Some questions

- What makes longitudinal data different, so that we need special methods to analyze it?

- What are clustered data?

- What are benefits of longitudinal models? (Or models for clustered data.)

- Why are longitudinal methods not used more?

Longitudinal designs

- Designed experiments and observational studies can be applied to cross-sectional or longitudinal settings. Here, they are defined for the latter.

- A controlled experiment involves an *intervention*, while an observational study does not.

- In many cases a controlled experiment will have one or more true treatment groups, along with a 'control' group that either receives some type of placebo, or does not receive any treatment.

See the course notes for more detail on designed experiments versus observational studies.

Time series and longitudinal data

- Time series methods (generally)…
  - focus on modeling one process over time (i.e., one observation taken at each time point, across time).
  - focus on predicting values of future occurrences.
- Generally, time series data can be found everywhere, including: stock prices, temperature, birth and mortality rates, health data for individuals (e.g., blood pressure), just to name a few areas.

- Longitudinal methods (generally)…
  - Involve measurements on multiple subjects.
  - Assume that the correlation structure is the same across subjects but that responses are independent between subjects.
- Often fewer time points for longitudinal data than time series data.
- Although analytical methods for time series and longitudinal data differ, they do have common elements, and the underlying processes that generate the data are often similar.

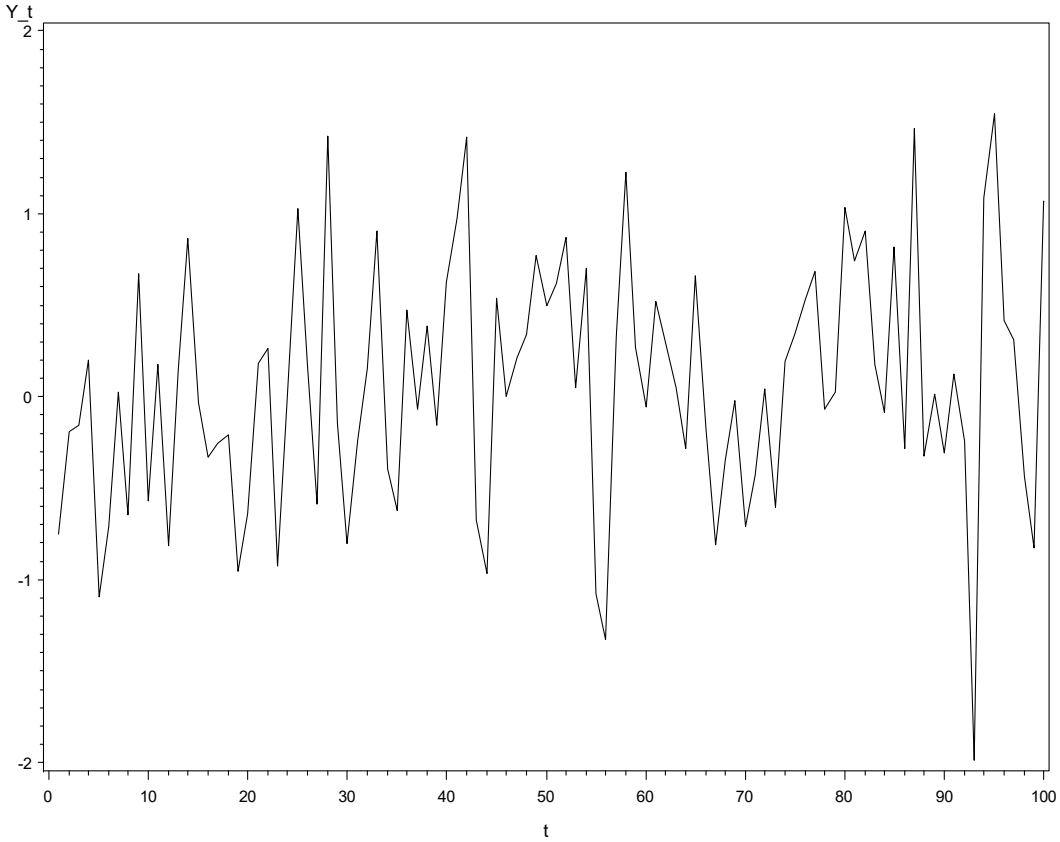
Time series data types and examples

Stationary processes

- A stationary process $\{Y_t\}$ has a constant mean (expected value) and finite 2nd moment for all times t, and the correlation between Y_t and Y_{t+h} does not depend on t, for all h.

- Below, data for stationary processes were simulated for the model $Y_t = \mu + \varepsilon_t$, where μ is the mean and ε_t are errors that are identically but not necessarily independently distributed.

<u>Example 1</u>: Stationary process with identically and independently (iid) distributed errors. For the simulated data to the right, μ=0 and ε_t ~ Normal with mean 0 and variance 0.46 for all *t*.



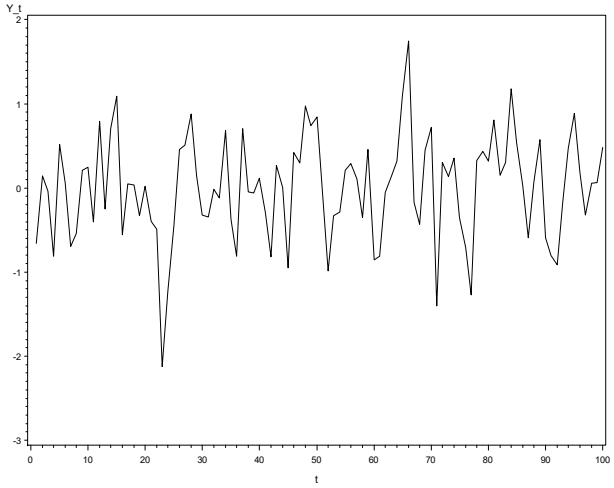
Example 2: Stationary process with correlated errors.

- Data below were generated using $\mu=0$ and errors that followed a first-order autoregressive [AR(1)] process: $\varepsilon_t = \phi\,\varepsilon_{t-1} + Z_t$ and $Z_t \sim iid$ for all t. (Specifically, Z_t ~Normal with mean 0 and variance 0.46.)

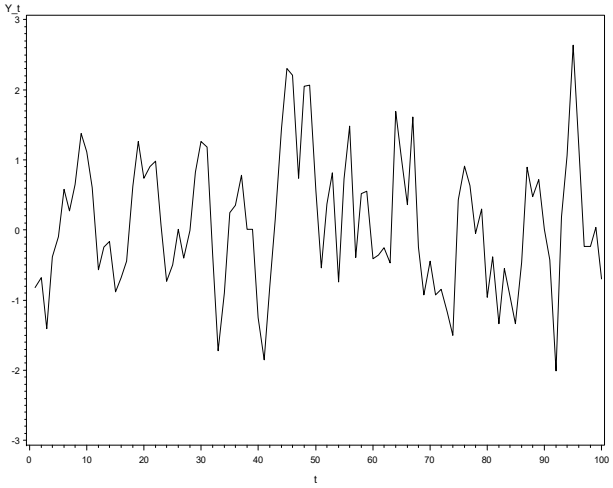
- A few notes on AR(1) processes:

  (i) Errors ε_t are identically distributed but not independent
  (ii) Must have $|\phi|<1$ for stationarity
  (iii) The higher the value of $|\phi|$, the higher degree of correlation between responses from day to day

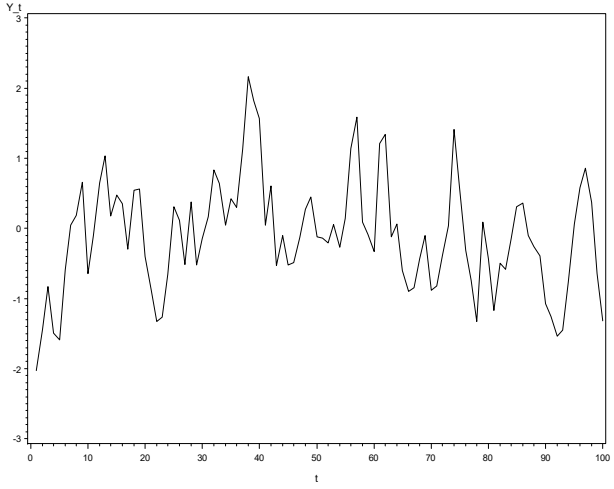
ϕ=0.25



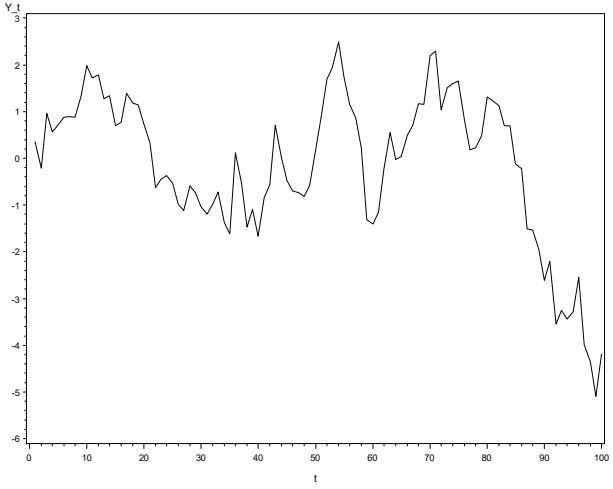
ϕ =0.5



ϕ =0.75

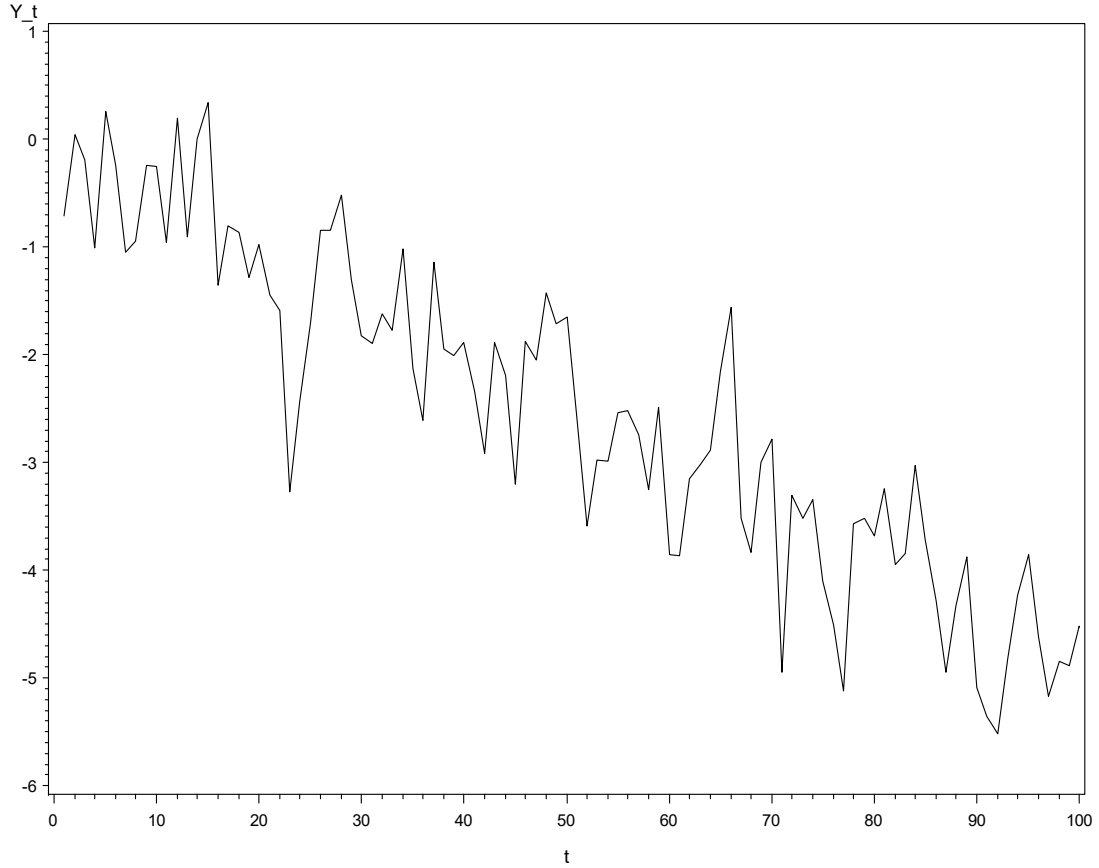


ϕ =0.99

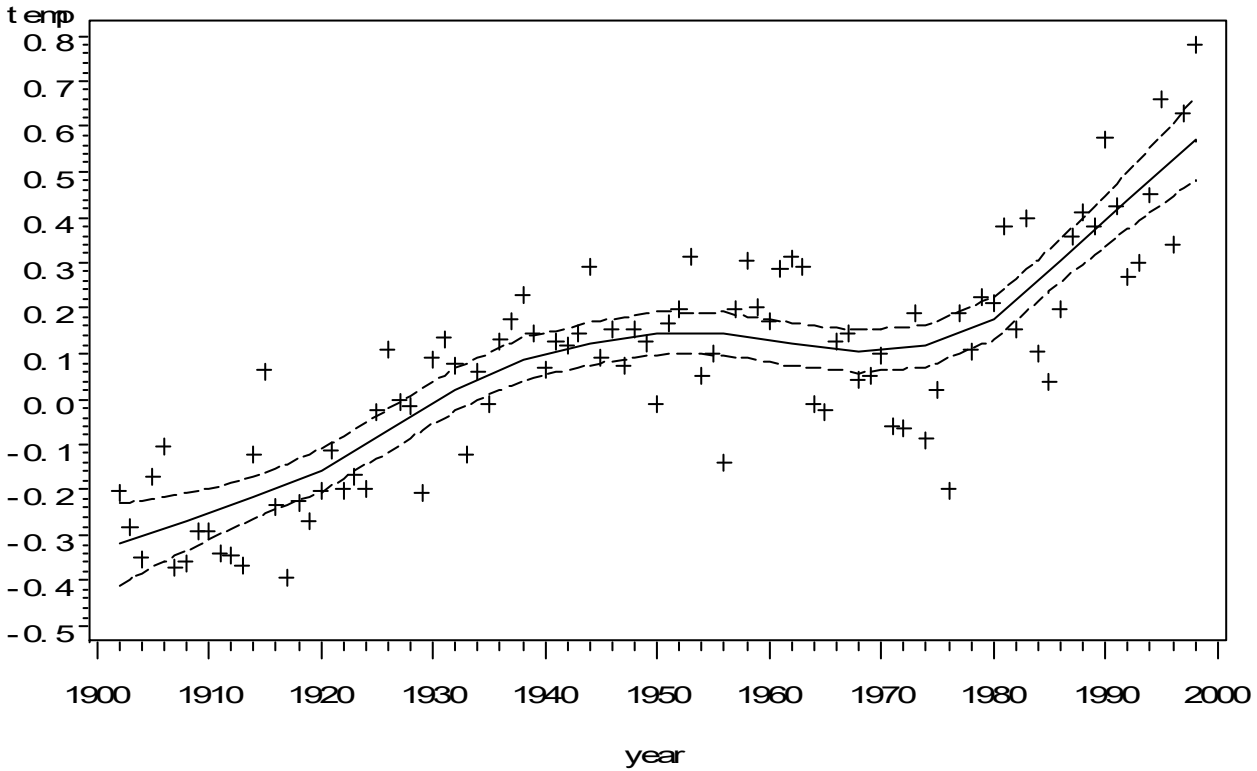


Processes with trend and correlated errors

- *Example 3*: AR(1) process with linear time trend. $Y_t = \beta_0 + \beta_1 t + \varepsilon_t$, β_0=0, β_1= –0.05, ε_t ~AR(1) (as in Ex. 2, last page, with ϕ =0.25)



- *Example 4*: Global temperature data, 20th century, with nonparametric regression fit (95% pointwise confidence bands for the mean in dashed lines).



Random walks – see course notes (includes Example 5)

Longitudinal data types and examples

Retrospective observational studies

Example 6: *Occupational Medicine – natural history of Beryllium disease.*

- Involved workers in production plants that use(d) Beryllium metal.
- Main aim: summarize the progression of illness for subject that developed Chronic Beryllium Disease (CBD) vs. those who only became sensitized (BeS), by comparing changes in pulmonary function (e.g., FEV_1, FVC), exercise physiology (e.g., $AADO_2$ Rest, Max VO_2) and blood and broncoalveolar lavage outcomes between these groups.
- Complex data:
  - Different tests were done on different days.
  - Each subject had different number of tests done over time.
  - Records spanned several decades.

Example 7: MDR-TB study.

- Involved obtaining records from subjects who were initially treated at National Jewish Health for multiple-drug resistant tuberculosis.
- Difficult retrospective study: many efforts for follow up.
- Analyses
  - Logistic regression for treatment success (evaluated based on sputum tests taken during or soon after discharge from the hospital)
  - Survival analysis for survival from TB (using longer-term data from medical records and follow up).
- Although these are longitudinal data, the logistic regression did not use repeated measures, and survival analysis is typically considered in a category separate from longitudinal data analysis.

Prospective observational studies

Example 8: Kunsberg/Air pollution study.

- Involved students attending the K-8 school at National Jewish Health.
- Health and behavioral variables were collected on subjects over time, some daily (e.g., daily albuterol use) and some more intermittent (e.g., personal exposure estimate and biomarkers from urine samples).
- Concurrently, air pollution measures from fixed monitors and personal monitors were taken. The relationship between health and environmental were examined. See Rabinovitch et al. (2004, JACI; 2006 and 2011, AJRCCM).

Example 9: COPDGene study.

Epidemiologic time series studies

<u>Example 10</u>: Relationship between hospital admission counts and PM_{10} in the San Luis Valley over 11 years.

- Involved determining the association between hospital counts at a medical center in Alamosa, Colorado that serves the greater San Luis Valley area, and concurrent PM_{10} concentrations – i.e., airborne coarse particulate matter.

- Although rural, airborne sand and dust particles in the valley can contribute to higher PM_{10} concentrations. (Don't forget, the Sand Dunes are down there!)

- To determine a more 'pure' relationship between health and PM_{10}, the model also accounted for temporal trends as well as other environmental factors such as meteorology. Both seasonal and long-term time trends were accounted for flexibly in the model by including spline terms.

Clinical trials

Example 11: One trial involved giving aspirin-allergic subjects an aspirin challenge, with eNO measurements coming just before the challenge, 1 day post, and 6 months post.

Example 12: Another FDA-funded trial involved a crossover design in which a dose-response curve was estimated based on subjects that took multiple doses of a drug designed to reduce eNO.

Designed experiment with an observational flavor

Example 12b: Measurements are being taken on kids in 3 consecutive Aprils (2017, 18 and 19), to help answer how school start times affect performance and health outcomes for K-12 kids (April 2017 is pre, others post). This research is likely to affect actual school start times!

Basic science experiments

- Help give a better understanding of what drives or modifies certain diseases at the cellular level.

- Often, blood or biopsy samples are taken from humans or animals in order to carry out the experiments. Sometimes cell cultures are extracted from a sample and placed into separate wells so that different treatments can be applied, after which cell counts or other measures are made to determine how the treatments affect cellular chemistry and activity. Measures may also be taken over time.

- Not hampered by issues that affect observational studies (easier to get data as planned).

- Experiments can often only be performed with a small number of experimental units.

Example 13: *Nuclear factor-Kappa B data,* based on article by Bai et al., to appear in *PLoS One*.

- Macrophage samples from a human subject were put into four separate cell cultures, each one incubated with one of 4 treatments (combinations of BAY – Y/N; TB: Y/N), then observed over time; this was repeated for a number of subjects. Thus, each subject sample had 16 measurements, over time and treatment ('doubly repeated measures' – we will examine in more detail later). Macrophages originate from WBCs; they are "big eaters" of cellular debris and pathogens.

- Based on outcomes being measured, one of the major findings of the experiment was that BAY treatment (an inhibitor of NFκB) reduced the viability of MTB in samples treated with MTB, with greater relative differences occurring over time.

Example 14: *Complement levels and Chronic fatigue syndrome data* (Sorensen et al., 2003).

- Involved measuring complement split products (biological markers) over time.
- In this case, groups involved those with or without CFS, and thus repeated measures only involved time.
- A special covariance structure was used to model the repeated measures since measurement times were unequally spaced.

Example 15: *Myostatin data* (Taylor et al., 2001).

- The myostatin protein is an inhibitor of skeletal muscle mass.
- The experiment involved a 2×3 factorial treatment structure in a completely randomized design to determine effects of myostatin (Y/N) and Time (1, 2, 3 days) on protein levels in muscle cells.
- Muscle cells were taken from 24 mice and grown in separate tissue culture wells that each had a specific treatment (presence or absence of myostatin) and time of measurement.
- Protein degradation was observed over time, and samples treated with myostatin had greater protein degradation than those that did not. Data appear in Strand et al. (*Journal of Stat. Software*, 2004). Although this experiment involves time, it is not a longitudinal experiment in that it did not involve repeated measures.

Recent or potential work

A. COPDGene analyses (papers recently published or 'in the works')

(1) Joint model paper: How do FEV1 and FEV1/FVC jointly progress over time?

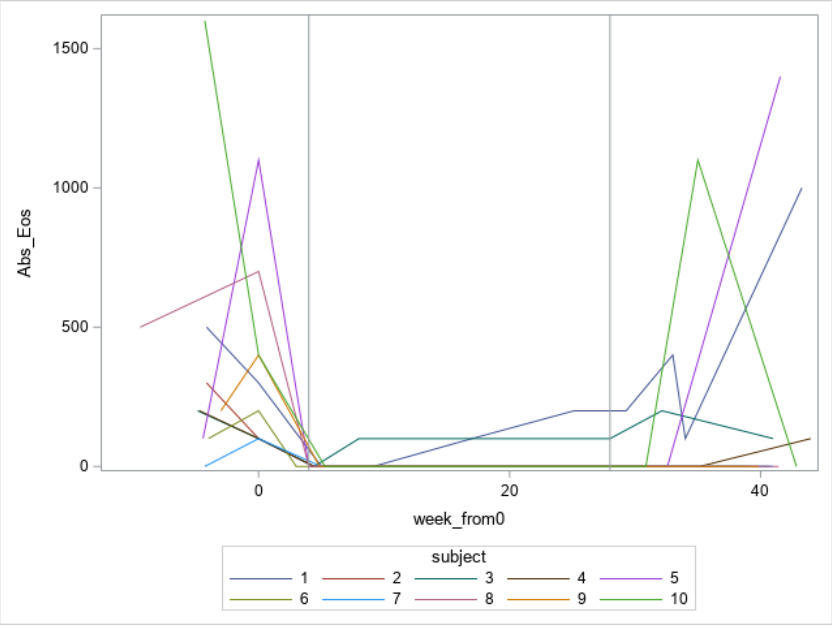
(2) Risk score paper: What is 10-year risk of mortality for subjects with COPD, based on certain characteristics?

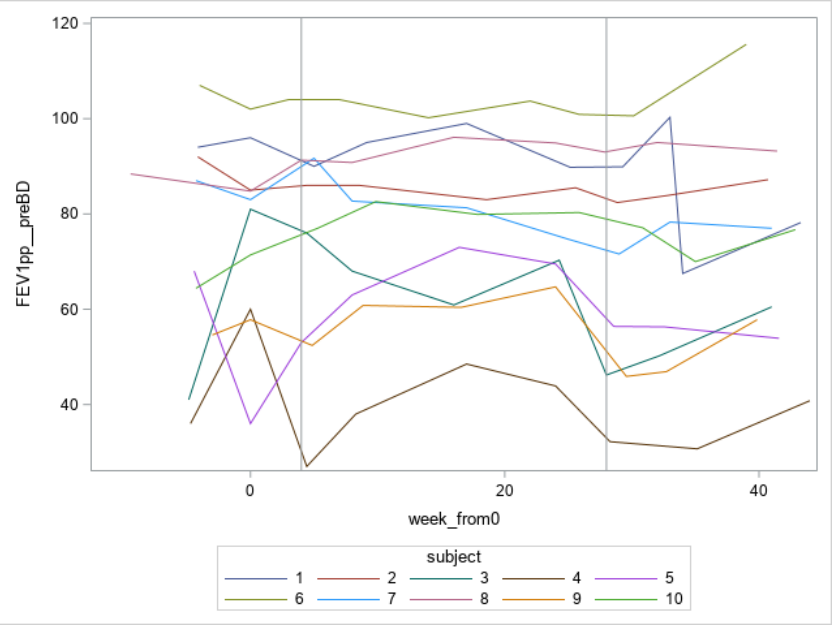
(3) How is progression of lung function modified by the number of exacerbations that you had previously?

(4) How much of the progression of COPD based on lung CT measures accounted for by changes in FEV1?

B. EGPA open label study (recent work)

How does taking a certain drug affect the need for prednisone (and other outcomes) in people with Eosinophilic granulomatosis with polyangiitis (**EGPA**; formerly Churg-Strauss; a rare disease that results in inflammation)? Open label study to see how outcomes within the treatment period compared with pre and post periods (40 weeks total); 10 subjects.





C. Another basic science experiment relevant to the flu (current)

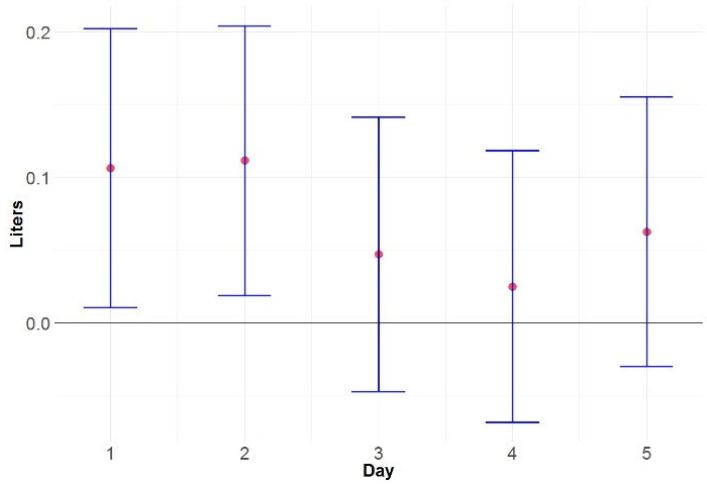
A current basic science experiment I have helped analyze suggest a new treatment for the flu that could help minimize outbreaks. In particular, a certain treatment applied to tissue cultures disrupted the spread of viruses from infected to non-infected cells.

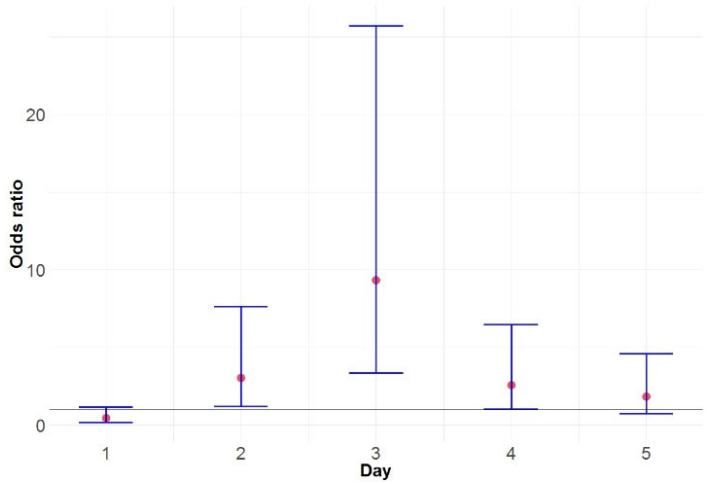
Responses were accrued for different treatment settings. Different experiments were performed ('replicates'). Within an experiment, responses were standardized to the control setting (i.e., 'response for treatment x' / 'response for control'). How do we account for the fact that responses for one experiment may generally be different than for another? (One experiment may yield higher responses, and another, lower, based on the sample(s) used.) I.e., how do we address the potential within-experiment correlation?

D. Sleep study (recently published)

How do health outcomes change for a week when subjects have shorter sleep, compared to a week when they have longer sleep (a designed experiment)? This experiment involved a crossover design; each subject had a 'long' sleep week and a 'short' sleep week. Within each of the crossover periods, there were repeated measures (each weekday). A few outcomes are shown below.

95% CI's by day, for FEV1 difference (Long–short period; left) and odds ratio for asthma symptoms (Short to Long, right. Results show effects of sleep occur within 1-3 days, with effects diminishing by later in the week.





E. Potential research

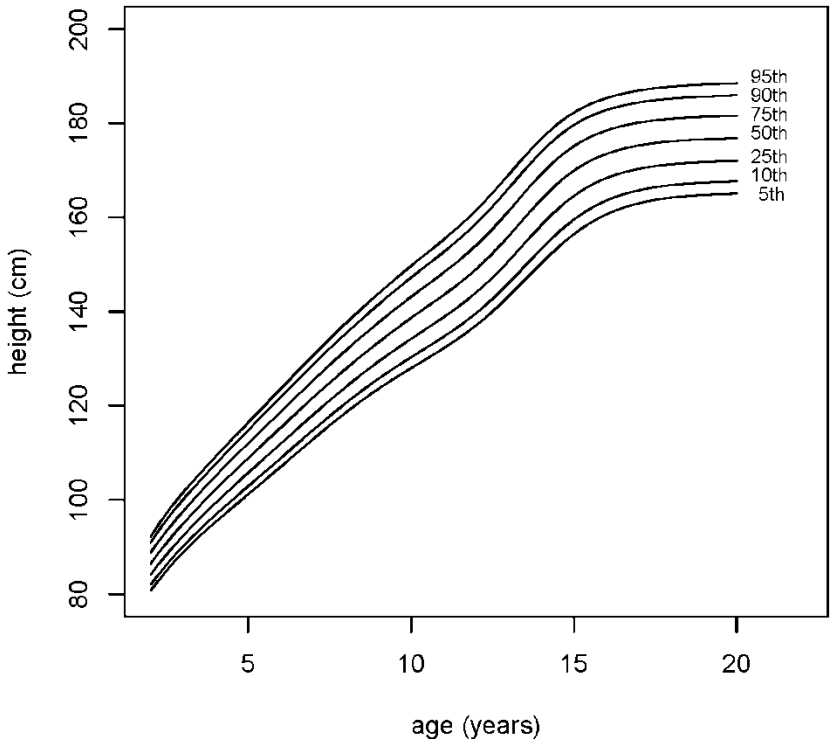
(1) What is the T-cell and antibody response for subjects with different diseases that are vaccinated against the COVID-19 virus? (Research motivated by the fact that breakthrough cases occur, which may be particularly problematic for those with diseases such as asthma or COPD. Research involves estimating the mean T-cell count over time after vaccination.

(2) How does using a nicotine help subjects with sarcoidosis? Clinical trial involves randomly assigning a nicotine or placebo patch to subjects, and monitoring over time.

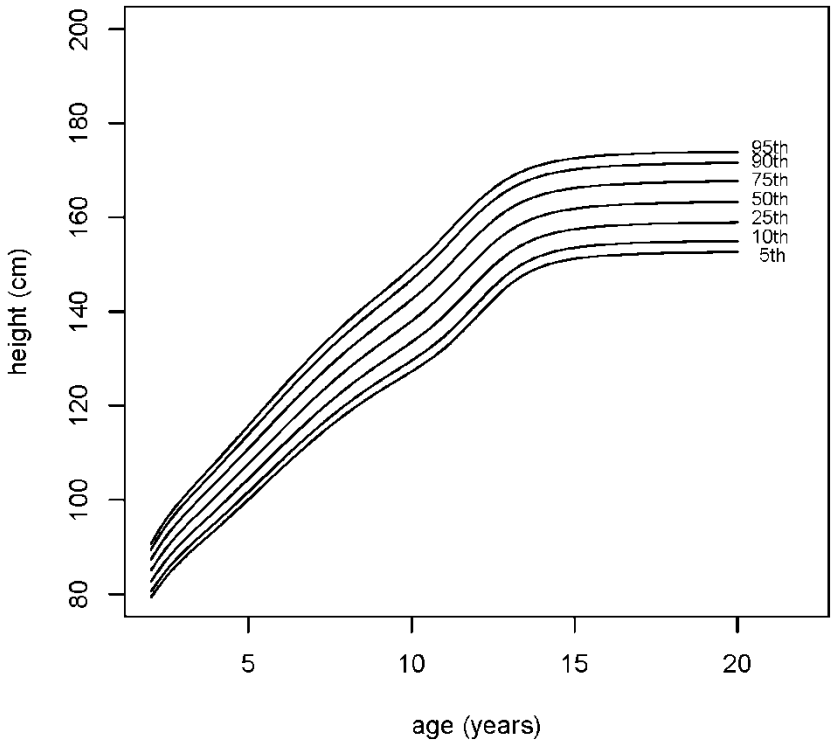
Growth curve data

Example 16: graphs for height as a function of age for boys and girls aged 2 to 20 years; constructed in R after obtaining growth data from the CDC (available at their website). For more information, please see http://www.cdc.gov/growthcharts/. These data show that girls approach their maximum height much more quickly than boys. The y-axis scales were made the same for easier comparison between graphs. Each curve is a percentile estimate as a function of age. We could create confidence bands for each percentile curve. If the curves are estimated using a lot of data, the widths of the bands should be narrow. Doctors look for dramatic changes between visits. The curves here may not be representative of all populations (e.g., differences due to race).

Boys


height (cm)
80
100
120
140
160
180
200
5
10
15
20
age (years)
95th
90th
75th
50th
25th
10th
5th

Girls


height (cm)
80
100
120
140
160
180
200
5
10
15
20
age (years)
95th
90th
75th
50th
25th
10th
5th

Formats for longitudinal data

- Covers…
  - how to set up data for ‘univariate’ versus ‘multivariate’ analysis
  - types of variables and notation for variables in longitudinal data versus cross-sectional data.

- Please review Section 4 in the Introduction chapter.

Clustered data

Example 19: After an exercise challenge performed on 20 subjects, resting heart rates are monitored at 5 minute intervals for one hour. How are data clustered?

Example 20: Families are selected to participate in a survey regarding health insurance. Each member of the family will be included in the study.

Example 21: arm length and leg length growth are measured for subjects once a year for 10 years, and then modeled with a linear mixed model.

Simple clustered/longitudinal analyses (that we've already done!)

- Experiments with pre-post measurements have 2 measurements on each subject over time. When there are only 2 measurements, the analysis simplifies when the difference is considered, as the analysis is reduced to one measurement per subject. Simple methods can then be used (e.g., paired *t*-test).

- Longitudinal models can still be beneficial here! But we'll discuss that later. For now, we consider simplified models.

- Let's take a closer look at the underlying models when we use a difference score or take the baseline-as-covariate approach.

- Change-score model:

  Y_{i1} = pre score
  Y_{i2} = post score
  $d_i = Y_{i2} - Y_{i1}$
  $d_i = \beta_0 + \beta_1 x_i + \varepsilon_i$.

- Baseline-as-covariate model:

  $Y_{i2} = \beta_0 + \beta_1 Y_{i1} + \beta_2 x_i + \varepsilon_i$.

  We allow the slope of the baseline value to be anything (based on fit).

Example for discussion: cholesterol data.

Any other type of simple clustering, with 2 responses per cluster can be analyzed similarly. (E.g., pairing by married couple, pairing by year of measurement.)

Usual assumptions for longitudinal models

- Assumption 1: Responses between subjects are independent.
  - If there are clear violations to the assumption, and data are available, then a random term could be added to deal with this non-independence.
  - For example, if a class is used for the sample, and there are several pairs of siblings in the class, a random term identifying family could be added to the model. (Lack of fit and lack of independence are related!)

- Assumption 2: There is a common covariance structure between all subjects, and the covariance parameters have the same value between subjects.

  - This assumption is usually not tested. However, to properly estimate covariance parameters, several subjects are needed (just as data for several subjects are needed to estimate a common population mean).

  - In some cases, homogeneous groups within the study may be identified (but heterogeneous between groups). With sufficient group sample sizes, group-specific covariance parameters can be put in the model and estimated.

Longitudinal designs and power – an initial glimpse

- Consider an experiment designed to compare two treatments. Two common approaches:
  - **A1**: Use independent samples (randomly assign some subjects one treatment, and some the other)
  - **A2**: Have all subjects have one treatment and then have them all take the other (e.g., use a crossover design to eliminate confounding effects related to time).
- For **A1**, we often use a 2-independent sample t-test, and for **A2**, a paired *t*-test. A study/experiment involving changes within subjects (e.g., analyzed with a paired *t*-test) is often more powerful than a study using independent samples.

- The general formula for the variance for the difference in means suggests why this may be expected (when correlations between responses within subjects are positive):

$$Var\left(\bar{Y}_1 - \bar{Y}_2\right) = Var\left(\bar{Y}_1\right) + Var\left(\bar{Y}_2\right) - 2Cov\left(\bar{Y}_1, \bar{Y}_2\right).$$

- Often there are many factors not of interest that distinguish the two independent samples, while for the paired data, the difference in responses is due more to the treatment alone and not to other factors, since we're using the same subjects.

- The same principle generalizing to multiple times and longitudinal data in general (e.g., air pollution study); subject serve as their own controls.

- But paired/longitudinal designs may not always be better. In some cases a short cross-sectional study/experiment involving many subjects may be more feasible and cost-effective.