UNIVERSITY OF DELAWARE

CENTER FOR BIOINFORMATICS & COMPUTATIONAL BIOLOGY

CBCB SEMINAR
2/24/2020

3:30 PM
DELAWARE BIOTECHNOLOGY INSTITUTE
Room 102

bioinformatics.udel.edu

BIOINFORMATICS SEMINAR

REBECCA HUBBARD

Associate Professor, Biostatistics
UNIVERSITY OF PENNSYLVANIA

AN INTRODUCTION TO EXTRACTION AND ANALYSIS OF EHR-DERIVED PHENOTYPES

A first step in most analyses conducted using electronic health records (EHR) data is characterization of patient phenotypes. However, EHR-based phenotyping is hampered by complex missing data patterns and heterogeneity across patients and healthcare systems in the amount and type of data that is available. As a result, not only are EHR-derived phenotypes expected to be imperfect, but they often feature exposure-dependent differential misclassification, which can bias results towards or away from the null. In this talk I will first review approaches to EHR-based phenotyping, highlighting how missing data affect phenotype estimation. I will then review some results on the implications of using EHR-derived phenotypes with differential misclassification for bias and type I error of subsequent association studies using these phenotypes as outcomes. Finally, I will present an approach to correcting for phenotyping error that does not require knowledge of sensitivity and specificity of the phenotype. The overall goal of this presentation is to improve awareness of phenotyping error, its implications for analyses, and available options for valid analysis of EHR-derived phenotypes.

BIOGRAPHY

Dr. Hubbard is an Associate Professor of Biostatistics in the Department of Biostatistics, Epidemiology and Informatics, a Senior Fellow of the Institute for Biomedical Informatics, and a Senior Scholar in the Center for Clinical Epidemiology and Biostatistics at the University of Pennsylvania. She completed an MSc in Epidemiology at the University of Edinburgh and an MSc in Applied Statistics at Oxford University before obtaining her PhD in Biostatistics at the University of Washington. Her research focuses on development and application of statistical methodology for studies using data from electronic health records (EHR) and medical claims data. Dr. Hubbard's research emphasizes statistical methods to overcome the messiness and complexity of real-world data by addressing challenges including informative observation schemes, phenotyping error, and error and missingness in confounders. Her methods have been applied to support the advancement of a broad range of research areas including health services research, cancer epidemiology, aging and dementia, and pharmacoepidemiology. Results of this work have been published in over 100 peer-reviewed papers in the statistical and medical literature. She has taught short courses for the American Statistical Association's Council of Chapters, the Deming Conference on Applied Statistics, and the Summer Institutes in Statistical Genetics and Statistics for Clinical Research at the University of Washington for over 10 years.

College of Engineering
CENTER FOR BIOINFORMATICS & COMPUTATIONAL BIOLOGY