

EPID 766: Analysis of Longitudinal Data from Epidemiologic Studies

Graduate Summer Session in Epidemiology

INSTRUCTOR: Dr. Daowen Zhang
Department of Statistics
North Carolina State University
Raleigh, NC 27695-8203
(919) 515-1933
Email: dzhang2@stat.ncsu.edu
Internet: www4.stat.ncsu.edu/~dzhang2

COURSE OBJECTIVES:

1. To understand the unique features of longitudinal studies and the methodological implications when analyzing data from such studies as opposed to a single event study, including timing and frequency of the measurements, duration of followup, strategies for sample size estimation, missing data problem, definition of outcomes, participant recruitment.
2. To appreciate the logistical and practical problems of field work associated with a longitudinal studies.
3. To understand statistical models, analyses, interpretation and software for longitudinal data.

COURSE COMPETENCIES:

1. To be able to identify the characteristics of a disease or a health outcome that warrant longitudinal investigation.
2. To be able to design a longitudinal study and to realize the advantages and disadvantages of various design alternatives based on the question of interest.
3. To be able to choose an appropriate statistical model and method for a longitudinal data set to conduct and interpret the statistical analysis.

GRADING: For those taking the course for a grade, grades will be based on the data analysis class project. Class will be divided to 4 to 5 groups depending on the class size. Each group will pick up a data set and 1) define a scientific question for a continuous and discrete outcome; 2) graph data; 3) conduct a linear mixed analysis and a GEE analysis; 4) present to the class. Students are encouraged to bring their own datasets.

TEXTBOOKS:

1. **Required:** No textbook is required. Lecture notes for this course can be downloaded from <http://www4.stat.ncsu.edu/dzhang2/epid766/index.html>. Other related materials such as SAS programs and data sets can be downloaded from this website too.

2. **Useful websit:** The website http://www.stat.ncsu.edu/~st732_info/Davidian contains the lecture notes, examples, SAS programs, ect, from my colleague Prof. Marie Davidian's course **Applied Longitudinal Data Analysis**. Take a look at this website and you may download the materials you find useful. This will be a very good source for your study and research.
3. **Recommended:** G.M. Fitzmaurice, N.M. Laird and J.H. Ware, *Applied Longitudinal Analysis*, Wiley, NJ, 2004.

TENTATIVE SCHEDULE FOR EPID 766

Date	Hour	Topic
Monday	8:30	Introduction and overview
	9:00	Review of studies: cross-sectional studies, prospective cohort studies, retrospective (case-control) studies; motivation of longitudinal studies; formulation of the scientific questions; examples; graphical descriptions; understanding patterns of variation in longitudinal data; within <i>vs.</i> between subject variation, etc.
	10:00	Break
	10:15	Features of longitudinal data; challenges in analyzing longitudinal data (issues in applying standard linear regression to longitudinal data); methods of analyzing continuous (normal) longitudinal data: two-stage, linear mixed model, generalized estimating equation (GEE).
	11:30	Data analysis projects – Assignment
Tuesday	8:30	Analyzing continuous (normal) longitudinal data using linear mixed models (Fixed effects <i>vs.</i> random effects, Interpretation of linear mixed models, Issues in choosing random effects and its variance-covariance matrix, Time-varying covariates, Estimates and significance tests)
	9:30	Break
	9:45	Linear mixed models (continued)
	10:45	Break
	11:00	Linear mixed models (continued)
Wednesday	8:30	Modeling issues: modeling raw data <i>vs.</i> modeling difference, use baseline outcome as a covariate, auto-regression type of modeling, etc.; what to do if mixed model fails: GEE as a rescue, missing data issue.
	9:30	Break
	9:45	Discussion I – HIV and CD4 examples
	10:45	Break
	11:00	Design issues: recruitment and retention of participants, selection bias, sample size estimation (its relationship with within and between subject variations).

Thursday	8:30	Discrete data: problems of applying logistic regression to longitudinal binary (binomial) data or Poisson regression to longitudinal count data – GEEs and generalized linear mixed models (GLMMs)
	9:30	Break
	9:45	GEEs and GLMMs (continued)
	10:45	Break
	11:00	GEEs and GLMMs (continued)
Friday	8:30	Discussion – Menstrual cycle: binary and continuous outcomes
	9:30	Break
	9:35	Class project presentations - 3 projects (20 minutes/project)
	10:40	Break
	10:50	Class project presentations - 2 projects
	11:30	Summary and evaluations

DATA ANALYSIS ASSIGNMENT FOR EPID 766

REQUIREMENTS

All individuals taking this course for credit are required to perform the data analysis assignment and to present their results on the last day of the class. Course grades will be based on the this presentation. Individuals who are not taking the course for credit are encouraged but not required to participate.

The class will be divided into 4-5 groups, based on the availability of data sets and interests. To improve efficiency in working in U-M computer systems, UM students are encouraged to distribute themselves among the groups.

Analysis will be preferably be conducted in SAS using Proc Mixed, Proc Genmod and Proc Nlmixed.

COMPUTER ACCESS

WEB site to download class data sets: <http://www4.stat.ncsu.edu/~dzhang2/epid766>

DATA SETS

Two data sets are available to the class in addition to the ones class participants may have brought with them. One data is from the Framingham study and the other is from a study on substance use and menstrual cycle function.

Please remember that all data sets are for use in this class only and should not be copied for personal use or for further analysis without explicit permission from the investigators.

ASSIGNMENT FOR EPID 766

- Monday 1) Define a scientific question from your data set for both a continuous and binary outcome
 2) Provide a description of the pattern of change in independent and dependent variables over time – use graphical representation of the data
- Tue & Wed 3) Run a linear mixed model analysis on the continuous outcome
- Thursday 4) Run a GEE model and GLMM model for the binary outcome
- Friday 6) Present to the class