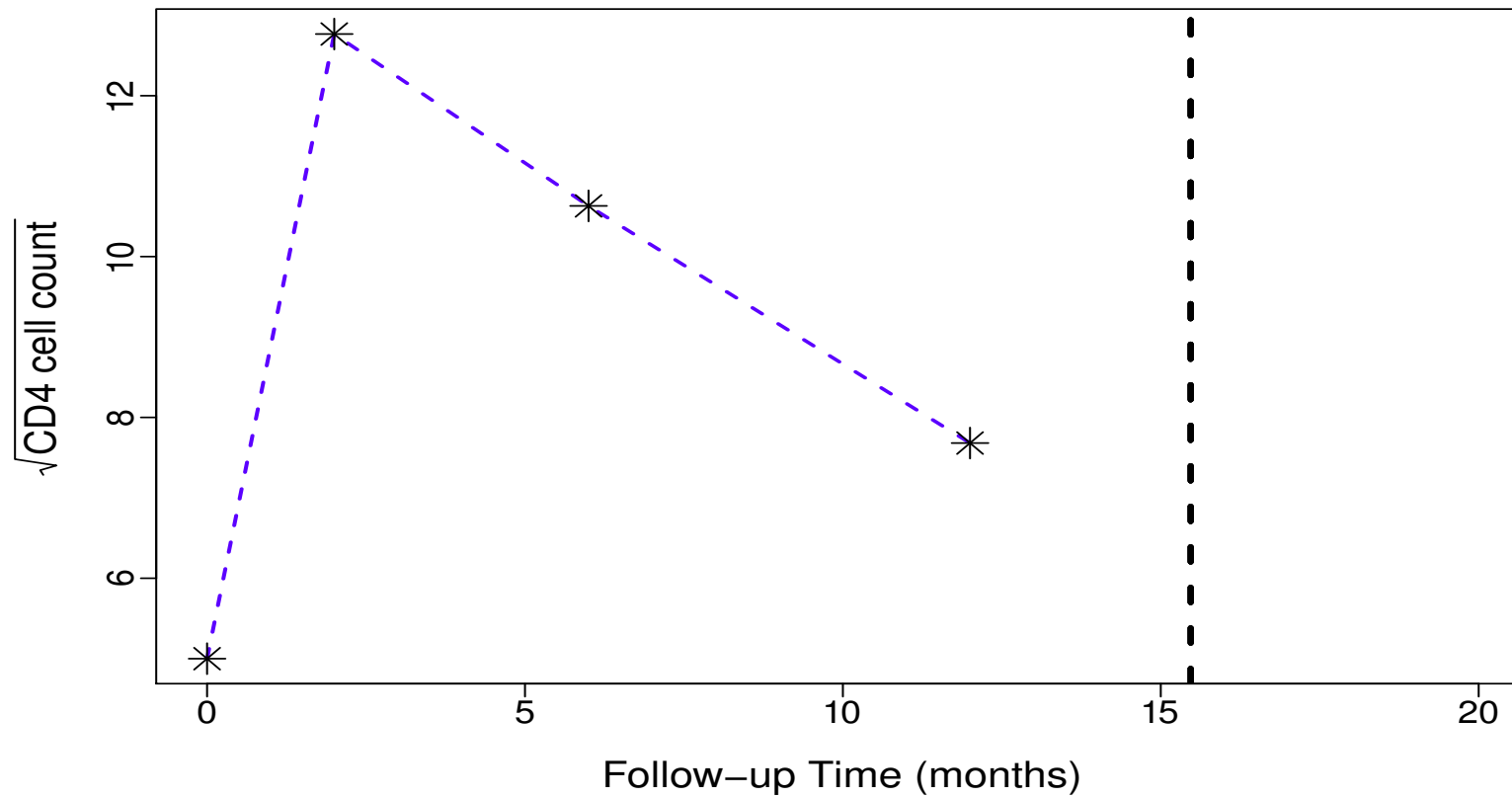
- ▷ Endogenous (aka internal): not Exogenous

3.3 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** \Rightarrow These are always **endogenous** covariates
 - ▷ measured with error (i.e., biological variation)
 - ▷ the complete history is not available
 - ▷ existence directly related to failure status

3.3 Time Dependent Covariates (cont'd)

Subject 127



3.4 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

3.4 Extended Cox Model (cont'd)

- Interpretation:

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

$\exp(\alpha)$ 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

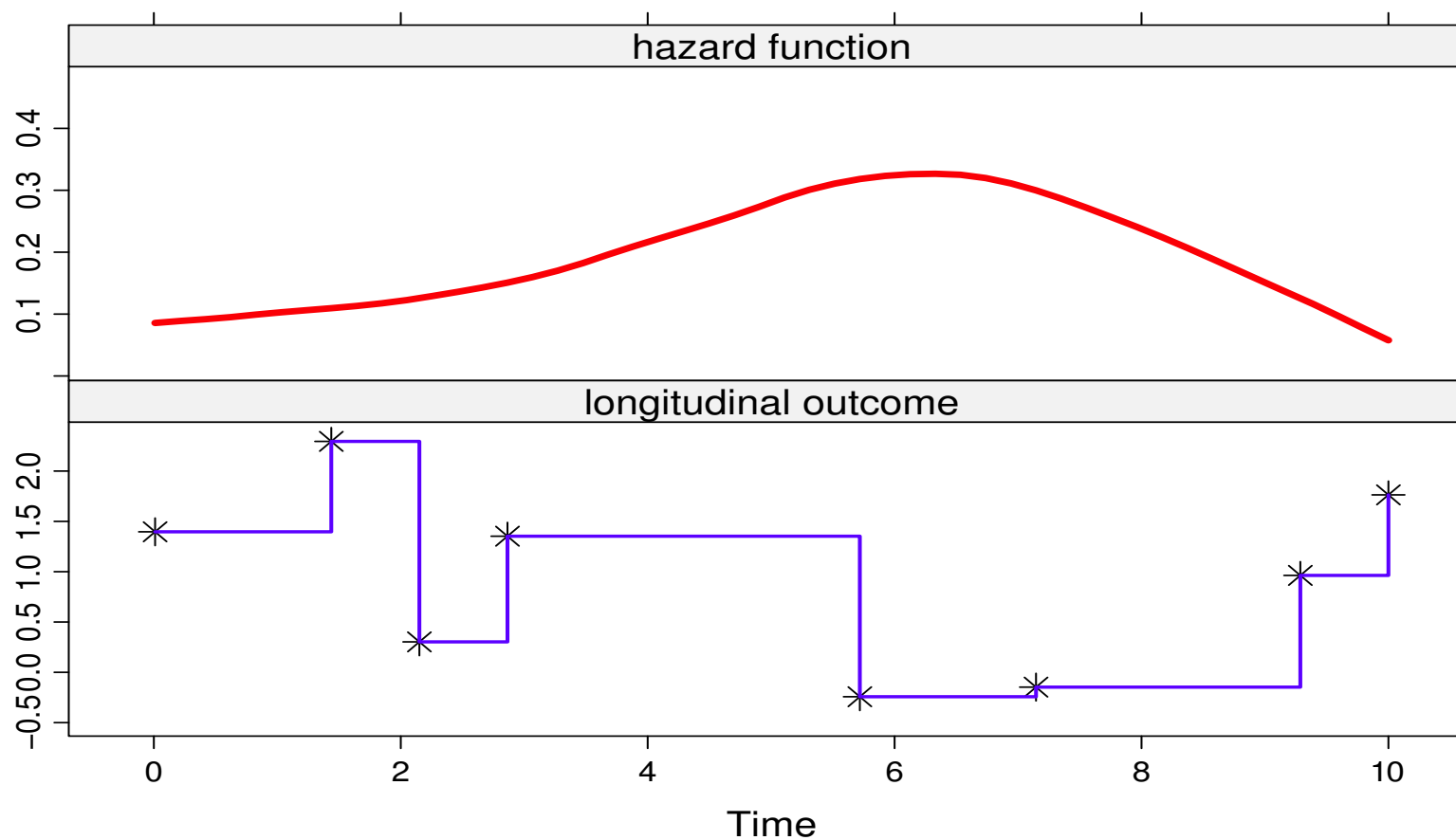
- Parameters are estimated based on the log-partial likelihood function

$$pl(\gamma, \alpha) = \sum_{i=1}^n \int_0^\infty \left\{ R_i(t) \exp\{\gamma^\top w_i + \alpha y_i(t)\} - \log \left[\sum_j R_j(t) \exp\{\gamma^\top w_j + \alpha y_j(t)\} \right] \right\} dN_i(t)$$

3.4 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

3.4 Extended Cox Model (cont'd)



3.4 Extended Cox Model (cont'd)

- 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

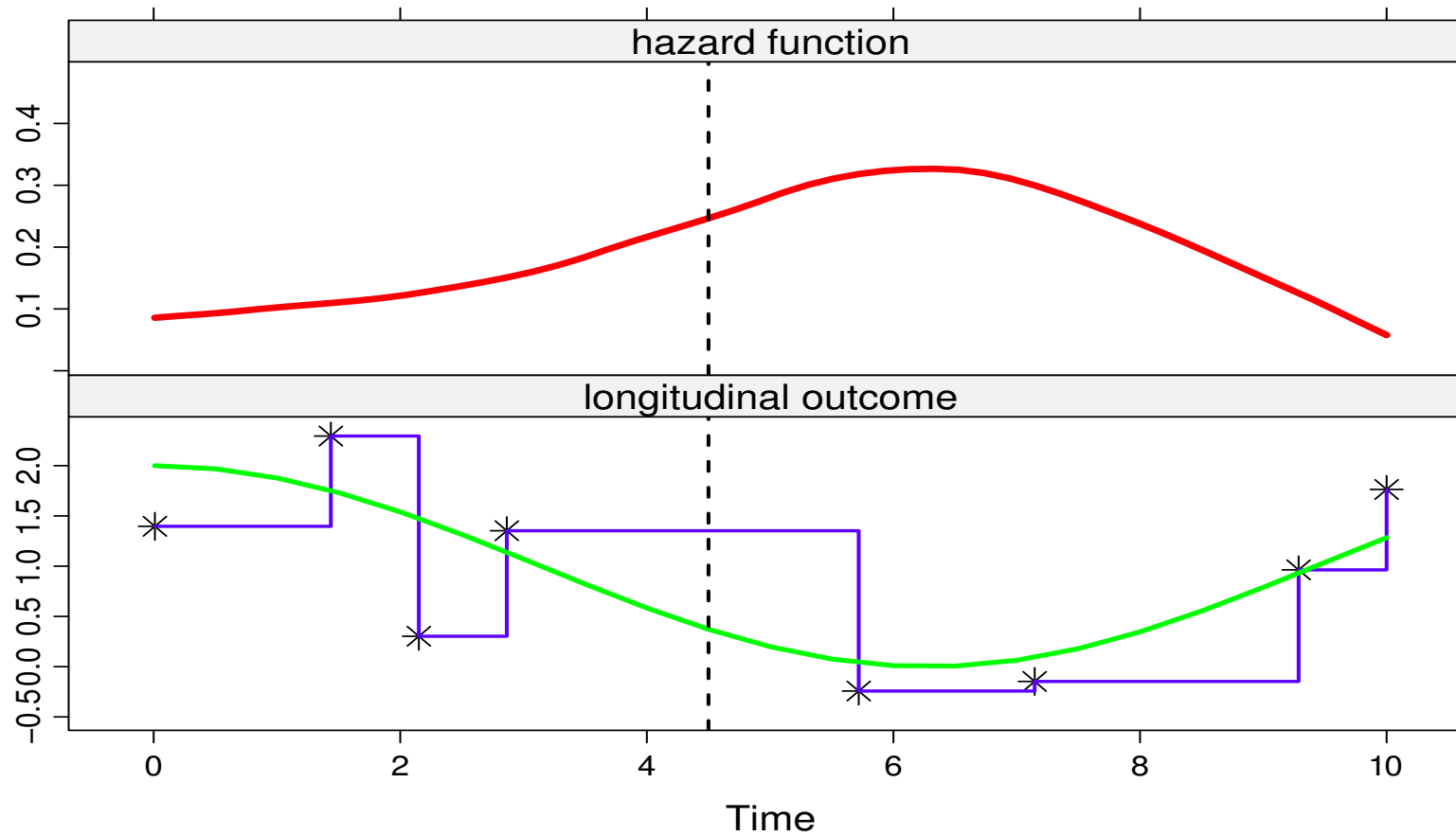
4.1 Joint Modeling Framework

- 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
 - ▷ δ_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

$$h_i(t | \mathcal{M}_i(t)) = h_0(t) \exp\{\gamma^\top w_i + \alpha m_i(t)\},$$

where

- ▷ $\mathcal{M}_i(t) = \{m_i(s), 0 \leq s < t\}$ longitudinal history
- ▷ α 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)

$$\begin{aligned}
 y_i(t) &= m_i(t) + \varepsilon_i(t) \\
 &= x_i^\top(t)\beta + z_i^\top(t)b_i + \varepsilon_i(t), \quad \varepsilon_i(t) \sim \mathcal{N}(0, \sigma^2),
 \end{aligned}$$

where

- ▷ $x_i(t)$ and β : 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
(Tsiatis & Davidian, Stat. Sinica, 2004)

$$p(y_i, T_i, \delta_i) = \int p(y_i | b_i) \{h(T_i | b_i)^{\delta_i} S(T_i | b_i)\} p(b_i) db_i,$$

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** \Rightarrow 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

$$p(y_i, T_i, \delta_i | b_i) = p(y_i | b_i) p(T_i, \delta_i | b_i)$$

$$p(y_i | b_i) = \prod_j p(y_{ij} | b_i)$$

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.

4.1 Joint Modeling Framework (cont'd)

- The survival function, which is a part of the likelihood of the model, depends on the whole longitudinal history

$$S_i(t | b_i) = \exp\left(-\int_0^t h_0(s) \exp\{\gamma^\top w_i + \alpha m_i(s)\} ds\right)$$

- Therefore, care in the definition of the design matrices of the mixed model
 - ▷ when subjects have nonlinear profiles \Rightarrow
 - ▷ use splines or polynomials to model them flexibly

4.1 Joint Modeling Framework (cont'd)

- Assumptions for the baseline hazard function $h_0(t)$
 - ▷ parametric \Rightarrow possibly restrictive
 - ▷ unspecified \Rightarrow within JM framework underestimates standard errors
- It is advisable to use parametric but flexible models for $h_0(t)$
 - ▷ splines

$$\log h_0(t) = \gamma_{h_0,0} + \sum_{q=1}^Q \gamma_{h_0,q} B_q(t, v),$$

where

- * $B_q(t, v)$ denotes the q -th basis function of a B-spline with knots v_1, \dots, v_Q
- * γ_{h_0} a vector of spline coefficients

4.1 Joint Modeling Framework (cont'd)

- It is advisable to use parametric but flexible models for $h_0(t)$
 - ▷ step-functions: piecewise-constant baseline hazard often works satisfactorily

$$h_0(t) = \sum_{q=1}^Q \xi_q I(v_{q-1} < t \leq v_q),$$

where $0 = v_0 < v_1 < \dots < v_Q$ denotes a split of the time scale

4.2 Estimation

- Mainly maximum likelihood but also Bayesian approaches
- The log-likelihood contribution for subject i :

$$\ell_i(\theta) = \log \int \left\{ \prod_{j=1}^{n_i} p(y_{ij} | b_i; \theta) \right\} \left\{ h(T_i | b_i; \theta)^{\delta_i} S_i(T_i | b_i; \theta) \right\} p(b_i; \theta) db_i,$$

where

$$S_i(t | b_i; \theta) = \exp \left(- \int_0^t h_0(s; \theta) \exp \{ \gamma^\top w_i + \alpha m_i(s) \} ds \right)$$

4.2 Estimation (cont'd)

- Both integrals do not have, in general, a closed-form solution \Rightarrow need to be approximated numerically
- Standard numerical integration algorithms
 - ▷ Gaussian quadrature
 - ▷ Monte Carlo
 - ▷ ...
- More difficult is the integral with respect to b_i because it can be of high dimension
 - ▷ Laplace approximations
 - ▷ pseudo-adaptive Gaussian quadrature rules

4.2 Estimation (cont'd)

- To maximize the approximated log-likelihood

$$\ell(\theta) = \sum_{i=1}^n \log \int p(y_i | b_i; \theta) \{h(T_i | b_i; \theta)^{\delta_i} S_i(T_i | b_i; \theta)\} p(b_i; \theta) db_i,$$

we need to employ an optimization algorithm

- Standard choices
 - ▷ EM (treating b_i as missing data)
 - ▷ Newton-type
 - ▷ hybrids (start with EM and continue with quasi-Newton)

4.3 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

$$\left\{ \begin{array}{l} y_i(t) = m_i(t) + \varepsilon_i(t) \\ \quad = \beta_0 + \beta_1 t + \beta_2 \{t \times \text{ddI}_i\} + b_{i0} + b_{i1} t + \varepsilon_i(t), \quad \varepsilon_i(t) \sim \mathcal{N}(0, \sigma^2), \\ \\ h_i(t) = h_0(t) \exp\{\gamma \text{ddI}_i + \alpha m_i(t)\}, \end{array} \right.$$

where

▷ $h_0(t)$ is assumed piecewise-constant

4.3 A Comparison with the TD Cox (cont'd)

	JM	Cox
	log HR (std.err)	log HR (std.err)
Treat	0.33 (0.16)	0.31 (0.15)
CD4 ^{1/2}	-0.29 (0.04)	-0.19 (0.02)

- Clearly, there is a considerable effect of ignoring the measurement error, especially for the CD4 cell counts

4.3 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

4.4 Joint Models in R

R> 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

```
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)
```

4.4 Joint Models in R (cont'd)

- R> The data frame given in `lme()` should be in the long format, while the data frame given to `coxph()` should have one line per subject*
 - ▷ the ordering of the subjects needs to be the same

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

- R> Argument `timeVar` specifies the time variable in the linear mixed model

* Unless you want to include exogenous time-varying covariates or handle competing risks

4.4 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

4.4 Joint Models in R (cont'd)

R> 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.,

```
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)
```

4.4 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
 - ▷ ...

4.4 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)

4.5 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

4.5 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

4.5 Connection with Missing Data (cont'd)

- Missing data mechanism:

$$p(T_i^* | y_i^o, y_i^m) = \int p(T_i^* | b_i) p(b_i | y_i^o, y_i^m) db_i$$

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

4.5 Connection with Missing Data (cont'd)

- Joint models belong to the class of *Shared Parameter Models*

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

the association between the longitudinal and missingness processes is explained by the *shared* random effects b_i

4.5 Connection with Missing Data (cont'd)

- The other two well-known frameworks for MNAR data are
 - ▷ Selection models

$$p(y_i^o, y_i^m, T_i^*) = p(y_i^o, y_i^m) p(T_i^* | y_i^o, y_i^m)$$

- ▷ Pattern mixture models:

$$p(y_i^o, y_i^m, T_i^*) = p(y_i^o, y_i^m | T_i^*) p(T_i^*)$$

- These two model families are primarily applied with discrete dropout times and cannot be easily extended to continuous time

4.5 Connection with Missing Data (cont'd)

- Example: In the AIDS data the association parameter α 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

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

4.5 Connection with Missing Data (cont'd)

	LMM (MAR) value (s.e.)	JM (MNAR) value (s.e.)
Inter	7.19 (0.22)	7.22 (0.22)
Time	-0.16 (0.02)	-0.19 (0.02)
Treat:Time	0.03 (0.03)	0.01 (0.03)

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

The End of Tutorial I!