EPID 766: Epidemiologic Research Using Healthcare Databases

COURSE INFORMATION

Class Time: Tuesdays/Thursdays 3:30-4:45pm
Location: McGavran-Greenberg Hall, Room 2308

Lead Instructor

Jennifer Lund
Department of Epidemiology
2102D McGavran-Greenberg
Phone: (919) 966-7440
Fax: (919) 966-2089
Email: Jennifer.Lund@unc.edu
Office hours: By appointment

Co-Instructor

M. Alan Brookhart, PhD
Email: abrookhart@unc.edu

Teaching Assistants

Alan Kinlaw
Tracy Kinsey
Office hours: 2304 McGavran-Greenberg: Mondays from 12:00-1:00pm
235 Rosenau: Wednesdays from 1:30-2:30pm

COURSE DESCRIPTION AND PREREQUISITES

Prerequisite, competency in data management with SAS (e.g., EPID 700, BIOS 511, or equivalent) and knowledge of basic epidemiology. The course will teach students how to use administrative healthcare utilization data for epidemiologic studies. Students will gain an understanding of how the data are generated through claims for reimbursement sent to payors from physicians, hospitals, and pharmacies. The common elements found in administrative healthcare utilization data will be described, including outpatient medical insurance claims, hospitalization information, and pharmacy claims. Students will learn about coding nomenclature, including ICD-9 CM diagnosis codes, CPT procedure codes, and various medication coding systems. Students will be given access to a sample of healthcare claims data and will learn how to use the data to identify populations of interest and conduct epidemiologic studies of the utilization and comparative effectiveness/safety of prescription drugs and healthcare services. Students will also learn about other sources of healthcare utilization data available at UNC including electronic medical records from the Carolina Data Warehouse, Medicare Parts A, B, and D data, the Surveillance, Epidemiology, and End Results program (SEER)-Medicare data, and international databases including the UK CPRD and the Danish medical registries.

DESIRED COURSE OUTCOMES

- To understand the contents and structure of typical healthcare insurance claims databases
- To be able to manipulate these data efficiently using SAS
- To be able to implement a basic comparative new user cohort study using healthcare claims data
- To develop improved research computing skills, including use of remote computing systems, knowledge of basic Linux/Unix commands, and improved SAS skills

HOMEWORK AND LABS

Students will be assigned 6 homework assignments during the course of the semester. Most of these assignments will be focused on developing a necessary aspect of an epidemiologic cohort study. Students should upload the SAS program and Word document to their individual folder within the "homework" directory on the server on the specified due date by 3:30pm. The subdirectories within each individual's folder should be named hwk1, hwk2, etc. All SAS programs should include a header (see file header.sas on Sakai for appropriate format) and should be richly commented so that the intent of the program is obvious to the instructors and TAs. Extensions on homework assigments, if needed, should be requested in advance by contacting the TAs, with the agreement on the honor code not to download or discuss the answer key until after you submit your homework to the "homework" directory. When submitting a homework after an agreed upon extension, please notify the TAs when you have uploaded it into your homework folder on the pharmacoepi2 server.

PROJECT

For the project, students will be expected to conduct a small study using the Truven Health data. Ideally, the project will be related to a student's research interest. For students without a developed research interest, the course instructor can suggest a project.

As a homework assignment, students will submit a 1-2 page proposal for their project. This project proposal should contain the following sections: 1) Background; 2) Specific Aims; 3) Methods. Some example project proposals will be circulated three weeks before the proposal is due.

The results of this research project should be submitted as a report on the final day of the class. The report should be similar to a brief scientific communication and should contain the following sections: 1) Introduction; 2) Methods; 3) Results; 4) Discussion/Conclusions. However, given that this not a fully developed research project, the introduction and discussion sections should be short (1-2 paragraphs each). The entire report should not exceed 2,500 words. The SAS code used to

generate the results should be submitted separately to the "homework" directory and will also be graded. As with the homework assignments, make sure to use good commenting.

For the lab component in weeks 11-14, groups of 2-3 students will give short 10 minute presentations on a database-related challenge and/or solution that they have encountered during the conduct of their projects. Each group will have 3-5 minutes to answer questions from the class. The purpose of these "working group" sessions are to: 1) recognize that we all face similar challenges in working with these data, 2) share tips and tricks with classmates, 3) get feedback and help on particular coding issues, and 4) seize the opportunity to hone your presentation skills.

Finally, students will present results of their final project in one of the last 4.5 days of the course. The presentations should be similar to those at the annual meeting of the International Society for Pharmacoepidemiology. Students will be allotted a 12 minutes time slot. They should plan to present for 8-10 minutes and reserve the final 2-4 minutes for questions. Students can work in pairs; however, projects done this way will be expected to be somewhat more substantial than projects done by individual students.

The project grade will be assigned based on the database "challenges" presentation (15%), final project presentation (25%), final project report (35%) and accuracy and clarity of the SAS program (25%).

GRADING

The final grade will be based on homework assignments (50%) and the final project and presentation (50%).

TOPICS TO BE ADDRESSED

- Introduction to epidemiologic research using health care data
- Orientation to pharmacoepi server environment
- US Healthcare system, types of healthcare data, and payment system
- Medical coding
- Relational databases in healthcare
- Introduction to Proc SQL in SAS
- Using outpatient and hospitalization claims to identify incidence and prevalence of diseases
- Using medical claims to estimate incidence of diseases and use of tests
- The new user design
- Treatment patterns after diagnosis: quality of care and patterns of prescribing
- Estimation of costs of care
- Estimation medication adherence and persistence

DATA PRIVACY AND DATA USE AGREEMENTS

The course is taught using proprietary data from Truven Health. Although direct patient identifiers have been removed from the data, the data contain dates of services and therefore it may be possible to identify individual patients. **For this reason, the data must remain on the server and cannot be copied onto individual computers for any reason.** Prior to being granted access to the data, students will be required to sign a contract that governs use of the data and our computing system. Deliberate violation of the terms of this agreement would have serious consequences for everyone at UNC using the Truven Health data for research (immediate termination of our data use agreement) and would expose you to potential civil and/or criminal penalties. Please talk to the course instructors or TAs if you have any questions about appropriate use of data.

COMPUTING SYSTEM

Students will be given temporary accounts that will allow them to use the research computing environment established by the Pharmacoepidemiology program area in the Department of Epidemiology and administered by the Sheps Center for Health Services Research. The students will be sharing this resource with programmers, faculty, and other graduate students using this computer for dissertation research. Students are requested to use the system only for class work. Student accounts will be closed at the end of the semester and all files saved on the computer will be deleted. If students want to keep copies of programs, they should download these to a USB drive before the last day of the semester. **Again, data cannot be taken off the system**. All use of the system is logged.

It is possible to gain access to the data after the end of the course for research purposes. Please talk to the course instructors to learn more about data access.

FINAL NOTE

This is a relatively new course that is technologically challenging. We have worked to improve the interface to our secure computing environment that will permit all of the students in the class to simultaneously access the database. However, it is possible that there will be technological/computing problems that will occur during the course of the semester. If problems arise, for example, if you have problems accessing data or running programs, please let us know as soon as possible. We will work to resolve the problems or develop work-around solutions. In the event of problems with the data or computing system that cannot be fixed, we will modify the expectations and/or due dates of assignments.