A novel model for estimating the number of consecutive negative samplings required to meet user-specified confidence of disease elimination

D. Polson

Boehringer Ingelheim Vetmedica, Inc., MO, USA

INTRODUCTION

Disease elimination protocols are expected to result in the targeted disease agent being no longer present in the animal population and flow. To judge the success of elimination efforts, it is common to conduct recurring sampling over time from targeted animal cohorts. Whereas sample size determination methods for disease detection from single samplings are generally well understood, the importance of recurring sampling in judging successful disease elimination is poorly understood and executed. As it relates to sampling, confidence in the success of elimination protocols is influenced not only by sample size for each sampling but also by repeated sampling. Production systems continuously produce new cohorts of animals (e.g., groups of weaned pigs) that grow through sequential production phases, typically being moved from one physical location to another location. Opportunities for measuring disease elimination success occur as each new cohort of animals exits a location and moves downstream to the next. To support this aspect of appropriate sampling protocols for judging disease elimination, a methodology, algorithm and model were developed to incorporate both aspects of confidence – sample size and number of samplings.

MATERIALS AND METHODS

Basic model user-defined input variables are: animal population size, animal prevalence, sample size (per sampling), assay sensitivity/specificity, number of sequential samplings and number of simulation runs. The model accommodates selection of: sampling with or without replacement, and fixed or stochastically-generated prevalence per model iteration. To test the model a population of 1,000 animals and sample sizes of 15, 30, 60 and 90 were used to generate sets of data at animal prevalence levels of 1%, 3%, 5% and 10%. A minimum of 200 model runs of 30 consecutive samplings per run were generated for each prevalence level. All sets of runs were generated using sampling without replacement and stochastically-generated iteration prevalence. For the purpose of these simulations the diagnostic assay sensitivity and specificity were both set at 100%. The detection threshold of interest was the sampling at which ≥95% of model runs were detected as positive for the specified prevalence level.

RESULTS AND DISCUSSION

Table 1 contains the results of the stochastic model simulation runs for the evaluated combinations of population prevalence and sample size per sampling event.

Table 1: Number of consecutive negative samplings required for greater than or equal to 95% of samplings to detect at least one or more positive samples per positive case

N samples	Prevalence threshold of positives per sampling			
	1%	3%	5%	10%
15	18	6	4	3
30	11	3	2	1
60	6	2	1	1
90	3	2	1	1

From Table 1, at the 1% population prevalence level, for sample sizes 15, 30, 60 and 90, the 95% detection threshold was achieved at the 18th, 11th, 6th and 3rd samplings, respectively. In contrast, at the 3% population prevalence level, for sample sizes 15, 30, 60 and 90, the 95% detection threshold was achieved at the 6th, 3rd, 2nd and 2nd samplings, respectively.

The detection rate threshold of 95% could be considered a reasonable proxy for a 95% level of detection confidence at the corresponding number of consecutive negative samplings required to equal or exceed that detection rate. It follows then that, for a basic interpretation of Table 1, each value indicates the number of consecutive negative samplings at which there is a 95% probability (and, by proxy, implied confidence) that the population prevalence is less than the evaluated level. For example, obtaining 11 consecutive negative case results where sampling 30 animals per sampling indicates there is a 95% probability that the prevalence of positives in the sampled population is less than 1%; whereas obtaining six consecutive negative case results where sampling 60 animals per sampling indicates there is a 95% probability that the prevalence of positives in the sampled population is less than 1%.

These specific simulations assumed "perfect" (100%) diagnostic sensitivity (Se) and specificity (Sp). However, because it is very unlikely that any diagnostic assay can achieve 100% Se and Sp, the model is capable of running simulations where the diagnostic Se and Sp are less than 100%. Clearly, all factors that influence diagnostic Se and/or Sp for tested samples (and, in turn, their corresponding cases) will influence the number of consecutive negative samplings required to, for a given level of confidence, judge a sampled population's prevalence to be below a required threshold. For example, a few key factors that influence the diagnostic Se of a tested sample are the inherent performance capability of the assay used; testing laboratory factors; and sample collection, handling and shipping factors (e.g., pooling).

CONCLUSION

This novel sampling model can be used to dynamically estimate the appropriate number of consecutive samplings at given sample sizes to collect for use in judging the success or failure of disease elimination protocols, as well as generate tables to be used as references for disease detection sampling that incorporate the influence of recurring samplings.

The model and related tables thus can be useful in situations where detecting positive samples and corresponding cases are important for informed decision-making, e.g., in cohorts of expected-negative replacement animals intended for entry into expected-negative farms sites; as well as in cohorts of expected-negative young animals produced by a farm and cohorts of sentinel animals placed among an existing population that is undergoing a disease elimination protocol.