JESSE GRONSBELL, University of Toronto
Leveraging electronic health records for data science

The adoption of electronic health records (EHRs) has generated massive amounts of routinely collected medical data with potential to improve our understanding of healthcare delivery and disease processes. However, the analysis of EHR data remains both practically and methodologically challenging as it is recorded as a byproduct of clinical care and billing, and not for research purposes. For example, outcome information, such as presence of a disease or treatment response, is often missing or poorly annotated in patient records, which brings challenges to statistical learning and inference. In this talk, I will discuss predictive modeling in settings with an extremely limited amount of outcome information and demonstrate the advantages of semi-supervised learning methods that incorporate large volumes of unlabeled data into model estimation and evaluation.