A Bayesian approach to sample size evaluation for estimating multiple herd prevalences.

Niels Peter Baadsgaard[1], Wesley O. Johnson[2], Ian Gardner[3]

[1]Danish Institute of Agricultural Sciences, Department of Animal Health and Welfare, Research Centre Foulum, P.O. Box 50, DK-8830, Tjele, Denmark
[2]Department of Statistics, UC Davis. CA. [3]Department of Medicine and Epidemiology, School of Veterinary Medicine, UC Davis. CA.

Abstract

Using a Bayesian approach for sample size calculation, the predictive distribution of the (future and as yet unobserved) data for a collection of herds is simulated, given available prior knowledge of test accuracy, a hypothesized set of prevalence values and selected set of sample sizes (3). The results from this study demonstrate that both sample size and test characteristics must considered when making multiple herd comparisons. In addition, we show that sample size and test characteristics also must be considered, if the objective is to classify herds into clusters with different disease prevalences.

Introduction

There are two objectives within the context of this study. First, our posterior inference for any given data set (simulated or real) concerns multiple herd comparisons by predicting their true prevalences, and to evaluate the impact of sample size and test characteristics on the 95% credibility intervals for the true prevalences. Second, we are concerned with the impact of sample size and test characteristics on the classification of the herds into clusters that are, say low medium and high-risk herds.

Materials

In 1999 a study was conducted to measure the prevalence of granulomatous mastitis in 76 Swedish pig herds (2). The sample size was 20. Herd sizes ranged from 60-400 sows. We used the distribution of the apparent prevalences from these herds at time of weaning as illustrative example for our model. We also use these data, and their background, to construct a simulation of a realistic situation of sample size determination.

Methods

The main idea presented is to simulate M data sets like the one that will ultimately be observed under certain pre-specified conditions on the true prevalences. The approach considered is iterative in nature. Using computer-programming jargon, there is an outside loop (performed using R), where the data are simulated. For each simulated data set, an inside loop is performed using WinBUGS (4) for obtaining the appropriate posterior probability for the given data set from the outside loop.

The iterations are identified by the notation m = 1,….,M. Let η and θ denote P(T+|D+) and P(T-|D-), respectively, and let pr_h denote the prevalence in herd *h*. For specified values of these quantities, define the corresponding expected apparent prevalence for herd *h* as:

$$E(app_h) = pr_h * \eta + (1 - pr_h) * (1 - \theta),\ h = 1,\ldots..,H. \qquad (1)$$

It is then possible to simulate herd apparent prevalences according to the formula

$$X_h^m \sim bin(E^{*m}(app_h), N_h)\ h=1,\ldots,H;\quad m = 1,\ldots,M. \qquad (2)$$

where N_h denotes the sample size for herd *h*, and $E^{*m}(app_h)$ denotes that the terms in (1) are replaced with values that have been simulated from the prior for the sensitivity and specificity; the values for the herd prevalence remain the same from one iteration to the next. All of this constitutes the outside loop.

Having simulated M possible collections of apparent prevalences from multiple herds under the above circumstances, the Gibbs sampler is used for each simulated data collection to estimate the true prevalences in the inside loop. The model used is the same as described in (1) and (2) above, except that here, any scientific information about the true prevalences in the herds to be evaluated is input in the form of a hierarchical prior on the prevalences.

Prior estimates of η and θ are available from another study (1). At each iteration of the Gibbs sampler, the true prevalences are sampled from the joint posterior. As part of the WinBUGS modeling, and for the purpose of handling the second objective of the study, we incorporate the latent categorical variable, Z_h, which identifies one of K possible clusters or latent classes to which the h'th herd might belong. That is (in WinBUGS notation): $Z_h \sim cat(Q)$, where *Q* is a vector of probabilities (which sum to one) of length *K* corresponding to the number of latent classes. That is, let $Q=[p_1,p_2,\ldots,p_K]$ where $\sum_k p_k = 1$; p_k is the true proportion of herds that are in latent class *k*. Our model specifies a hierarchical prior on the prevalence for each cluster. For example, all the prevalences corresponding to herds from the *k*'th cluster are assumed to be independent and identically distributed (conditional on parameters μ_k and τ_k) as Beta($\mu_k * \tau_k$, $(1-\mu_k) * \tau_k$). It is natural to regard the means of the clusters as ordered according to $\mu_1 < \mu_2 \ldots < \mu_K$. We can test that all the μ_k 's are the same by considering the parameter $\delta = \sum_k abs(\mu_k - \bar{\mu})/K$. If the posterior probability that this number is small enough (say less than 0.1) is larger than say 0.95, we could conclude that there was only one cluster. Assessing precisely how many clusters will be pursued in the subsequent paper. Our piors on the μ_k 's and τ_k's will be relatively non-informative for the purposes of this presentation, namely $\mu_k \sim beta(1,1)$ and $\tau_k \sim gamma(0.01,0.01)$, independently.

Results

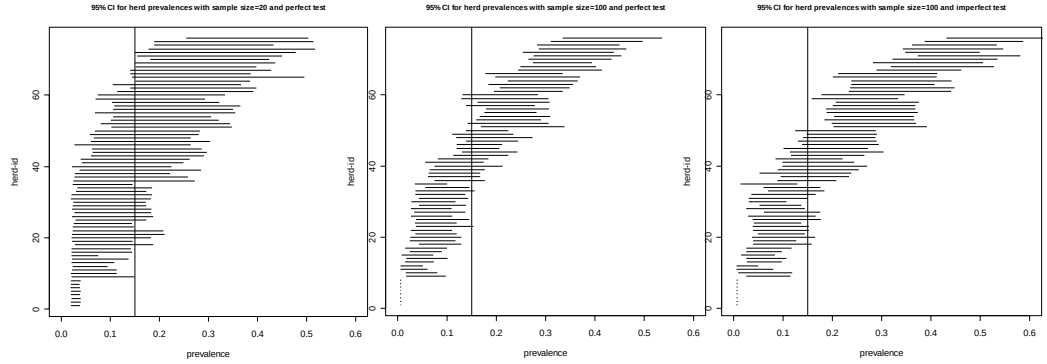

Figure 1. 95% CI for true herd prevalences for sample sizes= 20 and 100, and perfect and imperfect test assumed for the latter. Vertical line indicates an arbitrary threshold 0.15 for classifying herds as being in the high-risk group.

Results are shown for sample size=20 (the actual sample size in the study) and perfect test, sample size=100 and perfect test, and sample size=100 and imperfect test (in the following denoted analysis I, II and III). Herd prevalences are displayed with the 95% CI in Fig.1. The number of herds exceeding a threshold level of 0.15 was estimated to 34 (32-35) for analysis I and II, and 38 (36-40) for analysis III. However, if we wanted to be 95% confident that the h'th herd was above the threshold the numbers of herds were 8 (3-12), 22 (17-24) and 25 (21-28) for analyses I, II and III, respectively.

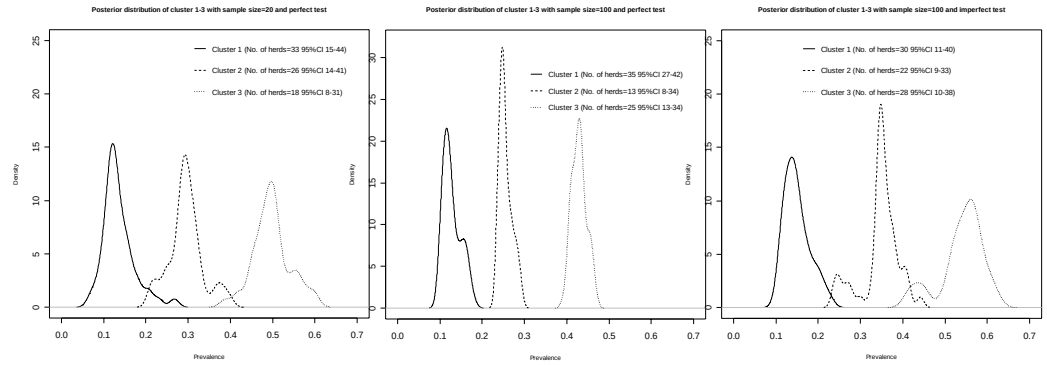

Figure 2. Posterior distributions of 3 clusters estimated for sample sizes=20 and 100, and perfect and imperfect test assumed for the latter.

Posterior distributions of cluster sizes μ_k are displayed in Fig. 2. Three clusters are easily identified, even for sample size=20. The chance of having only one cluster in the data (δ) was estimated to 0.32 (0.19-0.38), 0.42 (0.37-0.45) and 0.26 (0.23-0.43) for analyses I, II and III, respectively.

Discussion

Fig. 1 clearly demonstrates the inevitable impact of sample sizes on the predictive distribution of the data. This is particularly the case if very small sample sizes are used, as in the study on granulomatous mastitis in sows. The widths of the 95% CI could have been reduced significantly by increasing the sample size to 100.

However, the second and third graphs in Figs 1 also point out that misclassification must be considered before making multiple herd comparisons or to making inferences on the location of the herds. Fig. 2 indicates that more than 3 clusters may be considered in the data. The separation among the clusters is more pronounced, when the impact of misclassification is accounted for. By including costs/utility in the analysis further investigation into the relationship between herd comparison/classification and sample size may be scrutinized.

References:

1.Baadsgaard, N.,P., Jørgensen, E., 2003. A Bayesian approach to the accuracy of clinical observations. PreventiveVeterinary Medicine 59: 189-206.

2.Hulten, F., Persson, A., Eliasson-Selling, L., Heldmer, E., Lindberg, M., Sjögren, U., Kugelberg, C., Ehlorsson, C.J., 2003. Clinical characteristics, prevalence, influence of sow performance, and assessment of sow-related risk factors for granulomatous mastitis in sows. American journal of veterinary research 64(4): 463-469

3.Johnson, W., Su, C-L, Gardner, I.A., 2003. Sample Size Calculations for Surveys to Substantiate Freedom of Populations from Infectious Agents. Biometrics. Accepted.

4. Spiegelhalter, D. J., Thomas, A., Best, N.G., Gilks, W.R., 2000. BUGS Bayesian inference using Gibbs Sampling, version 1.3. Cambridge: MRC Biostatistics Unit.