UNC
GILLINGS SCHOOL OF
GLOBAL PUBLIC HEALTH

James E. Grizzle Distinguished Alumni Award Recipient Lecture

Biostatistics Seminar

Qingxia (Cindy) Chen, Ph.D.
Associate Professor
Department of Biostatistics
Vanderbilt University School of Medicine

Too Many Covariates and Too Few Cases?

Prior research indicates that 10-15 cases or controls, whichever fewer, are required per parameter to reliably estimate regression coefficients in multivariable logistic regression models. This condition may be difficult to meet even in a well-designed study when the number of potential confounders is large, the outcome is rare, and/or interactions are of interest. Various propensity score approaches have been implemented when the exposure is binary. Recent work on shrinkage approaches like lasso were motivated by the critical need to develop methods for the p>>n situation, where p is the number of parameters and n is the sample size. Those methods, however, have been less frequently used when p is O(n), and in this situation, there is no guidance on choosing among regular logistic regression models, propensity score methods, and shrinkage approaches. To fill this gap, we conducted extensive simulations mimicking our motivating clinical data, estimating vaccine effectiveness for preventing influenza hospitalizations in the 2011-2012 influenza season. Penalized logistic regression models that penalize all but the coefficient of the exposure may be the preferred approach in these types of studies.

Thursday, August 20, 2015

3:30-4:30 PM

Blue Cross & Blue Shield Auditorium (0001 MHRC)