



JOHNS HOPKINS  
BLOOMBERG  
SCHOOL of PUBLIC HEALTH

*Department of Biostatistics*

## **BIostatISTICS SEMINAR,**

### **Fitting Nearly-true Models to Subsampled Data**

**Thomas Lumley  
Professor of Biostatistics  
University of Auckland**

**Abstract:**

One problem with electronic health record data is that the quality is not always above average: validation can be useful. With Pam Shaw at Penn and Bryan Shepherd at Vanderbilt I have been looking at validation sampling as an approach to compensating for measurement error. It's possible to use either weighted estimators or semiparametric model-based estimators, and potentially to compare them. Comparisons in the statistical literature usually assume that the parametric or semiparametric model being fitted is true. Occasionally, there are comparisons when the model is obviously wrong. I have been investigating what happens when the model is nearly true, in the sense that you can't reliably find anything wrong with it using the available data; mathematically, this is represented by contiguous sequences of misspecified distributions. It turns out that the precision advantage of the efficient estimator over weighted estimators does not hold up for nearly-true models. This finding probably has important implications, but it's not clear what they are.

Johns Hopkins Bloomberg School of Public Health, Department of Biostatistics  
Thursday, March 28, 2019, 12:15:1-15, Room W5030 (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**