

**University of California, Irvine  
Statistics Seminar**

***A Novel Statistical Modeling Framework to Estimate Clinical  
Outcome Risks Using Electronic Health Records with Only  
Aggregate Patient Level Data***

**Baolin Wu  
Professor of Biostatistics  
UC Irvine**

**4 p.m., Thursday, November 30, 2023  
6011 Donald Bren Hall**

When conducting analysis of electronic health records (EHR), oftentimes the data utilized is patient level data which readily allows for statistical analyses that properly adjust for potential confounding factors or other important characteristic differences to accurately estimate the effect sizes of various risk factors. However, in cases where only summary data (aggregate patient level data such as sample mean and variance for a given patient group) is provided, it becomes challenging to account for risk factors to estimate adjusted exposure effects as individual level data is not available.

Motivated by a collaboration project here at UCI that studied the impact of COVID on various clinical practices utilizing the hospital data from the Vizient clinical database, we propose novel statistical models that take aggregate patient level data to appropriately estimate exposure effects. Specifically, we study methods that estimate the impact of risk factors on disease incidence rate and exposure effects on quantitative outcomes. We conduct extensive simulation studies to investigate the operating characteristics of the proposed methods and demonstrate their favorable performance via simulations and applications to real world aggregate data.

Our results showcase the importance of proper statistical models to unbiasedly estimate the impact of COVID on patients.