Module 14

Joint Modeling of Longitudinal and Survival Data

Benjamin French, PhD
Department of Biostatistics, Vanderbilt University

SISCER 2022
July 25–26, 2022

Learning objectives

- This module will focus on combining a model for repeated measures with a model for survival times using a joint regression model
  - Chapters 1–4 of Rizopoulous (2012)
  - See also Asar et al. (2015)
- Case studies will be used to discuss analysis strategies, the application of appropriate analysis methods, and the interpretation of results, with implementation in R, particularly the `JM` package
- Some theoretical background and technical details will be provided; our goal is to translate statistical theory into practical application
- At the conclusion of this course, you should be able to apply appropriate exploratory and regression techniques to summarize and generate inference from longitudinal and survival data

Overview

Motivation and examples

Longitudinal data analysis

Survival data analysis

Joint regression models

Overview

Motivation and examples

Longitudinal data analysis

Survival data analysis

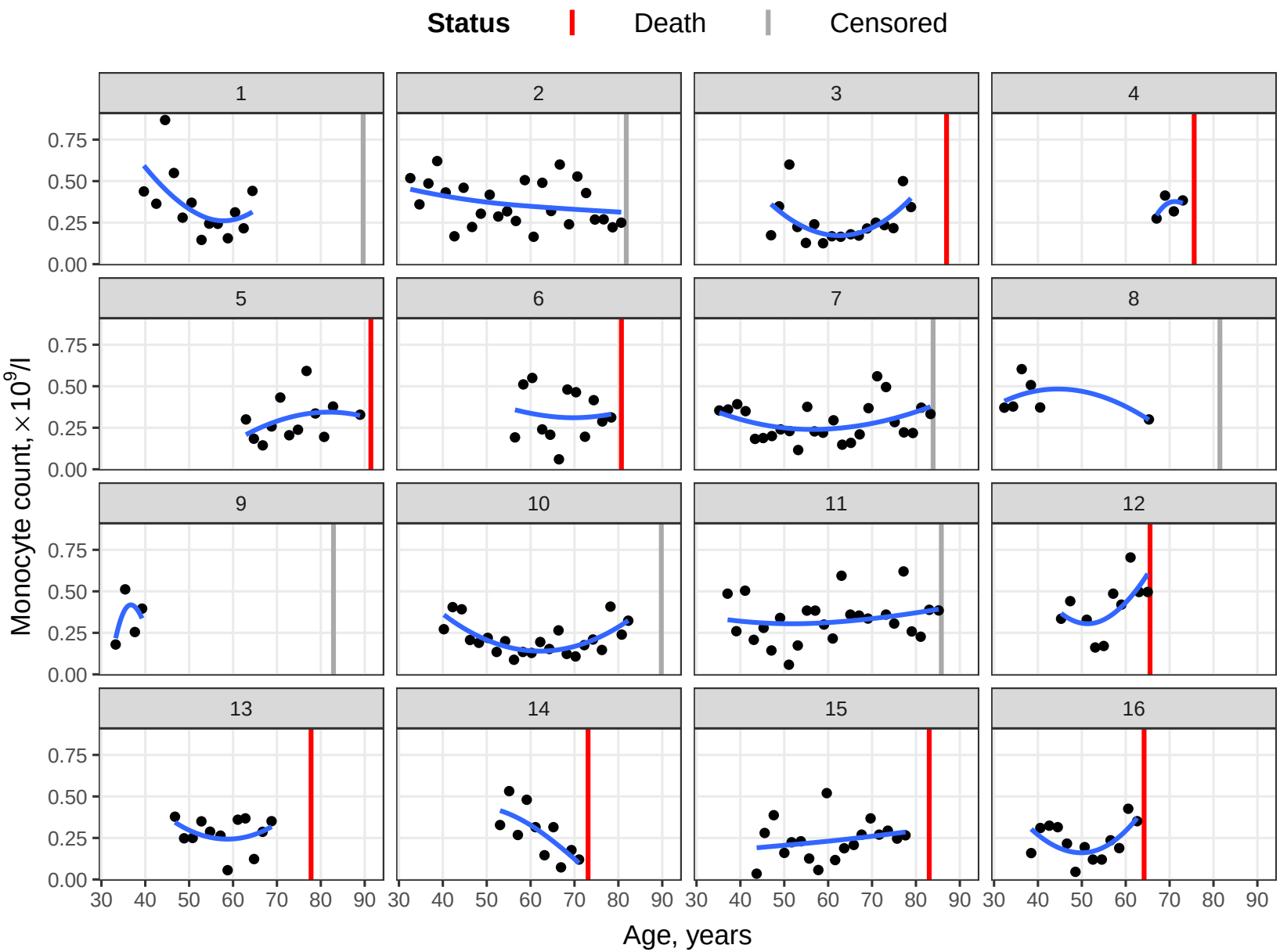
Joint regression models

Example 1

Longitudinal changes in peripheral monocytes (Yoshida et al., 2019)

- **Adult Health Study**
  - Subset of Life Span Study of atomic bomb survivors
  - Biennial clinic examinations since 1958
  - Detailed questionnaire and laboratory data
- DS02R1 radiation doses estimated from dosimetry system
- **Outcomes of interest**
  - Monocyte count (longitudinal) as a measure of inflammation
  - Time to death due to any cause (1958–2010)
- **Research questions**
  - What is the association between radiation and monocyte counts?
  - What is the association between monocyte counts and mortality?
  - Others?

AHS data

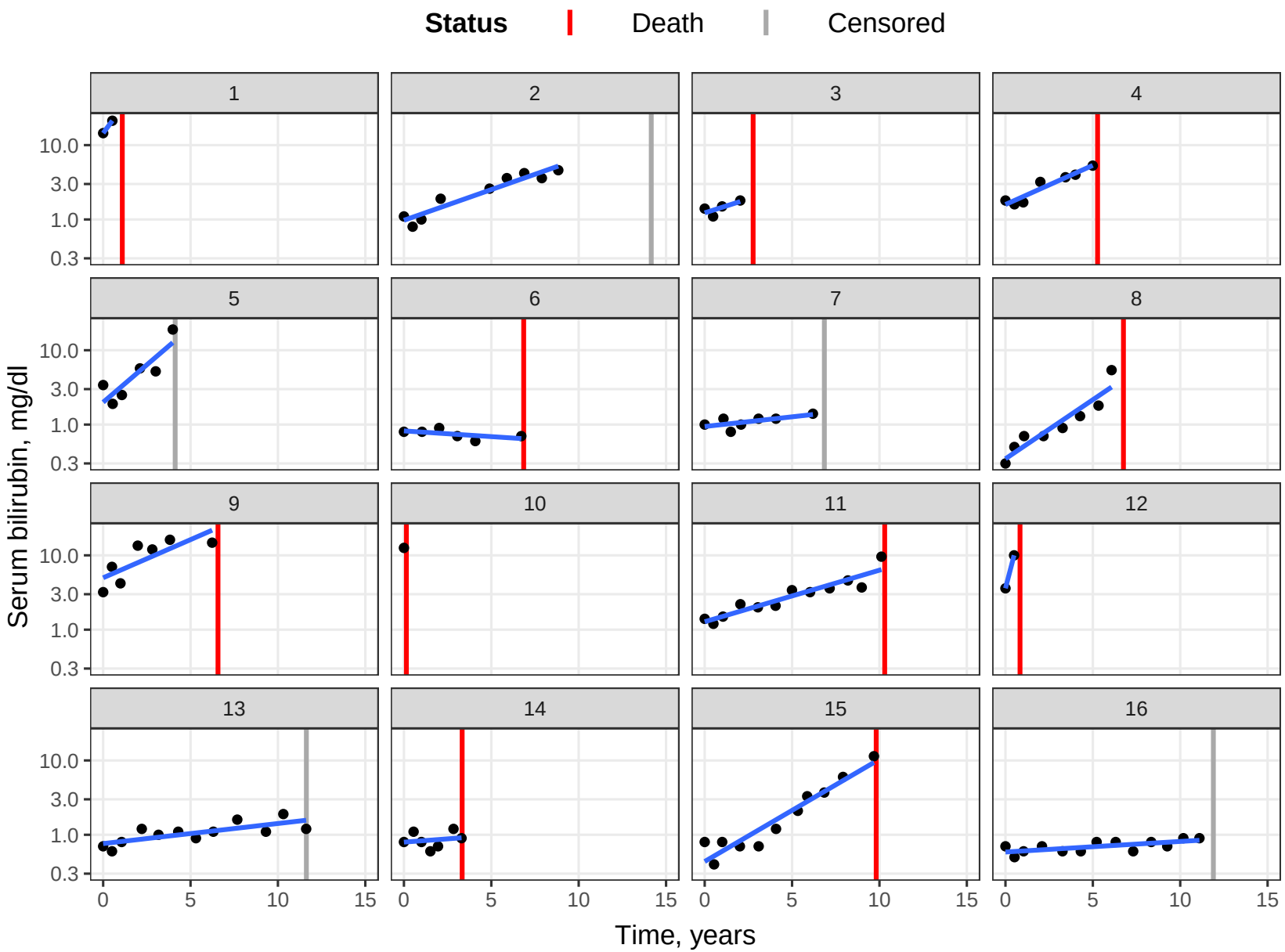

Status
Death
Censored
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
0.00
0.25
0.50
0.75
30
40
50
60
70
80
90
Monocyte count, $\times 10^9$/l
Age, years

Example 2

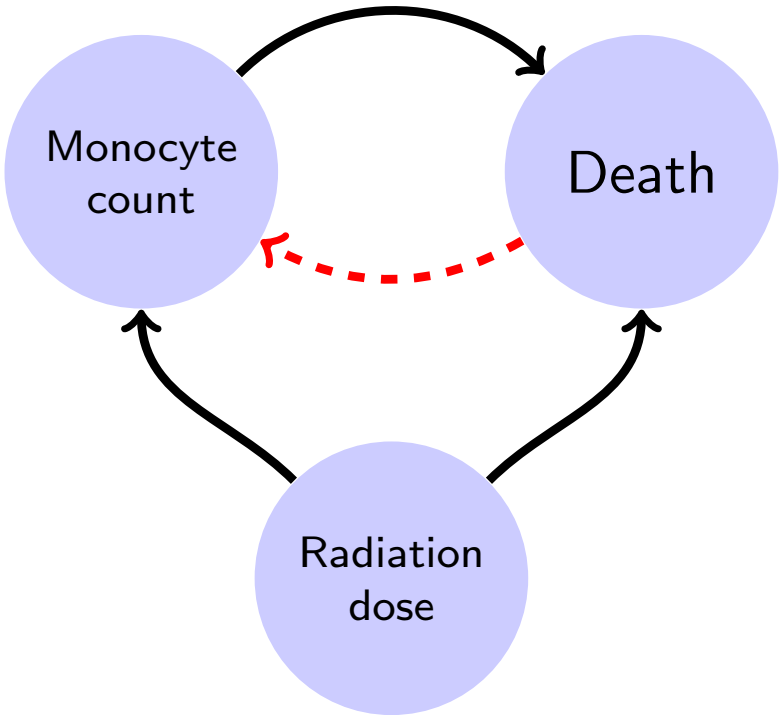
Mayo Clinic trial in primary biliary cirrhosis (Murtaugh et al., 1994)

- **Primary biliary cirrhosis**
  - Chronic and fatal but rare liver disease
  - Inflammatory destruction of small bile ducts within the liver
  - Patients referred to Mayo Clinic, 1974–1984
- 158 patients randomized to treatment with D-penicillamine; 154 randomized to placebo
- **Outcomes of interest**
  - Serum bilirubin levels (longitudinal) as a measure of liver function
  - Time to death and/or time to liver transplantation
- **Research questions**
  - What is the association between treatment and serum bilirubin levels?
  - What is the association between serum bilirubin levels and mortality?
  - Others?

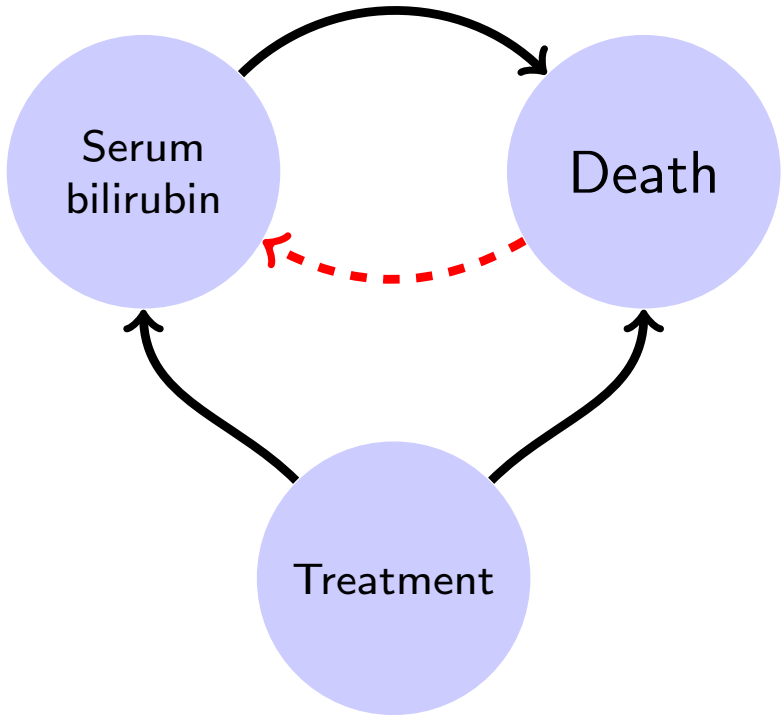
PBC data


Status
Death
Censored
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
10.0
3.0
1.0
0.3
0
5
10
15
Serum bilirubin, mg/dl
Time, years

Analysis framework



(a) AHS data



(b) PBC data

- Treatment can be associated with both bilirubin and mortality
- Bilirubin levels (and trends) can be associated with mortality
- Occurrence of death precludes observation of bilirubin levels

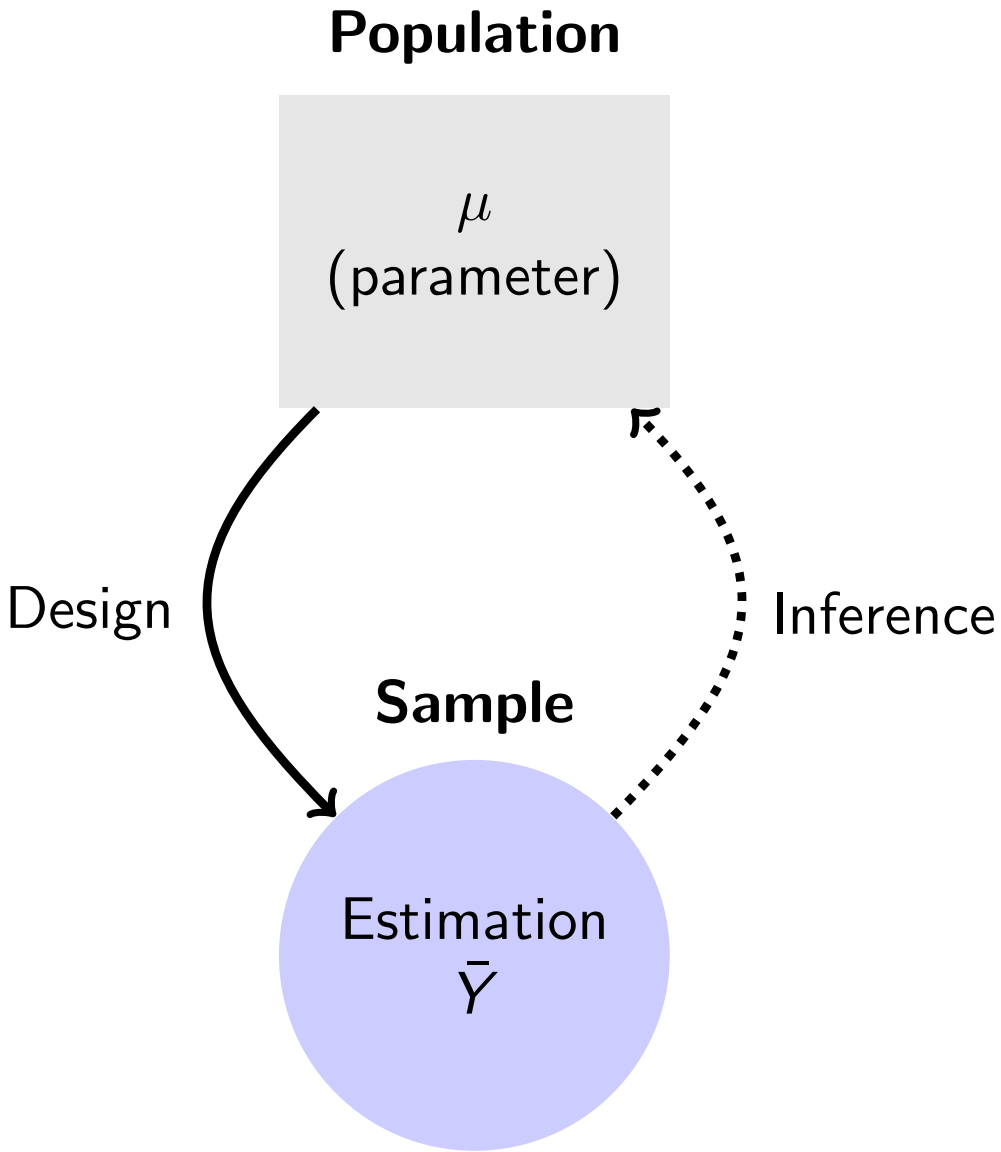
Analysis choices

Analysis methods should be selected based on the scientific question

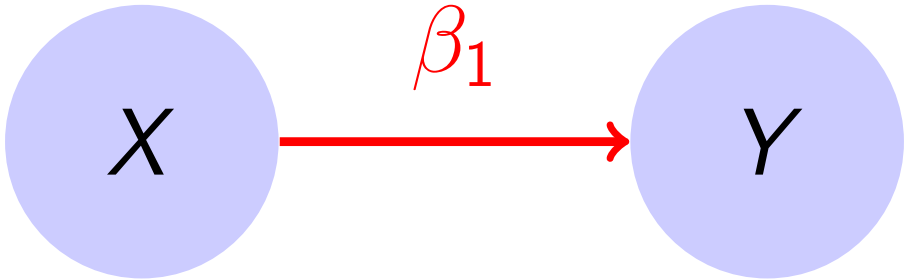
- Focus on the association between treatment and bilirubin levels
  - Standard longitudinal data analysis (separate analysis)
  - Formulate a regression model for repeated measures of bilirubin level
  - Ignores the impact of death on our ability to collect data $(-)$
- Focus on the association between bilirubin levels and mortality risk
  - Standard survival data analysis (separate analysis)
  - Formulate a Cox regression model for time to death or transplantation
  - Potentially adjust for risk differences due to treatment; mediation
  - Treats bilirubin levels as fixed, but these are measured with error $(-)$

$(-)$s motivate application of joint regression models (joint analysis)

Statistics


Population
μ
(parameter)
Design
Inference
Sample
Estimation
$\bar{Y}$

Regression



$$\mathrm{E}[Y \mid X = x] = \beta_0 + \beta_1 x$$

Estimation

- Coefficient estimate $\hat{\beta}_1$
- Standard error for $\hat{\beta}_1$

Inference

- Confidence interval for β_1
- Hypothesis test for $\beta_1 = 0$

Overview

Motivation and examples

Longitudinal data analysis

Survival data analysis

Joint regression models

Analysis approaches

Must account for correlation due to repeated measurements over time

- Failure to account for correlation $\Rightarrow$ incorrect standard estimates, resulting in incorrect confidence intervals and hypothesis tests
- **Approaches**: Include all observed data in a regression model for the mean response and account for longitudinal correlation
  - **Generalized estimating equations** (GEE): A marginal model for the mean response and a model for longitudinal correlation

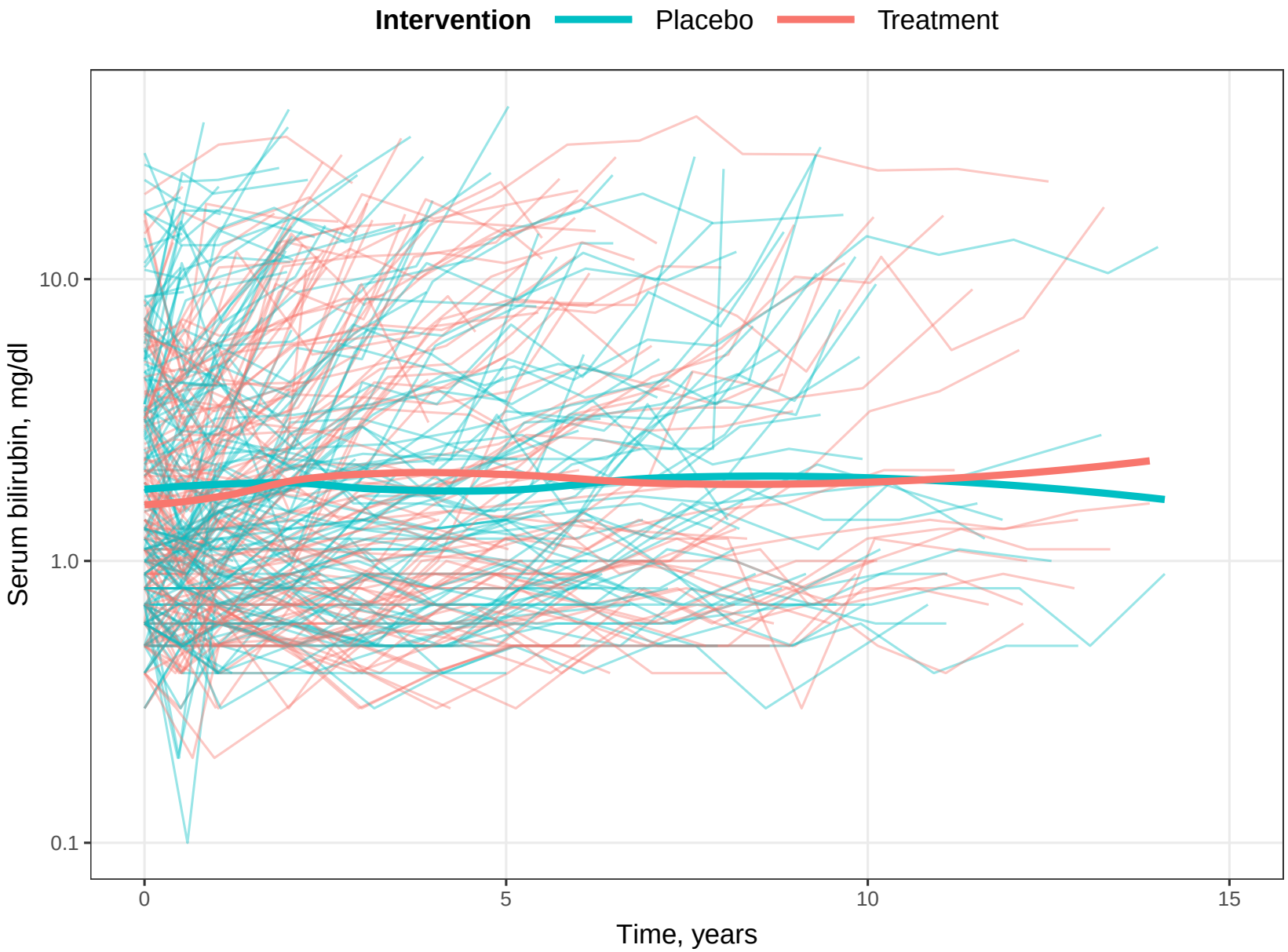
$$g(\mathrm{E}[Y_{ij} \mid x_{ij}]) = x_{ij}\beta \quad \text{and} \quad \mathrm{Corr}[Y_{ij}, Y_{ij'}] = \rho(\alpha),\ j \neq j'$$

  - **Generalized linear mixed-effects models** (GLMM): A conditional model for the mean response given subject-specific random effects, which induce a (possibly hierarchical) correlation structure

$$g(\mathrm{E}[Y_{ij} \mid x_{ij},\ b_i]) = x_{ij}\beta + z_{ij}b_i \quad \text{with} \quad b_i \sim N(0, D)$$

NB: Differences in interpretation of β between GEE and GLMM

Repeated measures



Mixed-effects models

⋆ Contrast outcomes both within and between individuals

- Assume that each subject has a regression model characterized by subject-specific parameters; a combination of
  - Fixed-effects parameters common to all individuals in the population
  - Random-effects parameters unique to each individual subject
- Although covariates allow for differences across subjects, typically cannot measure all factors that give rise to subject-specific variation
- Subject-specific random effects induce a correlation structure

(Laird and Ware, 1982)

Set-up

For subject i the mixed-effects model is characterized by

$$
\begin{aligned}
y_i &= \{y_{i1}, y_{i2}, \ldots, y_{im_i}\}^{\mathsf{T}} \\
\\
\beta &= \{\beta_0, \beta_1, \beta_2, \ldots, \beta_p\}^{\mathsf{T}} && \text{Fixed effects} \\
x_{ij} &= \{1, x_{ij1}, x_{ij2}, \ldots, x_{ijp}\} \\
X_i &= \{x_{i1}, x_{i2}, \ldots, x_{im_i}\}^{\mathsf{T}} && \text{Design matrix for fixed effects} \\
\\
b_i &= \{b_{i0}, b_{i1}, b_{i2}, \ldots, b_{iq}\}^{\mathsf{T}} && \text{Random effects} \\
z_{ij} &= \{1, z_{ij1}, z_{ij2}, \ldots, z_{ijq}\} \\
Z_i &= \{z_{i1}, z_{i2}, \ldots, z_{im_i}\}^{\mathsf{T}} && \text{Design matrix for random effects}
\end{aligned}
$$

for $i = 1, \ldots, n$; $j = 1, \ldots, m_i$; and $q \leq p$

Linear mixed-effects model

Consider a linear mixed-effects model for a continuous outcome y_{ij}

1. Model for response given random effects

$$y_{ij} = x_{ij}\beta + z_{ij}b_i + \epsilon_{ij}$$

with

- x_{ij}: vector a covariates
- β: vector of fixed-effects parameters
- z_{ij}: subset of x_{ij}
- b_i: vector of random-effects parameters
- ϵ_{ij}: observation-specific measurement error

2. Model for random effects

$$\begin{aligned} b_i &\sim N(0, D) \\ \epsilon_{ij} &\sim N(0, \sigma^2) \end{aligned}$$

with b_i and ϵ_{ij} assumed to be independent

Choices for random effects

Consider the linear mixed-effects models that include

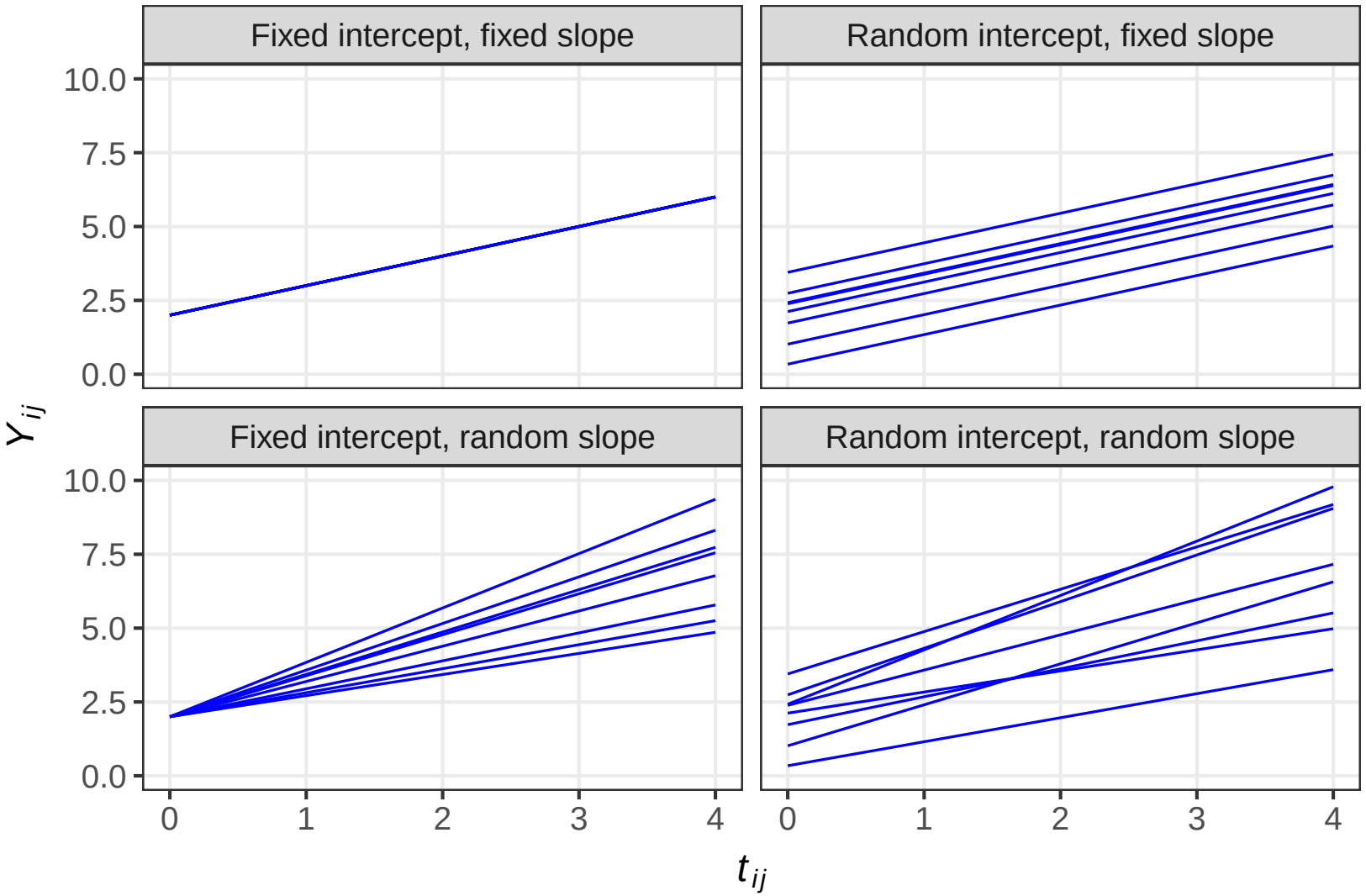
- **Random intercepts**

$$\begin{aligned} y_{ij} &= \beta_0 + \beta_1 t_{ij} + b_{i0} + \epsilon_{ij} \\ &= (\beta_0 + b_{i0}) + \beta_1 t_{ij} + \epsilon_{ij} \end{aligned}$$

- **Random intercepts and slopes**

$$\begin{aligned} y_{ij} &= \beta_0 + \beta_1 t_{ij} + b_{i0} + b_{i1} t_{ij} + \epsilon_{ij} \\ &= (\beta_0 + b_{i0}) + (\beta_1 + b_{i1}) t_{ij} + \epsilon_{ij} \end{aligned}$$

Choices for random effects



Choices for random effects: D

D quantifies random variation in trajectories across subjects

$$D = \begin{bmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{bmatrix}$$

- $\sqrt{D_{11}}$ is the typical deviation in the level of the response
- $\sqrt{D_{22}}$ is the typical deviation in the change in the response
- D_{12} is the covariance between subject-specific intercepts and slopes
  - $D_{12} = 0$ indicates subject-specific intercepts and slopes are uncorrelated
  - $D_{12} > 0$ indicates subjects with high level have high rate of change
  - $D_{12} < 0$ indicates subjects with high level have low rate of change

$(D_{12} = D_{21})$

Induced correlation structure

What is the correlation between measurements on the same subject?

- **Random intercepts model**
  - Assuming $\text{Var}[\epsilon_{ij}] = \sigma^2$ and $\text{Cov}[\epsilon_{ij}, \epsilon_{ij'}] = 0$

$$\begin{aligned} y_{ij} &= \beta_0 + \beta_1 t_{ij} + b_{i0} + \epsilon_{ij} \\ y_{ij'} &= \beta_0 + \beta_1 t_{ij'} + b_{i0} + \epsilon_{ij'} \end{aligned}$$

$$\begin{aligned} \text{Var}[Y_{ij}] &= \text{Var}_b[\text{E}_Y(Y_{ij} \mid b_{i0})] + \text{E}_b[\text{Var}_Y(Y_{ij} \mid b_{i0})] \\ &= D_{11} + \sigma^2 \end{aligned}$$

$$\begin{aligned} \text{Cov}[Y_{ij},\ Y_{ij'}] &= \text{Cov}_b[\text{E}_Y(Y_{ij} \mid b_{i0}),\ \text{E}_Y(Y_{ij'} \mid b_{i0})] \\ &\quad + \text{E}_b[\text{Cov}_Y(Y_{ij},\ Y_{ij'} \mid b_{i0})] \\ &= D_{11} \end{aligned}$$

Induced correlation structure

- **Random intercepts model** (continued)

$$\begin{aligned}
\text{Corr}[Y_{ij}, Y_{ij'}] &= \frac{D_{11}}{\sqrt{D_{11}+\sigma^2}\sqrt{D_{11}+\sigma^2}} \\
&= \frac{D_{11}}{D_{11}+\sigma^2} \\
&= \frac{\text{`Between'}}{\text{`Between'} + \text{`Within'}} \\
&\geq 0 \text{ (and } \leq 1)
\end{aligned}$$

  - Any two measurements on the same subject have the same correlation; does not depend on time nor the distance between measurements
  - Longitudinal correlation is constrained to be positive ($D_{11} \geq 0$, $\sigma^2 \geq 0$)

Induced correlation structure

- **Random intercepts and slopes model**
  - Assuming $\text{Var}[\epsilon_{ij}] = \sigma^2$ and $\text{Cov}[\epsilon_{ij}, \epsilon_{ij'}] = 0$

$$
\begin{aligned}
y_{ij} &= (\beta_0 + \beta_1 t_{ij}) + (b_{i0} + b_{i1} t_{ij}) + \epsilon_{ij} \\
y_{ij'} &= (\beta_0 + \beta_1 t_{ij'}) + (b_{i0} + b_{i1} t_{ij'}) + \epsilon_{ij'}
\end{aligned}
$$

$$
\begin{aligned}
\text{Var}[Y_{ij}] &= \text{Var}_b[\text{E}_Y(Y_{ij} \mid b_i)] + \text{E}_b[\text{Var}_Y(Y_{ij} \mid b_i)] \\
&= D_{11} + 2D_{12} t_{ij} + D_{22} t_{ij}^2 + \sigma^2
\end{aligned}
$$

$$
\begin{aligned}
\text{Cov}[Y_{ij}, Y_{ij'}] &= \text{Cov}_b[\text{E}_Y(Y_{ij} \mid b_i), \text{E}_Y(Y_{ij'} \mid b_i)] \\
&\quad + \text{E}_b[\text{Cov}_Y(Y_{ij}, Y_{ij'} \mid b_i)] \\
&= D_{11} + D_{12}(t_{ij} + t_{ij'}) + D_{22} t_{ij} t_{ij'}
\end{aligned}
$$

Induced correlation structure

- **Random intercepts and slopes model** (continued)

$$\text{Corr}[Y_{ij}, Y_{ij'}]$$
$$= \frac{D_{11} + D_{12}(t_{ij} + t_{ij'}) + D_{22} t_{ij} t_{ij'}}{\sqrt{D_{11} + 2D_{12} t_{ij} + D_{22} t_{ij}^2 + \sigma^2}\sqrt{D_{11} + 2D_{12} t_{ij'} + D_{22} t_{ij'}^2 + \sigma^2}}$$

  - Any two measurements on the same subject may not have the same correlation; depends on the specific observation times

Likelihood-based estimation of β

Requires specification of a complete probability distribution for the data

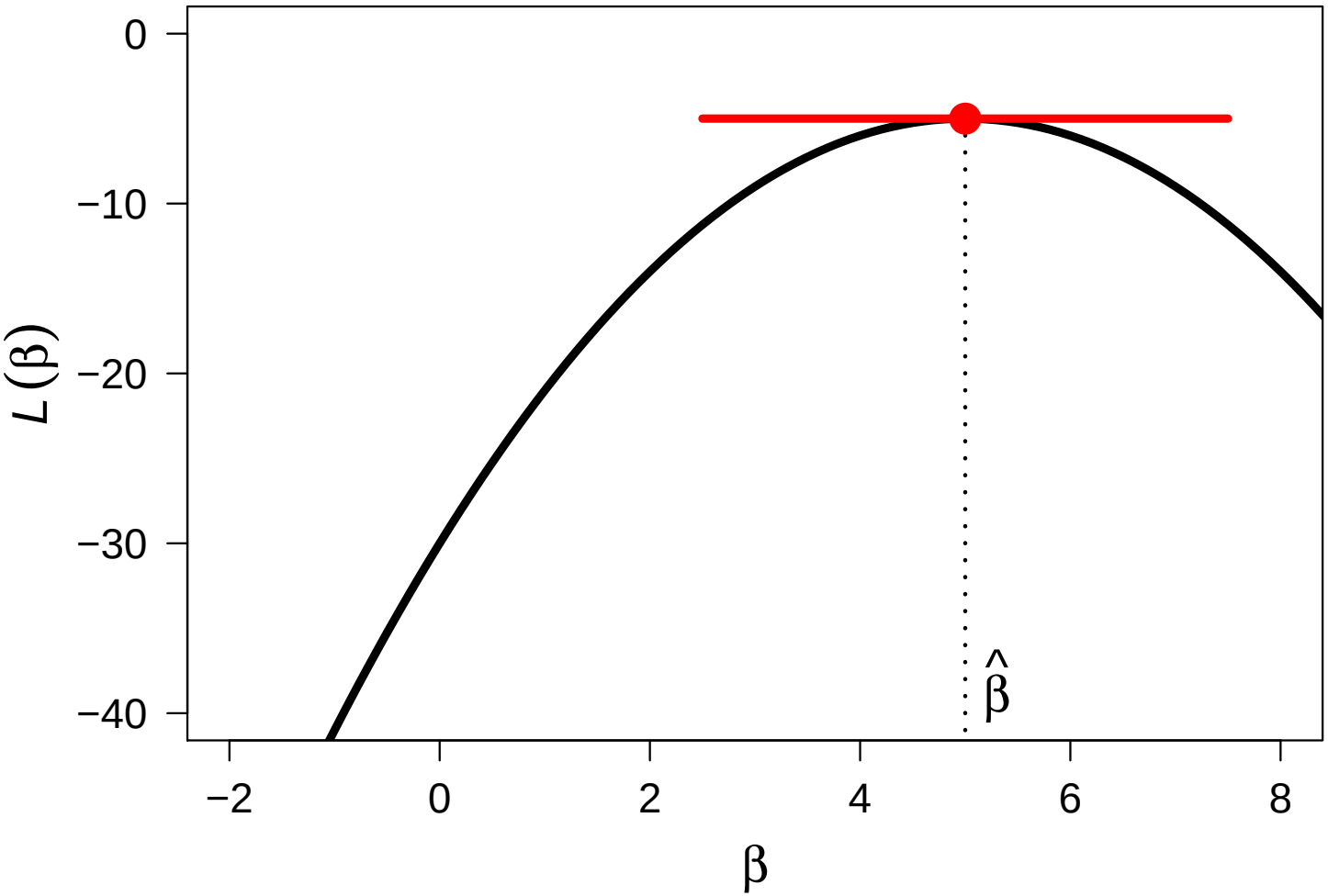
- Likelihood-based methods are designed for fixed effects, so integrate over the assumed distribution for the random effects

$$\mathcal{L}(\beta, \sigma, D) = \prod_{i=1}^{n} \int f_Y(y_i \mid b_i, \beta, \sigma) \times f_b(b_i \mid D) db_i$$

where f_b is typically the density function of a Normal random variable

  - For linear models the required integration is straightforward because y_i and b_i are both normally distributed (easy to program)
  - For non-linear models the integration is difficult and requires either approximation or numerical techniques (hard to program)

Likelihood-based estimation of β



Likelihood-based inference for β

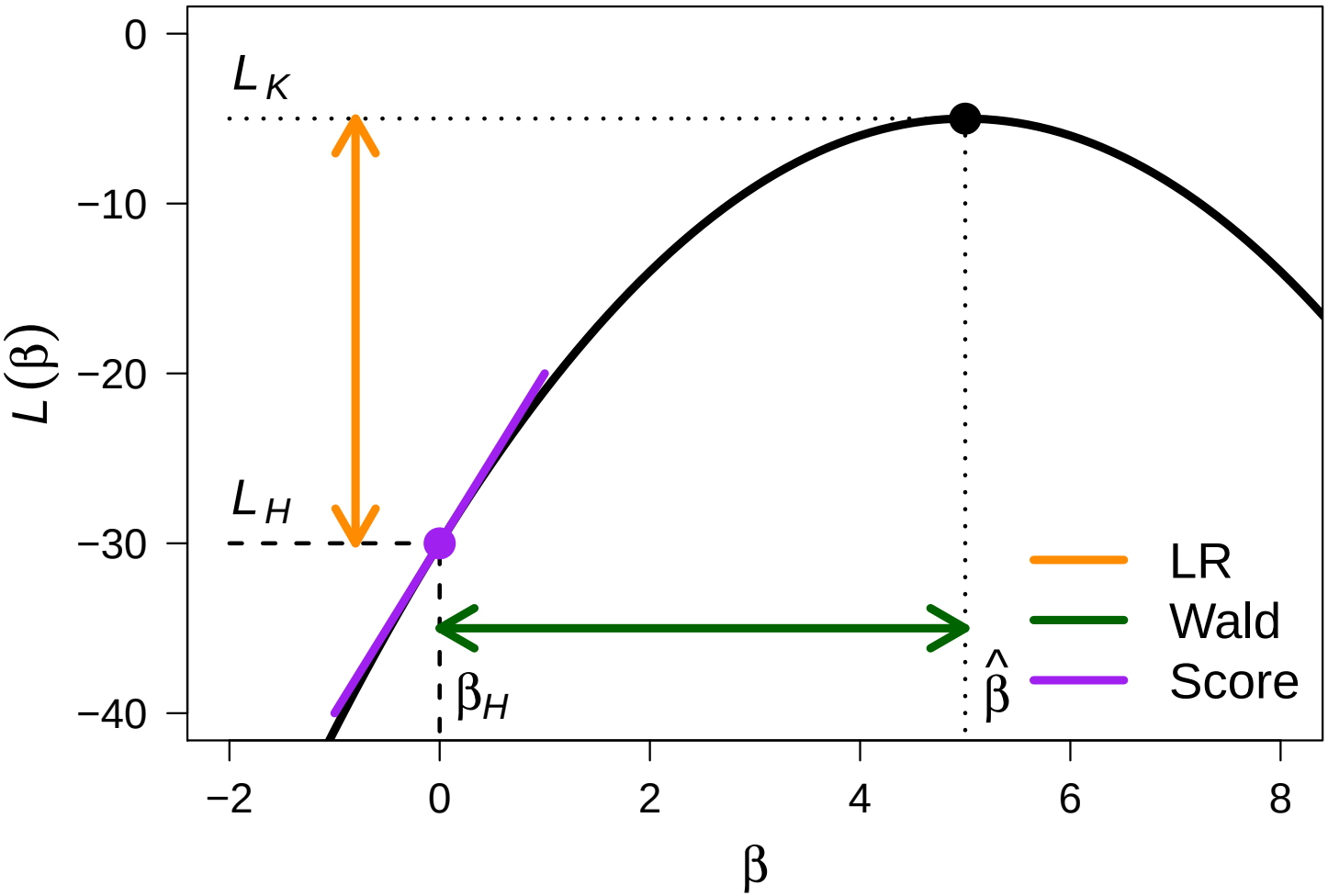
Consider testing fixed effects in nested linear mixed-effects models

$$H\colon\ \beta = \begin{bmatrix} \beta_1 \\ 0 \end{bmatrix} \quad \text{versus} \quad K\colon\ \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix},$$

i.e., $H\colon\ \beta_2 = 0$

- Likelihood ratio test is valid with maximum likelihood estimation
- Likelihood ratio test may not be valid with other estimation methods
- Wald test (based on coefficient and standard error) is generally valid

Likelihood-based inference for β



Effect modification

- Association of interest varies across levels of another variable, or another variable modifies the association of the variable of interest
- Modeling of effect modification is achieved by interaction terms

$$\mathrm{E}[Y \mid x, t] = \beta_0 + \beta_1 x + \beta_2 t + \beta_3 x \times t$$

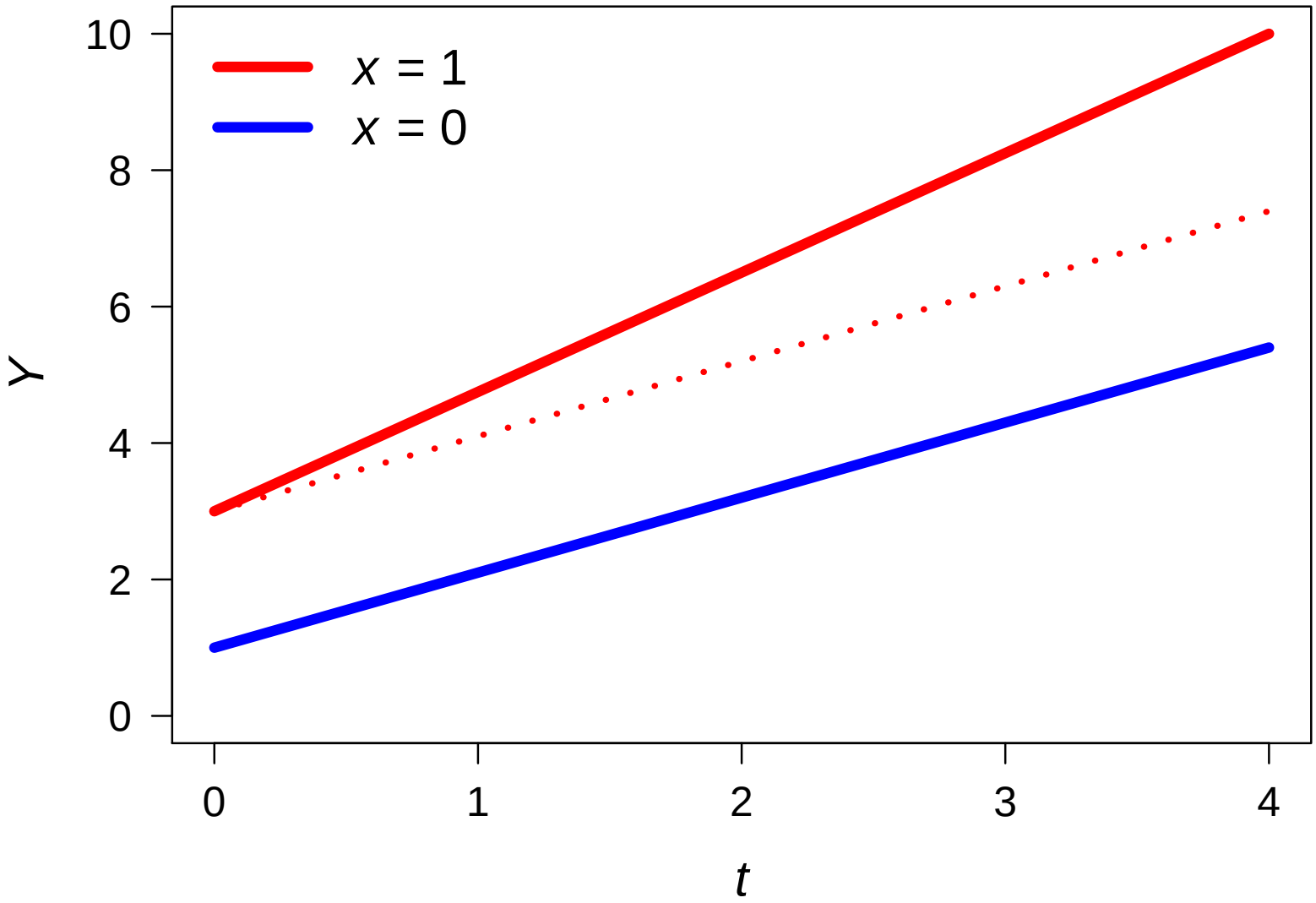
with

  - A binary variable x for drug: 0 for placebo, 1 for treatment
  - A continuous variable t for time since randomization

- Wish to examine whether treatment modifies the association between time since randomization and serum bilirubin

$$\begin{aligned} \text{Placebo: } \mathrm{E}[Y \mid x = 0, t] &= \beta_0 + \beta_2 t \\ \text{Treatment: } \mathrm{E}[Y \mid x = 1, t] &= \beta_0 + \beta_1 + \beta_2 t + \beta_3 t \\ &= (\beta_0 + \beta_1) + (\beta_2 + \beta_3) t \end{aligned}$$

Effect modification



Effect modification

- Contrasts for t (time) depend on the value for x (drug)

$$\begin{aligned} & \mathrm{E}[Y \mid x, t+1] - \mathrm{E}[Y \mid x, t] \\ = \; & \{\beta_0 + \beta_1 \cdot x + \beta_2 \cdot (t+1) + \beta_3 \cdot x \cdot (t+1)\} \\ & - \{\beta_0 + \beta_1 \cdot x + \beta_2 \cdot t + \beta_3 \cdot x \cdot t\} \\ = \; & \beta_2 + \beta_3 x \end{aligned}$$

- β_2 compares the mean bilirubin level between two placebo-treated populations whose time since randomization differs by 1 year ($x = 0$)
- $\beta_2 + \beta_3$ compares the mean bilirubin level between two drug-treated populations whose time since randomization differs by 1 year ($x = 1$)
- Hence β_3 represents a difference evaluating whether the association between time and serum bilirubin differs between treatment groups
- A hypothesis test of $\beta_3 = 0$ can be used to evaluate the difference

Likelihood-based inference for D

Consider testing whether a random intercept model is adequate

$$H\colon\ D = \begin{bmatrix} D_{11} & 0 \\ 0 & 0 \end{bmatrix} \quad \text{versus} \quad K\colon\ D = \begin{bmatrix} D_{11} & \\ D_{12} & D_{22} \end{bmatrix},$$

i.e., $H\colon\ D_{12} = D_{22} = 0$

- Adequate covariance modeling is useful for the interpretation of the random variation in the data
- Over-parameterization of the covariance structure leads to inefficient estimation of fixed-effects parameters β
- Covariance model choice determines the standard error estimates for $\hat{\beta}$; correct model is required for correct standard error estimates
- Generally recommend against this inferential procedure
  - Specification for the covariance structure should be guided by *a priori* scientific knowledge and exploratory data analysis

Assumptions

Valid inference from a linear mixed-effects model relies on

- **Mean model**: As with any regression model for an average outcome, need to correctly specify the functional form of $x_{ij}\beta + z_{ij}b_i$
  - Included important covariates in the model
  - Correctly specified any transformations or interactions
- **Covariance model**: Correct covariance model (random-effects specification) is required for correct standard error estimates for $\hat{\beta}$
- **Normality**: Normality of ϵ_{ij} and b_i is required for normal likelihood function to be the correct likelihood function for y_{ij}
- n sufficiently large for **asymptotic inference** to be valid

$\star$ These assumptions must be verified to evaluate any fitted model

Missing data

- Missing values arise in longitudinal studies whenever the intended serial observations collected on a subject over time are incomplete
  - Collect fewer data than planned $\Rightarrow$ decreased efficiency (power)
  - Missingness can depend on outcome values $\Rightarrow$ potential bias
- Important to distinguish between missing data and unbalanced data, although missing data necessarily result in unbalanced data
- Missing data require consideration of the factors that influence the missingness of intended observations
- Also important to distinguish between intermittent missing values (non-monotone) and dropouts in which all observations are missing after subjects are lost to follow-up (monotone)

Pattern	t_1	t_2	t_3	t_4	t_5
Monotone	3.8	3.1	2.0		
Non-monotone	4.1		3.8		

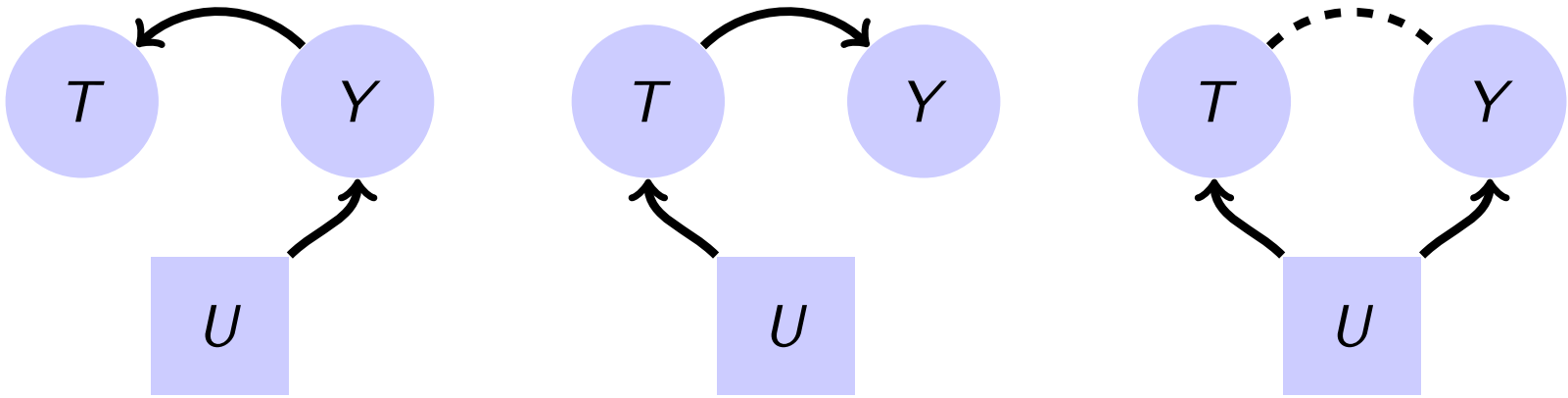
Mechanisms

Partition the complete set of intended observations into the observed and missing data; what factors influence missingness of intended observations?

- **Missing completely at random** (MCAR)
  Missingness does not depend on **either** the observed or missing data
  - Example: Administrative censoring at a fixed calendar time
  - Mixed-effects models are valid
- **Missing at random** (MAR)
  Missingness depends **only** on the observed data
  - Example: Individuals with no current weight loss in a weight-loss study
  - Mixed-effects models are valid (with additional assumptions)
- **Missing not at random** (MNAR)
  Missingness depends on **both** the observed and missing data
  - Example: Subjects in a prospective study based on disease prognosis
  - Mixed-effects models are not valid

(Rubin, 1976; Ibrahim and Molenberghs, 2009)

Models under MNAR



(a) Selection (b) Pattern-mixture (c) Shared-parameter

(a) Subject-specific random effects or latent characteristics influence the outcome, which subsequently determines the propensity to drop out

(b) Subject-specific characteristics initially determine propensity to drop out, with consequential variation in the outcome between drop-out cohorts

(c) Outcome and drop-out processes jointly respond to unobserved subject-specific characteristics, observation of which would convert the model into one in which the outcome and drop-out time are independent

⋆ Joint regression models and shared-parameter models are analogous

Summary

- Mixed-effects models assume that each subject has a regression model characterized by subject-specific parameters; a combination of
  - Fixed-effects parameters common to all individuals in the population
  - Random-effects parameters unique to each individual subject
- Estimation and inference can focus both on average outcome levels and trends, and on heterogeneity across subjects in levels and trends
- Subject-specific random effects induce a correlation structure
- Must consider the potential for bias due to missing data (drop-out)
  - Ignores the impact of death on our ability to collect data $(-)$
- Further reading: Verbeke and Molenberghs (2000)

Overview

Motivation and examples

Longitudinal data analysis

Survival data analysis

Joint regression models

Censoring

Analysis methods must account for censoring of event-time outcomes

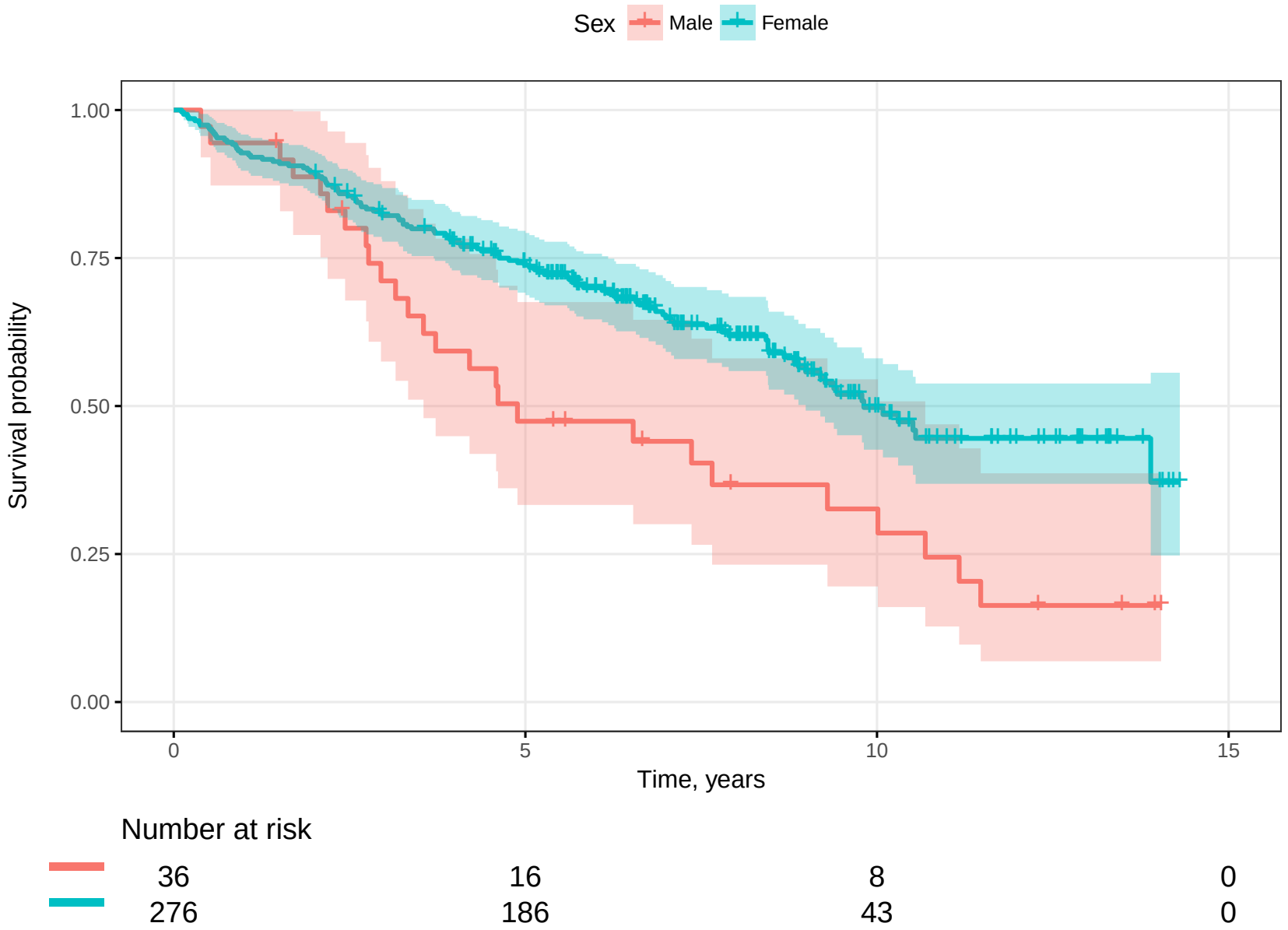
- **Types of censoring**
  - Location of true event time w.r.t. censoring time: right, left, interval
  - Probabilistic relationship between true event time and censoring time: informative, non-informative (similar to MNAR and MAR)
- **Implications**
  - Standard approaches (e.g., t test, linear regression) cannot be used
  - Inference can be sensitive to the assumed distribution of event times

⋆ **Our focus**: Non-informative right censoring

- Event of interest is not fully observed for all subjects
  - Do not experience the event before the end of the study period
  - Are (randomly) lost to follow-up during the study period

  $\Rightarrow$ event time (or, survival time) is not known for these subjects
- But, subjects contribute time at risk up until their censoring time

Censoring


Sex
Male
Female
1.00
0.75
0.50
0.25
0.00
Survival probability
0
5
10
15
Time, years
Number at risk
36
16
8
0
276
186
43
0

Set-up

- **Notation**
  - True event time T_i
  - Censoring time C_i
- **Observed data**
  - Observed event time $T_i^\star = \min(T_i, C_i)$
  - Event indicator $\delta_i = 1$ if event observed, $\delta_i = 0$ if censored
- **Objective**: Generate inference for T_i using $\{T_i^\star, \delta_i\}$

Hazard functions

- **Probability density function** for the true event time T: $f_T(t)$
- **Cumulative distribution function**: Probability of failure by time t

$$F_T(t) = \mathrm{P}[T \leq t] = \int_0^t f_T(s)ds$$

- **Survival function**: Probability of survival beyond time t

$$S_T(t) = \mathrm{P}[T > t] = 1 - F_T(t) = \int_t^\infty f_T(s)ds$$

- **Hazard function**: Probability of failure in an very small time period between time t and $t + \Delta t$ given survival up until time t

$$h_T(t) = \frac{f_T(t)}{S_T(t)} = \lim_{\Delta t \to 0} \frac{\mathrm{P}[t \leq T < t + \Delta t \mid T \geq t]}{\Delta t} = -\frac{\partial}{\partial t} \log S_T(t)$$

as a measure of risk; larger hazard $\Rightarrow$ larger risk of failure

Hazard models

- **Relative risk model**: Multiplicative effect of covariates on hazard

$$\begin{aligned} h_i(t \mid w_i) &= h_0(t) \exp\{w_i \gamma\} \\ \log h_i(t \mid w_i) &= \log h_0(t) + \gamma_1 w_{i1} + \gamma_2 w_{i2} + \cdots + \gamma_p w_{ip} \end{aligned}$$

with

  - $h_i(t \mid w_i)$: hazard of an event for subject i at time t
  - $h_0(t)$: common baseline hazard function
  - $w_i = \{w_{i1}, w_{i2}, \ldots, w_{ip}\}$: time-independent covariates for subject i
  - $\gamma = \{\gamma_1, \gamma_2, \ldots, \gamma_p\}^{\mathsf{T}}$: regression parameters of interest

- **Excess relative risk model**: Additive effect of covariates on hazard

$$\begin{aligned} h_i(t \mid w_i) &= h_0(t)\{1 + w_i \gamma\} \\ \log h_i(t \mid w_i) &= \log h_0(t) + \log\{1 + \gamma_1 w_{i1} + \gamma_2 w_{i2} + \cdots + \gamma_p w_{ip}\} \end{aligned}$$

which is a standard model for estimating radiation effects

Hazard models

- **Relative risk model**: Multiplicative effect of covariates on hazard

$$h(t \mid w_1) = h_0(t) \exp\{\gamma_1 w_1\}$$

$$\begin{aligned} h(t \mid w_1 = w) &= h_0(t) \exp\{\gamma_1 \times w\} \\ h(t \mid w_1 = w+1) &= h_0(t) \exp\{\gamma_1 \times (w+1)\} \end{aligned}$$

$$\frac{h(t \mid w_1 = w+1)}{h(t \mid w_1 = w)} = \exp(\gamma_1)$$

so that $\exp(\gamma_1)$ is a hazard ratio quantifying the impact of a one-unit increase in w_1 on the hazard of an event

  - Does not depend on time t

Estimation and inference

Cox regression model: Unspecified baseline hazard function

- Semi-parametric: No assumption for distribution of event times
- Assumes proportional hazards across covariate levels
- Estimates and standard errors obtained from (log) partial likelihood

$$\ell(\gamma) = \sum_{i=1}^{n} \delta_i \left[w_i\gamma - \log \left\{ \sum_{T_{i'} \geq T_i} \exp(w_{i'}\gamma) \right\} \right]$$

which is a measure of how well the model orders (ranks) the subjects w.r.t. their survival time

  - 'Partial' likelihood because $h_0(t)$ is not involved in estimation

(Cox, 1972)

Time-dependent covariates

- Often interested in the association of a time-dependent covariate
  - AHS: Association between monocyte counts and mortality
  - PBC: Association between serum bilirubin levels and mortality
- Standard Cox model is appropriate for time-independent covariates
  - AHS: City, sex, birth cohort, radiation dose
  - PBC: Treatment, sex, baseline age, baseline serum bilirubin
- Cox model can be extended for a certain type of covariate
  - External or exogenous time-dependent covariate $(+)$
  - Internal or endogenous time-dependent covariate $(-)$

Time-dependent covariates

- **Exogenous**: Future path of the covariate up until any time $t > t'$ is not affected by the occurrence of an event at time t'

$$\mathrm{P}[\mathcal{Y}_i(t) \mid \mathcal{Y}_i(t'), T_i \geq t'] = \mathrm{P}[\mathcal{Y}_i(t) \mid \mathcal{Y}_i(t'), T_i = t']$$

  with $0 < t' \leq t$ and $\mathcal{Y}_i(t) = \{y_i(s), 0 \leq s < t\}$ denotes the history
  - Allows the covariate to be associated with the failure rate
  - But its future values are the same whether a failure occurs or not
  - Examples: Season of the year, treatment regimen, air pollution level
- **Endogenous**: Not exogenous
  - Typically arise as time-dependent measurements on study subjects
  - Subject must survive in order for the covariate to exist
  - Examples: Monocyte count (AHS), serum bilirubin level (PBC)

Extended Cox model

- Model formulated using counting processes (Andersen and Gill, 1982)

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

with

  - $\{N_i(t), R_i(t)\}$: event process for subject i
  - $N_i(t)$: number of events for subject i by time t
  - $R_i(t)$: at-risk indicator for subject i at time t
  - $h_i(t)$: intensity process for $N_i(t)$
  - $h_0(t)$: common baseline intensity function
  - $w_i = \{w_{i1}, w_{i2}, \ldots, w_{ip}\}$: time-independent covariates for subject i
  - $y_i(t)$: time-dependent covariate for subject i at time t
  - $\alpha, \gamma = \{\gamma_1, \gamma_2, \ldots, \gamma_p\}^\mathsf{T}$: regression parameters of interest

- $\exp(\alpha)$ represents the relative increase in risk of an event at time t that results from a simultaneous one-unit increase in $y_i(t)$

Extended Cox model

- Estimates and standard errors obtained from (log) partial likelihood

$$\ell(\alpha, \gamma) = \sum_{i=1}^{n} \int_0^{\infty} \left[R_i(t) \exp\{w_i\gamma + \alpha y_i(t)\} - \log \sum_{i'} R_{i'}(t) \exp\{w_{i'}\gamma + \alpha y_{i'}(t)\} \right] dN_i(t)$$

- Valid estimation and inference requires several assumptions
  - Existence of covariate is not related to failure status (exogenous)
  - Covariate levels are fixed and known in between measurement times; French and Heagerty (2009): $y_i(t)$ exists only at measurement times
  - Covariate values are free of measurement error

⋆ Cox regression analysis of an endogenous time-dependent covariate can provide spurious results

Summary

- Analysis methods must account for censoring of event-time outcomes
  - Do not experience the event before the end of the study period
  - Are (randomly) lost to follow-up during the study period
- Focus on multiplicative effect of covariates on hazard (risk)
- Semi-parametric Cox regression model under proportional hazards
- Extended Cox model for exogenous time-dependent covariates
  - Existence of covariate is not related to failure status
  - Covariate levels are constant in between measurement times
  - Covariate values are free of measurement error
- Further reading
  - Kalbfleisch and Prentice (2002)
  - Cologne et al. (2012): Choice of primary time scale

Overview

Motivation and examples

Longitudinal data analysis

Survival data analysis

Joint regression models

Review

Mixed-effects models for longitudinal data analysis (separate analysis)

- Estimate population-level and subject-specific levels and trends
- Quantify heterogeneity across subjects in outcome levels and trends
- Subject-specific random effects induce a correlation structure
- Must consider the potential for bias due to missing data $(-)$

Cox regression models for survival data analysis (separate analysis)

- Focus on multiplicative effect of covariates on event hazard
- Account for non-informative right censoring of event-time outcomes
- Appropriate only for exogenous time-dependent covariates $(-)$
- Assume that covariate values are free of measurement error $(-)$

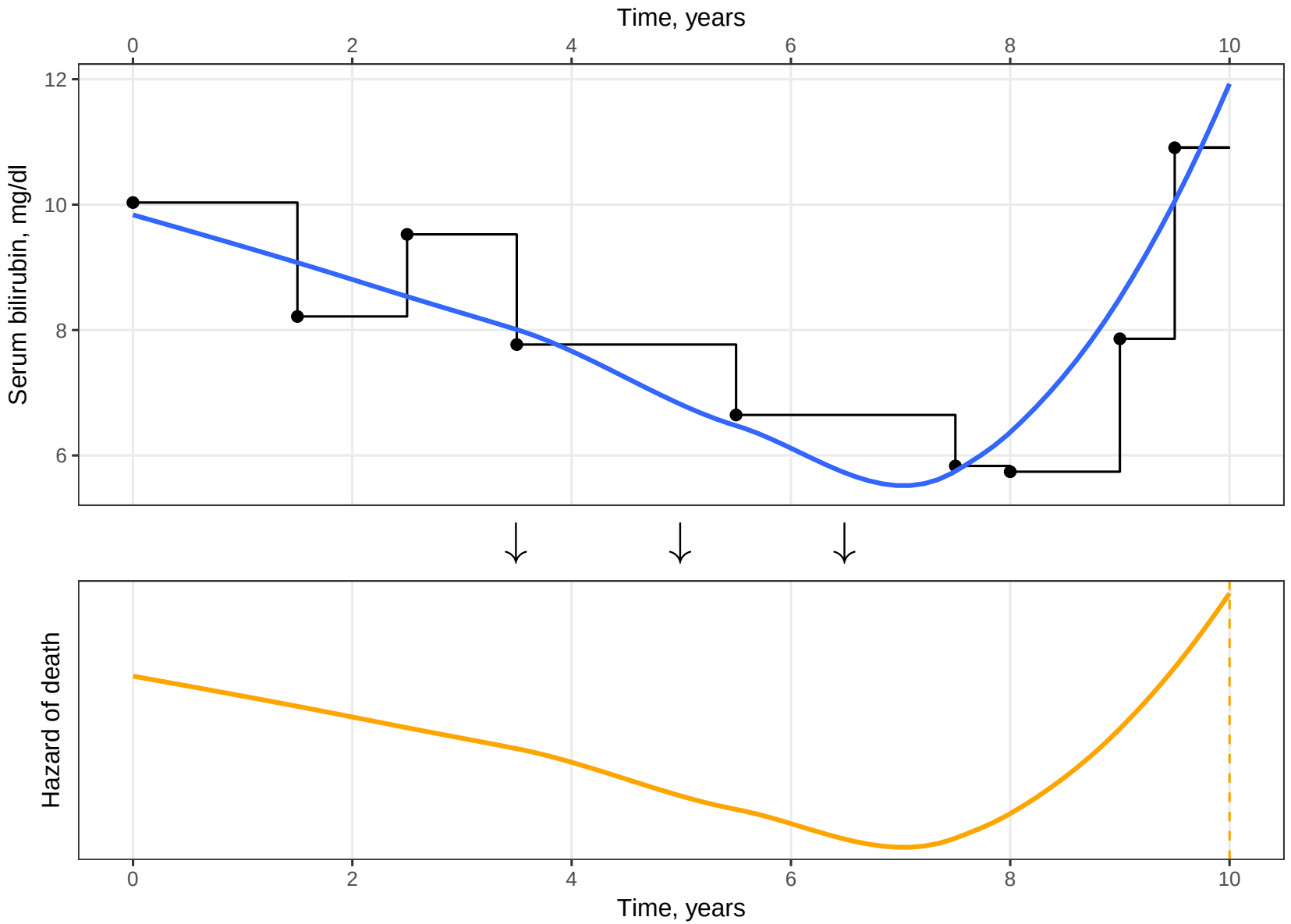
$\star$ Motivates development of joint regression models (joint analysis)

Joint models

$$\boxed{\text{Longitudinal model} + \text{Survival model} = \text{Joint model}}$$

1. Describe evolution of marker levels over time for each subject
   - Subject-specific random effects
   - Flexible adjustment for temporal trends
   - Incorporates random error term for measurement error
2. Associate the subject-specific evolutions with event hazard
   - Allow censoring of event-time outcome
   - Accommodate marker as an endogenous time-dependent covariate
   - Marker is not assumed constant in between measurement times
3. Specify a dependence structure between the two models

Joint models



Set-up

Step 1: Describe evolution of marker levels over time for each subject

$$\begin{aligned} y_i(t) &= x_i(t)\beta + z_i(t)b_i + \epsilon_i(t) \\ &= m_i(t) + \epsilon_i(t) \end{aligned}$$

using a standard linear mixed-effects model with

- True marker value $m_i(t) = x_i(t)\beta + z_i(t)b_i$
- Fixed-effects parameters β
- Random-effects parameters $b_i \sim N(0, D)$
- Measurement error $\epsilon_i(t) \sim N(0, \sigma^2)$

NB: Covariate vectors and error terms are functions of time t

Set-up

Step 2: Associate the subject-specific evolutions with event hazard

$$\begin{aligned} h_i(t \mid \mathcal{M}_i(t), w_i) &= \lim_{\Delta t \to 0} \mathrm{P}[t \leq T < t + \Delta t \mid T \geq t, \mathcal{M}_i(t), w_i]/\Delta t \\ &= h_0(t) \exp\{w_i \gamma + \alpha m_i(t)\} \end{aligned}$$

using a standard relative risk model with

- Marker history $\mathcal{M}_i(t) = \{m_i(s),\, 0 \leq s < t\}$
- Common baseline hazard function $h_0(t)$
- Parameters γ for time-independent covariates w_i
- $\exp(\alpha)$ represents the relative increase in risk of an event at time t that results from a simultaneous one-unit increase in $m_i(t)$

Set-up

Step 3: Specify a dependence structure between the two models

1. Define a joint distribution for markers y_i and event times $\{T_i^\star, \delta_i\}$

$$f(y_i, T_i^\star, \delta_i) = \int f_Y(y_i \mid b_i)\, h(T_i^\star \mid b_i)^{\delta_i}\, S(T_i^\star \mid b_i)\, f_b(b_i)\, db_i$$

with

- f_Y: density function for markers y_i
- f_b: density function for random effects b_i
- $h(t)$: hazard function for event times $T_i^\star$
- $S(t)$: survival function for event times $T_i^\star$
- δ_i: event indicator (1, event; 0, censored)
- b_i: random effects for inter-dependencies

(Tsiatis and Davidian, 2004)

Set-up

Step 3: Specify a dependence structure between the two models

2. Assume full conditional independence given random effects
   - Repeated measurements (longitudinal) are mutually independent

$$y_{ij} \perp\!\!\!\perp y_{ij'} \mid b_i, j \neq j'$$

     - ⋆ Random effects explain the correlation among repeated measures
   - Longitudinal outcome and event-time outcome are independent

$$y_i \perp\!\!\!\perp \{T_i^\star, \delta_i\} \mid b_i \, \forall \, i$$

     - ⋆ Random effects explain the association between the longitudinal and event-time outcomes (recall shared-parameter models)
   - ⋆ Random effects explain all the inter-dependencies in the data

NB: Conditional independence is difficult to evaluate using observed data

Assumptions

- Censoring and observation-time processes are non-informative; study withdrawal or appearance at study visits for data collection
  - Can depend on observed history (covariates and previous responses)
  - Cannot depend on unobserved characteristics associated with prognosis

  Tan et al. (2014): Dependence on an unobserved latent variable
- Survival function depends on the complete history of the marker

$$\begin{aligned} S_i(t \mid \mathcal{M}_i(t), w_i) &= \mathrm{P}[T_i > t \mid \mathcal{M}_i(t), w_i] \\ &= \exp\left(\int_0^t h_0(s) \exp\{w_i\gamma + \alpha m_i(s)\} ds \right) \end{aligned}$$

  so that development of the mixed-effects model should consider
  - Correct specification for subject-specific random effects
  - Flexible adjustment for temporal trends (polynomials, splines)
  - Time interactions with time-independent covariates

Assumptions

Baseline hazard function $h_0(t)$

- Unspecified baseline hazard function in standard Cox regression
  - Avoid an assumption for the distribution of event times $(+)$
  - In joint model can result in under-estimation of standard errors $(-)$
- Parametric but with different levels of flexibility
  - **Fully parametric**: Risk function for a parametric distribution
    - ★ Exponential: $h_0(t) = \lambda$
    - ★ Weibull: $h_0(t) = \lambda a t^{a-1}$
  - **Somewhat parametric**: Parametric but flexible risk function
    - ★ Piecewise constant step function
    - ★ Regression splines

    but avoid an over-specified model with too many parameters; between $1/20$ and $1/10$ of the total number of events

Estimation and inference

Given the set-up and assumptions. . .

- Likelihood estimation for $\{\alpha, \beta, \gamma, \sigma, D\}$ given data $\{y_i, T_i^\star, \delta_i\}$

$$\ell(\alpha, \beta, \gamma, \sigma, D) = \sum_{i=1}^{n} \log \int f_Y(y_i \mid b_i)\, h(T_i^\star \mid b_i)^{\delta_i}\, S(T_i^\star \mid b_i)\, f_b(b_i)\, db_i$$

  with densities f_Y and f_b, hazard $h(t)$ and survival $S(t)$ functions
  - Integration requires numerical approximation
    - ★ Laplace approximation
    - ★ Adaptive Gaussian quadrature
  - Maximization requires optimization algorithms
    - ★ Expectation-maximization algorithm
    - ★ Newton-Raphson algorithm
- Likelihood inference facilitates use of likelihood ratio test; Wald tests are also available for regression parameters

jointModel command

```
?jointModel
?jointModelObject

ml4 <- lme(log(serBilir) ~ year + drug:year, random = ~ year | id,
           data=pbc2)

ms1 <- coxph(Surv(years, status=='dead') ~ drug + sex + I(age-50),
             data=pbc2.id, x=TRUE)

mj1 <- jointModel(lmeObject=ml4, survObject=ms1, timeVar='year',
                  method='weibull-PH-aGH')
```

jointModel arguments

- `lmeObject`
  - Linear mixed-effects model for longitudinal marker
  - Multi-record data, ordered in 'long' format (`pbc2`)
  - No additional correlation structure beyond random effects
- `survObject`
  - Standard Cox regression model for censored survival outcome
  - Single-record data in same order as for `lmeObject` (`pbc2.id`)
  - `x=TRUE` so that design matrix is returned
- `timeVar`
  - Time variable in `lmeObject`
  - `lmeObject` and `survObject` must have same time scale

jointModel arguments

- `method`: Specifies the baseline hazard function, parameterization of relative risk model, and procedure for numerical integration
  - `weibull-PH-aGH` (default)
  - `weibull-PH-GH`
  - `weibull-AFT-aGH`
  - `weibull-AFT-GH`
  - `piecewise-PH-aGH`
  - `piecewise-PH-GH`
  - `spline-PH-aGH` (allows strata)
  - `spline-PH-GH` (allows strata)
  - `Cox-PH-aGH`
  - `Cox-PH-GH`

  PH: proportional hazards; AFT: accelerated failure time
  GH or aGH: standard or adaptive Gauss-Hermite quadrature

Results summary

- Coefficients (SEs) from mixed-effects model and joint model

Variable	Mixed model	Joint model
Year	0.176 (0.018)	0.182 (0.018)
Year-by-treatment	0.003 (0.024)	0.005 (0.024)

- Coefficients (SEs) from extended Cox model and joint model

Variable	Cox model	Joint model
Treatment	−0.026 (0.173)	−0.034 (0.185)
Sex	0.180 (0.235)	0.163 (0.250)
Age	0.068 (0.009)	0.065 (0.009)
Serum bilirubin	1.465 (0.094)	1.358 (0.101)

Model diagnostics

Evaluate adequacy of the fitted joint model using residual analysis

- **Longitudinal submodel**
  - Subject-specific (or, conditional) residuals

$$r_i(t) = y_i(t) - \{x_i(t)\hat{\beta} - z_i(t)\hat{b}_i\}$$

    with focus on constant variance and normality
  - Marginal (or, population-averaged) residuals

$$r_i = y_i - X_i\hat{\beta}$$

    with focus on specification of the mean model
- **Survival submodel**: Martingale residuals

$$r_i(t) = N_i(t) - \int_0^t \hat{h}_0(s)R_i(s)\exp\{w_i\hat{\gamma} + \hat{\alpha}\hat{m}_i(s)\}ds$$

with focus on specification of the model for $m_i(t)$

Advanced models

Facilitate flexible modeling to fully utilize longitudinal information

- **Interaction effects** (§5.1.1)

$$h_i(t) = h_0(t) \exp\{w_i\gamma + [w_{i1} \times m_i(t)]\alpha\}$$

Example: Interaction between treatment and bilirubin on mortality

- **Time-dependent slopes** (§5.1.3)

$$h_i(t) = h_0(t) \exp\{w_i\gamma + \alpha_1 m_i(t) + \alpha_2 m_i'(t)\}$$

Example: Effect of level and slope of serum bilirubin on mortality

Advanced topics

- Discrete markers (§5.7)
- Competing risks (§5.5.1)
- Recurrent events (§5.5.2)
- Mediation (Richiardi et al., 2013)
- Prediction (§7)

Analysis goals

- **Estimation**
  - Characterize association between exposure and outcome
  - Control for observed confounders, modeling effect modification
  - What is the association between serum bilirubin and mortality?
- **Prediction**
  - Predict average outcome from covariate values
  - Model selection focuses on maximizing prediction accuracy
  - Can serum bilirubin be used to predict mortality risk?

(Shmueli, 2010; French et al., 2016)

Dynamic prediction

Update predictions over time based on new longitudinal data

- Prediction of conditional survival probability

$$\pi_i(u \mid t) = \mathrm{P}[T_i^\star \geq u \mid T_i^\star > t, \mathcal{Y}_i(t), w_i, \mathcal{D}_n; \theta]$$

  at time $u > t$ given survival to time t

- Prediction of longitudinal outcome

$$\omega_i(u \mid t) = \mathrm{E}[Y_i(u) \mid T_i^\star > t, \mathcal{Y}_i(t), w_i, \mathcal{D}_n; \theta]$$

  at time $u > t$ given observed responses $\mathcal{Y}_i(t)$

given sample data $\mathcal{D}_n = \{T_i^\star, \delta_i, y_i; i = 1, \ldots, n\}$

Summary

- Longitudinal model + Survival model = Joint model
  - Describe evolution of marker levels over time for each subject
  - Associate the subject-specific evolutions with event hazard
  - Specify a dependence structure between the two models
- Combine a linear mixed-effects model for longitudinal marker with a relative risk model for a censored survival outcome
- Advantages over separate longitudinal and survival analyses
  - Incorporates random error term for measurement error
  - Accommodate marker as an endogenous time-dependent covariate
  - Marker is not assumed constant in between measurement times
  - Flexible modeling to fully utilize longitudinal information
- Model fitting can be complex and computationally intensive; analysis methods should be selected based on the scientific question
- Software: R, `JM::jointmodel`; Stata, `stjm`

Resources

Textbooks

1. Kalbfleisch JD, Prentice RL. The Statistical Analysis of Failure Time Data, 2002.
2. Rizopoulous D. Joint Models for Longitudinal and Time-to-Event Data with Applications in R, 2012.
3. Verbeke G, Molenberghs G. Linear Mixed Models for Longitudinal Data, 2000.

Resources

Articles

1. Andersen PK, Gill RD. *Ann Stat* 1982; **10**:1100–20.
2. Asar Ö, et al. *Int J Epidemiol* 2015; **44**:334–44.
3. Cologne J, et al. *Epidemiology* 2012; **23**:565–73.
4. Cox DR. *J R Stat Soc Ser B* 1972; **34**:187–220.
5. French B, Heagerty PJ. *Biometrics* 2009; **65**:415–22.
6. French B, et al. *Stat Methods Med Res* 2016; **25**:255–71.
7. Ibrahim JG, Molenberghs G. *Test* 2009; **18**:1–43.
8. Laird NM, Ware JH. *Biometrics* 1982; **38**:963–74.
9. Murtaugh PA, et al. *Hepatology* 1994; **20**:126–34.
10. Richiardi L, et al. *Int J Epidemiol* 2013; **42**:1511–9.
11. Rubin DB. *Biometrika* 1976; **63**:581–92.
12. Shmueli G. *Stat Sci* 2010; **25**:289–310.
13. Tan KS, et al. *Stat Med* 2014; **33**:4770–89.
14. Tsiatis AA, Davidian M. *Stat Sin* 2004; **14**:809–34.
15. Yoshida K, et al. *Br J Haematol* 2019; **185**:107–15.