



JOHNS HOPKINS  
BLOOMBERG  
SCHOOL of PUBLIC HEALTH

*Department of Biostatistics*

## BIOSTATISTICS SEMINAR,

### **Bridging the Gap Between Noisy Healthcare Data and Knowledge: Causality and Portability**

**Xu Shi,**

**Department of Biostatistics, Harvard TH Chan School of Public Health**

#### **Abstract:**

Routinely collected healthcare data present numerous opportunities for biomedical research but also come with unique challenges. For example, critical issues such as data quality, unmeasured and mismeasured confounding, high-dimensional covariates, and patient privacy concerns naturally arise. In this talk, I present tailored causal inference methods and automated data quality control pipeline that aim to overcome these challenges and make the transition from data to knowledge. I detail the challenge of inconsistent “languages” used by different healthcare systems and coding systems. In particular, different healthcare providers may use alternative medical codes to record the same diagnosis or procedure, limiting the transportability of phenotyping algorithms and statistical models across healthcare systems. I formulate the idea of medical code translation into a statistical problem of inferring a mapping between two sets of multivariate, unit-length vectors learned from two healthcare systems, respectively. The statistical problem is particularly interesting because the training data is corrupted by a fraction of mismatch in the response-predictor pairs, whereas classical regression analysis tacitly assumes that the response and predictor are correctly linked. I propose a novel method for mapping recovery and establish theoretical guarantees for estimation and model selection consistency.

Johns Hopkins Bloomberg School of Public Health, Department of Biostatistics  
Monday, April 22, 2019, 12:15:1-15, Room W2008 (Refreshments 12:00pm)

For disability access information or listening devices, please contact the Office of Support Services at 410-955-1197 or on the web [www.jhsph.edu/SupportServices](http://www.jhsph.edu/SupportServices), EO/AA

**Dept. of Biostatistics, 615 N. Wolfe Street, Suite E3527 Baltimore, MD 21205**