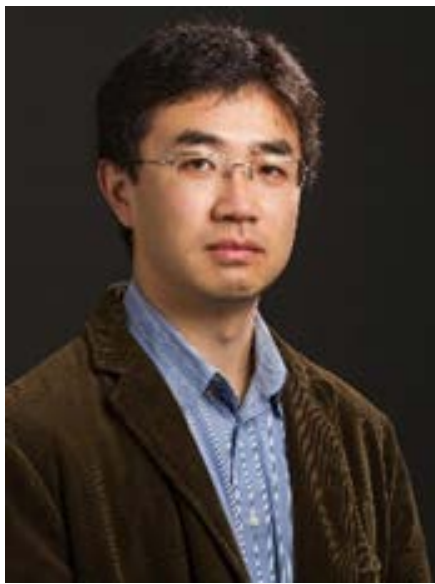


# BIOSTATISTICS SEMINAR



**Shuangge Ma**  
PhD, Professor of  
Biostatistics,  
Yale School of  
Public Health

## **Pan-disease clustering analysis of the trend of period prevalence**

For all diseases, prevalence has been carefully studied. In the “classic” paradigm, the prevalence of different diseases has usually been studied separately. Accumulating evidences have shown that diseases can be “correlated”. In this study, we take advantage of the uniquely valuable Taiwan National Health Insurance Research Database (NHIRD), and conduct a pan-disease analysis of period prevalence trend. The goal is to identify clusters within which diseases share similar period prevalence trends. For this purpose, a novel penalization pursuit approach is developed. In data analysis, the period prevalence values are computed using records on close to 1 million subjects and 14 years of observation. For 405 diseases, 35 nontrivial clusters (with sizes larger than one) and 27 trivial clusters (with sizes one) are identified. A closer examination suggests that the clustering results have sound interpretations.

**Thursday April 11, 2019**

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

**Blue Cross and Blue Shield of North Carolina Foundation Auditorium**



**UNC**  
GILLINGS SCHOOL OF  
GLOBAL PUBLIC HEALTH