Estimating the true prevalence with stratified (multistage) surveys using imperfect diagnostic test systems: is that possible?

Doherr MG(1), Dürr S(1), Fahrion A(1), Hartnack S(1), Toft N(2)

(1) DCR-VPH, Vetsuisse Faculty, University of Bern, CH-3001 Bern, Switzerland
(2) Faculty of Life Science, University of Copenhagen, DK-1870 Frederiksberg C, Denmark

Quite frequently, estimating the prevalence of disease in a given target population is based on a structured stratified, clustered or multistage sample of herds, flocks or farms, and animals selected from those primary sampling units. There are well-described frequentists' statistical methods to calculate sample sizes for various population structures, and to derive overall prevalence estimates with correct (adjusted for the sampling structure) confidence intervals. These methods, either implemented in user-build spreadsheets or specific survey analysis modules of commercial software packages, typically produce stratum-specific and weighted (by stratum-size) overall prevalence estimates with corrected ("exact") statistically based confidence intervals. However, imperfect test characteristics, yielding a certain proportion of false-positive and false-negative test results, are not considered in these approaches. The easiest approach to correct for imperfect tests, the Rogan Gladen estimator, becomes unreliable in certain combinations of test characteristics and observed apparent prevalences. Recently developed and increasingly used Bayesian models with routines implemented in MCMC simulation environments such as WinBugs (http://mathstat.helsinki.fi/openbugs/) will generate test characteristic-adjusted prevalence as well as updated test characteristic estimates. Those models require careful parametrization and interpretation of results, and a good "gold standard" for comparison (i.e. knowning what is the truth) is rarely available, especially when using field data. Having used "best estimates" as prior inputs and documented that the Bayesian model seems robust (insensitive to the width and shape of the input distributions and to different sets of starting values) is frequently accepted as evidence that the model must be appropriate. In the context of a course we developed a routine (programmed in the r language) that will generate a data set that represents a population of herds and animals within herds that belong to several different strata and, in addition to a true disease status indicator (pos/neg) can have up to three different diagnostic test states (each either pos/neg and based on imperfect test characteristics). This artificial population with known characteristics can then be sampled using different survey sampling protocols and sample sizes. Given that the true population herd- and animal-level prevalence is known for all population strata, prevalence estimation biases of the different approaches can easily be calculated and compared. Preliminary analyses using a "2 Dependend Test 2 Population" Bayesian model indicate that the magnitude of bias in all point estimates increases with the level of covariance between tests. The magnitude of bias of the test characteristic estimates appears to be independent of the subpopulation prevalences , but substantially effected by the (preset) test characteristic values. Sensitivities and specificities were generally overestimated. During the presentation, further results will be presented and critically discussed.