



THE UNIVERSITY OF CHICAGO

Department of Statistics

MASTER'S THESIS PRESENTATION

QIAN CHEN

Department of Statistics
The University of Chicago

Association Studies with Imperfect Phenotypes:
Evaluation of Methods

THURSDAY, February 18, 2016, at 8:30 AM
Eckhart 110, 5734 S. University Avenue

ABSTRACT

Many algorithms have been developed recently to study phenotypes by combining information from multiple structured and unstructured data extracted from electronic medical records (EMRs). These algorithms have been shown to be quite accurate, but they typically do not provide perfect classification due to the difficulty in inferring meaning from the text. Some algorithms can produce for each patient a probability that the patient gets the disease, which can be thresholded to define case-control status. And this estimated case-control status has been used to replicate known genetic associations in EMR-based studies. However, using the estimated disease status in place of true disease status usually results in diminishing test power and bias odds ratio estimates. We implemented a novel algorithm proposed by Sinnott et al (2014) to directly model the probability of being a case, and compared with the method using estimated case-control status to demonstrate how the novel method improves test power and reduces bias in simulation studies.

For information about building access for persons with disabilities, please contact Laura Rigazzi at 773.702-0541 or send an email to lrigazzi@galton.uchicago.edu. If you wish to subscribe to our email list, please visit the following web site: <https://lists.uchicago.edu/web/arc/statseminars>.