Week 1: A ”Gentle” Introduction to Joint Modelling in Health Research

Laurent Briollais[1,2]

[1] Lunenfeld-Tanenbaum Research Institute
[2] Dalla Lana School of Public Health, University of Toronto, Toronto, Canada

May 6, 2020

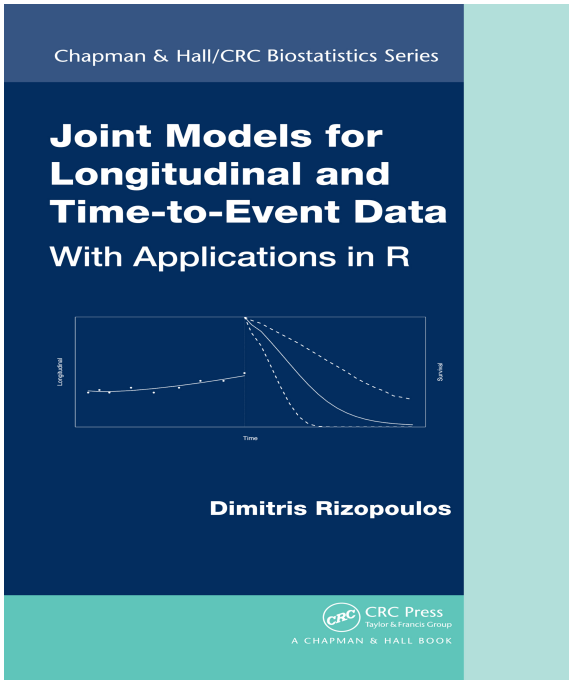
Outline

1. Joint Modeling (JM): General concepts
2. Original applications of JM in Health Research
3. The longitudinal component of the JM
4. The survival component of the JM

1) Joint Modeling (JM): General concepts enumerate

References

- **Joint Models for Longitudinal and Time-to-Event Data With Applications in** *R* from Dimitris Rizopoulos. Chapman & Hall/CRC biostatistics series. CRC Press, Taylor & Francis Group. 2012.



Review papers

- Ibrahim JG, Chu H, Chen LM. **Basic concepts and methods for joint models of longitudinal and survival data.** J Clin Oncol. 2010 Jun 1;28(16):2796-801.

- Papageorgiou G, Mauff K, Tomer A, and Rizopoulos D. **An Overview of Joint Modeling of Time-to-Event and Longitudinal Outcomes.** Annual Review of Statistics and Its Application. First published as a Review in Advance on August 15, 2018.

Tutorial

Journal of Statistical Software

October 2017, Volume 81, Issue 3. *doi: 10.18637/jss.v081.i03*

Tutorial in Joint Modeling and Prediction: A Statistical Software for Correlated Longitudinal Outcomes, Recurrent Events and a Terminal Event

Agnieszka Król
ISPED, INSERM U897
Université de Bordeaux

Audrey Mauguen
ISPED, INSERM U897
Université de Bordeaux

Yassin Mazroui
Université Pierre et Marie Curie
& INSERM, UMR S 1136, Paris

Alexandre Laurent
ISPED, INSERM U897
Université de Bordeaux

Stefan Michiels
INSERM U1018, CESP
& Gustave Roussy, Villejuif

Virginie Rondeau
ISPED, INSERM U897
Université de Bordeaux

Abstract

Extensions in the field of joint modeling of correlated data and dynamic predictions improve the development of prognosis research. The R package **frailtypack** provides estimations of various joint models for longitudinal data and survival events. In particular, it fits models for recurrent events and a terminal event (`frailtyPenal`), models for two survival outcomes for clustered data (`frailtyPenal`), models for two types of recurrent events and a terminal event (`multivPenal`), models for a longitudinal biomarker and a terminal event (`longiPenal`) and models for a longitudinal biomarker, recurrent events and a terminal event (`trivPenal`). The estimators are obtained using a standard and penalized maximum likelihood approach, each model function allows to evaluate goodness-of-fit analyses and provides plots of baseline hazard functions. Finally, the package provides individual dynamic predictions of the terminal event and evaluation of predictive accuracy. This paper presents the theoretical models with estimation techniques, applies the methods for predictions and illustrates **frailtypack** functions details with examples.

Keywords: dynamic prediction, frailty, joint model, longitudinal data, predictive accuracy, R, survival analysis.

Joint Modeling: General concepts

- JMs are applicable in settings where subjects are followed-up over time, e.g., to monitor progress of a disease or medical condition.
- The progression is typically evaluated via repeated measurements of a biomarker.
- It is often of scientific interest to determine the effect of such a biomarker on the time to an event of interest (e.g., death).
- The biomarker measures are called ”endogenous”: their value at any given time point is dependent upon the occurrence of the event prior to that time point.
- They are usually measured with error and their value is only known for the specific time points at which they are measured.

- Methods for analysis of time-to-event outcomes with a time varying covariate (extended Cox model) assume the values of the time-varying covariate are constant in between measurements

  $\Rightarrow$ results in biased estimates and standard errors (Prentice 1982).

- Linear mixed model for longitudinal data and the Cox proportional hazards model for time-to-event data do not consider dependencies between these two different data types (longitudinal and time-to-event data).

- In the follow-up studies, there is always missing data when subjects drop out or do not adhere to the scheduled visiting times.

- When the probability of a subject dropping out depends on their unobserved longitudinal measurements, this dropout process is defined as nonrandom or informative.

  $\Rightarrow$ This process cannot then be ignored, and valid inferences may only be made based on a joint distribution of the longitudinal measurements and the missingness process.

- In both cases of endogeneity and informative missingness, the JM framework provides a solution:
  $\Rightarrow$ A relative risk model for the time-to-event outcome depends on the true underlying value of the longitudinal outcome and where estimation is based on the joint distribution of the two outcomes.

In summary, main motivations for JM

1. Scientific interest
2. Endogeneity
3. Informative missingness
4. Another one to be discussed later is related to dynamic predictions

Different types of JMs

- Basic formulation of JM: association between a single time-to-event outcome and a single continuous longitudinal outcome.
- Other applications: multiple correlated longitudinal outcomes and/or multiple correlated time-to-event outcomes.
- In this "classical" model, the longitudinal data, such as circulating tumor cells, immune response to a vaccine, a genetic biomarker, or a health outcome, can be important predictors or surrogates of a time to event, such as relapse-free survival or overall survival.

Different types of JMs

- Trivariate JM can include a longitudinal outcome, a recurrent time-to-event outcome event and a terminal time-to-event.
- In genetic applications, we could be interested in imposing specific correlation structures, e.g. familial dependences.
- These models are well suited for the analysis of time-varying and complex dynamics of covariates on the disease, e.g. cancer treatment biomarkers and cancer progression.
- One use of JM that is of particular clinical interest is personalized predictive modeling and the development of effective and personalized treatment and prevention such as cancer screening.

Why should we use JM?

- JM are increasingly used in clinical trials because they provide more efficient estimates of the treatment effects on the time to event.
- They provide more efficient estimates of the treatment effects of the longitudinal marker.
- They reduce bias in the estimates of the overall treatment effect, that is, the treatment effect on survival and the longitudinal marker.

2) Original applications of JM in Health Research

Three Illustrative Examples of JM Application

1. The early development of joint models for longitudinal and survival data was primarily motivated from HIV/AIDS clinical trials, in particular, joint modeling of survival data and longitudinal CD4 counts (DeGruttola and Tu, 1994).

2. Primary biliary cirrhosis (PBC) data. PBC is a chronic liver disease that leads to cirrhosis and eventually death.

3. Evaluation of chimio-therapy in cancer treatment.

Example 1: AIDS

- 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 (ddI) and zalcitabine (ddC)
- Outcomes of interest
  - time to death
  - randomized treatment: 230 patients ddI and 237 ddC
  - CD4 cell count measurements at baseline, 2, 6, 12 and 18 months
  - prevOI: previous opportunistic infections

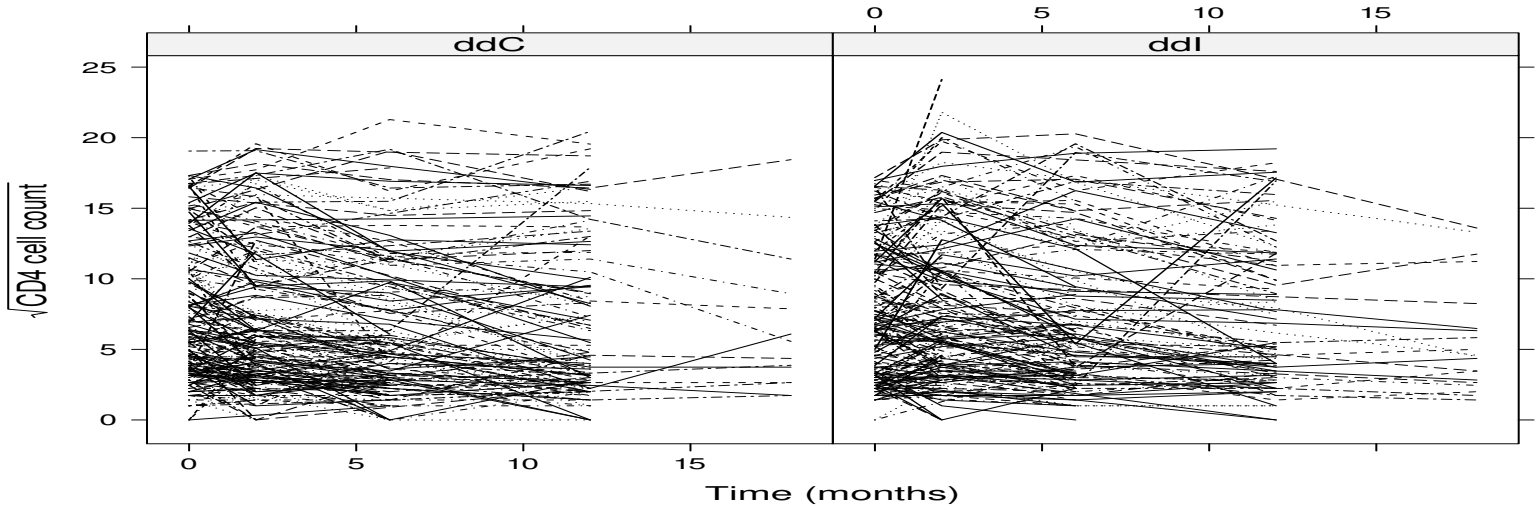
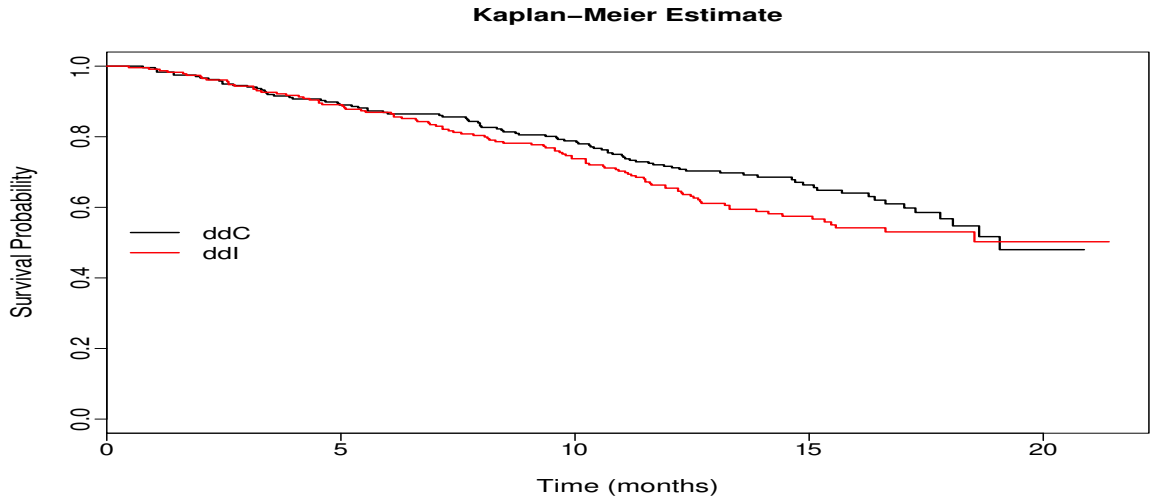


Figure: CD4 cell counts longitudinal profiles



AIDS: 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?

Example 2: Primary Biliary Cirrhosis (PBC)

- 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)

Primary Biliary Cirrhosis

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

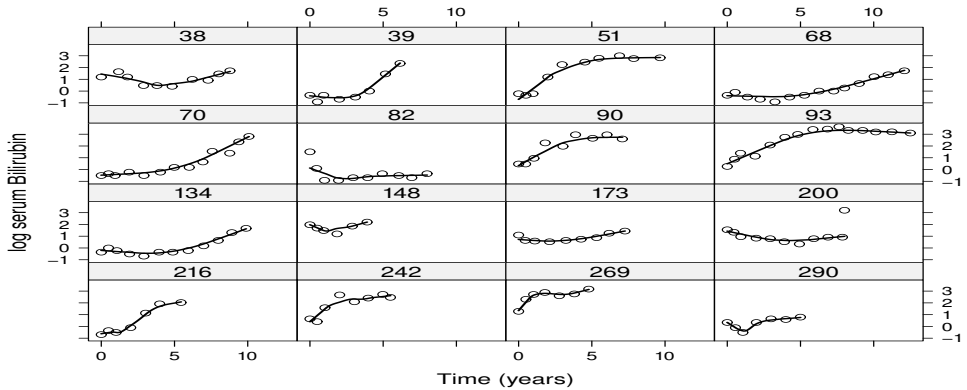


Figure: Longitudinal profiles of serum bilirubin levels

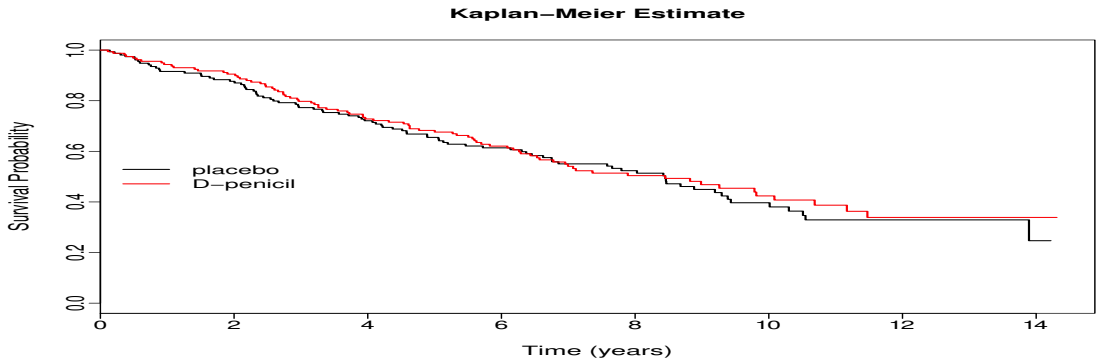


Figure: Overall survival curves

PBC: 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?

Example 3: Cancer treatment randomized clinical trials (RCT)

Careful evaluation of cancer treatment needs to account for:

- Longitudinal evolution pattern of tumor size of target lesions
  $\Rightarrow$ Longitudinal outcome
- Non-target lesions progressions and appearance of new lesions
  $\Rightarrow$ Recurrent time-to-event outcome
- Terminal event of interest: Overall survival (OS), progression free survival (PFS), etc...
  $\Rightarrow$ Recurrent time-to-event outcome

Some history about RECIST

The development of anticancer agents has necessitated a 'common language' to measure tumor response to therapy

- 1979 - WHO criteria
- 2000, 2009 (v1.1) - RECIST (Response Evaluation Criteria in Solid Tumors)
- 2009 - irRC (Immune Related Response Criteria)
- 2017 - iRECIST for clinical trials testing immunotherapeutics

Statistical perspective to cancer clinical trial evaluation

- Model the full continuous longitudinal tumor size data
- Assess OS (or other endpoints) 'jointly' with longitudinal tumor size data
- Incorporate genetic/genomic information to stratify patients with different responses
- Perform individualized predictions of OS conditional on genomic and tumor size information
- Elucidate causal pathway(s) 'driving' the response to treatment

The PRIME Study: Phase III Trial on Untreated Metastatic Colorectal Cancer

VOLUME 28 · NUMBER 31 · NOVEMBER 1 2010

JOURNAL OF CLINICAL ONCOLOGY

ORIGINAL REPORT

Randomized, Phase III Trial of Panitumumab With Infusional Fluorouracil, Leucovorin, and Oxaliplatin (FOLFOX4) Versus FOLFOX4 Alone As First-Line Treatment in Patients With Previously Untreated Metastatic Colorectal Cancer: The PRIME Study

Jean-Yves Douillard, Salvatore Siena, James Cassidy, Josep Tabernero, Ronald Burkes, Mario Barugel, Yves Humblet, György Bodoky, David Cunningham, Jacek Jassem, Fernando Rivera, Ilona Kocákova, Paul Ruff, Maria Błasińska-Morawiec, Martin Šmakal, Jean-Luc Canon, Mark Rother, Kelly S. Oliner, Michael Wolf, and Jennifer Gansert

See accompanying editorial on page 4668 and article on page 4706

The PRIME Study

Figure: 4 groups of patients

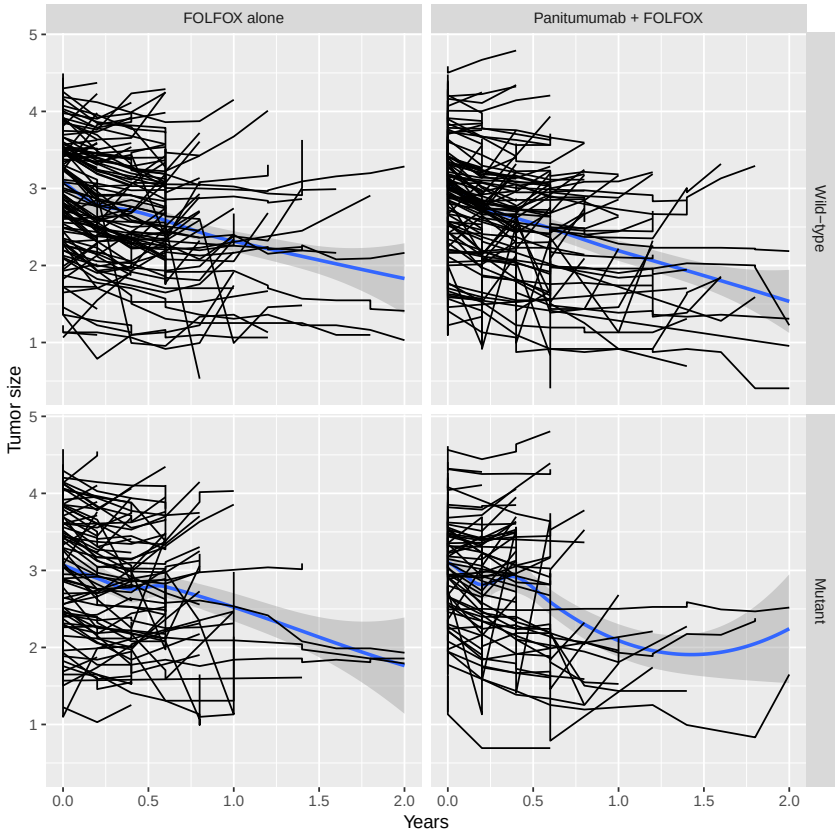


Figure: Tumour evolution

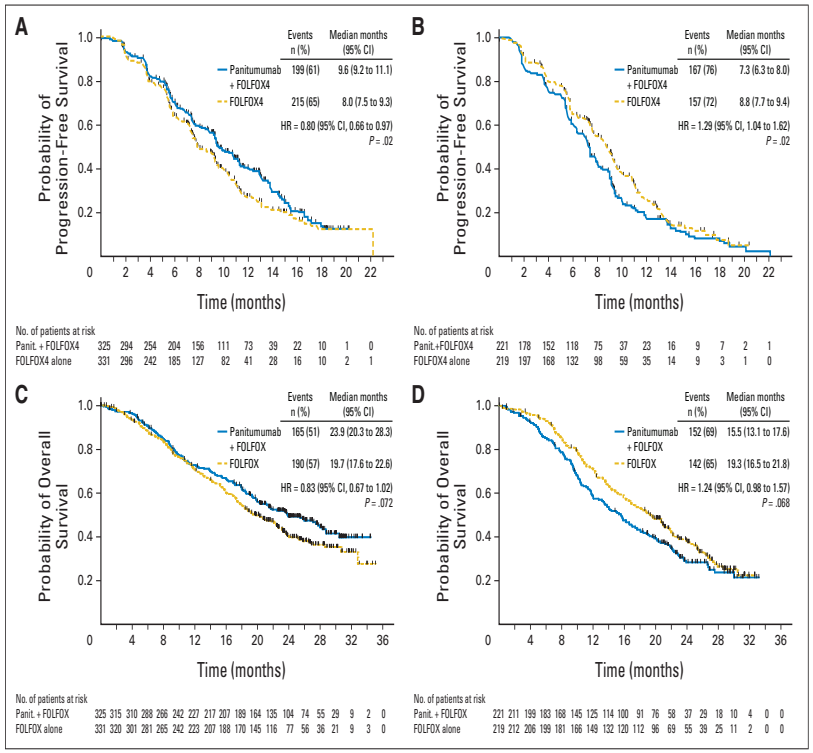


Fig 2. Progression-free survival in patients with (A) wild-type (WT) *KRAS* and (B) mutant (MT) *KRAS*. Overall survival in patients with (C) WT *KRAS* and (D) MT *KRAS*. FOLFOX4, infusional fluorouracil, leucovorin, and oxaliplatin; Panit., panitumumab; HR, hazard ratio.

Figure: OS

The PRIME Study: Research questions

- How about we evaluate OS given PFS or even better based on tumor evolution?
- Does the occurrence of new lesions impact the final conclusion on OS?
- Can we better stratify patients based on an improved model?

Summary of the 3 examples' research questions

- How strong is a treatment effect on survival, or on the longitudinal biomarker or on both?
- What's the association between the longitudinal biomarker and survival?
- Can the biomarker discriminate low/high risk patients?
- Can the biomarker history on each patient improve dynamic prediction of survival outcomes?
- Understand causal relationships between treatment, biomarker and survival.

3) The longitudinal component of JM

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
- Longitudinal studies allow to investigate
  - How treatment means differ at specific time points, e.g., at the end of the study (cross-sectional effect)
  - How treatment means or differences between means of treatments change over time (longitudinal effect)
- Measurements on the same subject are correlated
  $\Rightarrow$ standard statistical tools lead to wrong inference.

The Linear Mixed Model (LMM)

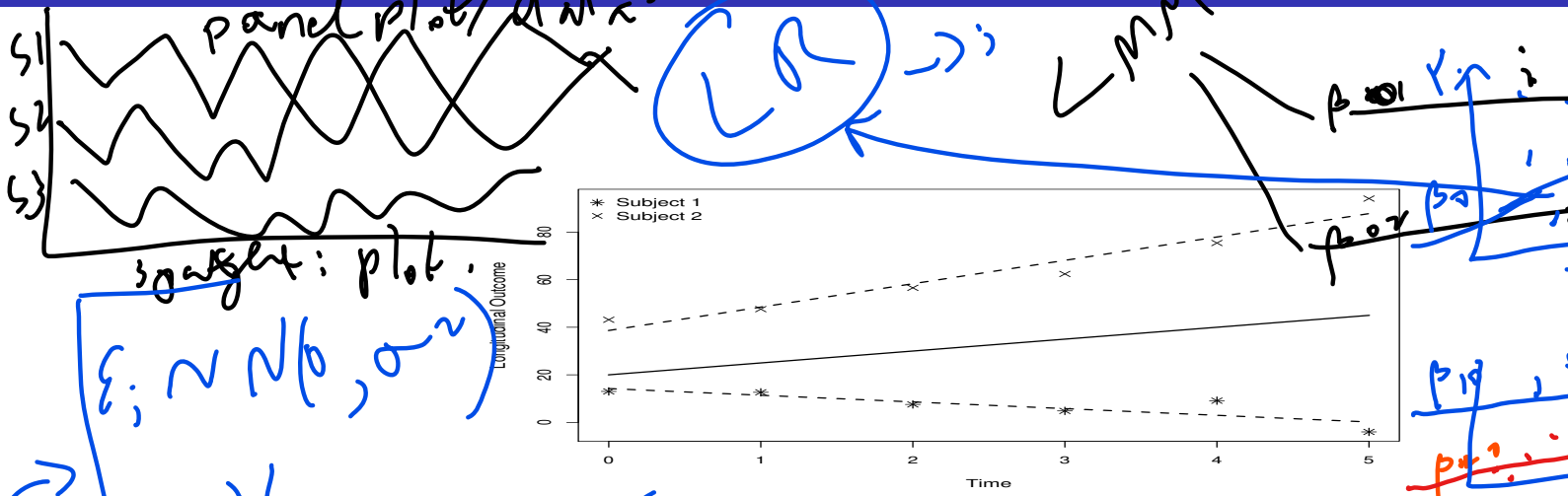
- The direct approach to model correlated data is multivariate regression

$$y_i = X_i^T \beta + \epsilon_i, \quad \epsilon_i \sim N(0, V_i),$$

where
y_i is the vector of responses for the ith subject,
X_i is the design matrix describing structural component,
V_i covariance matrix describing the correlation structure and can be assumed compound symmetry, autoregressive process, etc...

- Alternative intuitive approach: Each subject in the population has his/her own subject-specific mean response profile over time

The Linear Mixed Model (LMM)



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

$$y_{ij} = \beta_{i0} + \beta_{i1} t_{ij} + \epsilon_{ij}, \quad \epsilon_{ij} \sim N(0, \sigma^2),$$

where

y_{ij} the jth response of the ith subject,

β_{i0} is the intercept and β_{i1} the slope for subject i.

The Linear Mixed Model (LMM)

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

$$\beta_i \sim N(\beta, D)$$

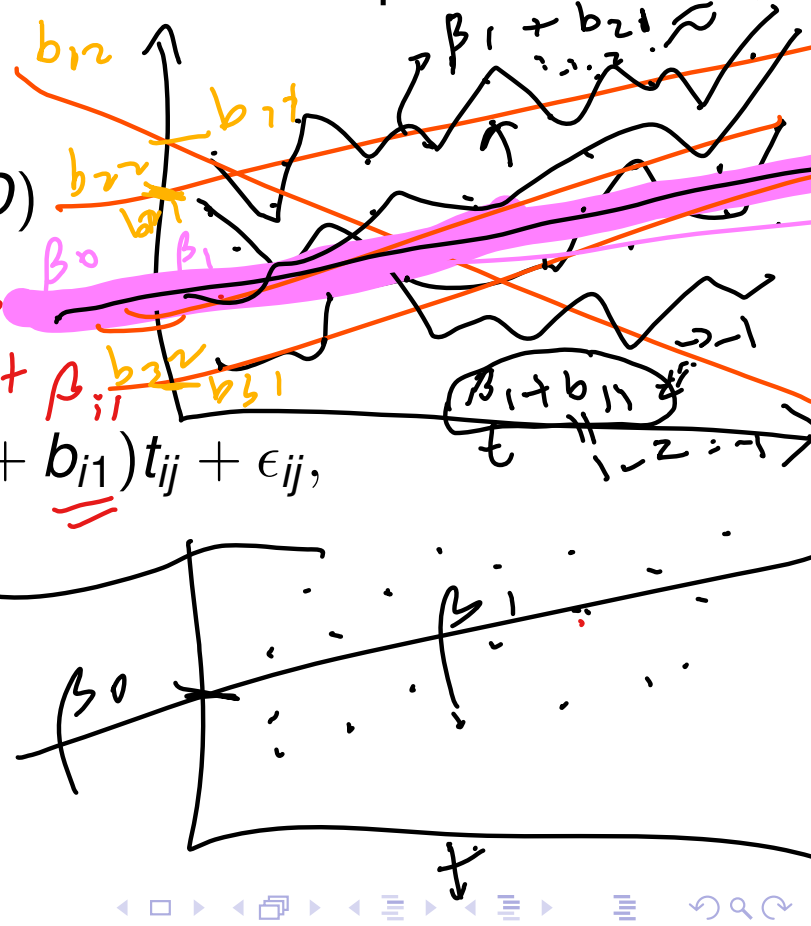
- We can reformulate the model as

$$y_{ij} = (\beta_0 + b_{i0}) + (\beta_1 + b_{i1})t_{ij} + \epsilon_{ij},$$

where

βs are known as the fixed effects,

b_is are known as the random effects.



The Linear Mixed Model (LMM)

Put in a general form, we can write the LMM as

$$\begin{cases} y_i = X_i\beta + Z_i b_i + \epsilon_i, \\ b_i \sim N(0, D), \quad \epsilon_i \sim N(0, \sigma^2 I_{n_i}), \end{cases}$$

with

X design matrix for the fixed effects β,

Z design matrix for the random effects b_i,

$b_i \perp\!\!\!\perp \epsilon_i$

The Linear Mixed Model (LMM)

- 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 ith subject deviates from those in the population
- Advantageous feature: population + subject-specific predictions
  - β describes the mean response changes in the population
  - $\beta + b_i$ describes individual response trajectories

The Linear Mixed Model (LMM)

How do the random effects capture correlation:

Given the random effects, the measurements of each subject are independent (conditional independence assumption)

$$p(y_i|b_i) = \prod_{j=1}^{n_i} p(y_{ij}|b_i)$$

Marginally (integrating out the random effects), the measurements of each subject are correlated

$$p(y_i) = \int p(y_i|b_i)p(b_i)db_i \Rightarrow y_i \sim N(X_i\beta, Z_iDZ_i^T + \sigma^2 I_{n_i})$$

The Linear Mixed Model (LMM)

Estimation

Fixed effects: For known marginal covariance matrix $V_i = Z_i D Z_i^T + \sigma^2 I_{n_i}$, the fixed effects are estimated using generalized least squares

$$\hat{\beta} = \left(\sum_{i=1}^{n} X_i^T V_i^{-1} X_i \right)^{-1} \sum_{i=1}^{n} X_i^T V_i^{-1} y_i$$

Variance Components: The unique parameters in V_i are estimated based on either maximum likelihood (ML) or restricted maximum likelihood (REML).

Example: The AIDS dataset

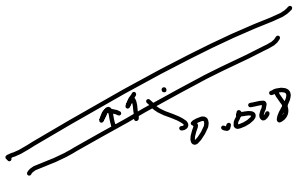
- We fit a LMM with different average longitudinal evolutions per treatment group (fixed part)
- Assume random intercepts & random slopes (random part)

$$\begin{cases} y_{ij} = \beta_0 + \beta_1 t_{ij} + \beta_2 \{ddI_i \times t_{ij}\} + b_{i0} + b_{i1} t_{ij} + \epsilon_{ij}, \\ b_i \sim N(0, D), \quad \epsilon_{ij} \sim N(0, \sigma^2 I_{n_i}), \end{cases}$$

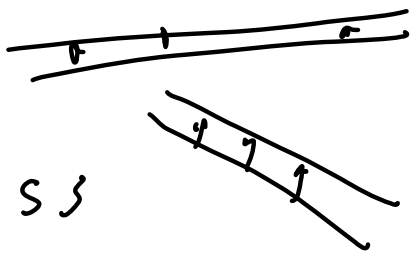
- Note: We did not include a main effect for treatment due to randomization

Example: The AIDS dataset

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



	Value	Std.Err.	t-value	p-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



The Linear Mixed Model (LMM)

We have seen two classes of models for longitudinal data, namely

Marginal Models

$$y_i = X_i\beta + \epsilon_i, \quad \epsilon_i \sim N(0, V_i),$$

Conditional Models

$$\begin{cases} y_i = X_i\beta + Z_i b_i + \epsilon_i, \\ b_i \sim N(0, D); \quad \epsilon_i \sim N(0, \sigma^2 I_{n_i}), \end{cases}$$

The Linear Mixed Model (LMM)

It is also possible to combine the two approaches and obtain a linear mixed model with correlated error terms

$$\begin{cases} y_i = X_i\beta + Z_i b_i + \epsilon_i, \\ b_i \sim N(0, D), \quad \epsilon_i \sim N(0, \Sigma_i), \end{cases}$$

where, as in marginal models, we can consider different forms for Σ_i.

The corresponding marginal model is of the form

$$y_i \sim N(X_i\beta, Z_i D Z_i^T + \Sigma_i)$$

The Linear Mixed Model (LMM)

Features

- Both b_i and Σ_i try to capture the correlation in the observed responses y_i
- This model does not assume conditional independence

Choice between the two approaches is to a large extent philosophical

- Random Effects: trajectory of a subject dictated by time-independent random effects $\Rightarrow$ the shape of the trajectory is an inherent characteristic of this subject
- Serial Correlation: attempts to more precisely capture features of the trajectory by allowing subject-specific trends to vary in time

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 prespecified follow-up times
  - Often subjects miss some of their planned measurements
- We can have different patterns of missing data

Subject	Visits				
	1	2	3	4	5
1	×	×	×	×	×
2	×	×	×	?	?
3	?	×	×	×	×
4	?	×	?	×	?

- ▷ Subject 1: Completer
- ▷ Subject 2: dropout
- ▷ Subject 3: late entry
- ▷ Subject 4: intermittent

Missing Data in Longitudinal Studies

Implications of data 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

Missing Data in Longitudinal Studies

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

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

- The observed data are denoted y_i^o (when $r_{ij} = 1$) and missing data y_i^m (when $r_{ij} = 0$), e.g. due to drop out.

- To describe the probabilistic relation between the measurement and missingness processes Rubin (1976, Biometrika) has introduced 3 mechanisms

Missing Completely At Random (MCAR)

The probability that responses are missing is unrelated to the observed response and missing response.

$$p(r_i|y_i^o, y_i^m) = p(r_i)$$

Examples

- subjects go out of the study after providing a pre-determined number of measurements
- laboratory measurements are lost due to equipment malfunction

Features of MCAR

- The observed responses can be considered a random sample of the complete data
- Can use any statistical procedure that is valid for complete data

Missing at random (MAR)

The probability that responses are missing is related to y_i^o but is unrelated to y_i^m

$$p(r_i|y_i^o, y_i^m) = p(r_i|y_i^o)$$

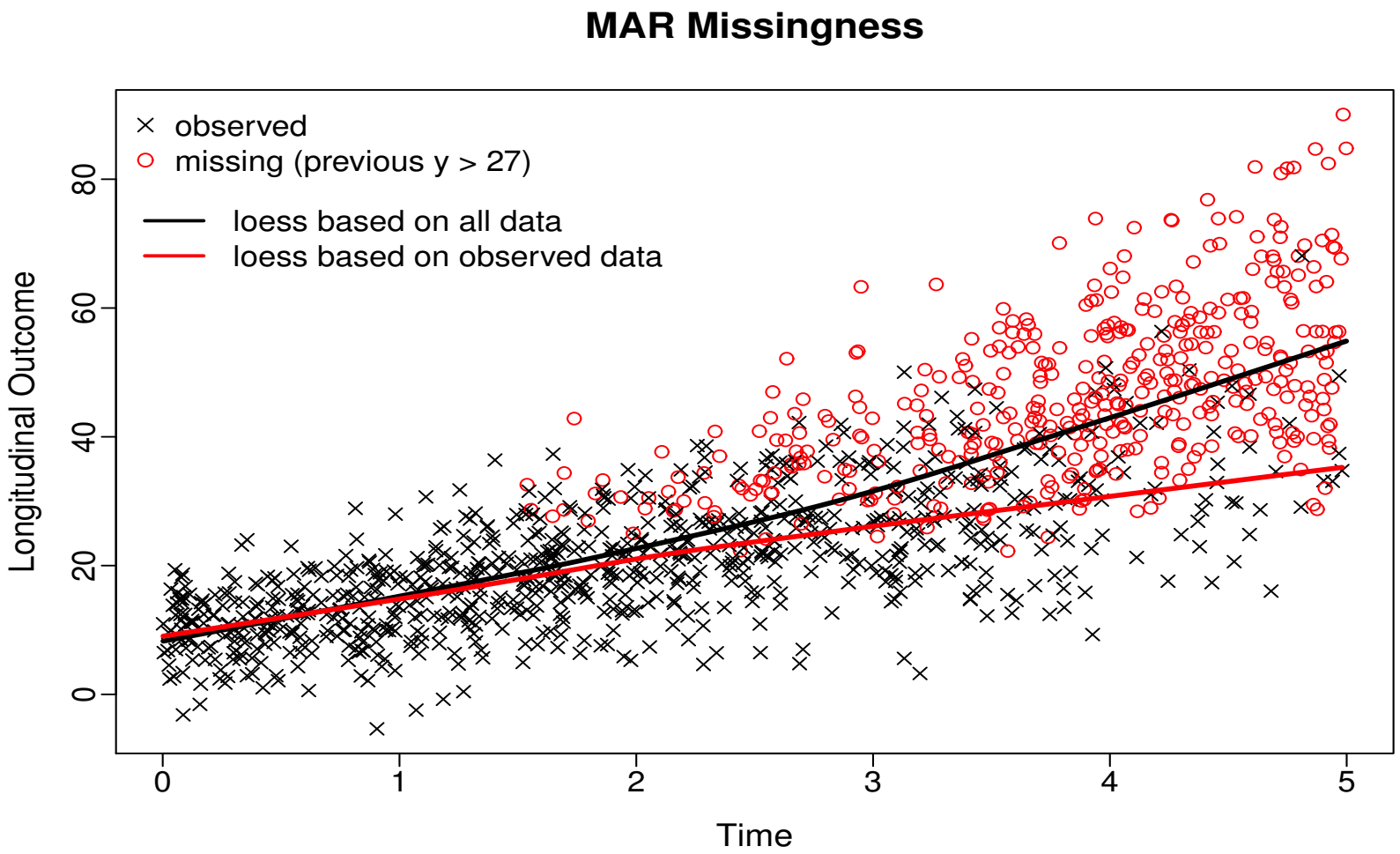
Examples

- 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

Features of MAR

- The observed responses cannot be considered a random sample of the complete data
- Not all statistical procedures provide valid results

Missing at random (MAR)



Missing Not At Random (MNAR)

The probability that responses are missing is related to y_i^m and possibly also to y_i^o

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

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

Missing Not At Random (MNAR)

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

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

Missing Not At Random (MNAR)

- Note that imputation techniques are usually not included in softwares running JMs.
- Missing data is always something to keep in mind either when interpreting the results and their potential biases or when trying to impute the missing values (response and/or covariates)
- JMs can account for some MNAR patterns: longitudinal data censored by time-to-event outcome

4) The survival component of JM

General features

- The most important characteristic that distinguishes the analysis of time-to-event outcomes from other areas in statistics is Censoring.
- It means that the event time of interest is not fully observed for all subjects under study and specific methods should be used to analyze the data.
- Here we will consider non-infornative right censoring.

General features

Notations (i denotes one individual with $i = 1, \cdots, N$):

- T_i^* is the true event time
- C_i is the censoring time (end of study or random censoring)
- The observed event time is: $T_i = \min(T_i^*, C_i)$
- The event indicator is: $\delta_i = 1$ if the event occurred; $\delta_i = 0$ if the individual is censored, i.e., $\delta_i = I_{\{T_i = T_i^*\}}$.
- Survival analysis makes inference on T_i^* by using information on $\{T_i, \delta_i\}$.

Basic functions

Hazard function: The instantaneous risk of an event at time t, given that the event has not occurred until t

$$h(t) = \lim_{dt \to 0} \frac{\Pr(t \leq T^* < t + dt | T^* \geq t)}{dt}, \quad t > 0$$

Survival function: The probability to be alive up to time t

$$S(t) = \Pr(T^* > t) = \exp\left\{ - \int_0^t h(s) ds \right\}$$

where $\int_0^t h(s) ds$ is known as the cumulative hazard function.

Basic functions

Consistent estimates for the survival and cumulative hazard functions that account for censoring are provided by the

- Kaplan-Meier estimator: $\hat{S}_{KM}(t) = \prod_{i:t_i \leq t} \frac{r_i - d_i}{r_i}$
- Nelson-Aalen estimator: $\hat{H}_{NA}(t) = \sum_{i:t_i \leq t} \frac{d_i}{r_i}$

  where r_i is the number of subjects still at risk at t_i and d_i is the number of events observed at t_i.

Cox regression model

Standard approach: focus on the effect of explanatory variables on the hazard function

$$h_i(t) = h_0(t) \exp(X_i^T \beta)$$

- $h_0(.)$ is the baseline hazard function
- X_i is the vector of covariates
- β is the vector of regression coefficients

$\Rightarrow$ We make no assumptions for the baseline hazard function

Cox regression model

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

$$pl(\beta) = \sum_{i=1}^{n} \delta_i \left[X_i^T \beta - \log\{ \sum_{j: T_j \geq T_i} \exp(X_i^T \beta) \} \right]$$

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

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

	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

Time Dependent Covariates

- Often interest in the association between a time-dependent covariate and the risk for an event.
- In the PBC study, are the longitudinal bilirubin measurements associated with the hazard for death?

- Two types of time-dependent covariates
  - Exogenous (i.e. external): the future path of the covariate up to time $t > s$ is not affected by the occurrence of an event at time point s
  - Endogenous (i.e. internal): not Exogenous

- In our motivating examples all time-varying covariates are biomarkers $\Rightarrow$ endogenous: e.g., existence and value directly related to failure status.

Time Dependent Covariates

Characteristics of ~~exogenous~~ endogen time dependent covariates

- Measured with error (i.e., biological variation)
- The complete history is not available
- Existence directly related to failure status

The Cox model

Can be extended to handle time-dependent covariates

$$h_i(t|Y_i(t), X_i) = h_0(t)R_i(t)\exp\{X_i^T\beta + \alpha y_i(t)\}$$

$$Y_i(t) = \{y_i(s), 0 \geq s < t\}$$

$R_i(t)$ denotes the at risk process (1 if subject i still at risk at t and 0 otherwise),

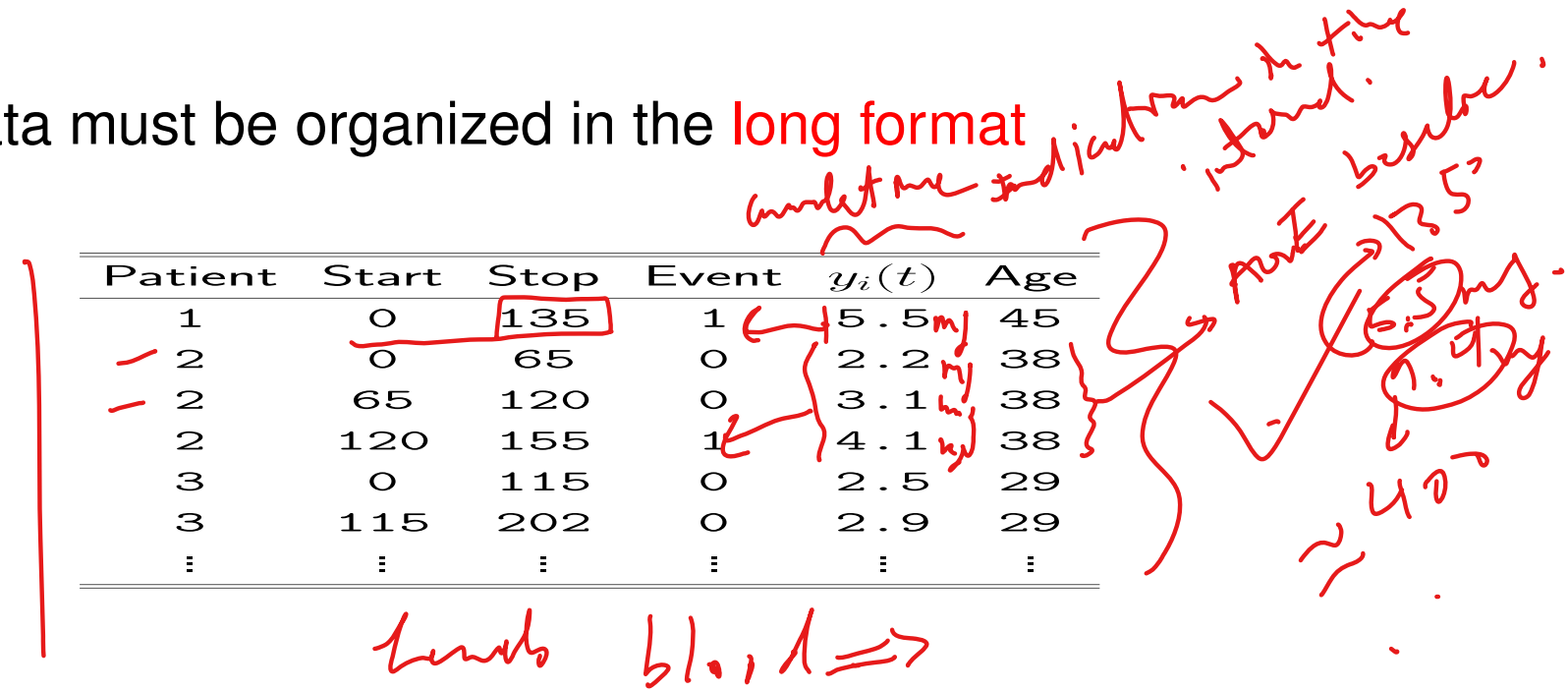
$y_i(t)$ denotes the value of the time-varying covariate at t.

$\exp(\alpha)$ 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

The Cox model: Time-dependent covariates

Typically, data must be organized in the long format

Patient	Start	Stop	Event	$y_i(t)$	Age
1	0	135	1	5.5	45
2	0	65	0	2.2	38
2	65	120	0	3.1	38
2	120	155	1	4.1	38
3	0	115	0	2.5	29
3	115	202	0	2.9	29
⋮	⋮	⋮	⋮	⋮	⋮

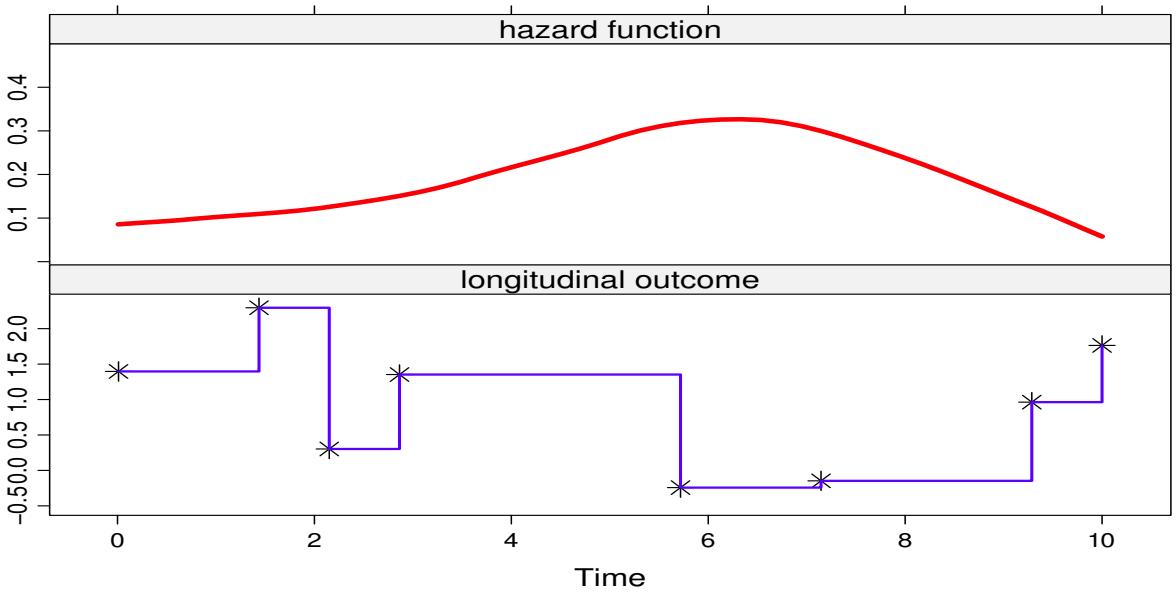


How does the extended Cox model handle time-varying covariates?
$\Rightarrow$ existence of the covariate is not related to failure status

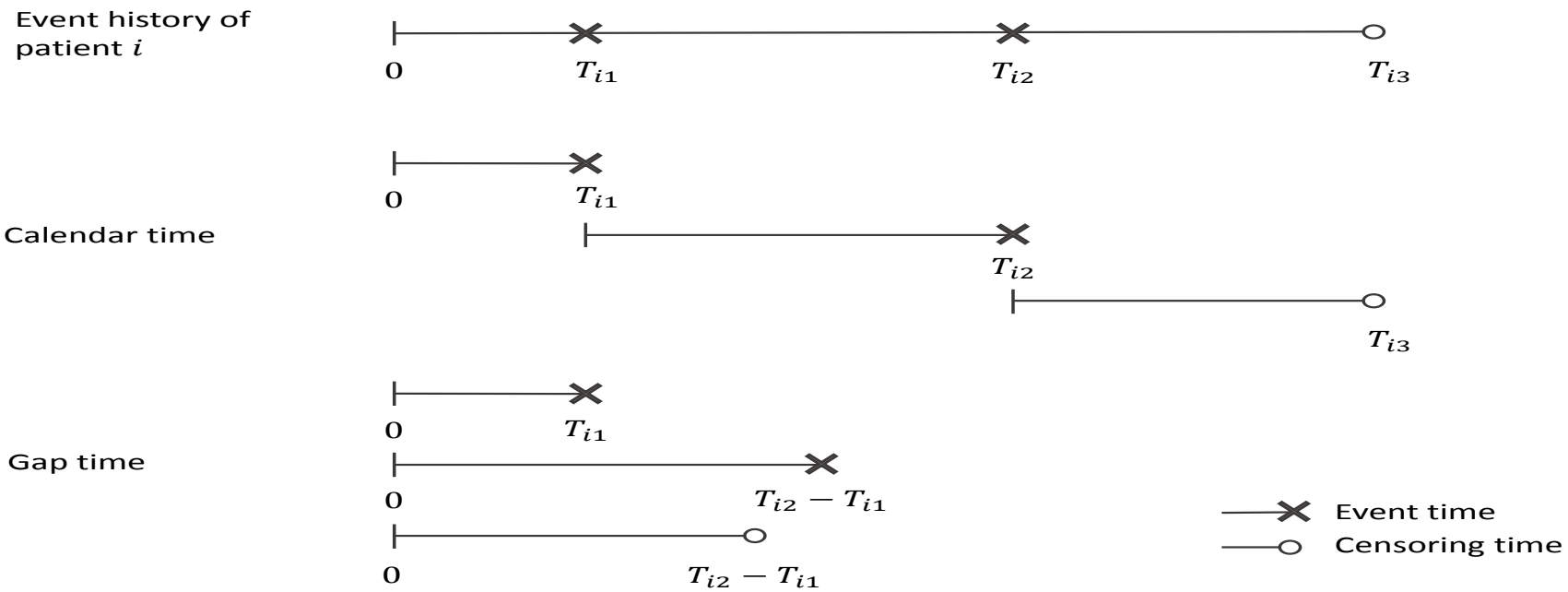
Therefore, the extended Cox model is only valid for exogenous time-dependent covariates $\Rightarrow$ Motivation for using JM

The Cox model: Time-dependent covariates

The Cox model also assumes the time-depend covariates are constant within interval.

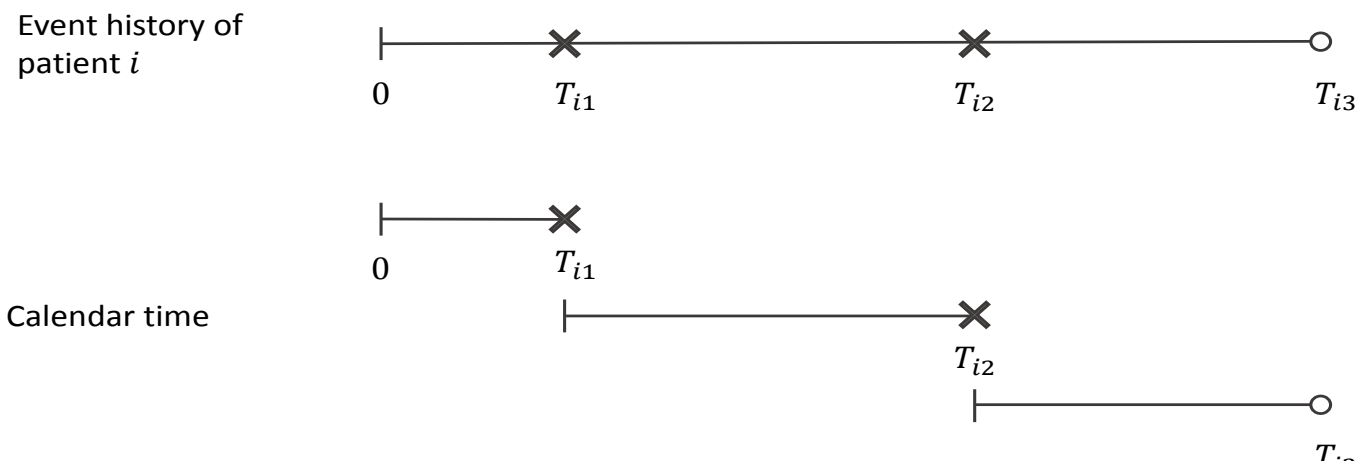


Time scale


Event history of patient i
0
T_{i1}
T_{i2}
T_{i3}
Calendar time
0
T_{i1}
T_{i2}
T_{i3}
Gap time
0
T_{i1}
0
$T_{i2} - T_{i1}$
0
$T_{i2} - T_{i1}$
Event time
Censoring time

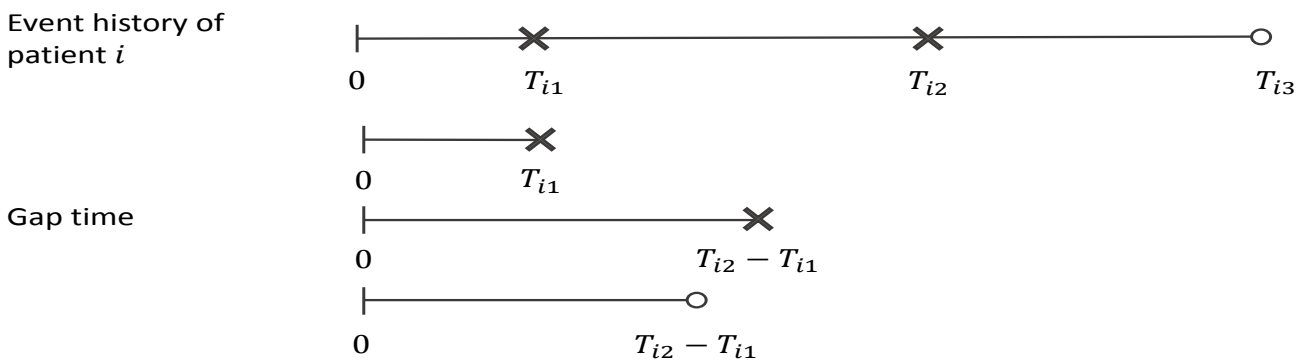
Time scale: Calendar time scale

- Time from the study start until the occurrence of an event
- The beginning of the at risk period corresponds to the time of the previous event $\Rightarrow$ left-truncation considered
- The counters are not reset to zero
- The order is important



Time scale: Gap time scale

- Time interval between two successive events
- Time reset to 0 after every event
- The order of events is not considered
- Transition probability between two states depends only on the time spent in the state



Choice of time-scale

Calendar time

- Consideration of progressive evolution of disease
- In clinical trials when the interest is to compare the effect of different treatments

Gap time

- Interest in delays between successive events (e.g. intensity of a process)
- When the occurrence of events does not affect importantly the condition of a subject

Analysis of clustered data/recurrent events

Marginal approach

- Model average number or rate of events marginally on inter-cluster dependencies but conditionally on covariates e.g.:
- Andersen-Gill model $\Rightarrow$ inter-cluster events are uncorrelated, given the covariates

Conditional approach

- Shared frailty models $\Rightarrow$ extension of Cox PH model
- Random effect, called frailty, describes the excess risk of a cluster
- Frailty term represents this part of heterogeneity that is unobserved or even unobservable

Shared frailty model

Notations

- Clustered data : observation $j = 1, \cdots, r_i$ from cluster $i = 1, \cdots, N$
- Recurrent events : event j of individual i
- We observe (T_{ij}, δ_{ij}) defined as before
- Calendar time scale: $T_{i1}, T_{i2}, \cdots, T_{ir_i}$
- Gap time scale: $S_{i1} = T_{i1}, S_{i2} = T_{i2} - T_{i1}, \cdots, S_{ir_i} = T_{ir_i} - T_{ir_{i-1}}$

The Cox model with Shared frailty

Formulation

$$h_{ij}(T|u_i) = u_i h_0(t) \exp(X_{ij}^T \beta)$$

- $h_0(t)$ is the baseline hazard function
- $X_{ij} = (X_{1ij}, \cdots, X_{pij})^T$ is the vector of covariates
- β is the vector of regression coefficients
- u_i is the frailty term shared by all individuals from the same cluster $\Rightarrow$ all unobserved risk factors
- when $u_i > 1$ the hazard increases (frail subjects experience the event earlier)

The Cox model with Shared frailty

Assumptions

- Dependence of the survival times within each cluster
- Independence of the survival times between clusters
- Proportionality of the hazards conditionally on the frailty
- Frailty distribution can be assumed Gamma, log-normal, etc...
- Estimation in maximum likelihood framework or Bayesian framework

General conclusion

Conclusion

- Motivations for JM
  - Scientific interest
  - Endogeneity
  - Informative missingness
  - Another one to be discussed later is related to dynamic predictions
- JMs have gained increased interest in public health fields
- Methodologies have been developed for JMs either based on frailty models or some extended survival analysis models (e.g. Cox regression model).