On the analysis of case-control studies in cluster-correlated data settings

Haneuse and Rivera-Rodriguez

Epidemiology, Jan 2018

* Describes sampling and inference issues for case-control studies with two different sampling strategies – standard case control and case-control with sampling stratified by cluster
  + Situations where case-control study is conducted in a setting where the underlying population exhibits clustering
* As opposed to the standard setting for GEE which assumes you first sample clusters and then sample individuals from those clusters
* Motivated by HIV study in clinics in Malawi
* Analysis: inverse probability weights with GEE
  + Weights:
    - Standard case control
      * Cases: Total # cases in pop / # cases sampled
      * Controls: Total # controls in pop / # controls sampled
    - Case-control within each clinic
      * Cases: Total # cases in clinic k / # cases sampled from clinic k
      * Controls: Total # controls in clinic k / # controls sampled from clinic k
  + Working correlation structure: independent; exchangeable (more efficient)
  + Robust sandwich estimator for inference
* Sampling strategies they tried
  + 4 balanced designs
    - SRS
    - Random sample within each clinic
    - Standard case-control
    - Case-control within each clinic
      * For each, an average of 20, 40, 60, and 80 cases + controls per clinic
  + 2 unbalanced designs stratified by clinic size with 10, 16/20, 40/34 drawn from small, medium and large clinics
* Results
  + Case control with exchangeable correlation structure generally more efficient
  + Tradeoffs in efficiency across covariates
  + Standard case-control generally has more power and efficiency than case-control within clinics
  + If using case-control within clinics, unbalanced designs are more powerful and efficient
* \*Suggests that could use conditional logistic for the designs with sampling stratified by clinic, but this would change the interpretation and could not estimate effects of clinic-invariant characteristics
* \*The authors are working on estimation/inference for generalized linear mixed model for case-control and case-control with cluster-stratified case-control