BIOS 7721

Joint Modeling of Longitudinal and Survival Data

Getting to Know You

- Meet with each of you individually (sign-up link in Canvas)
- Use discussion boards to meet each other
- Will try to join the class zoom link 10 minutes ahead and stay on for 10 minutes after
- **Office Hours**
  - Fridays 10-11?
  - Or by appointment

Class Info

Requirements

- **Homework assignments (45%)**
  - HW 1: Longitudinal data analysis & survival analysis (15%)
  - HW 2: Time-dependent covariates & two-stage models (15%)
  - HW 3: Joint modeling (15%)
- **Class participation (25%)**
  - In-class quizzes (20%)
  - Participating in class & Discussion boards (5%)
- **Final project (30%)**
  - Option 1: Research Paper
  - Option 2: Simulation Study

Class Info

Class Participation (25%)

Quizzes

- Every class the last ~10 minutes will be reserved for organized class participation
- Will be put into breakout rooms to answer 3-4 questions
- Individually submit the answers to these questions on Canvas by the end of the week
- These quizzes will be graded for completion

Ad-hoc

- Discussion boards (questions, answers, etc.)
- In-class participation (questions during presentations, etc.)

Class Info

Final Project (30%)

Option #1: Research Paper

- Read research paper
- Prepare a 10 min presentation (record and post on Canvas)
- Watch two presentations and comment
- Submit a max 2 page report

Option #2: Simulation Study

- Simulate data from a joint model
- Submit a max 2 page report
- Strongly recommended for students pursuing a PhD

Course Competencies

- Overview of the theory and application of JMs for longitudinal and survival data
  - Identify when it is appropriate to use a JM
  - Using the JM package in R to fit and interpret shared random effects JMs
- Skills as a statistical researcher
  - Simulate survival and longitudinal data
  - Read a research paper, understand it, and discuss it

Introduction to Joint Modeling

Day 1

- Why do we need joint models?
- When do we need joint models?
- How do joint models work?

Breakout session #1

1. What are two reasons we need joint models for longitudinal and survival outcomes?
2. What is informative dropout in a longitudinal study?
3. What could be a consequence of informative dropout when estimating a treatment effect?
4. What is your favourite Denver food/restaurant?

Example: PBC Study

- Primary biliary cirrhosis (PBC) is a chronic liver disease that leads to cirrhosis and eventually death
- 10-year study conducted by Mayo clinic *(Murtagh et al., Hepatology, 1994)*
  - 158 randomized to treatment, 154 to placebo
- Longitudinal biomarker measurements of serum bilirubin at times 0, 6m, 1y, 2y, etc.
- Outcomes of interest:
  - **Survival** (Time to death)
  - **Longitudinal serum bilirubin levels**
  - **Association between them**

Terminology

- **Biomarker:** measure of a biological process
  - e.g., serum bilirubin
- Inherent features of biomarkers:
  - Measured with error
  - Measurements taken on the same individual are correlated
  - Value of the biomarker may be related to prognosis
- **Baseline:** time 0 of a follow-up study
  - e.g., time of enrollment, time of treatment, time of diagnosis

Follow-up Studies

- Follow-up studies involve following individuals for a period of time
- **Outcomes** collected:

1. **Longitudinal biomarkers**
   - Repeated measurements
   - Binary or continuous
2. **Time to a terminating clinical event (Survival outcome)**
   - Time to the event
   - Indicator of experiencing the event or **censored**

Research Questions

Goal:

1. **Longitudinal marker:** Estimate the biomarker profile/trajectory
2. **Survival outcome:** Estimate the risk of the clinical event
3. **Association:** Estimate the relationship between the biomarker profile/trajectory and the clinical event

Current Toolkit

- Methods for separately analyzing these explicit outcomes are available in the literature

- **Longitudinal Data Analysis**
  - Mixed effects models
  - GEE
- **Survival Analysis**
  - Nonparametric methods (Kaplan-Meier, Nelson-Aalen)
  - Relative risk models (Cox models)

Outcomes in Follow-up Studies

- **Explicit outcomes**
  - Longitudinal responses (e.g., markers, blood values)
  - Time-to-event or survival outcomes
- **Implicit outcomes**
  - Missing data
    - Missing observations
    - Dropout/censoring
  - Random visit times

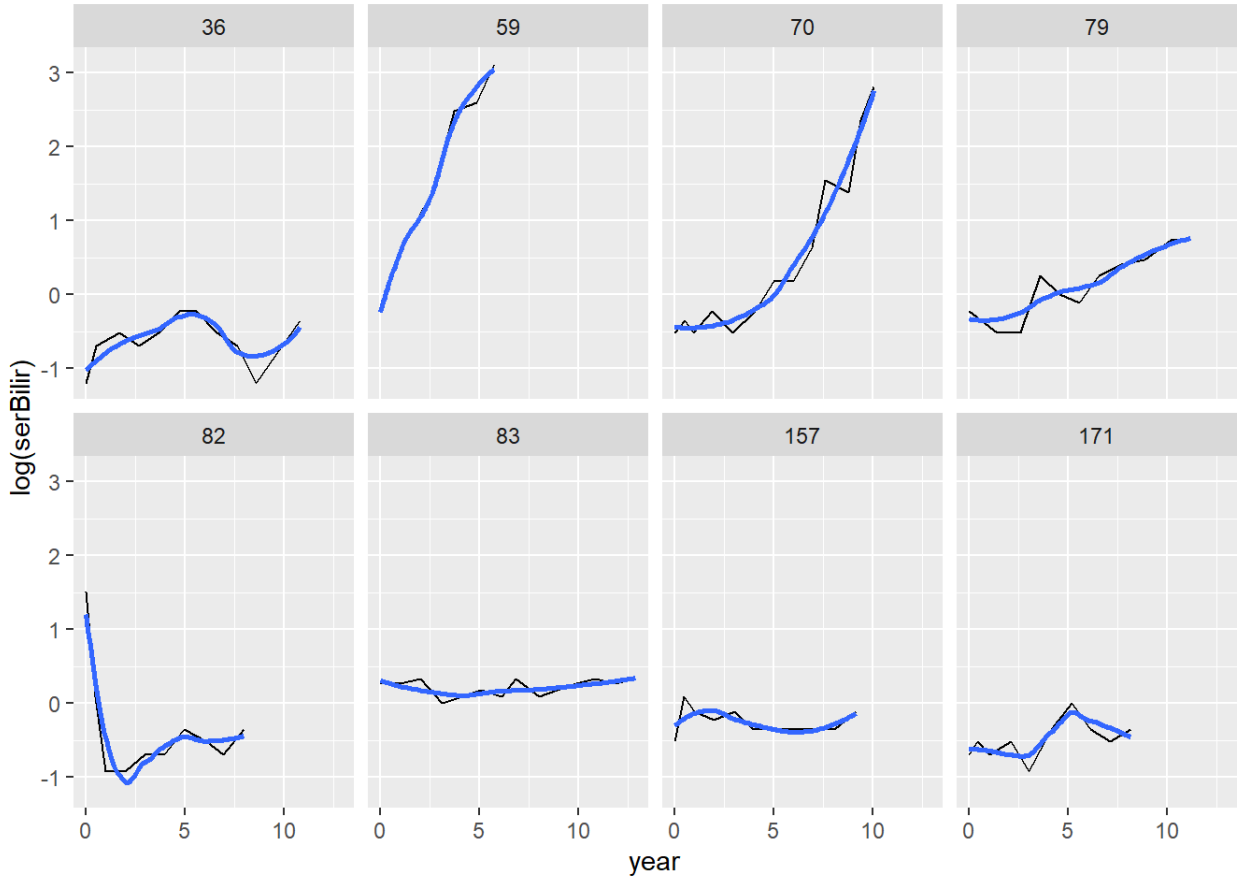
PBC Study

1. Longitudinal biomarkers (serum bilirubin)
2. Time to a terminating clinical event (time to death)

- Q1: Describe the evolution/change in the biomarker over time.
- Q2: Describe how changes in the biomarker influence a patient's risk of death.

PBC Study

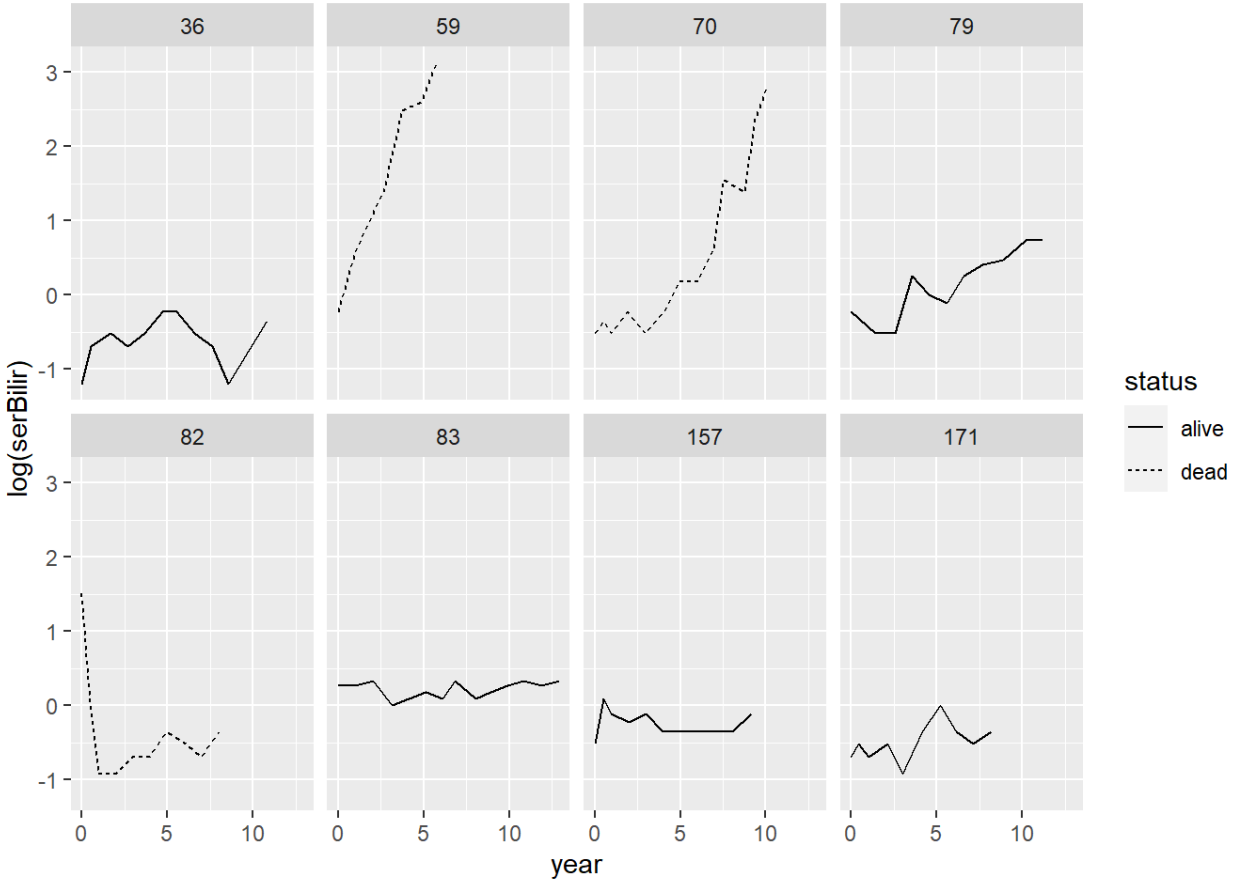
1. Longitudinal biomarkers (serum bilirubin)



Q1: Describe the evolution/change in the biomarker over time.

PBC Study

1. Longitudinal biomarkers (serum bilirubin)
2. Time to a terminating clinical event (time to death)



Terminology

- **Informative dropout:** the dropout probability depends on the unobserved measurements
  - **Informative censoring**
  - **Non-ignorable**
  - **Missing not at random (MNAR)**

- **Random dropout:** dropout process depends on the observed measurements (i.e., those collected prior to dropout)
  - **Non-informative**
  - **Missing at random (MAR)**

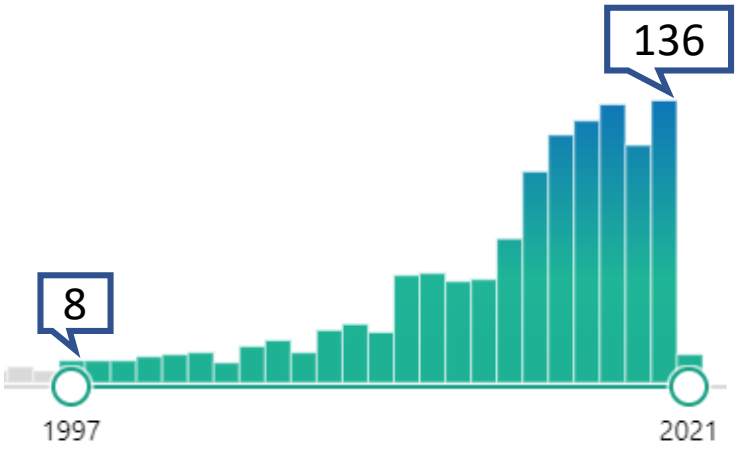
Follow-up Studies

1. Longitudinal biomarkers (serum bilirubin)
2. Time to a terminating clinical event (time to death)

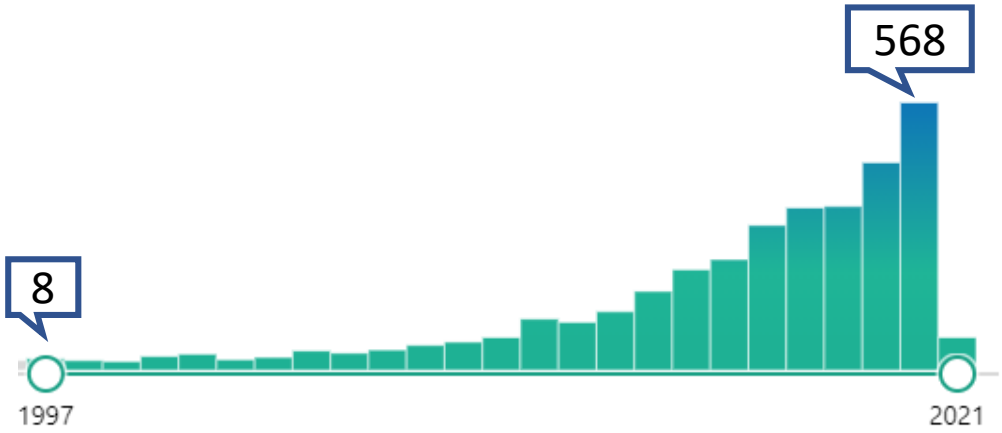
- Q1: Describe the evolution/change in the biomarker over time.
  - Informative dropout
- Q2: Describe how changes in the biomarker influence a patient's risk of death.
  - Baseline value versus longitudinal trajectory

- Joint models can handle both of these!

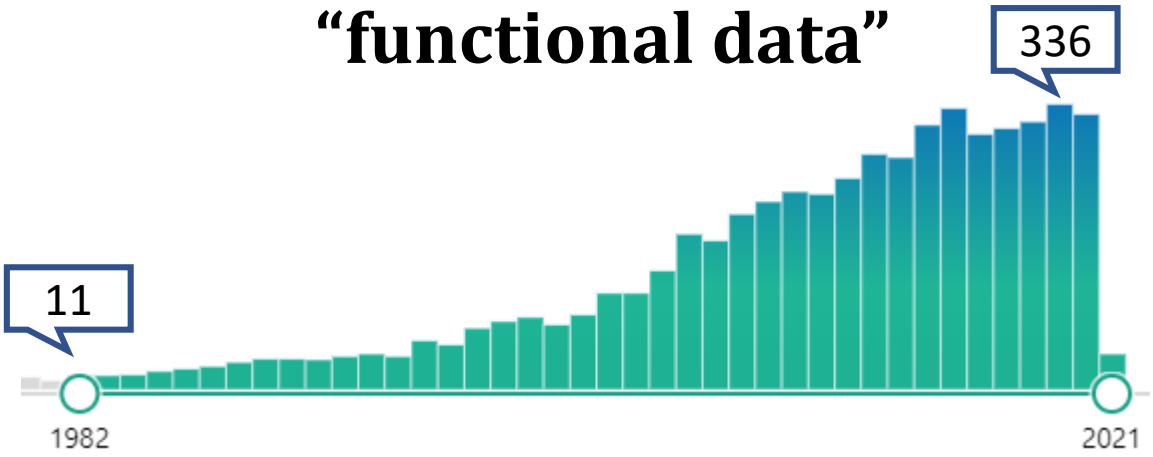
“joint model”



“causal inference”



“functional data”



Images generated from PubMed search

Why do we need joint modeling?

- **Types of data collected:**
  - New technologies lead to more biomarkers are being used and collected
  - EHR leads to easier connection of biomarker and outcome data
- **Types of research questions:**
  - Clinical decision making is moving towards personalized medicine
  - Identify if biomarkers are surrogates
  - Up-to-date personalized predictions

Developments in Joint Modeling

- Previously, focused on separate analysis per outcome or naïve analysis
  - Lack of data linkage
  - Lack of methodology
  - Lack of software
- Now, development of methods and software for many different joint modeling approaches
  - Can handle different types of outcomes
  - Has reasonable computing time
  - Can be flexible
  - Can answer the questions of interest

When to use joint modeling?

- Outcome processes are correlated. Main settings:

1. The focus is on the **survival outcome**, and we want to account for the effect of a time-dependent covariate measured with error
   - versus Cox model with a time-dependent covariate
2. The focus is on the **longitudinal outcome** and we want to correct for nonrandom/informative dropout
   - versus a mixed model

Research Questions

- 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 of the marker different between males and females?
- Handling implicit outcomes
  - Focus on single longitudinal outcome but with dropout or random visit times
- Focus on multiple outcomes
  - How strong is the association between the longitudinal evolution of marker and the risk of death?
- Prediction
  - Taking into account a person's changing marker value in producing an updated prediction

PBC Study

- **Research questions:**
  - To understand within-subject patterns of bilirubin change
  - Can bilirubin discriminate between patients of low and high risk?
  - How strong is the association between bilirubin and the risk of death?
  - Can observed serum bilirubin levels be used to predict a patient's survival probability?

How do joint models work?

- Let Y_1 and Y_2 be two outcomes of interest that we are interested in jointly modeling
- Both can be measured longitudinally
  - E.g., Serum bilirubin and albumin in PBC patients
- One can be longitudinal and one survival
  - E.g., Serum bilirubin and time to death in PBC patients
- What are some approaches to construct the joint density $p(y_1, y_2)$ of (Y_1, Y_2)?

How do joint models work?

What are some approaches to construct the joint density $p(y_1, y_2)$ of (Y_1, Y_2)?

1. Multivariate models
2. Conditional models:

$$p(y_1, y_2) = p(y_1)p(y_2|y_1)$$

3. Copulas:

$$p(y_1, y_2) = c\{F(y_1), F(y_2)\}p(y_1)p(y_2)$$

4. **Random Effects Joint Models** are most popular

Random Effects Joint Models

- **Intuition:** Latent underlying process
- Involves specifying:
  - Longitudinal submodel
  - Survival submodel

shared random effect models
p(y1|b); p(y2|b) both depend on b

- Assume that random effects (b) underly both the longitudinal and survival processes
  - "Shared random effects"
- Random effects induce the dependence between the two processes

Random Effects Joint Models

Random Effects Joint Models

- **Random Effects Joint Models** specifies

$$p(y_1, y_2) = \int p(y_1, y_2|b)p(b)db$$

$$= \int p(y_1|b)p(y_2|b)p(b)db$$

- Unobserved random effects b explain the association between Y_1 and Y_2
- **Conditional independence assumption** $Y_1 \perp\!\!\!\perp Y_2 \mid b$

How do joint models work?

$$p(y_1, y_2) = \int p(y_1, y_2|b)p(b)db$$

$$= \int p(y_1|b)p(y_2|b)p(b)db$$

- A mixed model for the longitudinal outcome
- A relative risk model for the event process
- Using “shared random effects” to join together the two probability distributions
- Random effects are usually assumed to be normally distributed

Recap: Why/When/How joint modeling?

- Applicable in settings where individuals are followed over time
- Collect information on two inter-linked processes:
  - Biomarker process **(longitudinal) -> mixed model**
  - Time to event process **(survival) -> Cox regression**
- **Goal:**
  1. **Longitudinal marker:** Estimate the biomarker profile/trajectory allowing for informative dropout, e.g., death
  2. **Survival outcome:** Estimate the relationship between the biomarker profile/trajectory (measured with error) and the clinical outcome
- **Random Effects Joint Model!**

Course Modules

- Longitudinal data
  - Linear mixed-effects models
  - Missing data mechanisms
- Survival data
  - Relative risk models
- Basic Joint Model
  - Shared random effects model
- Software for fitting a joint model
- Missing Data
- Extensions of joint models
- Dynamic Prediction

Breakout session #1

1. What are two reasons we need joint models for longitudinal and survival outcomes?
2. What is informative dropout in a longitudinal study?
3. What could be the consequence of informative dropout when estimating a treatment effect?
4. What is your favourite Denver food/restaurant?