Bias in Longitudinal Cohort Studies

Denis Haine, Ian Dohoo, Simon Dufour

# Motivation

Write a paper for Frontiers in Veterinary Sciences Special Issue: Quantifying and Addressing Bias Associated with Imperfect Observation Processes in Epidemiological Studies.

# Proposal

* Stable population over the follow-up of time t, no elimination of disease;
* No clustering (single-level model);
* Compute disease incidence;
* For the following test characteristics:
  + Se from 0.6 to 1.0, increasing by 0.05,
  + Sp from 0.6 to 1.0, increasing by 0.05;
* For the following disease characteristics:
  + prevalence of 5, 15, 25, and 50%,
  + incidence of 0.01, 0.05 and 0.1/animal-time t,
  + i.e. 12 different contexts according to prevalence and incidence of disease;
* For each of these 12 contexts, simulate 1,000 cohort studies, of 1,000 observations each;
* The same test is used to identify observations at risk at beginning of the cohort, and to identify cases at the end of the cohort (although we could test the effect of using different tests at beginning and end of the cohort according to computation and time constraints).

For each cohort, the true disease status is known at beginning (S1) and end (S2) of the follow-up. True S1 status is used to identify observations at risk at beginning of the cohort; true S2 is used to identify the true outcome. On each dataset, new S1' and S2' are generated by applying the misclassification resulting from the test characteristics to the S1 and S2 samples. Incidence is computed by using S1' and S2 (selection bias), S1 and S2' (misclassification bias), and S1' and S2' (total bias).