

**Department of Epidemiology and Biostatistics
Biostatistics Seminar**

Thursday, November 14, 2013
12:00pm-1:00pm -- WG73
Bring your own lunch
11:45am - Water and light refreshments

“Approaches to Disclosure Limitation”

Daniell Toth¹

Bureau of Labor Statistics

Abstract: Statistical agencies often collect sensitive data for release to the public. In order to protect confidentially, they usually must use one or more disclosure limitation methods before releasing the data to the public. In this talk, we briefly describe several different methods currently used to protect sensitive data and some of the methodology used to evaluate them. We focus on the case where the intent of data collection is to provide information only at certain aggregated levels in the tabular form. A new disclosure limitation method is proposed that focuses on replacing the full set of micro-data with synthetic data for use in publicly released tables. This synthetic dataset is obtained by replacing each unit's values with a weighted-average of values from the surrounding area. This is done by repeatedly sampling from the set of original micro-data. We will show that, under proper conditions, the synthetic dataset is produced in a way to give approximately unbiased estimates for an aggregate cell as the number of units in the cell increase. The method is applied to the U.S. Bureau of Labor Statistics' Quarterly Census of Employment and Wages data, which is released to the public quarterly in tabular form, aggregated across varying scales of time, area, and industry. The application of disclosure limitation approaches to the biomedical data will be open for discussion.

KEY WORDS: cell suppression; contingency tables; synthetic data; confidentiality; sensitive data; nearest neighbor.

Notes from Seminar Organizer (Sun): All please think about what we can learn from this talk and work from the BLS, at the interface of Open Science and HIPPA Patient Privacy Act.

¹ Daniell Toth is Senior Research Mathematical Statistician, Office of Survey Methods Research, Bureau of Labor Statistics, Suite 1950, Washington, DC 20212 (email: toth.daniell@bls.gov)