

# BIOSTATISTICS SEMINAR



**Jinbo Chen, PhD,**  
Professor of  
Biostatistics,  
Department of  
Biostatistics,  
Epidemiology and  
Informatics,  
Perelman School of  
Medicine, University of  
Pennsylvania

## **Addressing Phenotyping Inaccuracy in the Analysis of EHR-based Case-Control Studies**

When conducting case-control studies using electronic health records (EHRs), the true condition of a patient must be derived using information available in the EHR. Separate rules are usually created for identifying cases and controls, with the stringency of rules set to ensure high positive predictive values (PPVs). While this practice is generally effective for identifying controls, such PPV-guided selection of cases can lead to biased estimates and small sample sizes in downstream analysis. To address shortcomings of this method, we present an estimating equation (EE) approach that achieves balance between the accuracy and sample size of the case pool. Our approach relies on training a phenotyping model in a validation subset from the case pool to distinguish between case and non-case status using low- or high-dimensional predictors. We assess the performance of our method when different methods are utilized for creating the phenotyping model and compare it to the “high-PPV” method. Our method can effectively correct bias due to the presence of non-cases in the case pool. Therefore it allows for flexible identification of cases from EHRs, promising higher study power and wider generalizability.

**Thursday April 1, 2021**

**3:30 pm - 4:30 pm**

**Zoom meeting: Please also find a link in the email invite, with the password.**  
<https://uncsph.zoom.us/j/95116832073?pwd=TnV4QUtGLzMwaFRBTlRsd2xmTjVMQT09>



**UNC**  
GILLINGS SCHOOL OF  
GLOBAL PUBLIC HEALTH