

Design and Analysis Strategies with “Secondary” Use Data

**Ran Tao, PhD**

**Assistant Professor**

**Department of Biostatistics**

**Vanderbilt University Medical Center**

**12:00 Noon Eastern time**, **Tuesday, April 4th**

**47 College Street, Room 106 A**

**ABSTRACT**

The growing availability of observational databases like electronic health records (EHR) provides unprecedented opportunities for secondary use of such data in biomedical research. However, these data can be error-prone and need to be validated before use. It is usually unrealistic to validate the whole database due to resource constraints. A cost-effective alternative is to implement a two-phase design that validates a subset of patient records that are enriched for information about the research question of interest. In this talk, I will discuss proper statistical approaches to analyze such two-phase studies, which can efficiently use the information in the unvalidated data in Phase I and address the potential biased validation sample selection in Phase II. I will demonstrate the advantages of the proposed methods over existing ones through extensive simulations and an application to an ongoing HIV observational study.