Spreadsheet Model for Estimating the Probability Herds are Free of Paratuberculosis After Successive Serial Tests

M.T. Collins

School of Veterinary Medicine, University of Wisconsin,
2015 Linden Drive West, Madison, WI 53706 USA

Abstract

Estimation of the probability herds are not infected with *M. avium* subsp. *paratuberculosis* is useful for justification of herd certification programs. Existing epidemiological models only allow such estimation after a single herd test. A simple spreadsheet model was created to permit estimation of the probability herds are not infected after successive tests. In addition, the model incorporates economic parameters such that both the benefit (probability herds are not infected) and the cost of attaining this probability can be evaluated. The purpose of the model is to provide a framework for rational discussion and evaluation of testing option when states, regions, or countries embark on design of herd certification programs. The model has been employed in design of the Dutch and U.S. programs

Introduction

Diagnostic tests for paratuberculosis are applied to individual animals or selected populations, e.g., herds or flocks (referred to as herds hereafter, for simplicity). Most often, the purpose of using paratuberculosis tests is to discover if the animal or herd is infected with *Mycobacterium avium* subsp. *paratuberculosis*. Herd certification-like programs, such as the Market Assurance Programs for cattle and sheep in Australia (Kennedy, D.J., et al, 1999), the Voluntary Johne's Disease Herd Status Program in the U.S.A. (USAHA National Johne's Disease Working Group Certification Subcommittee, 99 A.D.), and the National Paratuberculosis Program of the Netherlands (Benedictus, G., et al, 1999; Vecchio, T.J., 1966; Sherman, D.M., et al, 1984), have the opposite objective, that is to classify herds for the probability they are NOT infected with *M. a. paratuberculosis*.

When diagnostic tests for paratuberculosis are applied to individual animals, the probability of correctly diagnosing each tested animal as infected or not infected with *M. a. paratuberculosis* is the positive or negative predictive value of the test, respectively. Predictive values are derived from three parameters: two that define the accuracy of the test employed, sensitivity and specificity, the third being the frequency of the infection or disease in the population of animals selected for testing, called prevalence . The population from which the test subject comes is often a herd, but it may be defined in other ways, such as all cattle showing clinical signs compatible with Johne's disease. Sherman demonstrated that 65% of cattle with clinical signs compatible with Johne's disease seen at a university teaching hospital actually were infected with *M. a. paratuberculosis* thus providing one estimate of paratuberculosis prevalence in so defined populations of cattle.

Obviously, the accuracy of a diagnostic test directly affects the predictive value. Thus, considerable effort is made to precisely and objectively measure the sensitivity and specificity of diagnostic tests. Disparity among published reports of the sensitivity and specificity of paratuberculosis diagnostics are in part due to differences in case definition for both the infected and infection-free study populations. However, in general there is remarkably good agreement among studies on the accuracy of the absorbed ELISA (Sweeney, R.W., et al, 1995; Ridge, S.E., et al, 1991; Cox, J.C., et al, 1990; Milner, A.R., et al, 1990; Sockett, D.C.,

et al, 1992a), fecal culture (Sockett, D.C., et al, 1992b; Whipple, D.L., et al, 1992; Whipple, D.L., et al, 1991), and PCR amplified DNA probes for IS900 (Sockett, D.C., et al, 1992b; Briscoe, J., 1980; Dierksheide, W.C., 1987).

Prevalence has a major influence on the predictive value of a diagnostic test . In fact, prevalence has a greater influence on the probability of a correct diagnosis than does the sensitivity or specificity of the test, over the range of values reported by different investigators. The difficulty is that infection prevalence and is usually not known precisely but rather must be estimated from clinical experience. Veterinary practitioners use a variety of factors, often intuitively, to estimate within herd prevalence of paratuberculosis such as frequency of animals showing clinical Johne's disease, age of animals showing signs of Johne's disease, degree to which herd management practices facilitate infection transmission, etc. Testing of all animals in a herd, so called whole herd testing, can be used to get a more objective estimate of the within herd prevalence, and formulas for calculating true prevalence from apparent (test) prevalence are well known (Obasanjo, I.O., et al, 1997; Yanagawa, T., et al, 1984; Martin, S.W., et al, 1987). However, while testing may seem to be the best measure of within herd prevalence, it is important to recognize that it may be biased. For example, if the herd owner has tested previously and has made culling decisions on the basis of these test results, test prevalence will not be an accurate reflection of true prevalence as the test-positive animals have been selectively removed from the herd. Regardless of the method, some estimate of within herd prevalence is necessary in order to express positive or negative diagnostic test results as a positive or negative predictive values, respectively.

When the diagnostic unit of concern is a herd, instead of an individual animal, the same concepts apply. Accuracy of the diagnostic test used, however, is defined as the rate of correct classification of truly infected and truly non-infected herds referred to as herd sensitivity and herd specificity, respectively. Ideally, these two very important parameters should be defined experimentally on large numbers of herds objectively characterized as infected or not infected by criteria independent from the test being evaluated, just as when test sensitivity and specificity is measured at the individual animal level. Such studies require testing many herds, and consequently such large numbers of animals, that the cost of such studies precludes them. Instead, formulae are used that permit estimation of herd sensitivity and herd specificity of diagnostic tests from their respective sensitivity and specificity estimated from studies on individual animals (Martin, S.W., et al, 1994).

Expressing herd test results as the probability of correct herd diagnosis, infected or not infected with *M. a. paratuberculosis*, can be done using an adaptation of the predictive value model by substituting herd sensitivity, herd specificity, and among herd paratuberculosis prevalence in place of sensitivity, specificity, and within herd prevalence, respectively (Martin, S.W., et al, 1987; Frehel, C., et al, 1988). Recent papers by Jordan (Jordan, D., 1996; Jordan, D., 1996; Donald, A.W., et al, 1994) and Donald et al. describe some of the limitations of this approach using diagnostic tests with relatively low sensitivity and specificity, such as those for paratuberculosis.

The reports of Jordan (1996) and Donald (1994) only dealt with herd classification (diagnosis) accuracy at a single point in time after a single herd test. For herd certification or market assurance programs, however, herds are tested repeatedly, often at yearly intervals. The assumption is that more testing and more negative results will, over time, increase the confidence in the non-infected status of herds. No models to date permit estimation of the probability that herds, tested multiple times, are not *M. a. paratuberculosis*-infected. Rational design of testing programs would be assisted by such a model because then the choice of test(s) by accuracy, number and age of animals to test, and cost of the testing program to the individual herd owner or to the state could be weighed against the desired level of confidence in the true non-infected status of the herd. The goal of this study was to create a simple model to fulfill this need by creating a model containing both epidemiological and economic parameters.

Materials and Methods

Summary of terms and definitions.
H_t = total number of herds, infected and non-infected in a specific geographic unit
H_i = number of infected herds
H_n = number of non-infected herds
h = mean herd size for geographic area
P = herd prevalence; percentage of herds that are infected

Subscripts are used to denote the stage in the serial testing scheme. For example, P_0 means herd prevalence at the start, before the first test, P_1 meaning after the first test, P_2 meaning after the second test, etc.

p = estimated average within herd infection prevalence for infected herds with subscripts to denote the stage of serial testing to which the within herd prevalence refers: p_1, p_2, p_3, p_4.
Se = diagnostic test sensitivity reported in published literature at the individual animal level.
Sp = diagnostic test specificity reported in published literature at the individual animal level.
n = the number of animals tested in a herd.

HSe = Herd sensitivity is the percentage of test-positive herds that are truly infected. The formula to calculate herd sensitivity reported by Martin (1992) is:

$$HSe = 1 - (1-Se)^{p*n}] * [Sp^{((1-p)*n)}]$$

where Se is the sensitivity of a test, Sp is the specificity of the test, p is the within herd prevalence of infection, and n is the number of animals tested in the herd.

The formula for herd sensitivity used by in the Australia Cattle Market Assurance Program is simpler:

$$HSe = 1 - (1-Se)^{n}$$

where the term definitions are the same, except that *n* refers to the number of infected animals in the herd, meaning (p*n). This clever simplification of the HSe formula is made possible by setting the test specificity at 1.00. Test specificity will be 1.00 if all positive immunological tests are confirmed by fecal culture and we assume that fecal culture is 100% specific.

Thus, for the serial herd testing model herein described, the formula for HSe is:

$$HSe = 1 - (1-Se)^{p*n}$$

HSp = Herd specificity. As stated previously, by assuming that fecal culture for *M. a. paratuberculosis* is 100% specific and having a positive diagnostic test by any method other than fecal culture confirmed by fecal culture, then Sp = 1.00 and therefore HSp is also 1.00. As discussed by Jordan (1996) HSp and HSe can be adjusted by changing the number of positive tests are required to classify a herd as infected; this is referred to as the cut-off for a positive herd diagnosis. Raising the cut-off increases HSp but lowers HSe. For the purposes of the serial herd testing model, in all instances a single positive test in animals tested from the herd, if confirmed by fecal culture, is sufficient to classify a herd as infected.

Herd predictive value
Martin (1992) reported that predictive values at the herd level can be calculated in the same way as at the individual animal level after appropriate substitution of terms mentioned previously. Thus the formulae for herd positive predictive value (HPPV) and herd negative predictive value (HNPV) are, respectively, as follows:

$$HPPV = \frac{P \times HSe}{P \times HSe + [(1 - P) \times (1\text{-}HSp)]}$$

$$HNPV = \frac{(1 - P) \times HSp}{P \times (1 - HSe) + (1 - P) \times HSp}$$

Since this model is intended to predict disease freedom when herd tests are negative, only HNPV is relevant.

G = gain in certainty; the difference in P-free before and after the respective herd test, expressed as a percentage. (Connell, F.A., et al, 1985)

C = cost of each test on a per herd basis with subscripts denoting the relevant round of testing. This is derived from the cost, to the herd owner, of each test multiplied by the number of tests performed for that specific round of herd testing.

C/G = cost per unit gain certainty P-free measures. This parameter is a method of measuring in economic terms the real value of the herd test relative to the information gained, i.e., increase in probability the herd is free of infection.

Simplifying assumptions and parameters that must be estimated.

1. The population of herds is finite, for example all herds in a specific state or country, and no new herds are introduced to the population.
2. The percentage of herds that are infected in the population (herd prevalence) prior to the first round of herd testing is a set value, e.g., 25%, and not one described as a probability distribution function. After each round of testing, all test-positive herds are "removed" from the population. They become part of the infected herd population and are no longer members of the test-negative group. For this reason, the percentage of truly infected herds in the test-negative population prior to the subsequent round of testing always is less than that of the previous testing round. Removal of test-positive herds with each round of testing is, perhaps, the most fundamentally important aspect of the model. Removal is not meant to imply slaughter but rather merely isolation in the sense of animal trade; such that animals from infected herds do not enter non-infected herds as in most all other disease control programs. Test-positive herds would presumably enter paratuberculosis control programs, but for the purposes of the serial testing model to predict probability of infection-freedom among test-negative herds, infected herds are ignored.
3. Within herd prevalence at the start of the model is a set value, not one described as a probability distribution function, and is the same for all herds regardless of any other herd characteristics such as breed, geographic location, or frequency of purchase of animals from outside sources. After each round of testing of all herds, average within herd prevalence decreases by one-half that of the preceding round of testing. For example, before round 1, 2, 3, and 4 of testing the mean within herd infection prevalence for infected herds would be 10%, 5%, 2.5% and 1.25%. There is no objective data to use for such estimates, but it is intuitively logical that the average infection rate among test-negative herds must lessen after removal of all test-positive herds (those presumably with higher infection rates) from the population following each round of testing.
4. The probability a test-negative herd is not *M. a. paratuberculosis*-infected is the estimated number of truly infected herds left in the test-negative population, after removal of all test-positive herds, divided by the total number of herds left in the population. This term is designated "P-free".

Model variables. Before deriving model results the following parameters must be entered in the spreadsheet model. Below are listed those parameters as well as specific values used in the example described in the remainder of this paper. For the purposes of illustration, I will use values based on the population of cattle in my home state of the USA, Wisconsin, and the model testing program of the VJDHSP (USAHA

National Johne's Disease Working Group Certification Subcommittee, 99 A.D.). Specifically,

$H_t = 27{,}000$
$P_0 = 30\%$
$p_0 = 10\%$
ELISA Se = 45%
(based on published literature and the fact that only 2nd lactation or older animals are being tested).
ELISA cost = $8.00
Fecal culture Se = 55%
(based on published literature and the fact that only 2nd lactation or older animals are being tested).
Fecal culture cost = $16.00
h = 65
(mean number of milking cows per herd Wisconsin, 1996; roughly 70% of the milking herd will be 2nd lactation or older).

Sequence of calculations. Calculations are carried out in a specific sequence to mimic the order of operations performed in a serial, e.g., annual, herd testing scheme. The sequence of calculations is as follows:

Round #1 of herd testing:

1. Calculate number of infected herds in the population: $H_t * P$
2. Calculate HSe.
3. Calculate the number of infected herds detected and removed ($H_i * HSe$)
4. Subtract the result of step#1 from step #2 to find the number of infected herds remaining in the population.
5. Subtract the number of infected herds detected (step #3) from the total number of herds at the start of testing (H_t) to derive the number of herds in the population after the first round of testing.
6. Calculate the herd prevalence after round #1 of testing (P_1) by dividing the results of step #4 by step #5.
7. The post-test HNPV is (1 - the result of step #6), herein referred to as P-free. This value then becomes the pre-test HNPV for the subsequent round of herd testing.
8. Calculate gain in certainty as percentage of herds infected after testing (and removal), P_1, minus the percentage of herds infected before testing (P_0).
9. Calculate the cost of testing to the herd owner: number of tests done x cost/test.
10. Calculate the cost of testing all herds in the specified state, country or region as the cost for each average herd owner times the number of herds tested.
11. Calculate the cost per unit gain in certainty: result of step#9 divided by (step #8 * 100). The multiplier of 100 is used to give cost per percentage unit gain in certainty.

Subsequent rounds of testing:

The same process is repeated but with new among and within herd prevalence estimates and the appropriate parameters for the prescribed testing regimen (Se, cost, number of tests/herd).

Figure 1 shows a schematic representation of the model. The pie charts illustrate the proportion of herds infected (black section of the pie) before testing is begun and after each round of testing. The diamonds between each pie chart represent the diagnostic test applied to herds. Results of each herd test will be either positive or negative. Test positive herds, if confirmed infected, are "removed" from the population in the sense that they can no longer trade cattle with the test-negative herds. The model permits calculation of the actual numbers of herds predicted to test positive or negative at each round of testing. Additionally, it calculates the proportion of herds in the population that are infected. In the illustration used, notice the large gain in certainty that herds are not infected (P-free) before and after the first round of ELISA testing applied to 30 cattle 2nd lactation or older from a herd of 65 adult cattle. Subsequent gains in P-free with additional herd tests are smaller.

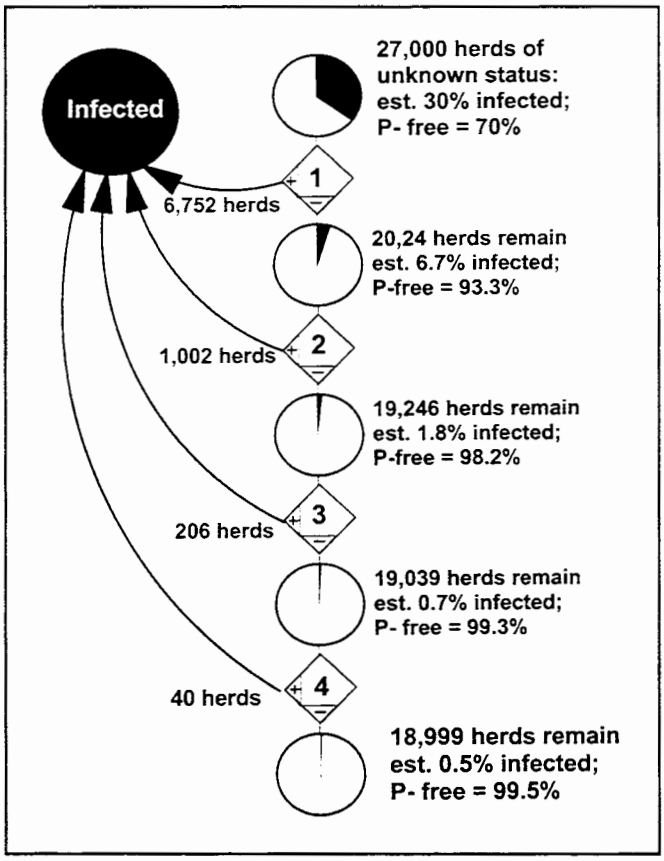

Figure 1. Schematic representation of the spreadsheet model of herd certification.

Figure 2 on the next page shows the actual Lotus 1-2-3 spreadsheet for these calculations.

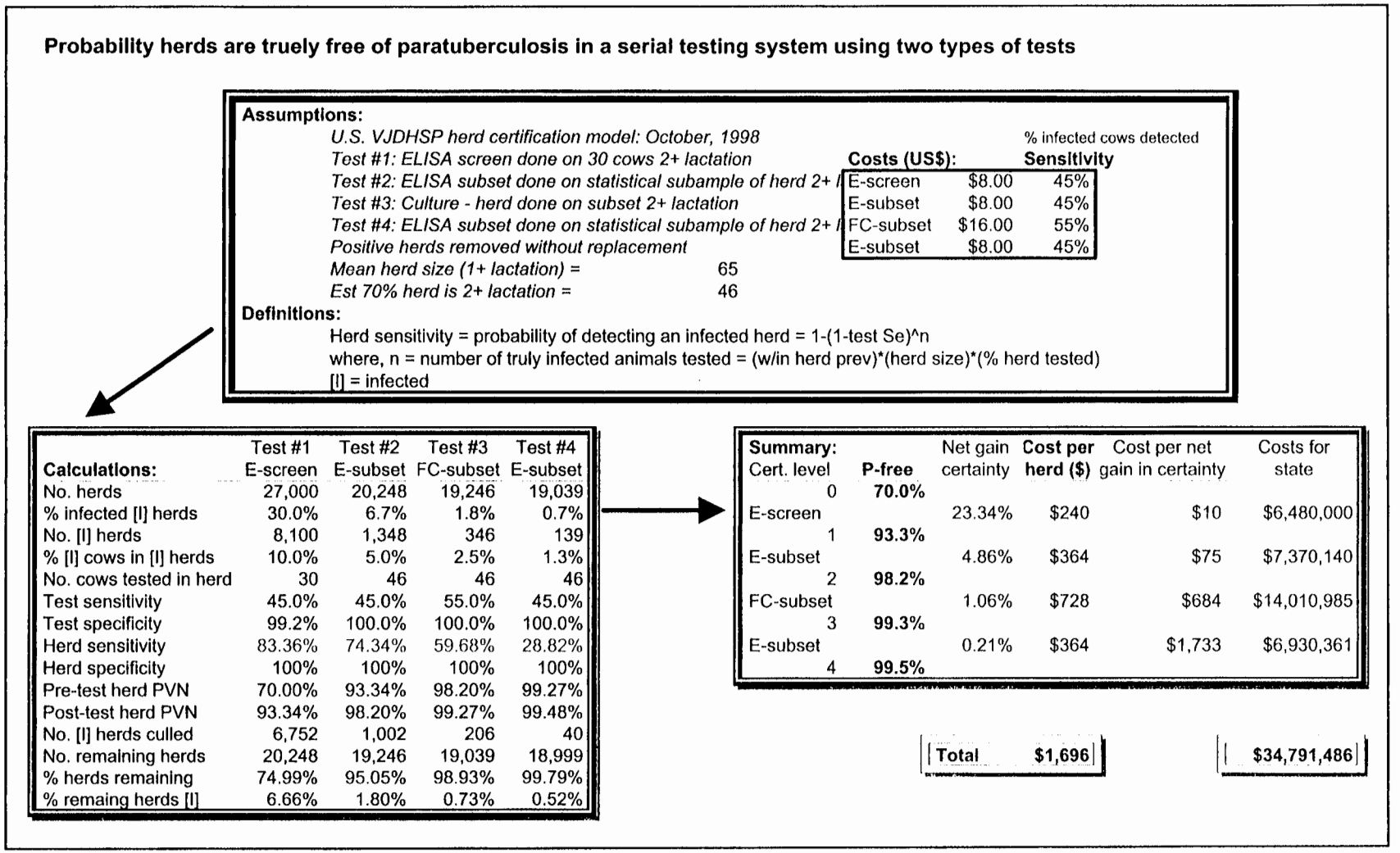

Figure 2. Sample of Lotus 1-2-3 spreadsheet showing assumptions, input variables, sequence of calculations, and results of the example herd classification system outlined in this paper and shown in Figure 1.

Discussion

The model is intentionally very simplified. This is in part due to the lack of objective data with which to mathematically describe parameters such as the distribution of within herd infection prevalence. The result is that the model has certain strengths as well as weaknesses and it is important that these are understood.

Limitations of the model:

1. The model is strongly influenced by the estimates of diagnostic accuracy at the individual animal level.
2. The model assumes that the spectrum of disease (proportion of cattle in each stage of paratuberculosis) is the same each year the test is applied. In other words, it assumes that the tests are independent each year and therefore that the sensitivity of the diagnostic test employed is constant across all years. While a weakness, there is no objective data on which to base an alternative assumption.
3. Herd prevalence distribution is ignored in the model and set at a fixed value representing the average of a Gaussian distribution. It is highly unlikely that this is the case in reality, but data to support an alternative assumption are lacking.
4. It is only a model. It is not expected to mirror reality.

Strengths of the model:

1. It is simple and easy to understand as it makes computations using actual herd numbers.
2. It permits estimation of the probability that herds in a state, country or region are free of paratuberculosis after any selected type or number of herd tests.
3. It calculates the costs of the testing program for individual herd owners and the entire population of herds in question, at each round of testing thereby permitting assessment of both the value (level of confidence the herds are non-infected = P-free) and the cost to attain that level of certainty. In voluntary state / national programs, much of the cost will be borne directly or indirectly by the herd owners. Thus, consideration of costs is imperative.
4. The model allows one to compare and contrast the costs and benefits of alternative herd certification testing strategies to support more rational design of the optimal program.

Conclusions from the model:

A. More testing, with removal of infected herds from the population, results in increasing confidence that the remaining test-negative herds are free of paratuberculosis.
B. Greater confidence in the true non-infected status of herds results from use of diagnostic tests with the highest sensitivity.
C. The cost of testing herds is significant, and almost the same level of confidence in the non-infected status of herds can be derived by tests of lesser sensitivity at half the cost.
D. After four rounds of testing, using existing tests for paratuberculosis, there is a high level of confidence that the remaining test-negative herds are free of infection (P-free >99%). However, some infected herds may still remain in the population.

To successfully combat the spread of paratuberculosis it is vital that states or countries find ways to limit the opportunities for entry of animals from infected herds into those that are as yet not infected with *M. a. paratuberculosis*. The first step toward this goal is classifying herds relative to the likelihood that they are infected. Herd certification / classification programs in place in Australia, The Netherlands, and various states in the U.S. are examples. Design of such programs is often hampered by lack of knowledge about the accuracy of diagnostics for paratuberculosis, particularly at the herd level. Additionally, while it is relatively simple to design herd classification programs that involve extensive herd testing multiple times with multiple tests, it is far more challenging to design programs that are **both**

scientifically sound and affordable. With all its limitations, the model herein described has been useful as a means of contrasting various testing strategies for herd classification. Notably it was used in design of the Voluntary Johne's Disease Herd Status Program in the U.S.A. and the National Paratuberculosis Program of the Netherlands.

References

Benedictus, G., Verhoeff, J., Schukken, Y.H., and Hesselink, J.W., 1999. in Proceedings of the Sixth International Colloquium on Paratuberculosis, edited by E.J.B. Manning and M.T. Collins (International Association for Paratuberculosis, Madison, WI) p. 2-15.

Briscoe, J., 1980. On the use of simple analytic mathematical models of communicable diseases. Int. J. Epidemiol. 9, 265-270.

Connell, F.A. and Koepsell, T.D., 1985. Measures of gain in certainty from a diagnostic test. Am. J. Epidemiol. 121, 744-753.

Cox, J.C., Drane, D., Ridge, S., and Milner, A.R., 1990, Abstracts of the Annual meeting of the Australian Society for Microbiology,Tasmania,July,1990 11, 241. (Abstract)

Dierksheide, W.C., 1987. Medical decisions: Interpreting clinical tests. ASM News 53, 677-680.

Donald, A.W., Gardner, I.A., and Wiggins, A.D., 1994. Cut-off points for aggregate herd testing in the presence of disease clustering and correlation of test errors. Prev. Vet. Med. 167-187.

Frehel, C., Rastogi, N., Benichou, J.-C., and Ryter, A., 1988. Do test tube-grown pathogenic mycobacteria possess a protective capsule? FEMS Microbiol. Letters 56, 225-230.

Jordan, D., 1996. Aggregate testing for the evaluation of Johne's disease herd status. Aust. Vet. J. 73, 16-19.

Kennedy, D.J. and Allworth, M.B., 1999. in Proceedings of the Sixth International Colloquium on Paratuberculosis, edited by E.J.B. Manning and M.T. Collins (International Association for Paratuberculosis, Madison, WI) p. 19-26.

Martin, S.W., Meek, A.H., and Willeberg, P., 1987. in Veterinary epidemiology - Principles and methods, edited by S.W. Martin, A.H. Meek, and P. Willeberg (Iowa State University Press, Ames, IA) p. 73-76.

Martin, S.W., Shoukri, M., and Thorburn, M.A., 1994. Evaluating the health status of herds based on tests applied to individuals. Prev. Vet. Med. 33-43.

Milner, A.R., Mack, W.N., Coates, K.J., Hill, J., Gill, I., and Sheldrick, P., 1990. The sensitivity and specificity of a modified ELISA for the diagnosis of Johne's disease from a field trial in cattle. Vet. Microbiol. 25, 193-198.

Obasanjo, I.O., Grohn, Y.T., and Mohammed, H.O., 1997. Farm factors associated with the presence of *Mycobacterium paratuberculosis* infection in dairy herds on the New York State Paratuberculosis Control Program. Prev. Vet. Med. 32, 243-251.

Ridge, S.E., Morgan, I.R., Sockett, D.C., Collins, M.T., Condron, R.J., Skilbeck, N.W., and Webber, J.J., 1991. Comparison of the Johne's absorbed EIA and the complement fixation test for the diagnosis of Johne's disease in cattle. Austr. Vet. J. 68, 253-257.

Sherman, D.M., Markham, R.J.F., and Bates, F., 1984. Agar gel immunodiffusion test for diagnosis of clinical paratuberculosis in cattle. J. Am. Vet. Med. Assoc. 185, 179-182.

Sockett, D.C., Carr, D.J., and Collins, M.T., 1992b. Evaluation of conventional and radiometric fecal culture and a commercial DNA probe for diagnosis of *Mycobacterium paratuberculosis* infections in cattle. Can. J. Vet. Res. 56, 148-153.

Sockett, D.C., Conrad, T.A., Thomas, C.B., and Collins, M.T., 1992a. Evaluation of four serological tests for bovine paratuberculosis. J. Clin. Microbiol. 30, 1134-1139.

Sweeney, R.W., Whitlock, R.H., Buckley, C.L., and Spencer, P.A., 1995. Evaluation of a commercial enzyme-linked immunosorbent assay for the diagnosis of paratuberculosis in dairy cattle. J. Vet. Diagn. Invest. 7, 488-493.

USAHA National Johne's Disease Working Group Certification Subcommittee, 99 A.D. in Proceedings of the Sixth International Colloquium on Paratuberculosis, edited by E.J.B. Manning and M.T. Collins (International Association for Paratuberculosis, Madison, WI) p. 33-43.

Vecchio, T.J., 1966. Predictive value of a single diagnostic test in unselected populations. New Engl. J. Med. 274, 1171-1173.

Whipple, D.L., Callihan, D.R., and Jarnagin, J.L., 1991. Cultivation of *Mycobacterium paratuberculosis* from bovine fecal specimens and a suggested standardized procedure. J. Vet. Diagn. Invest. 3, 368-373.

Whipple, D.L., Kapke, P.A., and Andersen, P.R., 1992. Comparison of a commercial DNA probe test and three cultivation procedures for detection of *Mycobacterium paratuberculosis* in bovine feces. J. Vet. Diagn. Invest. 4, 23-27.

Yanagawa, T. and Gladen, B.C., 1984. Estimating disease rates from a diagnostic test. Am. J. Epidemiol. 119, 1015-1023.