Using Administrative Data for Health Services Research (M19-5254)

Course Instructor:
Anne Mobley Butler, PhD
Assistant Professor
Division of Infectious Diseases
Division of Public Health Sciences
email: anne.butler@wustl.edu

Teaching Assistants:
John Sahrmann, MS    Katelin Nickel, MPH
Senior Statistical Data Analyst    Senior Programmer Analyst
Division of Infectious Diseases    Division of Infectious Diseases
email: jsahrmann@wustl.edu    email: katelin.nickel@wustl.edu

Class Time: Thursday 10:00AM – 1:00PM (Spring 1 and 2)

Class Location: 2nd floor, Taylor Avenue Building (TAB)

PREREQUISITES
- Basic background in epidemiology and biostatistics (M19-501 Introductory Clinical Epidemiology and M19-511 Introductory Biostatistics for Clinical Research required perquisites required, or equivalents).
- Competency in data management with programming language (e.g., R, SAS, Stata).

COURSE DESCRIPTION & OBJECTIVES
The objective of this advanced graduate course is to prepare students to understand and use large administrative healthcare databases to perform epidemiologic / health services research on the utilization and comparative effectiveness of healthcare services. Lectures will cover the translation of clinical care into healthcare utilization data, review various types of national and state administrative databases, describe methods for administrative database research, and emphasize key issues related to data security and confidentiality. We will consider the strengths and limitations of observational studies using large databases to augment evidence from randomized clinical trials. Students will develop a research proposal in their own area of interest and complete a small research project that uses administrative data; this project will require programming with R statistical software to identify study populations and perform analyses; interpreting results; and presenting to the class. Students will further gain experience with healthcare database research by reviewing journal articles weekly.
DESIRED COMPETENCIES

• Appreciate the origin, contents, structure, and completeness of typical healthcare databases.
• Communicate the strengths and limitations of observational studies using administrative data.
• Identify data sources for administrative database research.
• Understand the validity of data elements in administrative data and methods to improve accuracy.
• Implement an administrative data research study including defining and identifying the eligible study population, creating an analytic dataset, performing common data management steps, and performing basic analyses.
• Improve proficiency with R statistical software.
• Draw appropriate inferences from analyses of administrative data, with an understanding of the strengths and limitations of various research methods and databases.

EXPECTATIONS

• Attendance in lectures, which will be as interactive as possible.
• Required readings should be read before class each week; they will allow the students to understand the topics in greater depth and enable more active participation in class. Suggested readings provide more detailed information about particular subject areas selected as resources or guidance materials for a specific database or topic of interest.
• Students are required to use R statistical software and write basic code.
• There will be a series of short assignments over the course of the semester with a project due at the end of the course.
• Final grades will be based on the assignments below.

REQUIRED SOFTWARE
R Studio (freely available at https://www.rstudio.com/)

RECOMMENDED TEXT
ICD-10-CM book (old) – strongly recommended. Available through Amazon for < $10

Online resources for R:
• UCLA Website for Statistical Analysis: https://stats.idre.ucla.edu/r/
• R for Data Science: https://r4ds.had.co.nz/index.html
• An Introduction to R from the R Core Team: https://cran.r-project.org/doc/manuals/R-intro.pdf
• Overview of R: https://www.statmethods.net/index.html
• List of free e-books for R: https://r-dir.com/learn/e-books.html

ASSIGNMENTS
The final grade will be based on quizzes (5%), programming exercises (30%), journal article review (5%), project proposal (10%), and final project and oral presentation (50%).

Quizzes
Students will complete a weekly short quiz designed to reinforce important topics. The quizzes will be available on Canvas.

Formal Review of Journal Article
Students will review a journal article related to administrative data research (e.g., comparative effectiveness research or health services research). Students will present the information in class (~10 minutes). Guidelines for review of article will be provided in class.

Programming Exercises
Students will be asked to complete a number of programming exercises to demonstrate mastery of R programming. The exercises must be turned in prior to class.

Project Proposal
Students will submit a 1- to 2-paragraph description of their proposed research project (see description below), including the primary research question, proposed study population, and database. Students will be provided timely feedback so that they can take comments into account before finalizing their project.

Final Project and Oral Presentation
Students will design and implement a research study, ultimately preparing a conference abstract and delivering an oral presentation. The students should choose a topic of sufficient personal interest, with the hope that the student can continue developing the study and write a full manuscript after completion of the course. Using HCUP data, students will develop and examine a research question about healthcare utilization. Students will identify a study cohort based on demographic and/or clinical criteria, select relevant data elements from the database, and perform statistical analyses to address the study question.
HCUP DATA PRIVACY & DATA USE AGREEMENTS

The course is taught using HCUP State Databases from the Healthcare Cost and Utilization Project (HCUP), which are maintained by the Agency for Healthcare Research and Quality (AHRQ). Although direct patient identifiers have been removed from the data, the data contain information that may make it 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.** All users of HCUP data must complete the HCUP Data Use Agreement (DUA) Training Course and sign an HCUP DUA before accessing the data. Please complete the ~15 minute training available at [https://hcup-us.ahrq.gov/tech_assist/dua.jsp](https://hcup-us.ahrq.gov/tech_assist/dua.jsp) Deliberate violation of the terms of this agreement will have serious consequences for researchers at Wash U using the HCUP data for research and will expose you to potential civil and/or criminal penalties. Please talk to the course instructor if you have any questions about appropriate use of data.

**COMPUTING SYSTEM**

Students will receive temporary access to the research computing environment. The students will be sharing this resource with other faculty and research teams, therefore students are requested to use the server for class work only. All use of the system is logged. 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 statistical software programs, they should download these programs before the last day of the semester. **However, as noted above, data CANNOT be removed from the server.**

**CHANGES TO SERVER**

Although this course has been taught for years, this spring we are changing the server access. These changes will lower costs for students, which were previously covered by a student course fee. We have worked hard to create and provide a new secure computing environment that will permit all of the students in the class to simultaneously access the database. However, we recognize that this is a technologically challenging course and it is possible that there will be technological/computing problems that will occur during the semester. If problems arise (e.g., problems accessing data or running programs), please let the instructor and TAs 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 expectations and/or due dates of assignments.

**MENTAL HEALTH**

Mental Health Services are available for full-time students enrolled on the Medical School campus. Students can self-refer to a counselor by making an appointment with Dr. Karen Winters through Student Health Services (SHS), telephone: 314-362-3523, and follow the prompts. There are also contractual mental health service providers who are available off-campus. More information regarding this coverage and a list of participating providers are accessible via [https://wusmhealth.wustl.edu/](https://wusmhealth.wustl.edu/) and then clicking on **Students** and scrolling down to **Mental Health Information** [https://wusmhealth.wustl.edu/students/mental-health-information/](https://wusmhealth.wustl.edu/students/mental-health-information/). Please do not hesitate to reach out to Dr. Winters, 314-362-3523, or to any of our off-campus providers [https://wusmhealth.wustl.edu/](https://wusmhealth.wustl.edu/).
CLASS SCHEDULE

LECTURE 1 (January 20th)
Introduction to Administrative Data Research (Anne Mobley Butler)
R Lab (Sarah Humble & Nicole Ackerman)

Required Readings:


- R for Data Science available at URL: https://r4ds.had.co.nz/index.html

Suggested Readings:


Student Presentation:

- None

LECTURE 2 (January 27th)
Discharge / Billing Data: Healthcare Cost and Utilization Project (HCUP) (Margaret Olsen)
R Lab – Using R to Analyze HCUP Data (John Sahrmann)

Required Readings:

• Andrews RM. Statewide hospital discharge data: collection, use, limitations, and improvements. *Health Serv Res* 2015;50(S1):1273-99.

• Lucyk K, Tang K, Quan H. Barriers to data quality resulting from the process of coding health information to administrative data: a qualitative study. *BMC Health Serv Res* 2017;17:766.


• CMS Form 1500

• UB04 Form

**Suggested Readings:**


**Student Presentation:**

• None

---

**LECTURE 3 (February 3rd)**

Research Studies Using HCUP Data (Karen Joynt Maddox)

R Lab (John Sahrmann)

**Homework #1 due before class**

**Required Readings:**

• To be determined

**Suggested Readings:**


• STROBE Statement—Checklist of items that should be included in reports of cohort studies

**Student Presentation:**

• Talutis SD, Chen Q, Wang N, Rosen AK. Comparison of Risk-Standardized Readmission Rates of Surgical Patients at Safety-Net and Non-Safety-Net Hospitals Using Agency for

LECTURE 4 (February 10th)
Validation studies (Anne Mobley Butler)
R Lab (John Sahrmann)
**Homework #2 due before class

Required Readings:

Suggested Readings:

Student Presentation:

LECTURE 5 (February 17th)
Claims Data: CMS Medicare Data (Anne Mobley Butler)
R Lab (John Sahrmann)
**Project proposal outline due before class

Required Readings:


Suggested Readings:


Student Presentation:


LECTURE 6 (February 24th)

Claims Data: Private Insurer Data (Anne Mobley Butler)
R Lab (John Sahrmann)

**Project Table 1 due before class

Required Readings:


Suggested Readings:

Student Presentation:

**LECTURE 7 (March 3rd)**

Veterans Administration Data (Benjamin [Charlie] Bowe)

R Lab (John Sahrmann)

**Homework #3 due before class**

Required Readings:

Suggested Readings:

Student Presentation:
LECTURE 8 (March 10th)
Medicaid data (Derek Brown)
R Lab (John Sahrmann)
**Homework #4 due before class

Required Readings:

Student Presentation:

LECTURE 9 (March 17th)
Propensity scores (Anne Mobley Butler)
R Lab (John Sahrmann)
**Homework #5 due before class

Required Readings:

Suggested Readings:


Student Presentation:
• To be determined

**LECTURE 10 (March 24th)**
Hierarchical models (Margaret Olsen)
R Lab (John Sahrmann)

Required Readings:


Student Presentation:

**LECTURE 11 (March 31st)**
Survey and other data – Medicare Current Beneficiary Survey (Kenton Johnston)
R Lab (John Sahrmann)

Required Readings:

Suggested Readings:
• Johnston KJ, Hockenberry JM. Are two heads better than one or do too many cooks spoil the broth? Health Serv Res 2016;51:2176-205.

Student Presentation:

LECTURE 12 (April 7th)
Thinking outside the box – additional data for enrichment (Anne Mobley Butler)
R Lab – work on final project

Required Readings:

Suggested Readings:

Student Presentation:

LECTURE 13 (April 14th)
Overview of Pharmacoepidemiology (Anne Mobley Butler)
R Lab – work on final project

Required Readings:

Suggested Readings:

Student Presentation:

LECTURE 14 (April 21st)
R Lab – work on final project
LECTURE 15 (April 28th)
**Student presentations of final project

LECTURE 16 (May 5th)
**Student presentations of final project
**Written final project due May 5th by midnight