

## **Utilize external big data to improved two-stage design method on zero-inflated model**

林惠文<sup>1</sup>

<sup>1</sup> 東吳大學

### **Abstract**

Analysis of big data using population-based medical and epidemiological studies is growing in popularity. However, health care databases or administrative data often lack personal information, and these missing variables are often important factors for studies. In addition, these databases are often involving random events with excessive zero count data. For these reasons, we propose a two-stage calibration zero-inflated method to handle excessive zero count data and missing confounding factors, and utilize external large data such as health care databases and administrative data to enhance testing power. Our method does not require model assumptions to still obtain robust estimates. In the simulation study, our method is unbiased and can effectively improve the test power by 30%. Finally, we illustrate an application of our method to explore the risk of Outpatient visit by using the National Health Interview Survey database.

Keyword: missing confounding, zero-inflated mode, two-stage model