Longitudinal data analysis Examples and Basics

Workshop: Analysis of Longitudinal Data

12th Nov 2024

Jaroslaw Harezlak
Armando Teixeira-Pinto

THE UNIVERSITY OF SYDNEY

Introduction

1. Repeated measures
2. Longitudinal studies
3. Different types of outcomes
4. Advanced designs
5. Summary

Introduction

Longitudinal Studies:

Studies in which individuals are measured **repeatedly** over time.

–This short course will cover the design, analysis and interpretation of longitudinal studies.

–The short course will emphasize model development, use of statistical software, and interpretation of results.

Features of Longitudinal Data

- **Defining feature:** Repeated observations on the units of analysis (e.g. individuals), allowing the direct study of change.
- Ordinarily, the measurements are **commensurate**, that is, the same variable is measured repeatedly.
- In most analyses, the primary goal is to describe how individuals change over time and how that change depends on the characteristics of participants or their environment.
- Repeated measurements on study participants are typically positively correlated.
- Therefore, methods for longitudinal data must account for this correlation.

Terminology

- **Clustered Data:** outcome variable measurements are correlated.

Examples: group randomization, family data, multiple rater data and longitudinal data

- **Longitudinal Studies:** outcome variable measured repeatedly over time

- **Repeated Measures Studies:** special case of longitudinal studies with measurements on a common set of time points

Example 1: Tx of HIV Infection with Saquinavir, Zidovudine, and Zalcitabine – clinical trial

Patients with HIV infection were randomly assigned to receive:

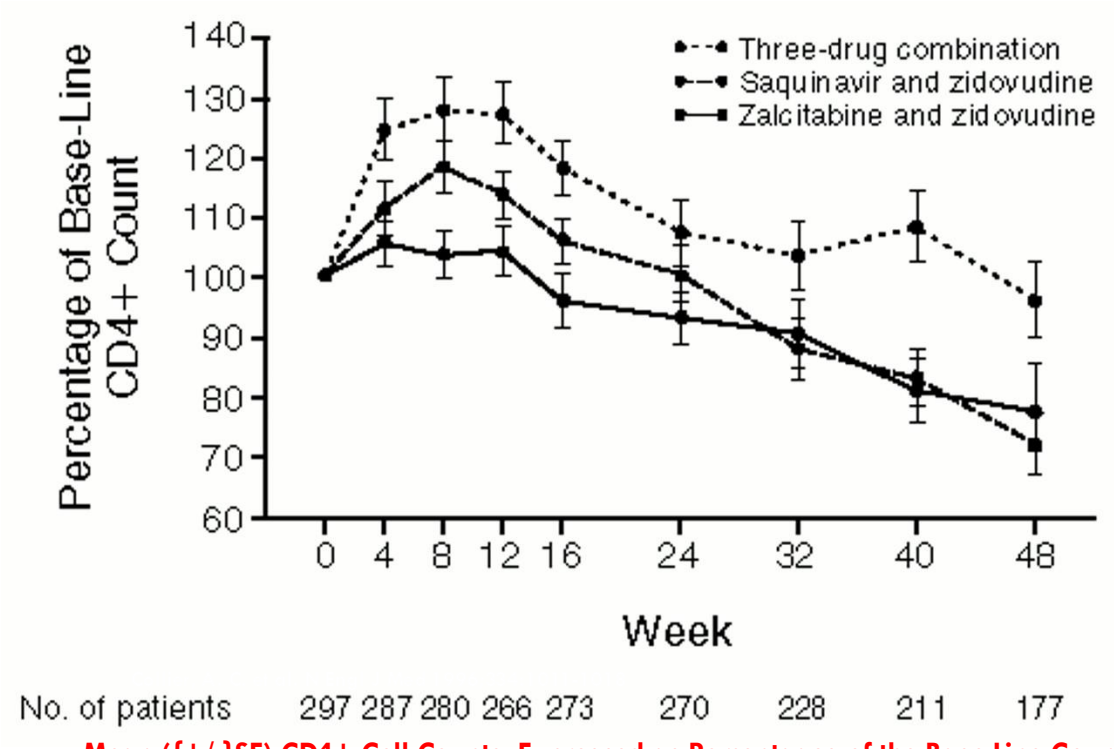
a. three drugs (zidovudine, saquinavir, and zalcitabine) or
b. two drugs (zidovudine plus saquinavir)
c. two drugs (zidovudine plus zalcitabine).

N=302 patients enrolled had CD4+ counts of 50 to 300 cells per cubic millimeter at baseline. The study lasted 24 weeks.

The primary end points indicative of efficacy were changes in absolute CD4+ cell counts and quantitative HIV titers in cultures of PBMCs during the first 24 weeks of treatment.

NEJM **Volume 334:1011-1018,** **April 18, 1996**

Example 1: Tx of HIV Infection with Saquinavir, Zidovudine, and Zalcitabine



Mean ({+/-}SE) CD4+ Cell Counts, Expressed as Percentages of the Base-Line Count, According to Study Week

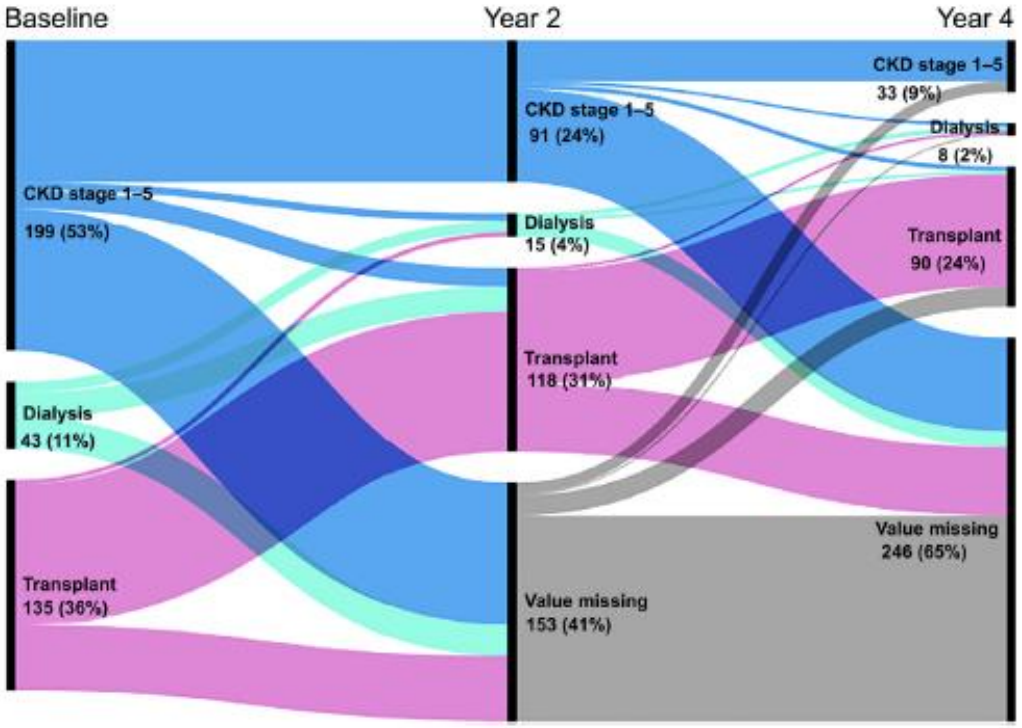
Example 2: Longitudinal cohort study of children with chronic kidney disease (CKD) – observational study

- Multi-center cohort study in Australia and New Zealand
- N=377 children (aged 6-18 years) with chronic kidney disease (CKD) observed over time: biennially for four years.
- CKD stages 1-5 (predialysis), dialysis, or transplant,

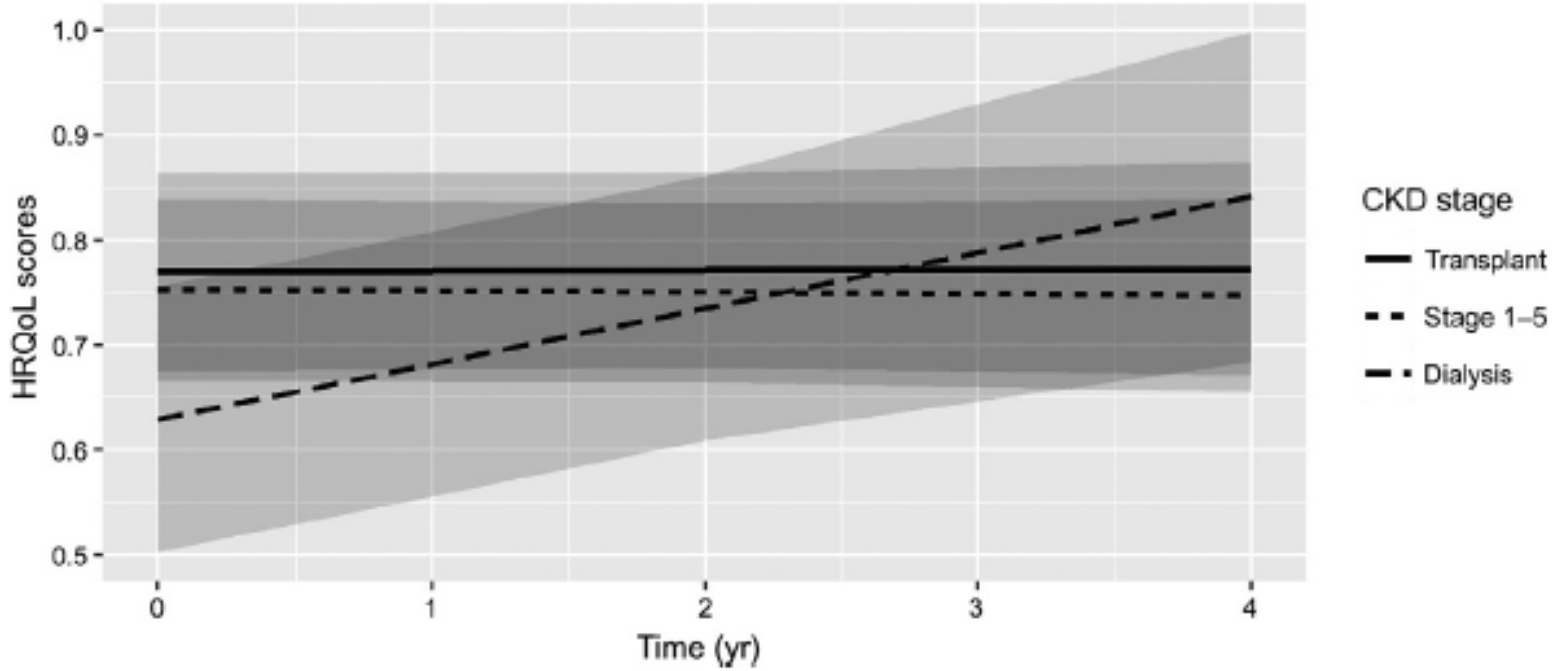
Outcome variable: Continuous health-related quality of life (HRQoL)

Guha, Chandana, et al. "Longitudinal assessment of the health-related quality of life of children and adolescents with chronic kidney disease." Kidney international 103.2 (2023): 357-364.

Example 2: Longitudinal cohort study of children with chronic kidney disease (CKD) – Data summary



Example 2: Longitudinal cohort study of children with chronic kidney disease (CKD) – Results



Example 3: Phase 3 double-blind, randomized, controlled trial of weight change - tirzepatide

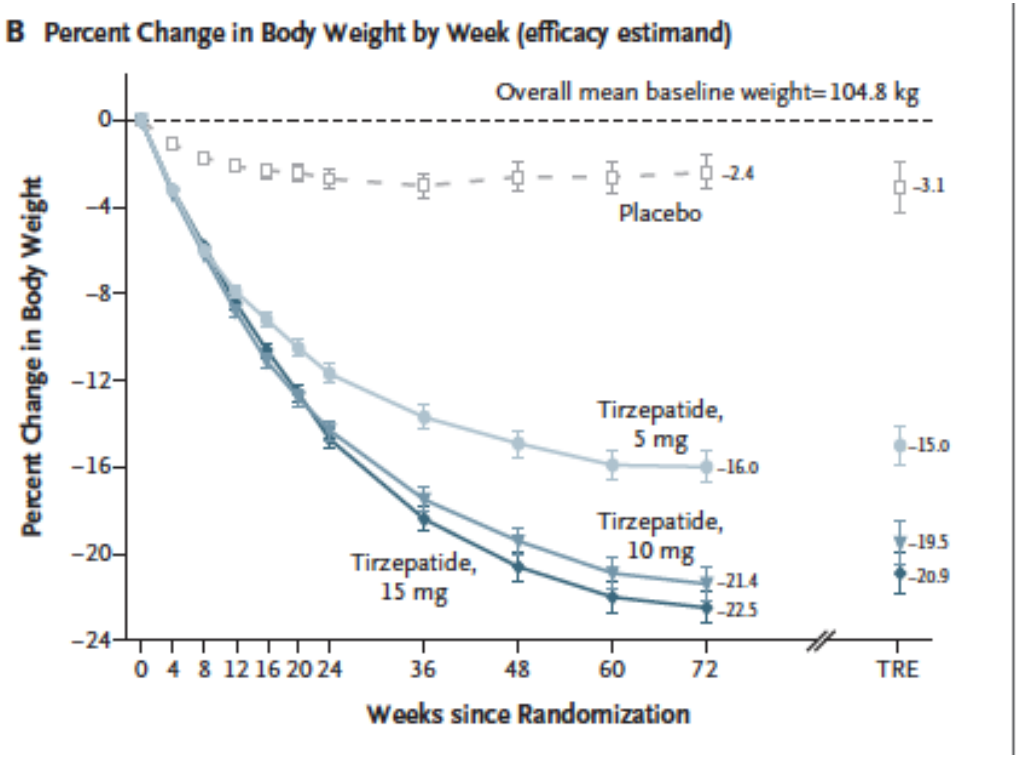
N = 2539 adults with a body-mass index (BMI) of 30 or more (or 27 or more and at least one weight-related complication, excluding diabetes)

Design: 1:1:1:1 ratio to receive once-weekly, subcutaneous tirzepatide (5 mg, 10 mg, or 15 mg) or placebo for 72 weeks, with 20-week dose-escalation period

Endpoints: Coprimary - the percentage change in weight from baseline and a weight reduction of 5% or more.

Jastreboff, Ania M., et al. "Tirzepatide once weekly for the treatment of obesity." New England Journal of Medicine 387.3 (2022): 205-216.

Example 3: Phase 3 double-blind, randomized, controlled trial of weight change - tirzepatide



Common Features of These Studies

Repeated measurements of study participants.

Two general types of design:

- **Parallel Design:**

  Groups of subjects defined by treatment or exposure category are followed over time. The main objective is to compare the trajectories of outcome variables between groups.

- **Crossover Design:**

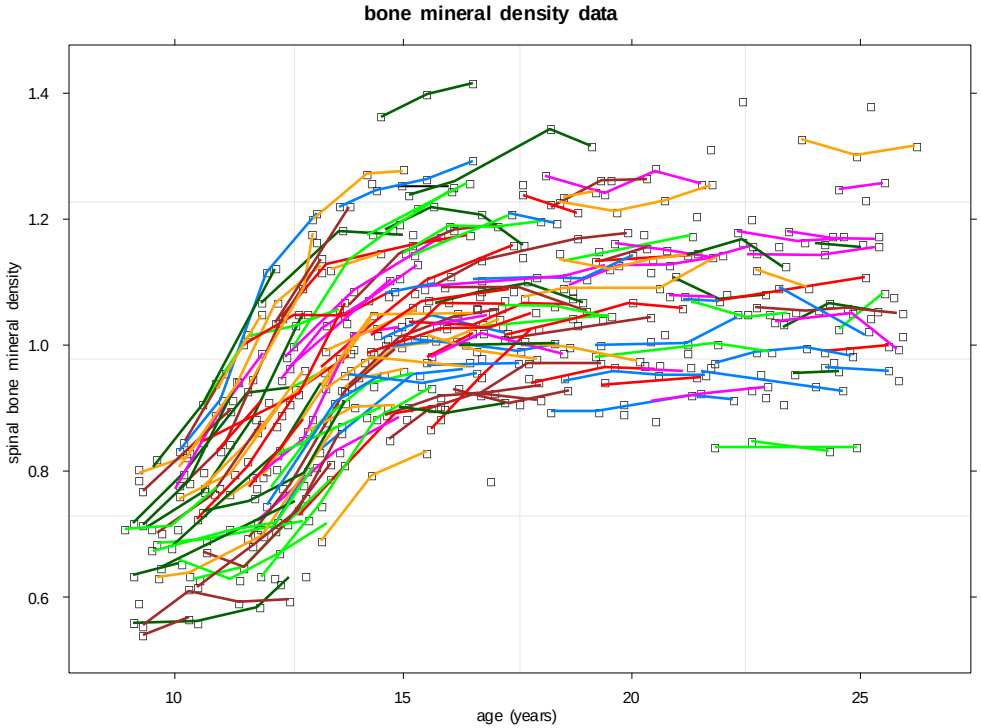
  Subjects are exposed to multiple treatments or exposures. The objective is to compare the responses of the same subjects to different conditions.

Complex designs

Accelerated longitudinal designs (ALD): Observational studies with short periods of observations per individual covering an overall large span

Multilevel designs: clustering of observations happening at different time scales and non-time-related correlations

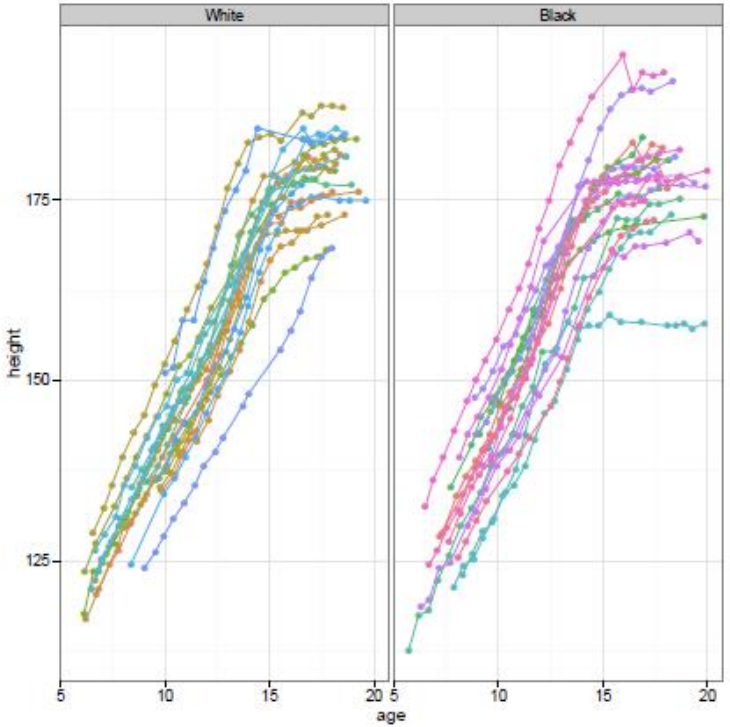
Example 4: Spinal Bone Mineral Density



Bachrach, L.K., Hastie, T., Wang, M.-C., Narasimhan, B. and Marcus, R. (1999). Bone mineral acquisition in healthy Asian, Hispanic, Black and Caucasian youth. A longitudinal study. J. Clin. Endocrin. Metab. 84, 4702–12.

Example 5:

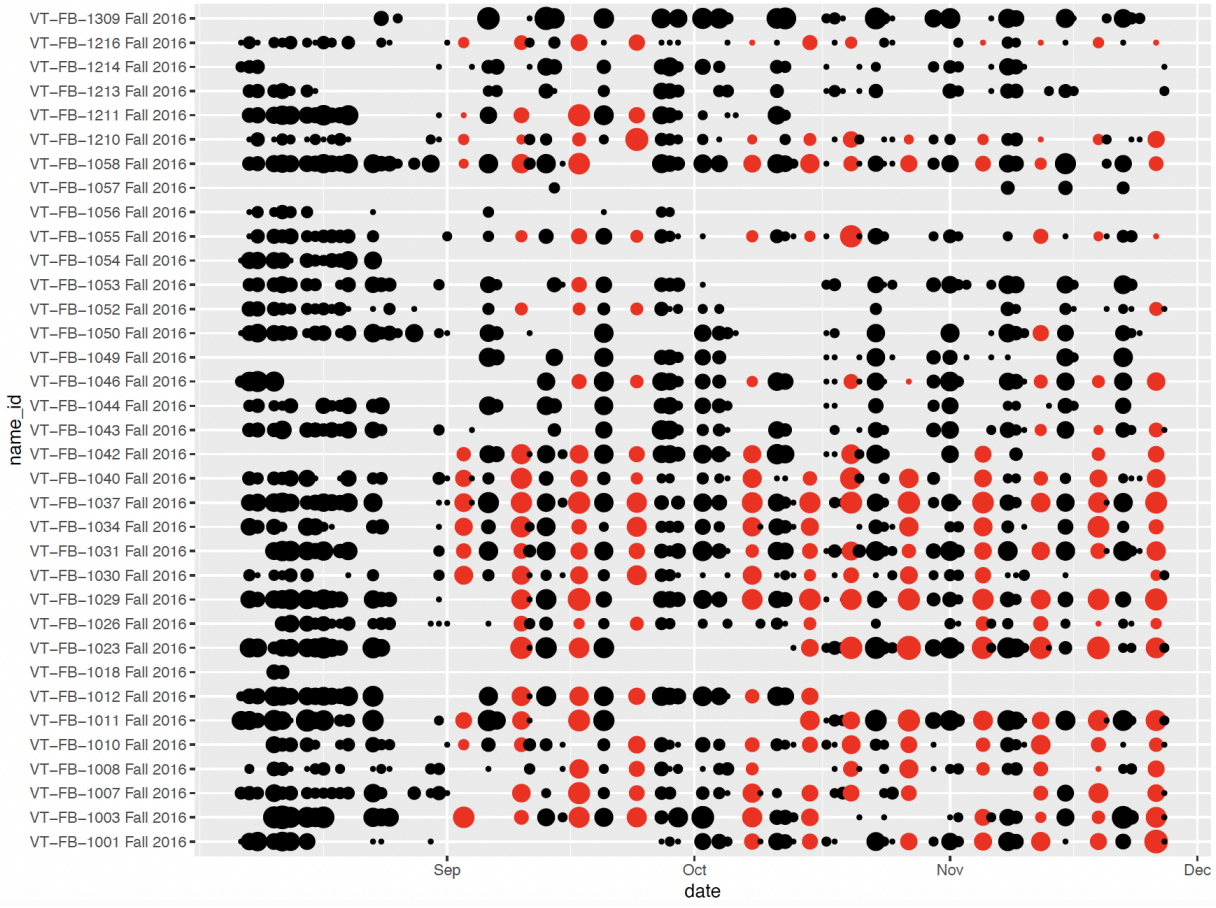
Children's growth data, Indianapolis, IN



Harezlak J., Ruppert D. Wand MP. *Semiparametric regression with R*. Vol. 109. New York: Springer, 2018.

Goal: Continuous data (longitudinal), model growth trajectories

Example 6: Exposure to the head impacts: American college football



General Data Structure

Y_{ij} = j^{th} observation of the i^{th} subject

In the balanced design, we have p times of measurement per subject

		Time				
		1	2	3	…	p
Subject	1	Y_{11}	Y_{12}	Y_{13}	…	Y_{1p}
	2	Y_{21}	Y_{22}	Y_{23}	…	Y_{2p}
		…				
	n	Y_{n1}	Y_{n2}	Y_{n3}	…	Y_{np}

General Data Structure

In the unbalanced design, subjects may be measured at different times. In that case, we must also have a measurement of the "metameter", the variable representing time, and will use that to model the data.

Subject 1: $(Y_{11}, \mathbf{X}_{11}), (Y_{12}, \mathbf{X}_{12}), (Y_{13}, \mathbf{X}_{13}), \ldots (Y_{1p}, \mathbf{X}_{1p})$

Subject i: $(Y_{i1}, \mathbf{X}_{i1}), (Y_{i2}, \mathbf{X}_{i2}), (Y_{i3}, \mathbf{X}_{i3}), \ldots (Y_{ip}, \mathbf{X}_{ip})$

The variable representing time is one of the covariates.

Special Case: Two groups

Two-Groups Parallel Design

		Time				
		1	2	3	…	p
Subjects						
Tx 1	1	Y_{11}	Y_{12}	Y_{13}	…	Y_{1p}
	2	Y_{21}	Y_{22}	Y_{23}	…	Y_{2p}
	…					
	m	Y_{m1}	Y_{m2}	Y_{m3}	…	Y_{mp}
Tx 2	m+1	$Y_{m+1,1}$	$Y_{m+1,2}$	$Y_{m+1,3}$	…	$Y_{m+1,p}$
	m+2	$Y_{m+2,1}$	$Y_{m+2,2}$	$Y_{m+2,3}$	…	$Y_{m+2,p}$
	…					
	m+n	$Y_{m+n,1}$	$Y_{m+n,2}$	$Y_{m+n,3}$	…	$Y_{m+n,p}$

Summary

In longitudinal studies the outcome variable can be:

continuous, binary or a count

The study may be a randomized trial or an observational study.

The data set can be unbalanced or incomplete.

Unbalanced: Subjects are measured at different occasions.

Incomplete: Some planned observations are not obtained.

In this short course we will develop a set of statistical tools that can handle all of these cases.

Emphasis on concepts, model building, software, and interpretation.

Organization of the Short Course

1) *Linear Models for Longitudinal Data*

More general approach for fitting linear models to unbalanced and incomplete longitudinal data.

Outcome: Continuous

Unbalanced and incomplete data

Class of models: Linear models

Software:

R – functions gls() and lme() in the library(nlme)

R – function lmer() in the library(lme4)

2) *Generalized Linear Models for Longitudinal Data*

Generalizations for different types of outcome data.

Outcomes: Continuous, binary, count

Class of models: Generalized Linear Models (e.g. logistic regression, Poisson regression)

Software: R – libraries: nlme, lme4 and geepack

Organization of the Course (cont.)

3) Generalized Additive Mixed Models (GAMM)

Extensions to allow fitting of nonlinear models to longitudinal data.

Outcome: Continuous, binary, count

Class of models: Generalized Linear Models (e.g. logistic regression, Poisson regression)

Software: R – library: mgcv

4) Advanced topics:

Dropouts, missing data

study design

model evaluation

Background Assumed

1) Samples and populations

2) Sample and population values

   Population values: parameters

   Sample values: estimates

3) Representation of Variables:

   Y : Outcome, response, dependent variable

   X: Covariates, independent variables

4) Basics of Regression Models

   $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i$

5) Inference

   Estimation, testing, and confidence intervals

6) Multiple linear regression/ANOVA

   Multiple logistic regression