Like    Share   12 people like this. Sign Up to see what your friends like.

# Analysis of Big Healthcare Databases

*hosted by the*

## Kentucky Chapter of the American Statistical Association

## Instructor: Rebecca Hubbard, University of Pennsylvania

## Saturday, October 17, 2020
## 9:00 am to 4:00 pm ET

**In light of the ongoing COVID pandemic, this short course will be presented as a live webinar.  You will receive a unique link approximately two days prior to the meeting.**

The widespread adoption of electronic health records (EHR) as a means of documenting medical care has created a vast resource for the study of health conditions, interventions, and outcomes in routine clinical practice. Conducting research using healthcare databases, including EHR and administrative claims data, facilitates the efficient creation of large research databases, execution of pragmatic clinical trials, and study of rare diseases. The recent release of the Framework for FDA's Real-World Evidence Program has further spurred interest in the use of healthcare data to improve the efficiency and generalizability of pre- and post-marketing studies of medical products. Despite the potential benefits, there are many challenges for research conducted using healthcare data. To make valid inference, statisticians must be aware of data generation, capture, and availability issues and utilize appropriate study designs and statistical analysis methods to account for these issues. In this course, we will discuss topics related to the design and analysis of research studies using big healthcare databases. We will cover issues related to the structure and quality of the data, including data types and methods for extracting variables of interest; sources of missing data; error in covariates and outcomes extracted from EHR and claims data; and data capture considerations such as informative visit processes and medical records coding procedures. In the second half of the course, we will discuss statistical approaches to address some of the challenges and unique features of healthcare databases, including missing data and error in automated algorithm-derived covariates and outcomes. R code will be provided for implementation of the presented methods, and hands-on exercises will be used to compare results of alternative approaches. The overarching objective of this course is to provide participants with an introduction to the structure and content of healthcare databases and equip them with a set of appropriate tools to investigate and analyze this rich data resource.

**About the instructor:** Rebecca Hubbard is an Associate Professor of Biostatistics in the Department of Biostatistics, Epidemiology and Informatics at the University of Pennsylvania. Her methodological research emphasizes development of statistical tools to support valid inference for EHR-based analyses, accounting for complex data availability and data quality issues, and has been applied across a variety of domains including studies of cancer epidemiology, aging and dementia, and pharmacoepidemiology. She has experience conducting analyses using data from a number of large healthcare databases including Medicare, PCORnet, Kaiser Permanente, Flatiron Health, and Optum. Results of this work have been published in over 100 peer-reviewed papers in the statistical and medical literature. She has taught short courses at venues including ENAR, the FDA, and the Deming Conference on Applied Statistics for over 10 years.

**Preliminary schedule:**

- 9:00 – 9:15  Introduction
- 9:15 – 10:00  Overview of the structure of EHR data
- 10:00 – 10:45  Extracting data elements from the EHR
- 10:45 – 11:00  *Break*
- 11:00 – 11:30  Tutorial 1
- 11:30 – 12:15  Missing data issues
- 12:15 – 1:15  *Lunch*
- 1:15 – 2:30  Correcting for bias due to EHR data errors
- 2:30 – 3:00  Tutorial 2
- 3:00 – 3:15  *Break*
- 3:15 – 4:00  FDA's Framework for Real World Evidence and Wrap-Up