

THESIS

2
2004
50534404



This is to certify that the
dissertation entitled

**PERFORMANCE OF BOVINE TUBERCULIN SKIN TESTS AND
FACTORS AFFECTING THE PROPORTION OF FALSE POSITIVE
SKIN TEST RESULTS IN MICHIGAN**

presented by

Bo Norby

has been accepted towards fulfillment
of the requirements for the

Ph.D. degree in Large Animal Clinical Sciences

Major Professor's Signature

12/10/03

Date

PLACE IN RETURN BOX to remove this checkout from your record.
TO AVOID FINES return on or before date due.
MAY BE RECALLED with earlier due date if requested.

DATE DUE	DATE DUE	DATE DUE
FEB 22 2017		

**PERFORMANCE OF BOVINE TUBERCULIN SKIN TESTS AND
FACTORS AFFECTING THE PROPORTION OF FALSE POSITIVE
SKIN TEST RESULTS IN MICHIGAN**

By

Bo Norby

A DISSERTATION

**Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of**

DOCTOR OF PHILOSOPHY

**Department of Large Animal Clinical Sciences
(Epidemiology)**

2003

ABSTRACT

PERFORMANCE OF BOVINE TUBERCULIN SKIN TESTS AND FACTORS AFFECTING THE PROPORTION OF FALSE POSITIVE SKIN TEST RESULTS IN MICHIGAN

By

Bo Norby

Dairy and beef cattle (n=494) from seven Michigan farms located in the northeastern part of the lower peninsula in Michigan were used to estimate the sensitivity (Se) of the caudal fold tuberculin test (CFT), the CFT and comparative cervical test used in serial interpretation (CFTCCT_{SER}), and gross necropsy using a 'gold standard' approach. Using all seven herds, the sensitivities of the CFT, the CFTCCT_{SER}, and gross necropsy were 93.0% (95% CI: 80.9% to 98.5%), 88.4% (95% CI: 74.9% to 96.1%), and 86.1% (95% CI: 72.1% to 94.7%), respectively. The Se of the two skin tests was slightly higher when two or more gross lesions were present, and the Se of gross necropsy was significantly higher (P = 0.049). In addition, four two-population-two test simulation-based Bayesian models allowing for correlation between the two tests were used to estimate Se and specificity (Sp) of the CFT and CFTCCT_{SER}, assuming that a gold standard was not available. Sensitivity and Sp of both the CFT and CFTCCT_{SER} were estimated with results of mycobacterial culture (CULT). There were no significant correlations between CFT and CULT, but a significant negative correlation (0.218 [PCI: 0.005 to 0.731]) between CFTCCT_{SER} and CULT. The estimates for Se and Sp of the CFT, CFTCCT_{SER}, and CULT were 85.4% (PCI: 56.3 to 97.5), 75.8% (PCI: 47.5% to 93.0%), and 62.3% (PCI: 40.1% to 89.9%), respectively. Herd-level sensitivity (HSe)

and specificity (HSp) and predictive values for 1) CFTCCT_{SER}, 2) the overall sensitivity and specificity of serial use of CFTCCT_{SER}, histologic exam, CULT and PCR, 3) where specificity of 2) was assumed to be 100%, and 4) which was as 3) but increasing sensitivity to 90%. An empirical distribution of number of herds and herd sizes for the northeastern lower peninsula was used for these simulations. Point estimates of HSe and HSp for the four scenarios were 75.0% and 94.0%, 71.2% and 100%, 78.3%, and 100%, and 83.3% and 100%. All the HSe estimates ranged from 0 to 100% and between 0.023 to 100% for HSp. Herd-level predictive values positive and negative were 14.3% and 99.6%, 100 and 99.5%, 100 and 99.6%, and 100 and 98.9% for the four scenarios, respectively. Increasing the sensitivity of scenario 4 to 90% only increased the HSe from 78.3% to 83.3%. Results from 4,989 bovine tuberculosis (bTb) negative herds tested with the CFT between 1996 and 2001 were used to investigate the association between the proportion of false positive (PFP) results on the CFT by herd and herd, geographic, and ecologic factors using multiple negative binomial regression. The final multiple negative binomial model contained herd type (beef vs. dairy), agricultural region, and the testing veterinarian as a random effect. The number of cases of PFP per 100 tested animals was 2.25 for beef and 4.38 for dairy cattle. Cattle located in agricultural regions 1, 2, 3, 5, and 6 had significantly higher PFP as compared to region 9. Using 3,301 records with spatial coordinates, one primary and 32 secondary clusters of herds with high PFP were identified. Additionally, PFP was not significant associated with the elevation of the herd and whether or not a wetland area was located close to the farm.

**This work is dedicated to the memory of my father
who passed away at a much too early age.**

ACKNOWLEDGEMENTS

It would like to acknowledge the people whose contributions have made this work possible:

First and foremost, I would like to thank my major professor, Dr. Paul Bartlett, who has been an outstanding mentor and who gave me the freedom to pursue work relating to my special interests and to whom I will always be in great debt.

Dr. John Kaneene, whose door was always open, and who has made me appreciate epidemiology even more.

The other members of my committee: Drs. Daniel Grooms, Fitzgerald, and Templeman who, beyond their duty, made an important contribution to my education.

Michigan Department of Agriculture (MDA) and United States Department of Agriculture (USDA) for providing the data for this project. In particular, I would like to thank Drs. Larry Granger and Colleen Bruning-Fann who helped me gather data, allowed me free access to test records in their possession, and helped me understand all aspects of the bovine tuberculosis eradication effort in Michigan.

My fellow graduate students at the Population Medicine Center who tolerated my questions and who always had time to listen. I wish you the very best in your future endeavors.

Also a special thank to Dr. Tim Hanson whose help was instrumental in development of Bayesian models, and to my external examiner Dr. Ian Gardner, who first sparked my interest the epidemiology of diagnostic tests.

My mother Inger, my brother Steen, and my wife Lauren without whose unconditional support, this endeavor would not have been possible.

Finally for financial support of this research, a special thanks to the USDA, the College of Veterinary Medicine, and Department of Large Animal Clinical Sciences at MSU.

TABLE OF CONTENTS

LIST OF TABLES	xii
LIST OF FIGURES	xiv
LIST OF ABBREVIATIONS	xv
CHAPTER 1: Introduction, Rationale, Problem Statement, General Methods, and Overview.	
Introduction	1
Rationale	2
Problem Statement	4
Study objectives	5
General methods	6
Overview	9
References	10
CHAPTER 2: Literature Review.	
Tuberculins	12
Tuberculin skin tests	13
Sensitivity and specificity of the CFT and CCT in the United States	15
Factors affecting the sensitivity and specificity of the tuberculin skin tests	19
Herd-level estimates of sensitivity and specificity	21
References	23

CHAPTER 3: The sensitivity of gross necropsy, caudal fold and comparative cervical tests for the diagnosis of bovine tuberculosis.

Abstract	26
Introduction	28
Materials and Methods	30
Summary of state-wide testing procedure	30
CFT	30
CCT	31
Data selection	31
Gross necropsy	32
Histopathology	32
Mycobacterial culture and identification	33
Polymerase chain reaction	33
Statistical analysis	34
Results	34
Discussion	40
References	45

**CHAPTER 4: Estimation of sensitivity and specificity
of bovine tuberculosis skin tests in Michigan
when a perfect reference test is not available.**

Abstract	48
Introduction	50
Materials and Methods	53
Animals	53
Diagnostic tests and testing procedure	54
Latent-class models	55
Prior elicitation	57
Sensitivity analysis	59
Results	60
Discussion	65
References	70
Appendix A	74

**CHAPTER 5: Herd-level sensitivity, specificity, and predictive values
of bovine tuberculosis skin tests in Michigan.**

Abstract	83
Introduction	85
Materials and Methods	87
Diagnostic tests and testing procedure	87
Analyses	89
Input distributions for Se, Sp, TP, and HTP	90
Herd size distributions and sampling technique	92
Results	95
Discussion	98
The ‘HerdTest’ model	101
Comments and recommendations to the eradication program In Michigan	103
References	105

CHAPTER 6: Herd and spatial factors affecting the proportion of false positive results on the caudal fold tuberculin test in Michigan cattle

Abstract	107
Introduction	109
Materials and Methods	110
Diagnostic tests and testing procedure	110
Statistical analysis	112
Spatial analysis	115
Ecological associations	117
Results	117
Descriptive statistics	117
Negative binomial regression model	122
Spatial analysis	125
Ecological associations	127
Discussion	128
References	133
SUMMARY	136
CONCLUSIONS	139
APPENDIX B	145
BIBLIOGRAPHY	170

LIST OF TABLES

TABLE	PAGE
Table 2-1. Calculation of sensitivity and specificity when a gold standard is available.	16
Table 3-1. Relationship between mycobacterial isolates from cattle and skin tests, gross necropsy, and the reference test (positive on either mycobacterial isolation or PCR).	37
Table 3-2. Apparent bTb prevalence in seven herds using the caudal fold, comparative cervical and reference tests (positive on culture and/or PCR).	38
Table 3-3. Sensitivity of CFT, CFTCCT _{SER} , and gross necropsy using bacteriologic culture and PCR as the reference test.	39
Table 3-4. Sensitivity of CFT, CFTCCT _{SER} , and gross necropsy in animals with lesions in one, or two or more lymph nodes.	40
Table 4-1. Apparent prevalence (APP; %) in old and young subpopulations based upon the caudal fold tuberculin test (CFT) and Mycobacterial culture (CULT).	59
Table 4-2. Testing Scenario #1: 2 × 2 contingency table showing the observed counts for the caudal fold tuberculin test (CFT) and Mycobacterial culture (CULT) in populations 1 and 2.	60
Table 4-3. Testing Scenario #2: 2 × 2 contingency table showing the observed counts for the caudal fold and comparative cervical tuberculin tests used in serial interpretation (CFTCCT _{SER}) and Mycobacterial culture (CULT) in populations 1 and 2.	60
Table 4-4. Prevalences (π_1, π_2), sensitivities (Se) and specificities (Sp) for the caudal fold tuberculin test (CFT) and mycobacterial culture (CULT). Point estimates for all parameters are presented as the median of the posterior densities and their respective 95% PCI.	63
Table 4-5. Prevalences (π_1, π_2), sensitivities (Se) and specificities (Sp) for the caudal fold tuberculin test and comparative cervical tuberculin tests used in serial interpretation (CFTCCT _{SER}) and mycobacterial culture (CULT) for all four models. Point estimates for all parameters are presented as the median of the posterior densities and their respective 95% PCI.	64

TABLE	PAGE
Table 4-6. Two 2×2 contingency tables showing responses of two binary diagnostic tests.	74
Table 4-7. Four 2×2 contingency tables showing responses of two binary tests classified by true disease status.	75
Table 4-8. Classification of probabilities as explained by sensitivities and specificities.	76/77
Table 5-1. Input variables for sensitivity (Se) Specificity (Sp), and within herd prevalence (TP) and herd prevalence (HTP) used for the shareware program 'HerdTest'.(Jordan and McEwen, 1998).	93
Table 5-2. Distributions of herds in 10 counties and all of Michigan in 1997.	94
Table 5-3. Herd-level sensitivity (HSe) and specificity (HSp) for four different scenarios with individual animal sensitivity (Se) and Specificity (Sp) used in 10 counties in the northern part of the lower peninsula in Michigan.	95
Table 5-4. Herd-level predictive value positive (HPVP) and negative (HPVN) for four different scenarios with individual animal sensitivity (Se) and Specificity (Sp) used in 10 counties in the northern part of the lower peninsula in Michigan.	96
Table 5-5. Spearman's rank correlation between herd-level input variables and herd-level output variables for Scenario 3. <i>p</i> -values are shown in brackets ($H_0: r_s=0$).	96
Table 6-1. Proportion of false positive (PFP) on caudal fold tuberculin test (CFT) in Michigan, by herd type and agricultural region.	119
Table 6-2. Herd size by herd type and agricultural region.	120
Table 6-3. Univariate analysis of factors associated with the proportion of false positive (PFP) on caudal fold tuberculin test (CFT) in Michigan.	124
Table 6-4. Multivariable analysis of factors associated with number of CFT suspect results.	125
Table 6-5. Univariate analysis of ecological factors associated with the proportion of false positive (PFP) on caudal fold tuberculin test (CFT) in Michigan.	127

LIST OF FIGURES

FIGURE	PAGE
Figure 5-1. Geographical location of the 10-county-area in northern Michigan.	90
Figure 5-2. Scenario 2 - Medians for herd level sensitivity (HSe), specificity (HSp), predictive value positive (HPVP), and negative (HPVN) for different fixed herd sizes using input values from scenario 2.	97
Figure 5-3. Scenario 3 - Medians for herd level sensitivity (HSe), specificity (HSp), predictive value positive (HPVP), and negative (HPVN) for different fixed herd sizes in scenario 3.	98
Figure 6-1. Observed and expected frequencies of false positive test results per farm. Neg-Bin: Negative Binomial.	113
Figure 6-2. Nine agricultural regions in Michigan.	114
Figure 6-3. Distribution of proportion of CFT false positive (PFP) by herd size.	121
Figure 6-4. Distribution of the mean proportion of CFT false positive test results (PFP) by veterinarian.	122
Figure 6-5. Spatial distribution of herds tested for bovine tuberculosis with the caudal fold tuberculin test in Michigan, including spatial clusters of herds with a high proportion of CFT false positive test results. • Individual herds. o Spatial clusters.	126

LIST OF ABBREVIATIONS

ABBREVIATION	MEANING
APHIS	Animal and Plant Health Inspection Service
bTb	Bovine tuberculosis, caused by <i>Mycobacterium bovis</i>
CCT	Comparative cervical tuberculin test
CFT	Caudal fold tuberculin test
CFTCCT _{SER}	CFT and CCT used in serial interpretation
CWP	Cell wall protein
CWP-PG	Cell wall protein peptoglycan
EM	Expectation maximization
GGE	Generalized estimation equations
HCSM	Heat concentrated synthetic medium (tuberculin)
HPVN	Herd-level predictive value negative
HPVP	Herd-level predictive value positive
HSe	Herd-level sensitivity
HSp	Herd-level specificity
<i>M. bovis</i>	<i>Mycobacterium bovis</i>
MDA	Michigan Department of Agriculture
MLE	Maximum likelihood estimation
NR	Newton-Raphson
OT	Koch's old tuberculin
PCR	Polymerase chain reaction

PPD	Purified protein derivative
PVN	Predictive value Negative
PVP	Predictive value positive
Se	Sensitivity
SICTT	Single intradermal comparative tuberculin test
SIT	Single intradermal test
SP	Synthetic peptide
Sp	Specificity
USDA	United States Department of Agriculture
VS	Veterinary Services

CHAPTER ONE

INTRODUCTION, RATIONALE, PROBLEM STATEMENT, GENERAL METHODS, and OVERVIEW

INTRODUCTION

In 1979, Michigan was declared free of bovine tuberculosis (bTb) caused by *Mycobacterium bovis*.(Corso et al., 1996) In 1994, bTb was identified in a single white-tailed deer,(Corso et al., 1996) and a bTb positive bovine was diagnosed in 1995. As of June, 2003, bTb has been diagnosed in 30 cattle herds, primarily in the Northeast corner of Michigan's lower peninsula. Bovine tuberculosis has furthermore been diagnosed in approximately 400 white-tailed deer, two elk, one domestic cat, four bobcats, 13 coyotes, two opossums, two raccoons, and two red foxes.(C. Bruning-Fann, personal communication 2002)

Bovine tuberculosis is an infection with *Mycobacterium bovis*, which affects a wide range of domestic animals, particularly cattle, goats, farmed and wild deer and several wildlife species.(Bruning-Fann et al., 2001) Bovine tuberculosis is also a concern to public health because humans may become infected by inhaling infected droplets or ingesting contaminated meat or milk. The disease is of substantial economic concern to livestock producers, because infected herds most commonly are depopulated for disease control purposes and national and international markets may be lost due to the quarantines and importation limitation. In addition, movement bans and rigorous testing schemes may be posed on herds in areas where bTb is present in the cattle population.

According to the United States Department of Agriculture (USDA), bTb is a reportable disease and a plan was implemented for testing all domestic cattle herds in Michigan with the caudal fold tuberculin test (CFT).(October 31, 2000, Amendment to the Animal Industry Act 466, State of Michigan.) Cattle with “suspect” responses on the CFT were retested with the comparative cervical tuberculin test (CCT). Animals classified as “reactors” on one CCT (or “suspect” on two successive CCT) were submitted to the Diagnostic Center for Population and Animal Health, Michigan State University, for euthanasia and necropsy. Diagnosis of bTb-infected animals was based on mycobacterial isolation and/or identification by polymerase chain reaction (PCR).

RATIONALE

The bovine tuberculosis (bTb) skin tests have been used successfully for control and eradication of bTb in the United States since the eradication program was begun in 1917.(Anonymous, 1992) In 1992, there were only 10 states that had not achieved bTb-free status.(Frey, 1995) Nevertheless, bTb has become endemic in the cattle population in certain areas in the United States, including the northern part of Michigan’s lower peninsula.(Kaneene et al., 2002) Two tuberculin skin tests are still being used in series to detect individual animals and herds that are infected with bTb. The caudal fold tuberculin test (CFT) is used to screen herds, and animals that respond to the CFT are retested with the comparative cervical test (CCT).(USDA and APHIS, 1999) Both the CFT and the CCT form the core of the bTb eradication program in Michigan.

It is important that state and federal veterinarians have knowledge of how well the tests they use for bTb eradication perform when a testing strategy is outlined, and it has been recommended that accuracy of the tuberculin skin tests be evaluated in the geographic area in which they are used and that the sensitivity and specificity should be evaluated in populations naturally infected with bTb.(Greiner and Gardner, 2000; O'Reilly, 1995) Furthermore, it may be helpful to regulatory veterinarians if they know of certain herd or ecologic/geographic factors that systematically alter the results on the tuberculin skin tests so that such factors can be taken into consideration when a herd has been tested for bTb and the status of the herd is being evaluated.

Several authors have reported estimates of the sensitivity (Se) and specificity (Sp) of these two skin tests (CFT and CCT) used in cattle.(Francis and Seller, 1978; Gaboric et al., 1996; Lepper et al., 1977; Leslie and Hebert, 1975; Leslie et al., 1975a; Leslie et al., 1975b; Pollock et al., 2000; Roswurm and Konyha, 1973; Whipple et al., 1995; Wood et al., 1991; Wood et al., 1992) However, the estimates vary considerably, and the variation in estimates of Se and Sp is most likely due to one or a combination of the following factors: 1) Differences in the concentration of the tuberculin used for the CFT, 2) different (“gold”) standards used for estimation of Se and Sp of the CFT, 3) bTb prevalence, 4) the amount of purified protein derivative (PPD) used for the skin tests, and 5) regional variation in cross reacting mycobacteria that cause false positive results on the CFT,(Cooney et al., 1997; Corner and Pearson, 1978a, 1978b), and 6) variation in the interpretation of the CFT by the testing veterinarians.(Kleeberg, 1960). In Michigan, mycobacterial culture (CULT) and polymerase chain reaction (PCR) performed on histologic suspect samples were considered to be the definitive test for diagnosis of bTb.

Although the performance of the skin tests have been estimated in the United States(Gaboric et al., 1996; Roswurm and Konyha, 1973; Whipple et al., 1995), it has not been determined exactly how the tests have performed during the current bTb outbreak in northeastern Michigan.

It has been stated that the tuberculin tests should be considered as herd tests rather than individual animals tests, because in an eradication program it is more important to detect infected herds than infected individual animals. Hence, the herd-level estimates of sensitivity (HSe) and specificity (HSp) of the serial testing program in Michigan may be of more relevance to regulatory veterinarians managing the eradication effort than Se and Sp. The herd-level performance of the tuberculin tests have not been estimated before.

PROBLEM STATEMENT

It should be clear that working estimates of both individual animal and herd-level estimates of the skin-tests as used in the eradication effort in Michigan would be of value to regulatory veterinarians both in evaluation of the current program and in planning for adjustments or evaluation of use of new tests used alone or in combination with the CFT and CCT in the future. Furthermore, factors affecting the outcome of the CFT and CCT may help guide regulatory veterinarians when such programs are developed and implemented.

STUDY OBJECTIVES

The overall objective of this dissertation is to procure tools that may be used by regulatory veterinarians to help better understand results obtained by the caudal fold and comparative cervical tuberculin tests. In order to accomplish this aim several research objectives needed to be addressed. They are:

- 1) Determine the sensitivity of the caudal fold tuberculin test (CFT) and the CFT and comparative cervical tuberculin tests used in serial interpretation (CFTCCT_{SER}) as well as gross necropsy using a gold standard approach.
- 2) Determine the sensitivity and specificity of the caudal fold tuberculin test (CFT) and the CFT and comparative cervical tuberculin tests used in serial interpretation (CFTCCT_{SER}) using methods that do not require the assumption of a gold standard.
- 3) Estimate the herd-level sensitivity and specificity of the of the caudal fold tuberculin test (CFT) and the CFT and comparative cervical tuberculin test when used in the bovine tuberculosis infected region of Michigan.
- 4) Determine herd, geographic, and ecologic factors that may affect the outcome of the CFT.

- 5) Determine if spatial clustering of herds with a high proportion of false positive results on the CFT were present in Michigan.

GENERAL METHODS

The objectives in this study were addressed using two data sets. The first two objectives were addressed using test results from 494 cattle in seven conveniently selected herds. The third objective was addressed by a simulation study, and the fourth and fifth objectives were addressed using CFT test results from 4,989 and 3,301 herds, respectively obtained from the Michigan Department of Agriculture.

The first dataset was a subset of 22 herds (2 dairy and 20 beef) that were identified as having at least one animal positive by bacterial isolation or PCR. Of these 22 herds, our study was limited to the 7 herds for which cattle were individually evaluated by each of six diagnostic tests: CFT, CCT, necropsy, histopathologic exam, PCR and bacteriologic culture. In the remaining 15 herds, only cattle that were suspects or reactors on the CCT were submitted for a full necropsy, histopathologic exam, PCR or culture, and hence they could not be used in this study. Of 516 cattle from the seven herds, 22 cattle had missing or erroneous laboratory records and were therefore excluded. Therefore, 494 cattle were included in this study.

The following information was available on all 494 individual animals: Herd ID, test results for all tests described above, testing veterinarian, age, and gender.

Mycobacterial isolation and PCR were used in parallel as the reference or definitive test for bTb. As such, cattle were designated as being truly infected with bTb if they were found positive either by culture or PCR assay.

The sensitivity of the CFT and CFTCCT_{SER} and gross necropsy was estimated using a gold standard method. Furthermore, the sensitivity and specificity of the CFT and CFTCCT_{SER} were estimated using latent-class models and Bayesian inference, assuming that a gold standard was not available.

The second data set consisted of all cattle herds that had a whole herd test for bTb between January 1996 and November 2001. During this period, 8,260 bTb negative cattle herds (4,223 beef and 4,037 dairy) underwent a 'whole herd test' with the CFT. Some herds were tested more than once during this period. These herds were all judged to have been negative for bTb because all cows were negative on the CFT or the CCT or cows positive on CCT were later concluded to be negative by necropsy, culture and PCR. For a herd to be included in this study, the testing private veterinary practitioner had to have tested at least 3 herds. Furthermore, the herd size in this study was limited to beef herds > 12 cattle and dairy herds > 22 cattle. After deletion of herds that did not fulfill these criteria, 4,989 herds were left over for analysis of associations between the proportion of false positive test results on the CFT in herds and herd factors and geographical location.

For the spatial analysis, a spatial coordinate was need for each farm. The exact spatial coordinates (latitude and longitude) of the tested herds were not available, so the geographic location of herds was obtained by mapping (geocoding) the address of the

herd to spatial data on roads in Michigan using ArcINfO (ESRI, Redlands, CA). The spatial reference data on roads in Michigan were obtained from the 2000 Census (<http://www.michigan.gov/cgi/>). Herds that had an address match $\geq 80\%$ to the reference address were considered an acceptable match. After geocoding, spatial coordinates (x,y coordinates, Michigan GeoRef) were obtained for 1,039 beef and 2,262 dairy herds, respectively. These herds ($n = 3,301$) were used in the spatial analysis, as well as to investigate associations between the proportion of false positive test results on the CFT in herds and elevation of the herd and whether a wetland area was located within a certain radius from the farm.

Associations between the proportion of false positive test results on the CFT in herds and herd factors, geographic, and ecologic factors were tested using a multiple factor negative binomial regression model treating the testing veterinarian as a random effect. The spatial scan statistic was used to identify possible clusters of herds with an unusually high proportion of false positive test results on the CFT.

OVERVIEW

A literature review focusing on the sensitivity and specificity of the caudal fold and comparative cervical tuberculin tests is presented in chapter two. Chapter three addresses objective 1 and presents estimates of the CFT, and CFT and CCT used in serial interpretation, and of gross necropsy using a gold standard approach. Similarly, chapter four addresses objective 3, and presents sensitivity and specificity estimates of the two skin tests using a novel Bayesian approach accounting for possible dependence between the tests. Objective 4 is addressed in chapter five presenting herd-level estimates for the two skin tests. Chapter six addresses objectives 4 and 5 by evaluating factors associated with the herd-level proportion of false positive results on the CFT.

REFERENCES

Anonymous 1992. Assessment of risk factors for *Mycobacterium bovis* in the United States (Fort Collins, CO, USDA, APHIS, VS).

Bruning-Fann, C.S., Schmitt, S.M., Fitzgerald, S.D., Fierke, J.S., Friedrich, P.D., Kaneene, J.B., Clarke, K.A., Butler, K.L., Payeur, J.B., Whipple, D.L., Cooley, T.M., Miller, J.M., Muzo, D.P., 2001, Bovine tuberculosis in free-ranging carnivores from Michigan. *J. of Wildlife Dis.* 37, 58-64.

Cooney, R., Kazda, J., Quinn, J., Cook, B., Müller, K., Monaghan, M., 1997, Environmental mycobacteria in Ireland as a source of non-specific sensitization to tuberculins. *Irish Vet. J.* 50, 370-373.

Corner, L.A., Pearson, C.W., 1978a, Pathogenicity in cattle of atypical mycobacteria isolated from feral pigs and cattle and the correlation of lesions with tuberculin sensitivity. *Austr. Vet. J.* 54, 280-286.

Corner, L.A., Pearson, C.W., 1978b, Response of cattle to inoculation with atypical mycobacteria of bovine origin. *Austr. Vet. J.* 54, 379-382.

Corso, B., Wagner, B., Norden, D. 1996. Assessing the risks associated with *M. bovis* in Michigan free-ranging deer. (Fort Collins, CO, Centers for Epidemiology and Animal Health, USDA, APHIS), pp. 1-2.

Francis, J., Seller, R.J., 1978, The sensitivity and specificity of various tuberculin tests using bovine PPD and other tuberculins. *Vet. Rec.* 103, 420-425.

Frey, G.H., 1995, Bovine tuberculosis eradication: The program in the United States, In: Thoen, C.O., Steele, J.H. (Eds.) *Mycobacterium bovis* infection in animals and humans. Iowa State University Press, Ames, IA, pp. 119-129.

Gaboric, C.M., Salman, M.D., Ellis, R.P., Triantis, J., 1996, Evaluation of five-antigen ELISA for diagnosis of tuberculosis in cattle and Cervidae. *JAVMA* 209, 962-966.

Greiner, M., Gardner, I.A., 2000, Epidemiologic issues in the validation of veterinary diagnostic tests. *Prev. Vet. Med.* 45, 3-22.

Kaneene, J.B., Bruning-Fann, C.S., Granger, L.M., Miller, R., Porter-Spalding, B.A., 2002, Environmental and farm management factors associated with tuberculosis on cattle farms in northeastern Michigan. *JAVMA* 221, 837-842.

Kleeberg, H.H., 1960, The tuberculin test in cattle. *JSAVMA* 31, 213-225.

- Lepper, A.W.D., Pearson, C.W., Corner, L.A., 1977, Anergy to tuberculin in beef cattle. *Aust. Vet. J.* 53, 214-216.
- Leslie, I.W., Hebert, C.N., 1975, Comparison of the specificity of human and bovine tuberculin PPD for testing cattle. 3 - National trial in Great Britain. *Vet Rec* 96, 338-341.
- Leslie, I.W., Hebert, C.N., Barnett, D.N., 1975a, Comparison of the specificity of human and bovine tuberculin PPD for testing cattle. 2 - Southeastern England. *Vet Rec* 96, 335-338.
- Leslie, I.W., Hebert, C.N., Burn, K.J., MacClancy, B.N., Donnelly, W.J.C., 1975b, Comparison of the specificity of human and bovine tuberculin PPD for testing cattle. 1 - Republic of Ireland. *Vet Rec* 96, 332-334.
- O'Reilly, L., 1995, *Tuberculin Skin Tests: Sensitivity and Specificity.*, First Edition. Iowa State Press, Ames, 355 p.
- Pollock, J.M., Girvin, R.M., Lightbody, K.A., Clements, R.A., Neill, S.D., Buddle, B.M., Andersen, P., 2000, Assessment of defined antigens for the diagnosis of bovine tuberculosis in skin test-reactor cattle. *Vet Rec* 146, 659-665.
- Roswurm, J.D., Konyha, 1973, The comparative cervical tuberculin test as an aid to diagnosing bovine tuberculosis. *Proc. Annu. Meet. U.S. Anim. Health Assoc.* 77, 368-389.
- USDA, APHIS 1999. Bovine tuberculosis eradication. Uniform methods and rules, effective January 22, 1999, (United States Department of Agriculture.), pp. 1-34.
- Whipple, D.L., Bolin, C.A., Davis, A.J., Jarnagin, J.L., Johnson, D.C., Payer, J.B., Saari, D.A., Wilson, A.J., Wolf, M.M., 1995, Comparison of the sensitivity of the caudal fold skin test and a commercial gamma-interferon assay for diagnosis of bovine tuberculosis. *AJVR* 56, 415-419.
- Wood, P.R., Corner, L.A., Rothel, J.S., Baldock, C., Jones, S.L., Cousins, D.B., McCormick, B.S., Francis, B.R., Creeper, J., Tweddle, N.E., 1991, Field comparison of the interferon-gamma assay and the intradermal tuberculin test for the diagnosis of bovine tuberculosis. *Austr. Vet. J.* 68, 286-290.
- Wood, P.R., Corner, L.A., Rothel, J.S., Ripper, J.L., Fifis, T., McCormick, B.S., Francis, B., Melville, L., Small, K., Witte, K.D., Tolson, J., Ryan, T.J., Lisle, G.W., Cox, J.C., Jones, S.L., 1992, A field evaluation of serological and cellular diagnostic tests for bovine tuberculosis. *Vet. Microbiol.* 31, 71-79.

CHAPTER TWO

LITERATURE REVIEW

The tuberculin skin tests have been used around the world for over a century in control and eradication of bovine tuberculosis (bTb) in cattle. Although changes in the production of tuberculins and in the implementation of the tests have occurred over the years, the tuberculin skin tests are still used with success today. The tuberculin skin tests have been reviewed recently by Monaghan *et al.* (1994). A review by Francis *et al.* (1978) focused in particular on the bovine purified protein derivative (PPD) and its use with the skin tests. A broader review of both *in vivo* and *in vitro* diagnostic tests for detection of *Mycobacterium bovis* infections in cattle was published by Adams (2001).

Tuberculins

Purified protein derivative (PPD) is the most commonly used tuberculin today, including in the United States, however heat concentrated synthetic medium (HCSM) tuberculin may be used in some countries. Originally Koch's old tuberculin (OT) was used in tuberculin tests. In the U.S. bovine tuberculin (PPD-B) is derived from *M. bovis* strain AN5 and avian tuberculin (PPD-A) is derived from *M. avium* strain D4. (Angus, 1978) The PPD tuberculins are a mixture of sugars, lipids, proteins, and nucleic acids and contain a variety of antigens that are common to several mycobacteria, (Monaghan *et al.*, 1994), hence cross reaction may occur to other mycobacteria when the skin test is performed. (Affronti, 1988) reviewed various attempts to isolate fractions from different

cellular components of the mycobacterium and their use as antigens in skin tests. Angus *et al* (1989) tested the use of *M. bovis* cell wall protein (CWP), a cell wall protein peptoglycan complex (CWP-PG) and a 13 amino acid synthetic peptide (SP) as potential test tuberculin. They found that the SP lacked sufficient tuberculin reactivity for use in cattle, and there was considerable response to cattle sensitized with *M. bovis* and *M. avium* for both CWP and CWP-PG. However, further research in this field is needed.

Tuberculin skin tests

Hypersensitivity to tuberculin usually occurs within 30 to 50 days after infection with tubercle bacilli.(Francis, 1947) The reaction to tuberculin injected intradermally is a delayed-type hypersensitivity reaction, which induces messenger RNA (mRNA) expression of gamma interferon (γ -INF), interleukin (IL) 2, IL4, IL10, and tumor necrosis factor-alpha (TNF α), and the reaction is maximal after approximately 72 hours.(Francis and Seller, 1978; Ng et al., 1995)

Two types of tuberculin skin tests are used in cattle; the single intradermal test (SIT) where bovine PPD (PPD-B) is injected intradermally and the single intradermal comparative tuberculin test (SICTT) using both avian PPD (PPD-A) and PPD-B. In the United States, the SIT is performed in the caudal fold of the tail and is referred to as the caudal fold tuberculin test (CFT). The SICTT is used as a follow up test on animal that were positive (suspect) on the CFT, and it is referred to as the comparative cervical tuberculin test (CCT).(USDA, 1999)

In the United States, the CFT is performed by USDA/APHIS/VS accredited veterinarians or public veterinarians by intradermal injection of 0.1 ml of USDA bovine

purified protein derivative (1 mg/ml PPD-B) into either side of the caudal fold. The injection site is read by visual observation and palpation 72 hours (plus or minus 6 hours) following injection. Cattle may not be subjected to CFT retest at intervals less than 60 days.(USDA, 1999) Suspects in herds containing only CFT suspects must be retested with the CCT within 10 days of the CFT injection, retested with the CFT after 60 days, or shipped to slaughter. The CCT may only be administered by state or federal veterinarians. Intradermal injection of biologically balanced PPD-B and PPD-A are performed at separate sites approximately 12.5 cm apart on the midcervical region. The skin thicknesses of the areas for injection of bovine and avian PPD are measured to the nearest 0.5 mm before the injection and again approximately 72 hours (plus or minus 6 hours) later. Comparative cervical test responses must be recorded on the CCT scattergram (VS FORM 6-22d), and the animal is classified as negative, suspect, or reactor.(USDA, 1999) Suspects on the CCT must be retested negative by the CCT after 60 days or shipped directly to slaughter. Reactors on the CCT must be shipped to slaughter and post mortem examination.(USDA, 1999)

When a whole herd is tested for bTb, all animals over the age of 24 months must be tested with the CFT.(USDA, 1999) However, in Michigan all animals 12 months or older were tested. A post mortem exam is performed on all animals that are reactors or suspects on two consecutive CCT tests. Animals receive humane euthanasia and all organs including medial retropharyngeal, submandibular, tracheobronchial, mediastinal, mesentary, ileocolic, and hepatic lymph nodes are examined for gross lesions consistent with bTb. Selected tissues are subject to histological exam and mycobacterial culture.(Schmitt et al., 1997) Tissues from animals that are suspect on histological exam

are tested with polymerase chain reaction (PCR).(Miller et al., 1997) If tissues are positive on PCR or mycobacterial culture, the animal is concluded to be infected with *M. bovis*. All other animals are considered to be negative for bTb. If an animal in a herd is positive for bTb, the herd is considered infected with bTb, and in most instances the herd is depopulated.

Sensitivity and specificity estimates for the CFT and CCT in the United States

Diagnostic, screening, and other tests are often evaluated by their sensitivity (Se) and specificity (Sp). However, the predictive values are of equal importance when the results of an individual test are evaluated. The sensitivity [P(T+|D+)] of a diagnostic test is defined as the conditional probability that a test positive animals (T+) is in fact truly infected (D+), i.e. the proportion of T+ animals that are truly infected. Likewise the probability that a test negative animal (T-) is in fact truly negative (D-), is termed the specificity [P(T-|D-)], i.e. the proportion of T- animals that are truly disease negative. On the other hand, the predictive values are the 'post test' probabilities. The predictive values are the probabilities that a disease positive (negative) animal is positive (negative) on a given test. Hence the predictive value positive (PVP) is the conditional probability of P(D+|T+), and the predictive value negative (PVN) is the conditional probability of P(D-|T-).(Greiner and Gardner, 2000)

If a perfect reference test, or a series of tests, (i.e. 'gold standard') correctly identifies all tested animals, the Se, Sp, PVP, and PVN of a new test or test with unknown accuracy, may be estimated as shown in Table 2-1. Sensitivity = $a/a+c$, Sp = $d/b+d$, PVP= $a/a+b$, and PVN = $d/c+d$

Table 2-1. Calculation of sensitivity and specificity when a gold standard is available.

		‘Gold Standard’		
		D+	D-	
New test	T+	a	b	a+b
	T-	c	d	c+d
		a+c	b+d	N=a+b+c+d

Often a gold standard is not available, which is the case when evaluating the true infection status of cattle in a population where infection with *Mycobacterium bovis* is present. However, the Se and Sp of a new test may also be evaluated against a test with known sensitivity and specificity, or the Se may be estimated if the reference test has a perfect specificity. (Gart and Buck, 1966; Staquet et al., 1981) New statistical methods, including maximum likelihood estimation (MLE), (Georgiadis et al., 1998; Hui and Walter, 1980; Singer et al., 1998; Walter and Irwig, 1988; Weng, 1996) and simulation based Bayesian inference (Johnson et al., 2001; Joseph et al., 1995) offer the researcher a way to estimate Se and Sp of two or more tests when a gold standard is not available. A review of these methods has been published by Enøe *et al.* (2000). Furthermore, these simulation based methods also allow the investigator to account for dependency between tests. (Black and B.A. Craig, 2002; Dendukuri and Joseph, 2001; Enøe et al., 2001; Georgiadis et al., 2003)

In the following review of published estimates of Se and Sp of tuberculin skin tests performed in the U.S., a direct comparison of these studies is not possible because the skin tests are evaluated against different 'levels' of gold standards. I was only able to identify two recent studies from 1995 and 1996, that estimate the sensitivity and specificity of the CFT,(Gaboric et al., 1996; Whipple et al., 1995) and an additional study estimating the Se and Sp of the CFT and CCT from 1973.(Roswurm and Konyha, 1973) The study by Whipple *et al.* (1995) also estimated the accuracy of a γ -INF assay and Gaboric *et al* (1996) additionally estimated the Se and Sp of a five-antigen enzyme-linked immuno sorbent assay (ELISA). The study by Roswurm *et al.* (1973) is particularly interesting because it is the only study performed in the U.S. that also estimates Se and Sp of the CCT.

Although the source is unknown, the USDA published estimates of the Se and Sp of the CFT and CCT in cattle.(Anonymous, 1992) The Se and Sp of the CFT and CCT were 82% and 96%, and 74% and 96%, respectively. It is likely that these results are partially derived from the work of Roswurm and Konyha (1973).

In the study by Whipple *et al.* (1995), 610 cattle from one dairy herd were tested with the CFT. They used four different reference test scenarios to determine the 'true' infection status of the cattle; 1) cattle positive by bacteriologic culture only, 2) histologic examination only, 3) bacteriologic culture and histologic examination (serial interpretation), and 4) bacteriologic culture or histologic examination (parallel interpretation). Using these four reference standards the Se of the CFT ranged from 80.4 to 84.4%.

Gaboric *et al.* (1996) used a sample of 4,504 cattle from 23 herds of which 4 dairy and 2 beef herds were considered infected with *M. bovis*. Of the 4,504 cattle, 4,167 (93%) were considered not to be infected. An animal was considered to be positive for bTb if it was from an infected herd and was positive on either bacteriologic culture or histologic examination. They estimated the Se and Sp of the CFT to be 84.9% and 89.2%, respectively. The Se and Sp of the 5-antigen ELISA are 65.6% and 56.4%, respectively. More interestingly, when the two tests were interpreted in parallel (i.e. an animal was positive on one or both of the tests) the Se and Sp were 95% and 50%, respectively. In serial interpretation (an animal was positive on both tests), the Se and Sp were 55.5% and 95.7%, respectively. This particular study shows that even though the sensitivity of the 5-antigen ELISA is fairly low (65.6%), use of the ELISA and CFT in parallel increases the combined Se dramatically.(Gaboric et al., 1996)

In both of the above studies, the CFT consisted of an intradermal injection of 0.1 ml USDA PPD-B (1 mg/ml).

Roswurm and Konyha (1973) estimated the Se and Sp of the CCT and Sp of the CFT in 3 bTb positive and 27 bTb negative publicly owned cattle herds from the U.S. The negative herds had at least a seven year history of being bTb free. For an animal to be considered positive for bTb, it had to originate from a herd in which *M. bovis* had recently been cultured, and have either 1) gross pathologic lesions typical of bTb, 2) histopathologic evidence of mycobacterial infection, or 3) have *M. bovis* cultured from infected body tissues. The Se of the CCT when administered 7 days after the CFT, was 74.4% and 88.5% when using CCT reactors only, and CCT reactors and CCT suspects in parallel to represent a positive test result, respectively. The Sp of the CFT was estimated

to be 94.1%. The Sp of the CCT performed on CFT suspect animals was 92% when using only CCT negative animals. If both CCT negative and suspect animals were considered to be test negative, the Sp of the CCT was 99.1%. The volume of PPD-B and PPD-A used for the testing in this study is somewhat unclear, however the PPD tuberculins were biologically balanced so they were of equal potency with PPD-B and PPD-A containing 0.996 mg and 1.128 mg of protein per ml, respectively.

These three studies used similar criteria to characterize a bTb positive animal and the Se and Sp estimates were very similar. However, it is likely that for all three studies some animals truly infected with *M. bovis* were not discovered. It is not possible to predict how this would have affected the Se estimates.

It is commonly considered that Se and Sp of diagnostic tests are constant parameters of a test. However, empirical evidence suggests that both Se and Sp may vary within and among populations of animals.(Greiner and Gardner, 2000) In cattle infected with bTb, it is known that the Se is lower in animals in the pre-allergic, inactive, and anergic stages.(Francis, 1947; Lepper et al., 1977) Hence in a population with a large proportion of such animals, the Se will be estimated lower than in a population with a low proportion of such animals. Whether this affected the estimates and uncertainty of Se and Sp presented above is unknown.

Factors affecting the sensitivity and specificity of the tuberculin skin tests

A logical question to ask is why do some bTb negative animals react positively on the CFT and CCT and why do some bTb positive animals react negatively on the skin

tests while others do not? Different factors may affect the test including: 1) the tester, 2) the tuberculin used, 3) the instrumentation, and 4) the animal.(Kleeberg, 1960)

Firstly, it takes 30-50 days from when an animal is infected until it will react to injected tuberculin.(Francis, 1947) Furthermore, some cattle may not react to tuberculin injection, including many cattle with generalized infection.(Lepper et al., 1977) This condition is known as anergy and is poorly understood.

A large area for variation in the outcome of a skin test lies with the testing veterinarian. Mistakes can be made in both injecting the PPD and reading the test. For example, the veterinarian performing the test may consistently inject the tuberculin subcutaneously rather than intradermally. The outcome of such incorrect injection of PPD, however, has not been documented.

Incorrect storage and handling of PPD may lead to a decrease in potency and hence an increase in false negative test results. Furthermore, bacterial contamination of the PPD may have a negative effect on the potency of the PPD.

Successful testing also relies on the choice of suitable instruments.(Kleeberg, 1960) It has been shown that multiple dose syringes may administer varying volumes.(Monaghan et al., 1994) For the CCT, it is imperative to use good calipers and thoroughly palpate the skin for nodules and other abnormalities where the PPD is injected. Furthermore, it is recommended to use coil spring activated calipers thus ensuring uniform pressure on the skin at each measurement.(Kleeberg, 1960)

Radunz and Lepper (1985) showed that repeat testing with a single comparative tuberculin test within 4 and 7 days of initial testing with the CFT decreased the response to the CCT in tuberculin-sensitized cattle. However, the sensitivity had returned to

sufficient level when re-tested at 60 days. Hence it is important that re-testing with the CCT is not performed too close in time to the CFT.

Furthermore, immunosuppression (i.e. in post partum cows) may affect the animals ability to mount an effective immune response and react to the tuberculin injection. Malnutrition may also affect the immune response however, a study on 20 bullocks did not support this claim.(Monaghan et al., 1994)

False positive reaction to the tuberculin skin tests also occur, and are most likely caused by cross-reacting mycobacteria other than *M. bovis*. However, because the specificity of the skin tests already is high, the hope for improvement is low.

Herd-level estimates of sensitivity and specificity

The Se and Sp of the tuberculin skin tests have been estimated in various studies around the world.(Gaboric et al., 1996; Lepper et al., 1977; Leslie and Hebert, 1975; Leslie et al., 1975a; Leslie et al., 1975b; Roswurm and Konyha, 1973; Whipple et al., 1995; Wood et al., 1991) However, it is to my knowledge unknown how the skin tests perform on the herd-level, i.e. herd-level sensitivity (HSe) and specificity (HSp). Herd-level test interpretation has recently been reviewed by Christensen and Gardner (2000). Factors influencing estimates HSe and HSp include animal-level Se and Sp, the sample size, the true prevalence of the disease within positive herds (TP), and the number of test-positive animals used to define a positive herd (cutoff).(Martin et al., 1992) Because Se, Sp and prevalence within a herd changes during a bTb eradication program, HSe and HSp will likely change too. During an eradication program, the within herd prevalence usually decreases, which leads to a decrease in HSe.(Christensen and Gardner, 2000)

However, the HSp should remain unchanged unless Sp changes, in which case HSp decreases with decreasing Sp and vice versa.

REFERENCES

- Adams, L.G., 2001, *In vivo* and *in vitro* diagnosis of *Mycobacterium bovis* infection. Rev. sci. tech. int. Epiz. 20, 304-324.
- Affronti, L.F., 1988, Mycobacterial antigens: reagents for tuberculin skin testing and serodiagnosis of tuberculosis. Plenum Press, New York, 1-37 pp.
- Angus, R.D., 1978, Production of reference PPD tuberculins for veterinary use in the United States. J. of Biol. Standard. 6, 221-227.
- Angus, R.D., Payeur, J.B., Elksen, L.A., 1989, Preliminary evaluation of an intradermal tuberculin test utilizing cell wall protein extracted from *M. bovis*. Proceedings of the U.S. Animal Health Association 93, 111-124.
- Anonymous 1992. Assessment of risk factors for *Mycobacterium bovis* in the United States (Fort Collins, CO, USDA, APHIS, VS).
- Black, M.A., B.A.Craig, 2002, Estimating disease prevalence in the absence of a gold standard. Statist. Med. 21, 2653-2669.
- Christensen, J., Gardner, I.A., 2000, Herd-level interpretation of test results for epidemiologic studies of animal diseases. Prev. Vet. Med. 45, 83-106.
- Dendukuri, N., Joseph, L., 2001, Bayesian approach to modeling the conditional dependence between multiple diagnostic tests. Biometrics 57, 158-167.
- Enøe, C., Andersen, S., Sørensen, V., Willeberg, P., 2001, Estimation of sensitivity, specificity and predictive values of two serologic tests for the detection of antibodies against *Actinobacillus pleuropneumoniae* serotype 2 in the absence of a reference test (gold standard). Prev. Vet. Med. 51, 227-243.
- Enøe, C., Georgiadis, M.P., Johnson, W.O., 2000, Estimation of sensitivity and specificity of diagnostic tests and disease prevalence when the true disease status is unknown. Prev. Vet. Med. 45, 61-81.
- Francis, J., 1947, Bovine tuberculosis, Including a contrast with human tuberculosis. Staple Press, London, 220 p.
- Francis, J., Seller, R.J., 1978, The sensitivity and specificity of various tuberculin tests using bovine PPD and other tuberculins. Vet. Rec. 103, 420-425.
- Gaboric, C.M., Salman, M.D., Ellis, R.P., Triantis, J., 1996, Evaluation of five-antigen ELISA for diagnosis of tuberculosis in cattle and Cervidae. JAVMA 209, 962-966.

- Gart, J., Buck, A., 1966, Comparison of a screening test and a reference test in epidemiologic studies. *Am J Epidemiology* 83, 593-602.
- Georgiadis, M.P., Gardner, I.A., Hedrick, R.P., 1998, Field evaluation of sensitivity and specificity of a polymerase chain reaction (PCR) for detection of *N. salmonis* in rainbow trout. *J. Aquat. Anim. Health* 10, 372-380.
- Georgiadis, M.P., Johnson, W.O., Gardner, I.A., Singh, R., 2003, Correlation-adjusted estimation of sensitivity of two diagnostic tests. *Appl. Statist* 52, 1-14.
- Greiner, M., Gardner, I.A., 2000, Epidemiologic issues in the validation of veterinary diagnostic tests. *Prev. Vet. Med.* 45, 3-22.
- Hui, S.L., Walter, S.D., 1980, Estimating the error rates of diagnostic tests. *Biometrics* 36, 167-171.
- Johnson, W.O., Gastwirth, J.L., Pearson, M., 2001, Screening without a "gold standard": The Hui-Walter Paradigm revisited. *Am. J. Epidemiol.* 153, 921-924.
- Joseph, L., Gyorkos, T., Coupal, L., 1995, Bayesian estimation of disease prevalence and the parameters of diagnostic tests in the absence of a gold standard. *Am. J. Epidemiol.* 153, 921-924.
- Kleeberg, H.H., 1960, The tuberculin test in cattle. *JSAVMA* 31, 213-225.
- Lepper, A.W.D., Pearson, C.W., Corner, L.A., 1977, Anergy to tuberculin in beef cattle. *Aust. Vet. J.* 53, 214-216.
- Leslie, I.W., Hebert, C.N., 1975, Comparison of the specificity of human and bovine tuberculin PPD for testing cattle. 3 - National trial in Great Britain. *Vet Rec* 96, 338-341.
- Leslie, I.W., Hebert, C.N., Barnett, D.N., 1975a, Comparison of the specificity of human and bovine tuberculin PPD for testing cattle. 2 - Southeastern England. *Vet Rec* 96, 335-338.
- Leslie, I.W., Hebert, C.N., Burn, K.J., MacClancy, B.N., Donnelly, W.J.C., 1975b, Comparison of the specificity of human and bovine tuberculin PPD for testing cattle. 1 - Republic of Ireland. *Vet Rec* 96, 332-334.
- Martin, S.W., Shoukri, M., Thorburn, M.A., 1992, Evaluating the health status of herds based on tests applied to individuals. *Prev. Vet. Med.* 14, 33-43.
- Miller, J., Jenny, A., Rhyan, J., Saari, D., Suarez, D., 1997, Detection of *Mycobacterium bovis* in formalin-fixed, paraffin-embedded tissues of cattle and elk by PCR amplification of an IS6110 sequence specific for *Mycobacterium tuberculosis* complex organisms. *J Vet Diagn Invest* 9, 244-249.

- Monaghan, M.L., Doherty, M.L., Collins, J.D., Kazda, J.F., Quinn, P.J., 1994, The tuberculin test. *Vet Micro* 40, 111-124.
- Ng, K.H., Watson, J.D., Prestidge, R., Buddle, B.M., 1995, Cytokine mRNA expressed in tuberculin skin test biopsies from BCG -vaccinated and *Mycobacterium bovis* inoculated cattle. *Immunol. Cell. Biol.* 73, 362-368.
- Radunz, B.L., Lepper, A.W.D., 1985, Suppression of skin reactivity to bovine tuberculin in repeat tests. *Aust. Vet. J.* 62, 191-194.
- Roswurm, J.D., Konyha, 1973, The comparative cervical tuberculin test as an aid to diagnosis of bovine tuberculosis. *Proc. Annu. Meet. U.S. Anim. Health Assoc.* 77, 368-389.
- Schmitt, S., Fitzgerald S., Cooley, T., Bruning-Fann, C., Berry, D., T., Minnis, R., Payer, J., Sikarskie, J, 1997, Bovine tuberculosis in free-ranging white-tail deer in Michigan. *J Wildlife Dis* 33, 749-758.
- Singer, R.S., Boyce, W.M., Gardner, I.A., Johnson, W.O., Fisher, A.S., 1998, Evaluation of bluetongue virus diagnostic tests in free-ranging bighorn sheep. *Prev. Vet. Med.* 35, 265-282.
- Staquet, M., Rozenzweig, M., Lee, M., Muggai, F.M., 1981, Methodology for assessment of new dichotomous diagnostic tests. *J. Chronic Dis.* 34, 599-610.
- USDA, APHIS. 1999. Bovine Tuberculosis Eradication: Uniform Methods and Rules, Effective January 22, 1999, (United States Department of Agriculture.), pp. 1-34.
- Walter, S.D., Irwig, L.M., 1988, Estimation of error rates, disease prevalence and relative risk from misclassified data: a review. *J. Clin. Epidemiol* 41, 923-937.
- Weng, T.S., 1996, Evaluation of a new diagnostic test against a reference test less than perfect in accuracy. *Commun. Statist. Simul. Comput.* 35, 533-555.
- Whipple, D.L., Bolin, C.A., Davis, A.J., Jarnagin, J.L., Johnson, D.C., Payer, J.B., Saari, D.A., Wilson, A.J., Wolf, M.M., 1995, Comparison of the sensitivity of the caudal fold skin test and a commercial gamma-interferon assay for diagnosis of bovine tuberculosis. *Am J Vet Res* 56, 415-419.
- Wood, P.R., Corner, L.A., Rothel, J.S., Baldock, C., Jones, S.L., Cousins, D.B., McCormick, B.S., Francis, B.R., Creeper, J., Tweddle, N.E., 1991, Field comparison of the interferon-gamma assay and the intradermal tuberculin test for the diagnosis of bovine tuberculosis. *Austr Vet J* 68, 286-290.

CHAPTER THREE

THE SENSITIVITY OF GROSS NECROPSY, CAUDAL FOLD AND COMPARATIVE CERVICAL TESTS FOR THE DIAGNOSIS OF BOVINE TUBERCULOSIS.

ABSTRACT

Bovine tuberculosis was diagnosed in 22 cattle herds in the Northeast corner of Michigan's lower peninsula. Of these 22 herds, 494 animals in 7 herds were examined by gross necropsy, histopathologic exam, mycobacterial culture, and polymerase chain reaction assay performed only on samples that were histologically compatible for bTb. Results of culture and polymerase chain reaction assay interpreted in parallel were used as the reference test for calculation of the sensitivity of 1) the caudal fold test, 2) the caudal fold and comparative cervical skin tests used in series, and 3) gross necropsy. *Mycobacterium bovis* was isolated from 43 animals. Using all seven herds, the sensitivities of the caudal fold test, the caudal fold and comparative cervical tests used in series, and gross necropsy were 93.02%, 88.37%, and 86.05%, respectively. When the data were stratified by low and moderate prevalence herds, the sensitivities were 83.33%, 75.0%, and 83.33 in low prevalence herds and 96.77%, 93.55%, and 87.10% in moderate prevalence herds. The sensitivities of the two skin tests were slightly higher when 2 or more gross lesions were present, and the sensitivity of gross necropsy was significantly higher ($P = 0.049$) than when one gross lesion was present. The sensitivity of the caudal fold test was found to be notably higher than most estimates in other studies, however a direct comparison was not possible because the amount of purified protein derivative and the reference methods were different in this study as compared to other published studies. Though the sensitivities are high, two of the seven herds (29%) would have had one or

more positive animals left in the herd if a test-and-removal program had been used. This suggests that when positive herds are identified, selective culling of skin test reactors is a less acceptable disease control strategy than is complete depopulation.

INTRODUCTION

In 1979, Michigan was declared free of bovine tuberculosis (bTb) caused by *Mycobacterium bovis*.(Corso et al., 1996) In 1994, bTb was identified in a single white-tailed deer,(Corso et al., 1996; Schmitt et al., 1997) and a bTb positive bovine was identified in 1995. As of June, 2002, bTb has been diagnosed in 22 cattle herds, primarily in the Northeast corner of Michigan's lower peninsula. Bovine tuberculosis has furthermore been diagnosed in approximately 400 white-tailed deer, two elk, one domestic cat, four bobcats, 13 coyotes, two opossums, two raccoons, and two red foxes (C. Bruning-Fann, personal communication 2002).

Up until January 2002, 732,661 cattle in 13,689 herds have been tested with a serial skin testing protocol involving the caudal fold test (CFT) of all cattle and re-testing of suspects with the comparative cervical test (CCT). Tuberculin tests have been used successfully in the United States since the bovine tuberculosis control program was introduced in 1917.(Anonymous, 1992) In the U.S., "The caudal fold test is the official tuberculin test for routine use in individual cattle, bison or goats and herds of such animals where the tuberculosis status of the animals is unknown".(USDA and APHIS, 1999; USDA, 1999)) In Michigan, the CFT was used as the screening test and a program was undertaken in 1998-2002 to test all cattle in the state over the age of 12 months. In herds that had one or more CCT reactor(s) after the initial herd test, the CCT reactors(s) were removed and the remaining herd retested with the CFT. With herds that had any epidemiologic association with cattle from bTb positive herds, all cattle present in the bTb exposed herd were tested with the CFT after removal of the bTb exposed cattle.

Cattle that responded to the CFT and were from herds of unknown bTb status were retested with the CCT. Cattle that responded to the CFT and were from herds of positive bTB status were classified as reactors and removed from the herd.

Several authors have reported estimates of sensitivity and/or specificity for the CFT .(Francis and Seiler, 1978; Lepper et al., 1977; Leslie and Hebert, 1975; Leslie et al., 1975a; Leslie et al., 1975b; Pollock et al., 2000; Roswurm and Konyha, 1973a; Roswurm and Konyha, 1973b; Whipple et al., 1995; Wood et al., 1991; Wood et al., 1992) These estimates range from 41.7% to 94.9%. However, it was not possible to use these reports from the literature to estimate how the CFT would perform in MI, because of the wide variation in: 1) background mycobacterial fauna present in various geographic location, 2) bTb prevalence, and 3) the amount of purified protein derivative (PPD) used for the skin testing in different studies. It has therefore been recommended that the accuracy of tuberculin tests should be evaluated in the geographic area intended for its use, and that the sensitivity and specificity should be evaluated in populations naturally infected with bTb and in similar populations free from disease, respectively.(O'Reilly, 1995)

The objectives of this study were to estimate the sensitivities of the CFT, the CFT and the CCT used in series (CFTCCT_{SER}), and gross necropsy (necropsy) as used under Michigan conditions. The apparent individual-animal prevalence (AP) using CFT, CFTCCT_{SER}, and gross necropsy in affected herds were also compared. Results from this study will hopefully aid USDA, foreign countries and the state departments of agriculture in evaluating the risk of bTb infection in exported Michigan cattle as well as evaluating the effectiveness of the current bTb eradication effort in Michigan.

MATERIALS AND METHODS

Summary of state-wide testing procedure.

A plan was implemented for testing all domestic cattle herds in Michigan with the CFT (October 31, 2000, Amendment to the Animal Industry Act 466, State of Michigan.). Cattle with “suspect” responses on the CFT, as described below, were retested with the CCT. Animals classified as “reactors” on one CCT (or “suspect” on two successive CCT) were submitted to the Michigan State University, Diagnostic Center for Population and Animal Health for euthanasia and necropsy. Diagnosis of bTb-infected animals was based on mycobacterial isolation and/or identification by polymerase chain reaction (PCR).

CFT.

Public or private practicing veterinarians performed the CFT using an intradermal injection of 0.1 ml of USDA bovine PPD (1 mg/ml PPD)(Angus, 1978; USDA, 1999) in the caudal fold on either side of the base of the tail. Approximately 72 hours after initial injection, the injection site was palpated and examined visually by the same veterinarian who gave the injection 72 hr previously. The test was considered positive (“suspect”) if any sign of inflammation was detected (C. Bruning-Fann, personal communication 2002).

CCT.

Within 7 days of the reading of a suspect CFT result, a USDA or state veterinarian administered the CCT. The skin thicknesses of the areas for injection of bovine and avian PPD were measured to the nearest 0.5 mm before the injection and again approximately 72 hours later. The CCT utilized cervical intradermal injections of 0.1 ml of USDA bovine PPD and 0.1 ml of USDA avian PPD (biologically balanced, 1mg/ml PPD) approximately 12.5 cm apart from each other. The difference ('after' minus 'before') in skin-thickness was plotted on a scattergram used to interpret the CCT results as "negative", "suspect" or "reactor" (USDA, VS Form 6-22D). The underlying principle of test interpretation is that bTb-infected cattle have a greater response to bovine PPD in comparison to their response to avian PPD. For our analyses, we considered CCT suspects and reactors as both having a positive CCT result. Individual animals positive on both CFT and CCT were euthanized and necropsied for further testing as described below.

Data selection.

As of December 31, 2001, 732,661 cattle in 13,689 herds in Michigan had been tested with the CFT and CCT. Twenty-two of these herds (2 dairy and 20 beef) were identified as having at least one animal positive by bacterial isolation or PCR. Of these 22 herds, our study was limited to the 7 herds for which cattle were individually evaluated by each of four diagnostic tests: necropsy, histopathologic exam, PCR and bacteriologic culture. In the remaining 15 herds, only cattle that were suspects or reactors on the CCT were submitted for a full necropsy, histopathologic exam, PCR or culture, and hence they

could not be used in this study. Of 22 herds, 20 were depopulated, and two herds elected a quarantine and testing program.

Gross necropsy.

For animals in the seven herds with complete data, a pathologist performed necropsy and histological examination on all cattle submitted to the Diagnostic Center for Population and Animal Health (DCPAH), Michigan State University. Animals received humane euthanasia and all organs including medial retropharyngeal, submandibular, tracheobronchial, mediastinal, mesentary, ileocolic, and hepatic lymph nodes were examined for gross lesions consistent with bTb. In this study, an animal was considered positive on necropsy if 1 or more lymph nodes or other tissues contained focal or multifocal abscesses or granulomas.

Histopathology.

Lymph nodes were separated into three groups; head, chest and abdomen. Samples for histopathologic exam were examined by DCPAH and the National Veterinary Services Laboratories (NVSL), Ames, Iowa. Formalin fixed tissues were embedded in paraffin and cut into 6 µm sections. Histopathologic slides were prepared using standard Hematoxylin eosin (HE) staining methods.(Prophet et al., 1992) Ziehl-Neelson staining for the detection of acid fast bacteria was performed on tissue that was suspect for bTb on regular HE stained samples. A sample was considered positive for bTb if there was evidence of granulomatous inflammation associated with focal necrosis

or mineralization or if identification of acid-fast bacteria was made on the Ziehl-Neelson stain.

Mycobacterial culture and identification.

Fresh samples were collected from all animals and cultured at NVSL. For six animals, tissues were also cultured at the Michigan Department of Community Health (MDCH). The culture procedures used at both laboratories have been described previously.(Schmitt et al., 1997) After a series of biochemical tests used for the identification of mycobacterial isolates, a genetic probe (Genetic probe, Accuprobes, Gen-probe, San Diego, CA) was applied to cultures of *Mycobacterium* spp. as soon as growth was evident on weekly examinations.(Fitzgerald et al., 2000) The genetic probe is specific to the *M. tuberculosis* complex in that it will react to *M. tuberculosis*, *M. africanum*, *M. microti*, and *M. bovis*, but will not react to *M. avium* spp. or environmental mycobacterial species.(Reisner et al., 1994)

Polymerase chain reaction assay.

Samples that were suspect of bTb on histological examinations were submitted for testing by polymerase chain reaction assay (PCR) as previously described.(Miller et al., 1997) In summary, formalin-fixed, paraffin-embedded tissue was used for DNA extraction. To identify *M. tuberculosis* complex mycobacteria, the primers amplified a 123-bp fragment of insertion sequence IS6110.(Eisenach et al., 1990) Positive samples had a band of the expected size as compared with control DNA.

Statistical analysis.

Mycobacterial isolation and PCR (Isol-PCR) were used in parallel as the reference or definitive test for bTb. As such, cattle were designated as being truly infected with bTb if they were found positive by culture, PCR examination, or both. The apparent individual animal prevalence $P(T+)$ was calculated as the proportion of tested animals that were test positive ($T+$). The AP for all seven herds was estimated for CFT, CFTCCT_{SER}, necropsy and Isol-PCR (reference). The sensitivity was calculated as the conditional probability of a test positive result given that the animal was truly bTb-infected ($P(T+|D+)$). Apparent prevalence and sensitivity were reported as point estimates with exact 95% confidence interval for binomial proportions. The chi-square test or Fisher's exact test was used to compare apparent prevalence among tests, and prevalence and sensitivity between high and low prevalence herds. The Fisher exact test was used if an observed cell value was lower than five (5). The McNemar test was used to compare sensitivity between tests within strata (all, moderate and low prevalence herds). Results are shown in aggregate and stratified by factors such as moderate/low prevalence, and number of gross lesions. Commercial software (Proc Freq, SAS Institute, Cary, NC) was used for data analyses.

RESULTS

Of 516 cattle from the seven herds, 22 cattle had missing or erroneous laboratory records and were therefore excluded. Therefore, 494 cattle were included in this study. Forty-three animals were infected according to the reference test (Isol-PCR), which was

used as the definitive test in the calculation of sensitivity. Of the 43 infected animals, seven were positive by culture only, five by PCR only and 31 were positive on both culture and PCR.

Table 3-1 shows the relationship between test results on the CFT, CFTCCT_{SER}, necropsy and the isolation of different mycobacterial species. *Mycobacterium* spp. were isolated from 47 of the 494 animals included in the study. *Mycobacterium bovis* was isolated from 38 animals of which two animals also had Runyon Group IV mycobacterium isolated (Table 3-1). In addition, *M. fortuitum*, *M. smegmatis* and mycobacteria from the *M. avium* and *M. terrae* complexes were isolated (Table 3-1).

The apparent prevalence based on the CFT, CFTCCT_{SER}, and necropsy is shown in Table 3-2. The overall apparent prevalence based on the reference test, CFT, CFTCCT_{SER}, and necropsy were 8.70%, 15.8 %, 9.11%, and 8.70, respectively. Overall for all seven herds, the prevalence by CFT was significantly higher ($P < 0.005$) than for the other three tests. Herds 2 and 7 had significantly higher bTb prevalence than did the other five herds, as measured by all three tests ($P < 0.001$) (Table 3-2).

The caudal fold test had the highest sensitivity (93.02%) of the three tests (Table 3-3). When stratified by low and moderate prevalence herds, the sensitivity for the CFT and gross necropsy were lower for the low prevalence herds than for the high prevalence herds, although this difference was not statistically significant. Likewise, the sensitivity of CFTCCT_{SER} was lower (75.0%) in low prevalence herds than it was in high prevalence herds (93.55%). When the sensitivities of the three tests were stratified by the number of affected lymph nodes in the animal, the sensitivity of the CFT, CFTCCT_{SER} were slightly higher when two or more affected lymph nodes were detected, and the

sensitivity of gross necropsy was significantly higher ($P = 0.049$) when two or more lesions were present (Table 3-4).

Herds 4, 6, and 7 each had one animal with an apparent false-negative test result on the CFT and herds 4 and 7 each had one apparent false-negative animal on the CFTCCT_{SER}.

Table 3-1. Relationship between mycobacterial isolates from cattle and skin tests, gross necropsy, and the reference test (positive on either mycobacterial isolation or PCR).

Isolate(s)	Number positive	Positive/negative on CFT	Positive/negative on CFTCT (of 36 ^a)	Positive/negative on gross necropsy	Positive/negative on reference test
<i>M. bovis</i>	36	35/1	33/2	32/4	36/0
<i>M. bovis</i> and Runyon Group IV	2	0/2	0/0	1/1	2/0
<i>M. fortuitum</i>	3	1/2	0/1	1/2	1/2 ^b
Runyon Group IV	2	0/2	0/0	0/2	0/2
<i>M. avium</i> comp. ^c	2	0/2	0/0	0/2	0/2
<i>M. terrae</i> comp. ^c	1	0/1	0/0	0/1	0/1
<i>M. smegmatis</i>	1	0/1	0/0	0/1	0/1
Total	47	36/11	33/3	34/13	39/8

^aOnly the 36 animals that were positive on the CFT were tested with the CFTCT_{SER}.

^b*M. bovis* positive on PCR.

^cComplex

Table 3-2. Apparent bTb prevalence in seven herds using the caudal fold, comparative cervical and reference tests (positive on culture and/or PCR).

Herd	Obs. ³	APP ⁴ CFT (%)	Exact 95% CI (%)	APP ⁴ CFTCT _{SER} (%)	Exact 95% CI (%)	APP ⁴ Necropsy (%)	Exact 95% CI (%)	APP ⁴ Reference test (%)	Exact 95% CI (%)
1	47	6.38	1.34-17.54	2.13	0.05-11.29	4.26	0.52-14.54	2.13	0.05-11.29
2	13	53.85	25.13-80.78	23.08	5.04-53.81	15.38	1.92-45.45	23.08	5.04-53.81
3	20	15.00	3.21-37.89	5.00	0.13-24.87	5.00	0.13-24.87	5.00	0.13-24.87
4	127	6.30	2.76-12.03	3.94	1.29-8.95	5.51	2.24-11.03	5.51	2.24-11.03
5	27	14.81	4.19-33.73	3.70	0.09-18.97	3.70	0.09-18.97	3.70	0.09-18.97
6	174	9.77	5.80-15.18	1.72	0.36-4.96	1.15	0.14-4.09	1.15	0.14-4.09
7	86	38.37	28.08-49.49	36.05	25.97-47.12	32.56	22.84-43.52	32.56	22.84-43.52
Total ¹	494	15.18 ^b	12.19-18.66	9.11 ^a	6.72-12.00	8.70 ^a	6.37-11.55	8.70 ^a	6.37-11.55
‡2, 7	99	40.40	30.66-50.74	34.34	25.09-44.56	30.30	21.47-40.35	31.31	22.36-41.41
*1,3,4,5,6 ²	395	8.86 ^b	6.25-12.11	2.78 ^a	1.40-4.93	3.29 ^a	1.76-5.56	3.04 ^a	1.58-5.25

¹In the row for overall prevalence by the four tests, estimates with different superscript are significantly different at p=0.05 (chi-square test). ²In the row for low prevalence herds by the four tests, estimates with different superscript are significantly different at p=0.05 (chi-square test). ³Number of animals. ⁴Apparent prevalence. ⁵Herds with moderate bovine tuberculosis apparent prevalence. * Herds with low bovine tuberculosis apparent prevalence

Table 3-3. Sensitivity of CFT, CFTCCT_{SER}, and gross necropsy using bacteriologic culture and PCR as the reference test.

All herds					
Test	(T+,D+)	(T-,D+)	(D+)	Sensitivity (%)	Exact 95% CI
CFT	40	3	43	93.02	80.94-98.54
CFTCCT _{SER}	38	5	43	88.37	74.92-96.11
Necropsy	37	6	43	86.05	72.07-94.70

Low prevalence herds					
Test	(T+,D+)	(T-,D+)	(D+)	Sensitivity (%)	Exact 95% CI
CFT	10	2	12	83.33	51.59-97.91
CFTCCT _{SER}	9	3	12	75.00	42.81-94.51
Necropsy	10	2	12	83.33	51.59-97.91

Moderate prevalence herds					
Test	(T+,D+)	(T-,D+)	(D+)	Sensitivity (%)	Exact 95% CI
CFT	30	1	31	96.77	83.30-99.92
CFTCCT _{SER}	29	2	31	93.55	78.58-99.21
Necropsy	27	4	31	87.10	70.17-96.37

T+, T-, D+, and D- denotes test (T+, T-) and infection (D+, D-) positive and negative outcome, respectively. (T+,D+), (T-,D+), and (D+) represent true positive, false negative, and the total number of true positive animals, respectively.

None of the sensitivity estimates within any of the three strata (All herds, low and moderate prevalence) were significantly different.

The estimates of sensitivity for CFT, CFTCCT_{SER}, and necropsy were not significantly different between the low and high prevalence herds.

Table 3-4. Sensitivity of CFT, CFTCCT_{SER}, and gross necropsy in animals with lesions in one, or two or more lymph nodes.

Screening tests	Number of lesions	(T+,D+)	(T-,D+)	(D+)	Sensitivity (%)	Exact 95% CI
CFT	1	25	2	27	92.59	75-71-99.09
	>1	15	1	16	93.75	69.77-99.84
CFTCCT _{SER}	1	23	4	27	85.19	66.27-95.81
	>1	15	1	16	93.75	69.77-99.84
Necropsy	1	21	6	27	77.78 ^a	57.74-91.38
	>1	16	0	16	100 ^b	79.41-100

Values with different superscript are significantly different.

DISCUSSION

In this study, bacteriologic culture and/or PCR were used as the reference test (i.e. definitive test) to detect animals truly infected with bovine tuberculosis. Both of these tests ultimately use a genetic probe that is specific to the *Mycobacterium tuberculosis* complex,(Eisenach et al., 1990; Reisner et al., 1994) which includes *M. tuberculosis*, *M. africanum*, *M. microti*, and *M. bovis*. Barring laboratory error, mycobacterial isolation and PCR are widely accepted as 100% specific tests.(Eisenach et al., 1990; Reisner et al., 1994) It was therefore assumed that all animals positive on the reference test (Isol-PCR) were truly bTb-infected animals; i.e. the specificity of Isol-PCR was assumed to be 100%. This assumption allowed us to calculate unbiased estimates of sensitivity for the

three screening tests; necropsy, CFT and CFTCCT_{SER}.(Staquet et al., 1981) Furthermore, it was assumed that the proportion of false negative animals on the screening test(s) were the same for truly infected animals that were positive or negative on the reference test. This assumption was necessary to presume that sensitivity estimates were unbiased. Both mycobacterial isolation and PCR examination may expectedly fail to detect infected animals if lesions were not present (early infection) or lesions were present in tissues that were not collected.(Miller et al., 1997; O'Reilly, 1995) Likewise, this may explain why some animals were positive on culture and negative on PCR and vice versa. Even though the reference test is assumed to have perfect specificity, the sensitivity of Isol-PCR was probably less than perfect and false negative results were likely to have occurred. Therefore specificity and predictive values were not calculated for the screening tests being evaluated in this study.

The bTb infected herds were stratified into moderate prevalence herds ($\geq 35\%$; herds 2 and 7) and low prevalence herds ($< 35\%$; herds 1, 3, 4, 5, and 6) and these two strata presented two distinct epidemiologic patterns. Herds with bTb low prevalence had few lesions per infected animal. This may indicate recent transmission of the disease to a small number of cattle, most likely from the wild deer reservoir. Herds with moderate bTb prevalence suggest that some cattle had been infected for a long period of time thereby allowing time and opportunity for cow-to-cow transmission to occur.

The sensitivity $[P(T+|D+)]$ of a diagnostic test is defined as the conditional probability that a test positive animals (T+) is in fact truly infected (D+), i.e. the proportion of T+ animals that are truly infected. It is commonly assumed that sensitivity and specificity of diagnostic tests are unchanging properties of the diagnostic test itself

and thus are independent of the circumstances of its application (i.e. the sensitivity and /or specificity do not change among different groups of animals). This may not always hold true, as animals in different stages of a disease or under different circumstances may yield different test results.(Donald, 1993; Donald et al., 1994; Greiner and Gardner, 2000) This may explain what was observed in comparing the test sensitivities in the low prevalence herds to the moderate prevalence herds. Also, the prevalence of cross-reacting organisms may vary greatly across different ecological environments. For these reasons, it is important that the performance of the various diagnostic tests for bTb be evaluated under the conditions that the tests will be used under.(O'Reilly, 1995)

Though not statistically significant, this study suggests that the sensitivity of the CFT might be higher than that of the CFTCCT_{SER} and gross necropsy (Table 3-3). Since the same amount of bovine PPD was used for both the CFT and CFTCCT_{SER} it would be expected that the sensitivity would be almost the same for the two test scenarios. This trend of higher sensitivity for the CFT may be a function of the location on body on which the test is performed, but more likely is due to the CCT being a function of the difference in reaction between the bovine and avian antigens, rather than just a function of the bovine antigen alone. Also, the CCT as performed by federal and state veterinarians, may be read much more carefully and quantitatively than is the CFT applied by private practitioners screening large numbers of cattle. However, further studies would be needed to investigate this phenomenon.

The sensitivity of the CFT in this study is notably higher than most estimates in other studies.(Francis and Seiler, 1978; Lepper et al., 1977; Pollock et al., 2000; Whipple et al., 1995; Wood et al., 1991; Wood et al., 1992) However, a direct comparison among

studies reported in the literature is not possible because different reference tests and preparations and concentrations of PPD were used. A study of the CCT (Roswurm and Konyha, 1973a; Roswurm and Konyha, 1973b) estimated the sensitivity of the CCT to be 74.36% and 88.46% when using 'reactor' and 'reactors' and 'suspects' as test positives, respectively. That protocol was similar to the protocol used in this study, as was their estimate of CFTCCT_{SER} sensitivity.

The sensitivity of gross necropsy is often used (together with other tests) as part of a reference test. (Lepper et al., 1977; Pollock et al., 2000) In this study, the reference test was assessed independently from gross necropsy findings, which allowed us to estimate the sensitivity of gross necropsy as a diagnostic assay. The sensitivity of necropsy (86.05%) was almost as high as CFTCCT_{SER} (88.37%). Furthermore, the sensitivity of gross necropsy increased to 100% when 2 or more lesions were present. Such animals with multiple lesion sites would expectedly shed greater numbers of infective *Mycobacterium bovis* and thus would be a more important source of disease transmission. The gross post-mortem necropsy exam can perhaps be viewed as the "upper detection limit" of any practical slaughter-based system for bTb.

False negative results were most infrequently found with the CFT, followed by the CFTCCT_{SER} and then by gross necropsy. Several explanations have been suggested for this inability of the skin tests to detect infected animals, including anergic cattle, poorly injected PPD, variance in potency of PPD, loss of potency of PPD over time, mistaken animal identification, and operator error. (Lepper et al., 1977; Monaghan et al., 1994)

The focus of our study was not to investigate the specificity (or proportion of false-positive results) of the screening tests, however it is worth noting that only one animal out of 11 animals infected with atypical mycobacteria reacted to the CFT (Table 3-1, column 3). This animal was infected with *M. fortuitum*. No false positive results were observed on the CFTCCT_{SER}. Possible explanations for why the CFT was positive and the CFTCCT_{SER} was negative on an animal infected with *M. fortuitum* might relate to a higher specificity of the CFTCCT_{SER} or the animal being co-infected with *M. bovis*. Also, a greater reaction on the avian component of the CCT (including cross-reactions) tend to make the CCT yield negative results. Six bTb-infected animals had no detectable lesions on necropsy. This may be due to early infection (insufficient time for lesion development) or inhibition of lesion growth by the hosts' immune system.

It is well accepted that the CFT and CFTCCT_{SER} are more suited to the detection of infected herds rather than the detection of infected cattle, i.e. some bTb-infected animals will escape detection when the skin tests are used to assess the infection status of individual animals. In this study, bTb positive animal would have been undetected in two of the seven infected herds (29%) if only the CFTCCT_{SER} reactor animals were culled rather than depopulating the entire herd. This finding supports a disease control policy of total herd depopulation over a test and slaughter approach, because findings in this study suggest the test and slaughter approach would fail to detect all infected animals in approximately 29% of the herds.

REFERENCES

- Angus, R., 1978, Production of reference PPD tuberculins for veterinary use in the United States. *J Biol Stand* 6, 221-227.
- Anonymous 1992. Assessment of risk factors for *Mycobacterium bovis* in the United States (Fort Collins, Colorado USA, USDA, APHIS, VS).
- Corso, B., Wagner, B., Norden, D. 1996. Assessing the risks associated with *M. bovis* in Michigan free-ranging deer. (Fort Collins, Co, Centers for Epidemiology and Animal Health, USDA, APHIS), pp. 1-2.
- Donald, A., 1993, Prevalence estimation using diagnostic tests when there are multiple, correlated disease states in the same animal or farm. *Prev. Vet. Med.* 15, 125-145.
- Donald, A.W., Gardner, I.A., 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.* 19, 167-187.
- Eisenach, K., Cave, M., Bates, J., Crawford, J., 1990, Polymerase chain reaction amplification of a repetitive DNA sequence specific for *Mycobacterium tuberculosis*. *J Infect Dis* 161, 977-981.
- Fitzgerald, S., Kaneene, J., Butler, K., Clarke, K., Fierke, J., Schmitt, M., Bruning-Fann, C., Mitchell, R., Berry, D., Payer, J., 2000, Comparison of postmortem techniques for the detection of *Mycobacterium bovis* in white-tailed deer (*Odocoileus virginianus*). *J Vet Diagn Invest* 12, 322-327.
- Francis, J., Seiler, R.J., 1978, The sensitivity and specificity of various tuberculin tests using bovine PPD and other tuberculins. *Vet Rec* 103, 420-425.
- Greiner, M., Gardner, I.A., 2000, Epidemiologic issues in the validation of veterinary diagnostic tests. *Prev. Vet. Med.* 45, 3-22.
- Lepper, A.W.D., Newton-Tabrett, D.A., Carpenter, M.T., Williams, O.J., 1977, The use of bovine PPD tuberculin in the single caudal fold test to detect tuberculosis in beef cattle. *Austr Vet J* 53, 208-213.
- Leslie, I.W., Hebert, C.N., 1975, Comparison of the specificity of human and bovine tuberculin PPD for testing cattle. 3 - National trial in Great Britain. *Vet Rec* 96, 338-341.
- Leslie, I.W., Hebert, C.N., Barnett, D.N., 1975a, Comparison of the specificity of human and bovine tuberculin PPD for testing cattle. 2 - Southeastern England. *Vet Rec* 96, 335-338.

- Leslie, I.W., Hebert, C.N., Burn, K.J., MacClancy, B.N., Donnelly, W.J.C., 1975b, Comparison of the specificity of human and bovine tuberculin PPD for testing cattle. 1 - Republic of Ireland. *Vet Rec* 96, 332-334.
- Miller, J., Jenny, A., Rhyan, J., Saari, D., Suarez, D., 1997, Detection of *Mycobacterium bovis* in formalin-fixed, paraffin-embedded tissues of cattle and elk by PCR amplification of an IS6110 sequence specific for *Mycobacterium tuberculosis* complex organisms. *J Vet Diagn Invest* 9, 244-249.
- Monaghan, M.L., Doherty, M.L., Collins, J.D., Kazda, J.F., Quinn, P.J., 1994, The tuberculin test. *Vet Micro* 40, 111-124.
- O'Reilly, L., 1995, Tuberculin Skin Tests: Sensitivity and Specificity., First Edition. Iowa State Press, Ames, 355 p.
- Pollock, J.M., Girvin, R.M., Lightbody, K.A., Clements, R.A., Neill, S.D., Buddle, B.M., Andersen, P., 2000, Assessment of defined antigens for the diagnosis of bovine tuberculosis in skin test-reactor cattle. *Vet Rec* 146, 659-665.
- Prophet, E., Mills, B., Arrington, J., Sobin, L., eds, 1992, Laboratory methods in histotechnology. American Registry of Pathology, Washington, DC.
- Reisner, E., Gatson, A., Woods, G., 1994, Use of gen-probe accupobes to identify *Mycobacterium avium* complex, *Mycobacterium tuberculosis* complex, *Mycobacterium kansasii* and *Mycobacterium gordonae* directly from BACTEC TB broth cultures. *J Clin Microbiol* 32, 2995-2998.
- Roswurm, J.D., Konyha, 1973, The comparative cervical tuberculin test as an aid to diagnosis of bovine tuberculosis. *Proc. Annu. Meet. U.S. Anim. Health Assoc.* 77, 368-389.
- Schmitt, S., Fitzgerald, S., Cooley, T., Bruning-Fann, C., Berry, D., Carlson, T., Minnis, R., Payer, J., Sikarskie, J., 1997, Bovine tuberculosis in free-ranging white-tail deer in Michigan. *J Wildlife Dis* 33, 749-758.
- Staquet, M., Rozenzweig, M., Lee, M., Muggai, F.M., 1981, Methodology for assessment of new dichotomous diagnostic tests. *J. Chronic Dis.* 34, 599-610.
- USDA, A. 1999. Bovine Tuberculosis Eradication: Uniform Methods and Rules, Effective January 22, 1999, (United States Department of Agriculture.), pp. 1-34.
- Whipple, D.L., Bolin, C.A., Davis, A.J., Jarnagin, J.L., Johnson, D.C., Payer, J.B., Saari, D.A., Wilson, A.J., Wolf, M.M., 1995, Comparison of the sensitivity of the caudal fold skin test and a commercial gamma-interferon assay for diagnosis of bovine tuberculosis. *Am J Vet Res* 56, 415-419.

Wood, P.R., Corner, L.A., Rothel, J.S., Baldock, C., Jones, S.L., Cousins, D.B., McCormick, B.S., Francis, B.R., Creeper, J., Tweddle, N.E., 1991, Field comparison of the interferon-gamma assay and the intradermal tuberculin test for the diagnosis of bovine tuberculosis. *Austr Vet J* 68, 286-290.

Wood, P.R., Corner, L.A., Rothel, J.S., Ripper, J.L., Fifis, T., McCormick, B.S., Francis, B., Melville, L., Small, K., De Witte, K., Tolson, J., Ryan, T.J., de Lisle, G.W., Cox, J.C., Jones, S.L., 1992, A field evaluation of serological and cellular diagnostic tests for bovine tuberculosis. *Vet Micro* 31, 71-79.

CHAPTER FOUR

ESTIMATION OF SENSITIVITY AND SPECIFICITY OF BOVINE TUBERCULOSIS SKIN TESTS IN MICHIGAN WHEN A PERFECT REFERENCE TEST IS NOT AVAILABLE.

ABSTRACT

Four-hundred-and-ninety-four cattle from seven herds were tested with the caudal fold and comparative cervical tuberculin tests to detect infection with *Mycobacterium bovis* as part of the bovine tuberculosis eradication program in Michigan. Bayesian inference was used to estimate the sensitivity and specificity of the caudal fold tuberculin test and the comparative cervical tuberculin test in four different 'two-population-two-tests' latent-class models. The two tuberculin skin tests could not be used in the same model, hence the sensitivities and specificities of the two tuberculin skin tests were evaluated against results of mycobacterial culture. Possible dependence between the caudal fold and comparative cervical tests, and mycobacterial culture was investigated using a conditionally independence model and three models allowing for dependence between sensitivities and specificities. One of the dependence models assumed perfect specificity for mycobacterial culture. There was a statistically significant negative correlation between the comparative cervical tests and mycobacterial culture, however the dependence between the caudal fold test and mycobacterial culture was not significant. Assuming conditional dependence between both skin tests and mycobacterial culture, the Bayesian estimates and posterior 95% credible intervals of the sensitivity and specificity of the caudal fold and comparative cervical tests were 0.854 (0.563-0.975) and 0.939 (0.899-0.982), and 0.758 (0.475-0.930) and 0.986 (0.962-0.999), respectively. The

estimates of sensitivity and specificity obtained in this study generally agree with those reported from other studies in the United States.

INTRODUCTION

Bovine tuberculosis (bTb) is an infection with *Mycobacterium bovis*, which affects a wide range of domestic animals, particularly cattle, goats, farmed and wild deer and several wildlife species.(Bruning-Fann et al., 2001) Bovine tuberculosis is also a concern to public health because humans may become infected by inhaling infected droplets or ingesting contaminated meat or milk. The disease is of substantial economic concern to livestock producers, because infected herds most commonly are depopulated for disease control purposes and national and international markets may be lost due to quarantines and importation limitation. In addition, movement bans and rigorous testing schemes may be imposed on herds in areas where bTb is present in the cattle population.

Begun in 1917, the US bovine tuberculosis eradication program has successfully used the tuberculin skin tests to eradicate bTb from most areas of the United States.(Frey 1995) In 1995, a bovine with bTb was identified in Michigan, whereupon it was decided to test all cattle in Michigan with the caudal fold tuberculin test (CFT) (October 31, 2000, Amendment to the Animal Industry Act 466, State of Michigan). Up until January 2002, 732,661 cattle in 13,689 herds have been tested with a serial skin testing protocol (CFTCCT_{SER}) using the CFT as a screening test for all animals over the age of 12 months, with CFT-positive animals ('suspects') being re-tested with the comparative cervical tuberculin test (CCT).(Norby et al., 2003)

Several authors have reported estimates of the sensitivity (Se) and specificity (Sp) of these two skin tests (CFT and CCT) used in cattle.(Francis and Seller, 1978; Gaboric et al., 1996; Lepper et al., 1977; Leslie and Hebert, 1975; Leslie et al.,t 1975; Leslie et al.,

1975; Pollock et al. 2000; Roswurm and Konyha, 1973; Whipple et al., 1995; Wood et al., 1991; Wood et al., 1992) However, the estimates vary considerably, most likely because of the difference in 1) geographic location of the cattle and hence the background mycobacterial fauna present, 2) bTb prevalence, 3) the amount of purified protein derivative (PPD) used for the skin tests, and 4) the reference tests ('gold standards') the skin tests were evaluated against, and whether or not the tests were used independently, in series, or in parallel. Most laboratories consider mycobacterial culture (CULT) to be the definitive test for diagnosis of bTb, although polymerase chain reaction (PCR) is gaining rapid acceptance.

Both CULT and PCR are considered to have almost perfect specificity (very close to 1.0) because they use a genetic probe for final identification.(Miller et al., 1997; Schmitt et al., 1997) Using CULT and/or PCR allows for estimation of Se of the skin tests, but not the Sp of the skin tests.(Gart and Buck 1996; Staquet et al., 1981) In this scenario, it is not possible to obtain estimates of Sp of the skin tests.

Using, novel statistical methodologies, it possible to estimate the disease prevalences, Se, and Sp when each animal's true disease status is unknown.(Enøe, et al., 2000) Prevalence, Se, and Sp may be estimated using a latent-class analysis, such as maximum likelihood estimation using the Newton-Raphson algorithm(Georgiadis et al., 1998; Hui and Walter, 1980; Tanner 1996), the expectation-maximization algorithm(Dempster et al., 1977; Singer et al., 1998; Weng, 1996) or simulation-based Bayesian inference.(Dendukuri and Joseph, 2001; Enøe et al., 2001; Georgiadis et al., 2003; Johnson et al., 2001; Joseph et al., 1995)

If animals in one population are tested with two conditionally independent dichotomous tests, there are three independent cells (3 degrees of freedom) available to estimate five parameters (prevalence, Se and Sp of test 1, and Se and Sp of test 2 based on a likelihood estimation approach. Hence the model is unidentifiable, i.e. there are an infinite number of possible solutions to the model. However, data on two populations are available, there are six independent cells and, with the addition of another prevalence parameter, there are now six parameters to estimate, and hence a unique solution is possible.(Enøe et al., 2000; Hui and Walter, 1980; Johnson et al., 2001; Singer et al., 1998; Walter and Irwig, 1988; Weng 1996)

However, if a 'conditional independence model' is used when the diagnostic tests are correlated, biased estimates of prevalence, sensitivity and specificity are to be expected.(Black and Craig, 2002; Dendukuri and Joseph, 2001; Gardner et al., 2000; Georgiadis et al., 2003; Torrance-Rynard and Walter, 1997; Vacek, 1985) Several authors have proposed models allowing for pairwise dependence between tests using a classical frequentist approach.(Qu and Hadgu, 1998; Qu et al., 1996; Yang and Becker, 1997) In these models, at least four diagnostic tests are required to model pairwise dependence for Se and Sp.(Walter and Irwig, 1988) Bayesian approaches to estimate Se and Sp of two dependent tests applied to subjects in one population have been described.(Black and Craig, 2002; Dendukuri and Joseph, 2001). Bayesian models using test results from two dependent tests applied to two populations have also been proposed.(Georgiadis et al., 2003). Again, all parameters in the likelihood functions for these models are generally unidentifiable, i.e. there are eight parameters to estimate, but only six degrees of freedom. However, this dilemma can be overcome by assigning

informative prior information to some of the parameters when such information is available.(Black and Craig, 2002; Dendukuri and Joseph, 2001; Georgiadis et al., 2003)

The objectives of this study were to estimate Se and Sp of the CFT (scenario 1) and the CFT and CCT used in serial interpretation (CFTCCT_{SER}) (scenario 2). Bayesian models were used to estimate Se and Sp of the two testing scenarios applying CULT as the second test for both scenarios. An expansion of the model proposed by Black and Craig (2002) was used to model possible dependence between tests. Results from this model were compared to a conditional independence model, a model assuming perfect specificity for CULT, and a model proposed by Georgiadis et al., (2003).

MATERIALS AND METHODS.

Animals

As of June 2003, 30 cattle herds in Michigan had been diagnosed with bovine tuberculosis since the start of the current outbreak in 1995. Cattle in seven of these herds were used in this study because all cattle in these herds were individually and independently evaluated by both CFTCCT_{SER} and CULT.(Norby et al., 2003) In the remaining 23 bTb positive cattle herds, only cattle that were ‘suspects’ or ‘reactors’ on the CCT were submitted for full necropsy and CULT; hence they were not eligible for this study.

Diagnostic tests and testing procedure.

The statewide testing procedure and implementation of the CFT and CCT, and CULT has been described elsewhere.(Norby et al., 2003) In short, the CFT was performed by public or private practicing veterinarians, using an intradermal injection of 0.1 ml of USDA bovine PPD (1mg/ml) in the caudal fold on either side of the base of the tail.(USDA and APHIS, 1999) The injection site was palpated approximately 72 hours later, and the test was considered positive (“suspect”) if there was any sign of inflammation.(Bruning-Fann, personal communication, 2002) If the CFT was “suspect”, a USDA or state veterinarian administered the CCT within 7 days of when the CFT was read. For the CCT, 0.1 ml of USDA bovine PPD and 0.1 ml of USDA avian PPD (biologically balanced, 1 mg/ml) were injected intradermally on the neck approximate 12.5 cm apart. The thickness of the skin was measured before injection and approximate 72 hours after injections. The difference (‘after’ minus ‘before’) in skin-thickness was plotted on a scattergram (USDA, VS, Form 6-22D) used to interpret the CCT results as “negative”, “suspect”, or “reactor”. In practice and in this study, both “suspect” and “reactors” were classified into one group considered “positive” on the CCT. Because the CCT was only performed on animals that were suspects on the CFT, it was not possible to estimate Se and Sp of the CCT used as an independent test. Necropsy was conducted at the Diagnostic Center for Population and Animal Health at Michigan State University, and mycobacterial culture and identification was performed on fresh samples at the National Veterinary Services Laboratories, Ames, Iowa. For six animals, tissues were also cultured at the Michigan Department of Community Health.(Norby et al., 2003) The necropsy and culture procedure has been described elsewhere.(Norby et al., 2003;

Schmitt et al., 1997) A very specific genetic probe is used with CULT to classify mycobacterial species into isolates belonging to the *Mycobacterium* complex, which include *M. bovis*, *tuberculosis*, *africanum*, and *microtii*.(Schmitt et al., 1997) Since it was unlikely that cattle in Michigan were infected with mycobacteria belonging to the *Mycobacterium* complex other than *M. bovis*, we assume that all isolates that were positive on culture are in fact *M. bovis*.

Latent-class models

The parameters of interest were the Se and Sp of the CFT and CFTCCT_{SER}. The technique used in this study was a two-population, two-test approach in which it was assumed that the disease prevalence was different in the two populations, but Se and Sp remained constant in both populations.(Georgiadis et al., 2003) As argued by Georgiadis et al. (2003), a two-population-two test scenario was employed because addition of a “second population” in the fixed effects model increases the degrees of freedom to 6 as compared to three degrees of freedom in a one-population-two test model. If the two tests are assumed to be conditionally independent, the model is then identifiable because there are six degrees of freedom available to estimated 6 parameters, i.e. 2 prevalences, two sensitivities and two specificities. If the two tests are not conditionally independent then two extra parameters (a total of eight) needs to be estimated, resulting in eight parameters.(Georgiadis et al., 2003) Even though the model is no longer identifiable, the two-population-two-test model decreases the reliance upon additional prior information.(Georgiadis et al., 2003)

The cattle in the seven herds were divided into two subpopulations(Enøe et al., 2001; Enøe et al., 2000; Willeberg et al., 1997) based on the age of the cattle; cattle \geq four years of age comprised population 1 and population 2 was $<$ four years of age. Se and Sp were evaluated using the test results of CULT as the second test following CFT (Scenario 1) and following CFTCCT_{SER} (Scenario 2). Originally, the seven herds were ‘naturally’ divided into two herds with moderately high and five with moderately low apparent prevalences(Norby et al., 2003), and the analyses were also performed on this dataset. Four different Bayesian models were used to estimate these parameters: 1) a conditional independence model using non-informative (uniform) prior information all parameters (M1a)(Enøe et al., 2000) and a conditional independence model using informative priors on prevalences and Sp of CULT (M1b)(Johnson et al., 2001), 2) an extension of a conditional dependence model as described by Black and Craig (2002) (M2), 3) a conditional dependence model assuming that Sp of CULT was perfect (M3), and 4) a conditional dependence model as described by Georgiadis et al. (2003) (M4).

The models proposed by Black and Craig (2003) and Georgiadis et al. (2003), respectively, are based on estimation of the latent cell counts in two two-by-two tables representing a cross classification of the results on the two tests in the two populations (see Table 4-7 in the appendix). In essence, the two models are identical, however, the model by Georgiadis et al. (2003) is a reparameterization of the model proposed by Black and Craig (2003), using a one-to-one transformation of four parameters (p_{1111} , p_{0010} , and the sensitivity and specificity of test 1) (appendix A). Hence, all the full conditionals were recognizable and could be sampled using Gibbs sampler(Gelfand and Smith, 1990; Tanner 1996). In contrast the full conditional densities, two parameters in M2 and one

parameter in M3 were unrecognizable and were sampled using Metropolis-Hastings (MH) algorithm (Chib and Greenberg, 1995). The remaining six parameters were sampled using Gibbs sampler (Appendix A). Bayesian inference was performed using Markov-Chain Monte Carlo (MCMC) simulation.

Posterior point estimates and 95% credible intervals (PCI) for the six parameters were defined as the median, lower 2.5% and upper 97.5% percentiles of the posterior distribution, respectively. (Gelman et al., 2000) The PCI is a 95% Bayesian (probability) interval where parameters have a 95% probability of being included in the PCI. (Gelman et al., 2000) A more detailed description of the models and Bayesian sampling methods are presented in Appendix A.

A 'burn-in' of 2000 MCMC cycles were used to assure that the models had reached convergence. (Tanner, 1996) This was monitored by plotting the parameter estimates for each cycle. Then, 10,000 MCMC cycles were used to estimate the posterior densities for prevalences, Se's, and Sp's.

All four models were coded in the IML procedure in SAS (SAS Institute, Cary, NC) (Appendix B).

Prior elicitation

Prior information was elicited as beta distributions ($\text{Beta}(\alpha, \beta)$) for the prevalences, sensitivities, and specificities. (Joseph et al., 1995; Mendoza-Blanco et al., 1996) Beta distributions were chosen because they are highly flexible and the full conditional distributions of the parameters are also then beta distributions, facilitating easy sampling and involve relatively simple calculations. (Enøe et al., 2000)

The conditional dependence models require specification of prior information on at least some of the six parameters, in contrast to the conditional independence model where no *a priori* knowledge is necessary about the parameters (the model is fully identifiable). Use of priors effectively places restrictions on the parameter space by only including regions where the parameter values are likely to be concentrated. We chose to place prior information on prevalences, and the Sp of CULT, because we believed that we had relatively good information regarding these parameters. Choosing a ‘most likely’ value (prior mean) and a value for the lower 5th percentile of the prior density for the parameter of interest, it was then possible to specify the hyperparameters for a beta distribution that fits those specifications.

The mycobacterial culture method uses a very specific genetic probe, and it was therefore assumed that the specificity of CULT was 1.0, or very close to 1.0. The most likely value for Sp of CULT was set to 0.999 and the 5th percentile to 99 providing Beta(360.4, 1.36) for M2 and M4. The priors on Se and Sp of either CFT or CFTCCT_{SER} were always non-informative [Beta(1,1)].

We did not have exact prior knowledge on the bTb prevalence in the two populations. Hence the most likely value was assumed to lie between the apparent prevalence (APP) as obtained by CULT and the CFT (Table 1). The most likely value was determined as the mean of the APP for CFT and CULT, and the lower 5th percentile was set at the APP for CULT, because the specificity of CULT is close to one and false positive results were unlikely. The priors for prevalence were Beta(13.63, 57.94) for the young animals and Beta(5.98, 66.68) for the older animals (Table 4-1).

Table 4-1. Apparent prevalence (APP; %) in old and young subpopulations based upon the caudal fold tuberculin test (CFT) and Mycobacterial culture (CULT).

	Age \geq 4 years		Age < 4 years	
	APP (%)	CI ¹	APP (%)s	CI ¹
CFT	22.5	16.8, 28.9	10.4	7.2, 14.4
CULT	13.8	9.3, 19.4	3.7	1.9, 6.5
Mean	18.15		7.05	

¹95% Confidence interval.

Sensitivity analysis

A sensitivity analysis was performed to investigate how sensitive the posterior inference was to the priors that were placed on prevalences and the Sp of CULT. (Enøe et al., 2003) Alternately, 1) non-informative priors were placed in succession on one of these three parameters, while the other two were held at their original beta distribution (constant). 2) Diffuse, but correctly specified priors were placed on Sp of CULT, by gradually lowering the 5th percentile to 0.85, and the priors on prevalence were held constant. 3) Mis-specified priors were placed on Sp of CULT, i.e., the prior median of this new prior was centered at 0.95, 0.90, and 0.85, and the priors on prevalence were held constant. 4) Diffuse, but correctly specified priors were placed on the two prevalences. The median of the prevalence in population one held at 0.1815 and the 5th percentile was lowered to 0.1 and 0.05 and increased to 0.25. For population 2, the median was held at 0.0705 and the 5th percentile lowered to 0.01 and 0.005 and increased to 0.14. 5) Mis-specified priors were placed on prevalences in the same manner as described for the specificity, hence the medians were 0.1 and 0.25, and 0.05 and 0.14 for populations 1 and 2, respectively.

RESULTS

Twenty-two of 516 cattle in the seven herds had erroneous records and were excluded from the dataset. Therefore, 494 cattle were included in the study. The mean age in population 1 was 70.6 months (std =19.0) and 15.7 months (std = 11.1) in population 2. The apparent prevalences (APP) based upon CFT and CULT are shown in Table 4-1. Cross-classification of observed counts for scenario 1 (CFT and CULT) and scenario 2 (CFTCCT_{SER} and CULT) are shown in tables 2 and 3, respectively.

Table 4-2. Testing Scenario #1: 2 × 2 contingency table showing the observed counts for the caudal fold tuberculin test (CFT) and Mycobacterial culture (CULT) in populations 1 and 2.

Population 1					Population 2				
		CULT					CULT		
		+	-			+	-		
CFT	+	24	20	44	CFT	+	11	20	31
	-	3	149	152		-	0	267	267
		27	169	196			11	287	298

Table 4-3. Testing Scenario #2: 2 × 2 contingency table showing the observed counts for the caudal fold and comparative cervical tuberculin tests used in serial interpretation (CFTCCT_{SER}) and Mycobacterial culture (CULT) in populations 1 and 2.

Population 1					Population 2				
		CULT					CULT		
		+	-			+	-		
CFT- CCT _{SER}	+	23	7	30	CFT- CCT _{SER}	+	10	5	15
	-	4	162	166		-	1	282	283
		27	169	196			11	287	298

The overall observed agreement for the 494 animals were measured as the proportion of cattle with concordant test results (451/494=0.913 and 477/494=0.966) and by the Kappa statistic ($\kappa=0.576$ and $\kappa = 0.777$) for scenario 1 and 2, respectively. The

proportion with concordant results for populations 1 and 2 were $173/196 = 0.883$, $\kappa = 0.609$ and $278/298=0.933$, $\kappa = 0.496$ for scenario 1 and $185/196=0.944$, $\kappa=0.774$ and $283/298=0.950$, $\kappa=0.759$ and for scenario 2.

The estimated prevalences, Se's and Sp's from all four Bayesian models for both testing scenarios are presented as posterior medians (point estimates) and PCI's in tables 4 and 5. For example, for M2 in scenario 1 (CFT and CULT), the point estimate for Se of the CFT was 0.854 and the 95% PCI was (0.563,0.975). There was no significant correlation between disease-positive animals for models M2, M3, or M4 in either of the two scenarios, as the 95% PCI's for correlation between disease positive animals included 0. However, the correlation between disease negative animals was significant for CFTCCT_{SER} and CULT and very close to being significant for CFT and CULT because the 95% CI almost included 0; the lower 2.5th percentile was -0.004 and -0.005 for models 2 and 4, respectively. Hence, for this study, the estimates of the conditional dependence model (M2) were selected as being the preferred method of analysis.

Sensitivity analyses were performed on M1 and M2 to assess the effect of less informative prior information on the Sp of CULT and the two estimates of subpopulations prevalence. First, the prior information on Sp2 was made more diffuse by gradually lowering the 5th percentile to 0.85, and by using a non-informative prior. The resulting point estimates for all parameters were within the range on the original 95% PCI for the model using original priors for all analyses in both scenarios, except when using a non-informative prior for scenario 2, in which case the point estimate using the flat prior was 0.001 outside of the original 95% PCI. All point estimates obtained with the sensitivity analyses were within 0.016 of the parameters estimates obtained by the

original prior for Sp of CULT only when the point estimate was changed from 0.999 to 0.95 and the 5th percentile was lowered to 0.80, did the point estimate of the sensitivity analyses fall outside the original 95% PCI. Similar results were seen when analogous adjustments were made to the priors of the two prevalence estimates. Changing the prior for the prevalence in either of the two populations had the greatest effect on the sensitivity estimates. Making the prior more diffuse had less effect on the sensitivity estimated than did changing the location of the peak of the prior. The largest differences in sensitivity estimates were most commonly seen when a non-informative prior was used or when the prior's peak was changed considerably for the two prevalence estimates (results not shown).

Originally, the seven herds were 'naturally' divided into two populations based on the herd apparent prevalence. Two herds had moderately high prevalence and 5 herds had low prevalence. (Norby *et al.*, 2003) The two prevalence estimates, Se 's, and Sp 's were also estimated using this scenario. New priors for the prevalences were created as described before and the same prior for Sp_2 was used. The sensitivity estimates were 0.896 and 0.843 for CFT and CFTCCT_{SER}, and 0.806 and 0.789 for CULT in the two populations. Estimates of posterior medians of specificity estimates were 0.862 and 0.969 for CFT and CFTCCT_{SER}, and 0.997 and 0.996 for CULT in the two scenarios respectively.

Table 4-4. Prevalences (π_1, π_2), sensitivities (Se) and specificities (Sp) for the caudal fold tuberculin test (CFT) and mycobacterial culture (CULT). Point estimates for all parameters are presented as the median of the posterior densities and their respective 95% posterior credible interval (PCI).

Test	Parameter	^a M1a flat prior		^b M1b		^c M2		^d M3		^e M4	
		50%	95% PCI	50%	95% PCI	50%	95% PCI	50%	95% PCI	50%	95% PCI
CFT	π_1	0.200	(0.124,0.281)	0.196	(0.138,0.259)	0.204	(0.142,0.281)	0.213	(0.151,0.291)	0.204	(0.139,0.285)
	π_2	0.059	(0.028,0.113)	0.061	(0.034,0.104)	0.063	(0.034,0.106)	0.068	(0.038,0.113)	0.062	(0.033,0.106)
CULT	Se	0.924	(0.803,0.992)	0.917	(0.799,0.986)	0.854	(0.563,0.975)	0.826	(0.559,0.961)	0.804	(0.529,0.948)
	Sp	0.947	(0.905,0.990)	0.946	(0.906,0.988)	0.939	(0.899,0.982)	0.944	(0.901,0.986)	0.934	(0.894,0.975)
CULT	Se	0.664	(0.446,0.938)	0.666	(0.460,0.917)	0.623	(0.401,0.899)	0.596	(0.401,0.854)	0.627	(0.404,0.918)
	Sp	0.997	(0.985,1)	0.998	(0.992,1)	0.997	(0.990,1)	N/A	N/A	0.998	(0.990,1)
	^f d_S	N/A	N/A	N/A	N/A	0.230	(-0.166,0.771)	0.277	(-0.150,0.778)	0.351	(-0.137,0.815)
	^g d_C	N/A	N/A	N/A	N/A	0.096	(-0.004,0.351)	N/A	N/A	0.091	(-0.005,0.324)

^aConditional independence model using flat priors. ^bConditional independence model using informative prior. ^cConditional

dependence model. ^dConditional dependence model assuming perfect Sp for CULT. ^eConditional dependence model as describe by

Georgiadis et al., (2003). ^f d_S correlation between disease positive animals, and ^gdisease negative. d_S is not estimable for models 1 and 1b and d_C is not estimable for model 1a, 1b, and 3.

Table 4-5. Prevalences (π_1, π_2), sensitivities (Se) and specificities (Sp) for the caudal fold tuberculin test and comparative cervical tuberculin tests used in serial interpretation (CFTCCT_{SER}) and mycobacterial culture (CULT) for all four models. Point estimates for all parameters are presented as the median of the posterior densities and their respective 95% posterior credible interval (PCI).

Test	Parameter	^a M1a flat prior			^b M1b			^c M2			^d M3			^e M4		
		50%	95% PCI	50%	95% PCI	50%	95% PCI	50%	95% PCI	50%	95% PCI	50%	95% PCI	50%	95% PCI	
	π_1	0.161	(0.108,0.225)	0.172	(0.127,0.224)	0.189	(0.134,0.268)	0.196	(0.138,0.279)	0.190	(0.132,0.273)					
	π_2	0.045	(0.023,0.077)	0.052	(0.030,0.080)	0.057	(0.031,0.094)	0.060	(0.035,0.099)	0.056	(0.031,0.097)					
CFT-	Se	0.890	(0.748,0.988)	0.869	(0.731,0.964)	0.758	(0.475,0.930)	0.715	(0.452,0.911)	0.732	(0.449,0.912)					
CCT _{SER}	Sp	0.987	(0.966,0.999)	0.989	(0.969,0.999)	0.986	(0.962,0.999)	0.987	(0.964,0.999)	0.983	(0.960,0.998)					
CULT	Se	0.818	(0.638,0.978)	0.798	(0.629,0.965)	0.695	(0.448,0.933)	0.684	(0.437,0.929)	0.701	(0.435,0.954)					
	Sp	0.995	(0.982,1)	0.997	(0.992,1)	0.997	(0.989,1)	N/A	N/A	0.997	(0.989,1)					
	^f d_S	N/A	N/A	N/A	N/A	0.388	(-0.130,0.793)	0.581	(-0.106,0.808)	0.448	(-0.106,0.823)					
	^g d_C	N/A	N/A	N/A	N/A	0.218	(0.005,0.731)	N/A	N/A	0.201	(0.004,0.684)					

^aConditional independence model using flat priors. ^bConditional independence model using informative prior. ^cConditional

dependence model. ^dConditional dependence model assuming perfect Sp for CULT. ^eConditional dependence model as describe by

Georgiadis et al., (2003). ^f d_S correlation between disease positive animals, and ^gdisease negative. d_S is not estimable for models 1 and 1b and d_C is not estimable for model 1a, 1b, and 3.

DISCUSSION

In this study, we estimated the sensitivity and specificity of the CFT and CFTCCT_{SER} as they were applied to cattle in seven herds in the northeastern part of the lower peninsula in Michigan. We used CULT as the second test in our two-population, two-tests scenarios to accomplish that goal. Because the correlation between the disease negative animals was significant for scenario 2 and very close to being significant for scenario 1, we decided that the results from the conditional dependence model (M2) generally produced the best estimates.

Our estimates of the Se of the CFT were very similar to other US estimates. (range:0.804-0.849).(Anonymous, 1992; Gaboric et al., 1996; Whipple et al., 1995) The estimates of Sp of the CFT are similar to estimates reported by the USDA(Anonymous, 1992), and higher than reported by Gaboric et al., (1996). However, the Se estimates for CFT and CFTCCT_{SER} are lower than estimated reported by Norby et al., (2003) in which independence between the skin tests was assumed and the CULT was assumed to be the definitive test.

The three dependence models performed similarly, with the largest variability observed between Se estimates for CFT and CFTCCT_{SER}. The point estimates for each of the parameters were within the 95% PCI of the other models for all the estimable parameters, for both scenarios. In general, the Se estimates were slightly lower for the dependence models as compared to the estimates from the independence model. However, the parameter estimates for prevalence and Sp's were very similar for all the models. Since the correlation between disease negative animals was very low (0.096) for

the CFT and CULT scenario and moderately low (0.218) for the CFTCCT_{SER} and CULT scenario, we would not expect that there would be a large difference between parameter estimates for the independence and dependence models. Similar observations were made in simulation studies by Georgiadis et al., (2003). The 95% PCIs for Se's in the three dependence models were as expectedly wider than for the independence model, reflecting the additional uncertainty of estimating more parameters with the same information.(Georgiadis et al., 2003)

For all four models, we also observed that the 95% PCI's were wider for the Se than for the Sp estimates. This was most likely explained by the low 'overall' disease prevalence, and consequently less information on which to base the Se estimates as compared to Sp estimates. Using prevalence estimates from M2 in scenario 1, there were only approximately $(0.204 \cdot 196 + 0.063 \cdot 298)$ 58 disease positive animals in our total sample. A larger sample size or higher prevalence would most likely have produced more stable parameter estimates with narrower PCI's.

The 95% PCIs for the prevalence estimates were of approximately equal width for all four models, and were virtually identical for the 3 dependence models. In both of the two scenarios, Se and Sp estimates for CULT were under our assumptions expected to be equal. However, the point estimates for Se of CULT was between 0.072 and 0.088 for the four models in scenario 1 as compared to scenario 2. This discrepancy may in part be explained by the uncertainty in the data and the low disease prevalence. The specificity estimates for CULT were similar for the two scenarios.

Se and Sp estimates of a diagnostic test may vary within herds or populations.(Donald et al., 1994; Greiner and Gardner, 2000) In our analyses, it was

possible that the performance of the skin tests and CULT may have varied within herds based on the infection status of the animals. Perhaps animals that are heavily infected react more readily to the skin tests than animals with more recent or less severe infection. The Se estimates of CFT were slightly lower and Se estimates were slightly higher for the CFTCCT_{SER} when the population was divided based on age rather than upon herd apparent prevalence. The Sp estimates were significantly higher (based on the 95% PCI) for the CFT and slightly higher for the CFTCCT_{SER} when comparing the results based on the two methods we used to obtain two populations. This suggests that Se and Sp may not be constant within subpopulations. Hence, the Se and Sp estimates were overall estimates for all the populations in the study, and extrapolation between populations should be done with caution.(Enøe et al., 2000)

Because a perfect reference test (gold standard) was unavailable in this study, we estimated Se and Sp of the CFT and CFTCCT_{SER} using Bayesian inference.(Enøe et al., 2000) In the two-population, two-test scenario proposed here, it is possible to estimate the six parameter (two prevalences, two Se's, and two Sp's) without prior knowledge on any of the parameters, if it is assumed that the tests are independent and that the Se's and Sp's are constant across populations, i.e. the model is identifiable. However, it was not certain that the skin tests and CULT were conditionally independent, and model(s) that accounted for possible dependence were needed. Therefore, two extra parameters needed, p_{1111} and p_{0010} , respectively(Black and Craig, 2002), and hence the two-population, two-test scenario was unidentifiable. Additional degrees of freedom can be acquired by increasing the number of populations under the assumption that Ses and Sps are constant across populations. However, it was not feasible to divide our initial

population of 494 cattle into more than two subpopulations because of the low number of truly infected animals. Furthermore, it is questionable whether an addition of degrees of freedom will improve the model.(Georgiadis et al., 2003)

We chose to use a Bayesian approach as compared to a maximum likelihood approach because the problem of an unidentifiable model could be overcome by using prior information on some of the parameters.(Black and Craig, 2002; Georgiadis et al., 2003) Our independence and dependence models performed well, and converged within a few hundred MCMC cycles. We preferred the approach described by Black and Craig (2002), because it is easily extended to multiple populations and tests. Model M3 can also be used in scenarios where it is reasonable to assume that one of the tests have perfect specificity.

Several factors may affect the estimates of Se and Sp, including the stage of disease, length of exposure, animal age, prevalence of infection in the herd, and immune status of the host.(Greiner and Gardner, 2000) For bTB, the Se of the skin tests may be lower for pre-allergic, inactive, and anergic stages of the infection.(Greiner and Gardner, 2000) When we divided our sample according to the age of animals, prevalence (as determined by M2-M4) was almost 0.15 higher in older cattle as compared to younger cattle. This may be attributed to the fact that older animals are more likely to have generalized and more active infections, whereas younger animals will be more likely to have acquired their infection more recently, and hence not showing as imminent reactions to the skin tests.(Norby et al., 2003) It is likely that animals from herds with a high prevalence may have older and more established infections because such herds may have been infected for a longer period of time. Our analysis does not account for either of

these biological factors, and it is hence possible that our sensitivity estimates for both skin tests and CULT maybe slightly overestimated, and specificity estimates slightly underestimated.

The sample of animals used in this study was not a random and blinded sample of bTb positive and negative animals.(Greiner and Gardner, 2000) However, the CFT and CCT were performed before it was known if these herds would be deemed to be truly infected. Because all our models performed very uniformly, the internal validity of our study was likely high. This does not automatically signify that the external validity of the study also is high. However, our estimates of Se and Sp of the skin tests are close to previously published results from the U.S.(Anonymous 1992; Gaboric et al., 1996; Whipple et al., 1995) It should be noted, however, that when disease prevalence decreases when an eradication program progresses, it is likely to negatively affect 'working' estimates of Se and Sp of the skin tests.(Greiner and Gardner, 2000)

REFERENCES

- Anonymous. 1992, Assessment of risk factors for *Mycobacterium bovis* in the United States. Fort Collins: USDA:APHIS:VS:CEAH.
- Black, M.A., and B.A. Craig. 2002, Estimating disease prevalence in the absence of a gold standard. *Statist. Med.* 21, 2653-2669.
- Bruning-Fann, C.S., S.M. Schmitt, S.D. Fitzgerald, J.S. Fierke, P.D. Friedrich, J.B. Kaneene, K.A. Clarke, K.L. Butler, J.B. Payeur, D.L. Whipple, T.M. Cooley, J.M. Miller, and D.P. Muzo. 2001, Bovine tuberculosis in free-ranging carnivores from Michigan. *J. of Wildlife Dis.* 37, 58-64.
- Casella, G., and R.L. Berger. 2001, *Statistical Inference*. Second ed: Duxbury Press.
- Chib, S., and E. Greenberg. 1995, Understanding the Metropolis-Hastings algorithm. *Am. Statistician* 49, 327-335.
- Dempster, A., N. Laird, and D. Rubin, 1977. Maximum likelihood from incomplete data via the EM algorithm. *J. Royal Statist. Soc. Ser. B* 39, 1-38.
- Dendukuri, N., and L. Joseph. 2001, Bayesian approach to modeling the conditional dependence between multiple diagnostic tests. *Biometrics* 57, 158-167.
- Donald, A.W., I.A. Gardner, and A.D. Wiggins, 1994, Cut-off points for aggregate herd testing in the presence of disease clustering and correlation of test errors. *Prev. Vet. Med.* 19, 167-187.
- Enøe, C., S. Andersen, V. Sørensen, and P. Willeberg. 2001, Estimation of sensitivity, specificity and predictive values of two serologic tests for the detection of antibodies against *Actinobacillus pleuropneumoniae* serotype 2 in the ascence of a reference test (gold standard). *Prev. Vet. Med.* 51, 227-243.
- Enøe, C., M.P. Georgiadis, and W.O. Johnson. 2000, Estimation of sensitivity and specificity of diagnostic tests and disease prevalence when the true disease status is unknown. *Prev. Vet. Med.* 45, 61-81.
- Francis, J., Seiler, R.J., 1978, The sensitivity and specificity of various tuberculin tests using bovine PPD and other tuberculins. *Vet. Rec.* 103, 420-425.
- Frey, G.H., 1995, Bovine tuberculosis eradication: The program in the United States, In: Thoen, C.O., Steele, J.H. (Eds.) *Mycobacterium bovis* infection in animals and humans. Iowa State University Press, Ames, IA, pp. 119-129.
- Gaboric, C.M., M.D. Salman, R.P. Ellis, and J. Triantis. 1996, Evaluation of five-antigen ELISA for diagnosis of tuberculosis in cattle and Cervidae. *JAVMA* 209, 962-966.

Gardner, I.A., H. Stryhn, P. Lind, and M.T. Collins. 2000, Conditional dependence between tests affects the diagnosis and surveillance of animal diseases. *Prev. Vet. Med.* 45, 107-122.

Gart, J., and A. Buck. 1996, Comparison of a screening test and a reference test in epidemiologic studies. *American Journal of Epidemiology* 83, 593-602.

Gelfand, A.E., and A.F.M. Smith. 1990, Sampling-based approaches to calculating marginal densities. *J. Am. Statist. Assoc.* 85, 398-409.

Gelman, A., J.B. Carlin, H.S. Stern, and D.B. ZRubin. 2000, Bayesian data analysis. Edited by C. Chatfield and J. V. Zidek. First CRC reprint ed, Chapman & Hall, Texts in statistical science series. Boca Raton: Chapman & Hall/CRC.

Georgiadis, M.P., I.A. Gardner, and R.P. Hedrick. 1998, Field evaluation of sensitivity and specificity of a polymerase chain reaction (PCR) for detection of *N. salmonis* in rainbow trout. *J. Aquat. Anim. Health* 10, 372-380.

Georgiadis, M.P., W.O. Johnson, I.A. Gardner, and R. Singh. 2003, Correlation-adjusted estimation of sensitivity of two diagnostic tests. *Appl. Statist* 52, 1-14.

Greiner, M., and I.A. Gardner. 2000, Epidemiologic issues in the validation of veterinary diagnostic tests. *Prev. Vet. Med.* 45:3-22.

Hui, S.L., and S.D. Walter. 1980, Estimating the error rates of diagnostic tests. *Biometrics* 36:167-171.

Johnson, W.O., J.L. Gastwirth, and M. Pearson. 2001, Screening without a "gold standard": The Hui-Walter Paradigm revisited. *Am. J. Epidemiol.* 153, 921-924.

Joseph, L., T. Gyorkos, and L. Coupal. 1995, Bayesian estimation of disease prevalence and the parameters of diagnostic tests in the absence of a gold standard. *Am. J. Epidemiol.* 141, 263-272.

Lepper, A.W.D., D.A. Newton-Tabrett, L.A. Corner, M.T. Carpenter, W.A. Scanlan, O.J. Williams, and D.M. Helwig. 1977, The use of bovine PPD tuberculin in the single caudal fold test to detect tuberculosis in beef cattle. *Australian Veterinary Journal* 53, 208-213.

Leslie, I. W., and C. N. Hebert. 1975, Comparison of the specificity of human and bovine tuberculin PPD for testing cattle. 3 - National trial in Great Britain. *Vet Rec* 96, 338-341.

Leslie, I. W., C. N. Hebert, and D. N. Barnett. 1975, Comparison of the specificity of human and bovine tuberculin PPD for testing cattle. 2 - Southeastern England. *Vet Rec* 96, 335-338.

Leslie, I. W., C. N. Hebert, K. J. Burn, B. N. MacClancy, and W. J. C. Donnelly. 1975, Comparison of the specificity of human and bovine tuberculin PPD for testing cattle. 1 - Republic of Ireland. *Vet Rec* 96, 332-334.

Mendoza-Blanco, J.R., X.M. Tu, and S. Iyengar. 1996, Bayesian inference on prevalence using a missing-data approach with simulation-based techniques: Applications to HIV screening. *Statist. Med.* 15, 2161-2167.

Miller, J., A. Jenny, J. Rhyan, D. Saari, and D. Suarez. 1997, Detection of *Mycobacterium bovis* in formalin-fixed, paraffin-embedded tissues of cattle and elk by PCR amplification of an IS6110 sequence specific for *Mycobacterium tuberculosis* complex organisms. *Journal of Veterinary Diagnostic Investigation* 9, 244-249.

Norby, B., Paul C. Bartlett, Scott D. Fitzgerald, Larry Granger, Colleen Bruning-Fann, Diana L. Whipple, and Janet B. Payeur. 2003, The sensitivity of gross necropsy, caudal fold and comparative cervical tests for the diagnosis bovine tuberculosis. *JVDI*, In Press.

Pollock, J.M., R.M. Girvin, K.A. Lightbody, R.A. Clements, S.D. Neill, B.M. Buddle, and P. Andersen. 2000, Assessment of defined antigens for the diagnosis of bovine tuberculosis in skin test-reactor cattle. *Vet.Rec.*146, 659-665.

Qu, Y., and A. Hadgu. 1998, A model for evaluating the sensitivity and specificity for correlated diagnostic tests in efficacy studies with an imperfect reference test. *J. Am. Statist. Ass.* 93, 920-928.

Qu, Y., M. Tan, and M.H. Kutner. 1996, Random effects models in latent class analysis for evaluating accuracy of diagnostic tests. *Biometrics* 52, 797-810.

Roswurm, J.D., and L.D. Konyha. 1973, The comparative-cervical tuberculin test as an aid to diagnosing bovine tuberculosis. Paper read at Annual Meeting of US Animal Health Association.

Schmitt, S., Fitzgerald, S., Cooley, T., Bruning-Fann, C., Sullivan, L., Berry, D., Carlson, T., Minnis, R., Payeur, J., Sikarskie, J., 1997, Bovine tuberculosis in free-ranging white-tailed deer from Michigan. *J. Wildlife Dis.* 33, 749-758.

Singer, R.S., W.M. Boyce, I.A. Gardner, W.O. Johnson, and A.S. Fisher. 1998, Evaluation of bluetongue virus diagnostic tests in free-ranging bighorn sheep. *Prev. Vet. Med.* 35, 265-282.

Staquet, M., M. Rozenzweig, M. Lee, and F.M. Muggai. 1981, Methodology for assessment of new dichotomous diagnostic tests. *J. Chronic Dis.* 34, 599-610.

Tanner, M.A. 1996, Tools for statistical inference: Methods for the exploration of posterior distributions and likelihood functions. 3rd edition. New York: Springer.

- Torrance-Rynard, V.L., and S.D. Walter. 1997, Effects of dependent errors in the assessment of diagnostic test performance. *Statist. Med.* 16, 2157-2175
- USDA, and APHIS. 1999, Bovine tuberculosis eradication. Uniform methods and rules, effective January 22, 1999, (United States Department of Agriculture.), pp. 1-34.
- Vacek, P.M. 1985, The effect of conditional dependence on the evaluation of diagnostic tests. *Biometrics* 41, 959-968.
- Walter, S.D., and L.M. Irwig. 1988, Estimation of error rates, disease prevalence and relative risk from misclassified data: a review. *J. Clin. Epidemiol* 41, 923-937.
- Weng, T.S. 1996, Evaluation of a new diagnostic test against a reference test less than perfect in accuracy. *Commun. Statist. Simul. Comput.* 35, 533-555.
- Whipple, D. L., C. A. Bolin, A. J. Davis, J. L. Jarnagin, D. C. Johnson, J. B. Payer, D. A. Saari, A. J. Wilson, and M. M. Wolf. 1995, Comparison of the sensitivity of the caudal fold skin test and a commercial gamma-interferon assay for diagnosis of bovine tuberculosis. *Am J Vet Res* 56, 415-419.
- Willeberg, P., J.M. Wedam, I.A. Gardner, J.C. Holmes, J.A. Mousing, J. Kyrval, C. Enøe, and S. Andersen. 1997, A comparative study of visual and traditional post-mortem inspection of slaughter pigs: Estimation of sensitivity, specificity and differences in non-detection rates. *Epidémiologie et Santé Animale* 31/32:04.20.1-04.20.3.
- Wood, P.R., L.A. Corner, J.S. Rothel, C. Baldock, S.L. Jones, D.B. Cousins, B.S. McCormick, B.R. Francis, J. Creeper, and N.E. Tweddle. 1991, Field comparison of the interferon-gamma assay and the intradermal tuberculin test for the diagnosis of bovine tuberculosis. *Austr. Vet. J.* 68, 286-290.
- Wood, P.R., L.A. Corner, J.S. Rothel, J.L. Ripper, T. Fifis, B.S. McCormick, B. Francis, L. Melville, K. Small, K. De Witte, J. Tolson, T.J. Ryan, G.W. Lisle, J.C. Cox, and S.L. Jones. 1992, A field evaluation of serological and cellular diagnostic tests for bovine tuberculosis. *Veterinary Microbiology* 31, 71-79.
- Yang, I., and M. P. Becker. 1997, Latent variable modeling of diagnostic accuracy. *Biometrics* 53, 948-958.

APPENDIX A

For all the models, let $\pi_i, i = \{1,2\}$ represent the true prevalence in populations 1 and 2, and let S_1, S_2 , represent the sensitivities, and C_1 and C_2 represent the specificities of tests 1 and 2, respectively. The results of the two tests, submitted to all animals in the two population, are presented as counts $\{n_{ijk}\}$ in two 2×2 tables (Table 4-6), where the first subscript i , denote the population, the subscript j denote the results of test 1 (T_1 : CFT or CFTCCT_{SER}), and k denoting the result of test 2 (T_2 : CULT), ($i = \{1,2\}$ and $j,k = \{0,1\}$). If the true disease status of each tested animal was known, the count in each cell in the two 2×2 tables can be divided into two counts, where $n_{ijk} = a_{ijk} + b_{ijk}$. Here, a_{ijk} and b_{ijk} represent animals that are truly disease positive and disease negative, respectively. The data can then be summarized in four separate 2×2 tables (Table 4-7).

Table 4-6. Two 2×2 contingency tables showing responses of two binary diagnostic tests.

		Population 1			Population 2				
		Test 2			Test 2				
		+	-		+	-			
Test 1	+	n_{111}	n_{110}	$n_{11\cdot}$	Test 1	n_{211}	n_{210}	$n_{21\cdot}$	
	-	n_{101}	n_{100}	$n_{10\cdot}$		-	n_{201}	n_{200}	$n_{20\cdot}$
		$n_{1\cdot 1}$	$n_{1\cdot 0}$	$n_{1\cdot\cdot}$			$n_{2\cdot 1}$	$n_{2\cdot 0}$	$n_{2\cdot\cdot}$

Table 4-7. Four 2×2 contingency tables showing responses of two binary tests classified by true disease status.

		Disease positive animals							
		Population 1			Population 2				
		Test 2			Test 2				
		+	-		+	-			
Test 1	+	a_{111}	a_{110}	$a_{11\cdot}$	+	a_{211}	a_{210}	$a_{21\cdot}$	
	-	a_{101}	a_{100}	$a_{10\cdot}$	-	a_{201}	a_{200}	$a_{20\cdot}$	
		$a_{1\cdot 1}$	$a_{1\cdot 0}$	$a_{1\cdot\cdot}$			$a_{2\cdot 1}$	$a_{2\cdot 0}$	$a_{2\cdot\cdot}$

		Disease negative animals							
		Population 1			Population 2				
		Test 2			Test 2				
		+	-		+	-			
Test 1	+	b_{111}	b_{110}	$b_{11\cdot}$	+	b_{211}	b_{210}	$b_{21\cdot}$	
	-	b_{101}	b_{100}	$b_{10\cdot}$	-	b_{201}	b_{200}	$b_{20\cdot}$	
		$b_{1\cdot 1}$	$b_{1\cdot 0}$	$b_{1\cdot\cdot}$			$b_{2\cdot 1}$	$b_{2\cdot 0}$	$b_{2\cdot\cdot}$

The prevalence in the two populations, Se 's and Sp 's of the two tests may be estimated if the true disease status is known. However, in this study, the true disease status is unknown (because a perfect reference test was not available), and it is only possible to observe the sums of disease positive and negative animals ($n_{ijk} = a_{ijk} + b_{ijk}$). Hence the 2×2 table for each of the two populations (Table 4-6) is the result of collapsing the two tables of latent data for each population (Table 4-7). The use of latent data is a key component of algorithms that rely upon data augmentation such as the expectation-maximization algorithm as well as some forms of Monte Carlo inference, including Gibbs sampler. (Tanner, 1996)

Conditional independence model (M1).

Let $p_{jkl} = P(T_1 = j, T_2 = k | D = l)$, where $l = \{0, 1\}$, represent the corresponding joint classification probabilities for the latent data of disease positive and negative animals in

the two populations, where the subscript l denotes the true disease status of an animal (0 = negative, and 1 = positive). (Black and Craig, 2002; Singer et al., 1998) The classification probabilities were assumed to be equal across populations. As the two tests were considered to be independent

$$\begin{aligned}
 p_{jkl} &= p(T_1 = j, T_2 = k | D = l) \\
 &= p(T_1 = j | D = l) p(T_2 = k | D = l) \\
 &= p_{j \cdot | l} p_{\cdot k | l}
 \end{aligned}$$

Hence, the classification probabilities are the product of the conditional probabilities

(Table 4-8)

$$\begin{aligned}
 p_{11|1} &= S_1 S_2 & p_{11|0} &= (1-C_1)(1-C_2) \\
 p_{10|1} &= S_1(1-S_2) & p_{10|0} &= (1-C_1)C_2 \\
 p_{01|1} &= (1-S_1)S_2 & p_{01|0} &= (1-C_2) \\
 p_{00|1} &= (1-S_1)(1-S_2) & p_{00|0} &= C_1 C_2
 \end{aligned}$$

Table 4-8. Classification of probabilities as explained by sensitivities and specificities.

		Disease positive animals							
		Population 1			Population 2				
		Test 2			Test 2				
		+	-		+	-			
Test 1	+	$p_{111 1}$	$p_{110 1}$	S_1	Test 1	$p_{211 1}$	$p_{210 1}$	S_1	
	-	$p_{101 1}$	$p_{100 1}$	$1-S_1$		-	$p_{201 1}$	$p_{200 1}$	$1-S_1$
		S_2	$1-S_2$	l			S_2	$1-S_2$	l

Table 4-8 (continued). Classification of probabilities as explained by sensitivities and specificities.

		Disease negative animals					
		Population 1			Population 2		
		Test 2			Test 2		
		+	-		+	-	
Test 1	+	p_{11110}	p_{11100}	$1-C_1$	p_{21110}	p_{21100}	$1-C_1$
	-	p_{10110}	p_{10010}	C_1	p_{20110}	p_{20010}	C_1
		$1-C_2$	C_2	I	$1-C_2$	C_2	I

Ses and Sps are computed using the law of total probability(Casella and Berger, 2001)

$$S_1 = P(T_1|D=1) = P(T_1=1, T_2=1|D) + P(T_1=1, T_2=0|D) = p_{11111} + p_{10111}$$

$$S_2 = P(T_2|D=1) = P(T_1=1, T_2=1|D) + P(T_1=0, T_2=0|D) = p_{11111} + p_{01111}$$

$$C_1 = P(T_1|D=0) = P(T_1=0, T_2=0|D) + P(T_1=0, T_2=1|D) = p_{00010} + p_{00110}$$

$$C_2 = P(T_2|D=0) = P(T_1=0, T_2=0|D) + P(T_1=1, T_2=0|D) = p_{00010} + p_{10010}$$

The parameters π_1 , π_2 , S_1 , S_2 , C_1 , and C_2 were estimated using simulation-based Bayesian inference. In Bayesian inference all the parameters are treated as random variables conditioned on the observed data. Hence, the joint posterior inference is determined by the combined input of prior information and the likelihood, where the priors express preceding uncertainty about the parameter estimates. Finally, the marginal densities for the six parameters can be obtained using Gibbs sampler. In short, the Gibbs sampler works as follows. First, the latent data (a_{ijk}) are sampled from independently distributed binomial distributions, where the latent data are two 2×2 tables of the subjects that are disease positive ($D=1$), i.e. a_{ijk} . The latent counts are sampled

independent from their conditional distributions given the observed counts $\{n_{ijk}\}$, the classification probabilities $\{p_{jkl}\}$ and the prevalence $\{\pi_i\}$, (Johnson et al., 2001) hence

$$a_{ijk} \sim \text{Binomial} \left(n_{ijk}, \frac{\pi_i P_{jk11}}{\pi_i P_{jk11} + (1 - \pi_i) P_{jk10}} \right), \text{ Note that } p_{jk11} = S_1 S_2, \text{ and so}$$

forth.

The values for the latent counts in the disease negative tables are simply

$$b_{ijk} = n_{ijk} - a_{ijk} \text{ (Black and Craig, 2002)}$$

Second, the full conditional distributions for the six parameters are sampled successively.

$$\pi_1 \sim \text{Beta} \left(a_{1..} + \alpha_{\pi_1}, b_{1..} + \beta_{\pi_1} \right)$$

$$\pi_2 \sim \text{Beta} \left(a_{2..} + \alpha_{\pi_2}, b_{2..} + \beta_{\pi_2} \right)$$

$$S_1 \sim \text{Beta} \left(a_{.1} + \alpha_{S_1}, b_{.1} + \beta_{S_1} \right)$$

$$S_2 \sim \text{Beta} \left(a_{.1} + \alpha_{S_2}, b_{.1} + \beta_{S_2} \right)$$

$$C_1 \sim \text{Beta} \left(b_{.0} + \alpha_{C_1}, b_{.0} + \beta_{C_1} \right)$$

$$C_2 \sim \text{Beta} \left(b_{.0} + \alpha_{\pi_2}, b_{.0} + \beta_{\pi_2} \right),$$

where α 's and β 's represent *a priori* information. The Gibbs sampler is then iterated to convergence.

Conditional dependence model (M2)

For the dependence model, we extended the approach of Black and Craig (2002) to include test results from two populations. However, we limited our model to the full conditional dependence model, and no restrictions were placed on the relationship between p_{1111} and S_1S_2 , and p_{0010} and C_1C_2 . Hence, we allowed for the correlation between disease positive and negative animals to range between -1 and 1 . The restrictions placed on the parameter space by introducing p_{1111} and p_{0010} makes it difficult to sample from the full conditionals $\{S_1, S_2, p_{1111}\}$ and/or $\{C_1, C_2, p_{0010}\}$. However, this is resolved by jointly updating the sets of parameters using the Metropolis-Hastings (MH) algorithm. (Black and Craig, 2002) In their model, Craig and Black, (2002) placed uniform priors on p_{1111} and p_{0010} , respectively, given S_1, S_2, C_1 , and C_2 , thereby forcing positive correlation between Se and Sp. Hence, the classification probabilities in the disease positive and negative table are the function of Se's and p_{1111} , and Sp's and p_{0010} , respectively. (Black and Craig, 2002)

Using a Dirichlet(1,1,1,1) prior for the four cells in the disease positive and negative tables, then

$$\begin{bmatrix} \dot{p}_{11|1} \\ \dot{p}_{10|1} \\ \dot{p}_{01|1} \\ \dot{p}_{00|1} \end{bmatrix} \sim \text{Dirichlet} \begin{bmatrix} a_{.11}+1 \\ a_{.10}+1 \\ a_{.01}+1 \\ a_{.00}+1 \end{bmatrix} \quad \begin{bmatrix} \dot{p}_{11|0} \\ \dot{p}_{10|0} \\ \dot{p}_{01|0} \\ \dot{p}_{00|0} \end{bmatrix} \sim \text{Dirichlet} \begin{bmatrix} b_{.11}+1 \\ b_{.10}+1 \\ b_{.01}+1 \\ b_{.00}+1 \end{bmatrix}$$

$$\dot{S}_1 = \dot{p}_{11|1} + \dot{p}_{10|1}, \dot{S}_2 = \dot{p}_{11|1} + \dot{p}_{01|1}, \dot{C}_1 = \dot{p}_{01|0} + \dot{p}_{00|0}, \text{ and } \dot{C}_2 = \dot{p}_{10|1} + \dot{p}_{00|1}$$

For the disease positive table, the cycle move from the previous MCMC cycle to the next cycle with the probability (acceptance rate)

$$P(\text{move}) = \min \left(1, \frac{\min(S_1, S_2) - S_1 S_2}{\min(\dot{S}_1, \dot{S}_2) - \dot{S}_1 \dot{S}_2} \prod_{i=1}^2 \left(\frac{S_k}{\dot{S}_k} \right)^{\alpha_{S_k} - 1} \left(\frac{1 - S_k}{1 - \dot{S}_k} \right)^{\beta_{S_k} - 1} \right)$$

For the disease negative table

$$P(\text{move}) = \min \left(1, \frac{\min(C_1, C_2) - C_1 C_2}{\min(\dot{C}_1, \dot{C}_2) - \dot{C}_1 \dot{C}_2} \prod_{l=1}^2 \left(\frac{C_k}{\dot{C}_k} \right)^{\alpha_{C_k} - 1} \left(\frac{1 - C_k}{1 - \dot{C}_k} \right)^{\beta_{C_k} - 1} \right)$$

The prevalences and latent counts, a_{ijk} , can then be sampled directly from

$$a_{111} \sim \text{Binomial} \left(n_{111}, \frac{\pi_1 p_{11|1}}{\pi_1 p_{11|1} + (1 - \pi_1)(1 - C_1 - C_2 + p_{00|0})} \right)$$

$$a_{110} \sim \text{Binomial} \left(n_{110}, \frac{\pi_1 (S_1 - P_{111})}{\pi_1 (S_1 - P_{111}) + (1 - \pi_1) (C_2 - P_{0010})} \right)$$

$$a_{101} \sim \text{Binomial} \left(n_{101}, \frac{\pi_1 (S_2 - P_{111})}{\pi_1 (S_2 - P_{111}) + (1 - \pi_1) (C_1 - P_{0010})} \right)$$

$$a_{100} \sim \text{Binomial} \left(n_{100}, \frac{\pi_1 (1 - S_1 - S_2 + P_{111})}{\pi_1 (1 - S_1 - S_2 + P_{111}) + (1 - \pi_1) P_{0010}} \right)$$

$$a_{211} \sim \text{Binomial} \left(n_{211}, \frac{\pi_2 P_{111}}{\pi_2 P_{111} + (1 - \pi_2) (1 - C_1 - C_2 + P_{0010})} \right)$$

$$a_{210} \sim \text{Binomial} \left(n_{210}, \frac{\pi_2 (S_1 - P_{111})}{\pi_2 (S_1 - P_{111}) + (1 - \pi_2) (C_2 - P_{0010})} \right)$$

$$a_{201} \sim \text{Binomial} \left(n_{201}, \frac{\pi_2 (S_2 - P_{111})}{\pi_2 (S_2 - P_{111}) + (1 - \pi_2) (C_1 - P_{0010})} \right)$$

$$a_{200} \sim \text{Binomial} \left(n_{200}, \frac{\pi_2 (1 - S_1 - S_2 + P_{111})}{\pi_2 (1 - S_1 - S_2 + P_{111}) + (1 - \pi_2) P_{0010}} \right)$$

$$\pi_1 \sim \text{Beta} \left(a_{1..} + \alpha_{\pi_1}, b_{1..} + \beta_{\pi_1} \right)$$

$$\pi_2 \sim \text{Beta} \left(a_{2..} + \alpha_{\pi_2}, b_{2..} + \beta_{\pi_2} \right)$$

The correlations between S_1 and S_2 (d_S) and C_1 and C_2 (d_C) were estimated and sampled in each cycle of the MH algorithm, where

$$d_S = \frac{p_{1111} - S_1 S_2}{\sqrt{(S_1(1-S_2)S_2(1-S_1))}} > 0, \text{ and } d_C = \frac{p_{0000} - C_1 C_2}{\sqrt{(C_1(1-C_2)C_2(1-C_1))}} > 0$$

Conditional dependence model (M3)

For this model we assumed that the specificity of CULT (T_2) was 1 (*see prior specification*). Then $P(T_2=0|D=0)=1$, which in turn implies that $P(D=1|T_2=1)=1$ as well.

In other words, if an animal was positive on T_2 , it was truly infected. This furthermore implies that the vector of latent counts for disease positive animals $\mathbf{a} = (a_{\cdot 11}, a_{\cdot 10}, a_{\cdot 01}, a_{\cdot 00})^T = (n_{\cdot 11}, a_{\cdot 10}, n_{\cdot 01}, a_{\cdot 00})^T$. Therefore, the vector for disease negative animals $\mathbf{b} = (b_{\cdot 11}, b_{\cdot 10}, b_{\cdot 01}, b_{\cdot 00})^T = (0, b_{\cdot 10}, 0, b_{\cdot 00})^T$. In fact, $p_{0110} = 0$ and $p_{1110} = 0$. Sampling of p_{1111} , S_1 , and S_2 was performed as described in model M2, however, $p_{0000} =$

$C_1 \sim \text{Beta}(b_{\cdot 0} + \alpha_{C_1}, b_{\cdot 0} + \beta_{C_1})$, and $p_{0110} = 1 - C_1$. a_{ijk} were sampled in the same

manner as for M2. The prevalences are sampled as before, however they simplify to

$$\pi_i \sim \text{Beta}(n_{i11} + a_{i10} + n_{i01} + a_{i00} + \alpha_{\pi_1}, b_{i10} + b_{i00} + \beta_{\pi_1})$$

Conditional dependence model (M4)

As the last model, a 'correlation-adjusted'-estimation-model using two tests was used. (Georgiadis et al., 2003) This model uses a different parameterization of the correlation between Se's and Sp's of the two tests than described for M2. We refer to the paper by Georgiadis et al. (2003) for further detail.

CHAPTER FIVE

HERD-LEVEL SENSITIVITY, SPECIFICITY, AND PREDICTIVE VALUES OF BOVINE TUBERCULOSIS SKIN TESTS IN MICHIGAN.

ABSTRACT

In managing the statewide bovine tuberculosis eradication program in Michigan, correct classification of infected herds is the most relevant actionable event for regulatory intervention because, with few exceptions, the entire herd is depopulated if one or more animals within a herd are diagnosed with bovine tuberculosis. A shareware program was used to estimate herd-level sensitivity (HSe), specificity (HSp) and predictive value positive (HPVP) and negative (HPVN) for the bovine tuberculosis tests as they were implemented in Michigan. Four different scenarios were simulated. 1) serial interpretation of the caudal fold and comparative cervical tuberculin tests, 2) serial interpretation of the tuberculin tests, and mycobacterial culture and polymerase chain reaction when the later two are used in parallel, 3) as for scenario 2, but specificity fixed to 1.0, and 4) sensitivity was increased to 0.9 (4a) or 0.95 (4b) and Sp=1.0. Estimates of test sensitivity (Se), specificity (Sp), true animal and herd prevalences were obtained from the literature and in collaboration with regulatory veterinarians. An empirical distribution of herd sizes in the northern part of the lower peninsula in Michigan was used to simulate the variation in herd sizes in the area of Michigan where bovine tuberculosis is most likely to occur. Furthermore, scenarios 2 and 3 were used to investigate the effect of herd size on herd-level parameters. In general HSe estimates were reasonably high, ranging between 0.712 and 0.840 for the four scenarios, however, individual estimates all ranged between 0.0 and 1.0. Increasing Se only increased HSe little. Herd-level specificity estimates were lowest for the two scenarios (0.693 and

0.940, respectively) where Sp was not assumed perfect (scenario 1 and 2). HPVP estimates were very low (0.042 and 0.143, respectively) for the first two scenarios also, but increased to 1.0 when perfect specificity was assumed. The HPVN remained high for all four scenarios, ranging between 0.995 and 0.997 with very narrow confidence intervals. The size of the tested herd has a major effect on HSe, which increases as the herd size increases. Both HSp and HPVP decreased with increasing herd size. However, when Sp was assumed perfect, only minor effect was seen on HSp and HPVN, but HSe remained low at low herd size. It is apparent that the tests used in identifying herds that are infected with bovine tuberculosis work fairly well on a herd basis. However, as the herd size decreases so does the HPVN and HSe. Hence, it is advisable to test herds with less than approximately 100 cattle more frequently and perhaps for an extended period of time as compared to larger herds.

INTRODUCTION

Five years after the last confirmed case, the Michigan cattle population was declared free from bovine tuberculosis (bTb) caused by *Mycobacterium bovis* in 1979.(Corso et al., 1996) The disease was not found in any animal species for 15 years, with exception of a hunter-killed white-tailed deer in 1975.(Stuth and Fay, 1975) This deer was considered to be an isolated case(Schmitt et al., 1997) and the disease was not diagnosed again until 1994 when bTb was identified in a single white-tailed deer.(Schmitt et al., 1997) The following year, bTb was diagnosed in a Michigan beef cow. As of June 2003, bTb has been diagnosed in approximately 400 white-tailed deer, two elk, one domestic cat, four bobcats, 13 coyotes, two opossums, two raccoons, and two red foxes (C. Bruning-Fann and S.D. Fitzgerald, personal communications 2002). Further, 30 cattle herds in the northern part of Michigan's lower peninsula have been found to have at least one bTb-infected animal. A plan was implemented for testing of all Michigan domestic cattle herds with the caudal fold tuberculin test (CFT) and the comparative cervical test (CCT) used in series, with the CCT being applied only to animals testing positive or "suspect" on the CFT (October 31, 2000, Amendment to the Animal Industry Act 466, State of Michigan.).

In managing a statewide bTb eradication program, correct classification of infected herds (herds with one or more infected animals) is the most relevant actionable event for regulatory intervention because, with few exceptions, the entire herd is depopulated if one or more animals are diagnosed with bTb. While the individual animal-level sensitivity (Se) and specificity (Sp) of the tuberculin skin tests have been

estimated under local area conditions in Michigan(Norby et al., 2003a; Norby et al., 2003b), the herd-level estimates of sensitivity (HSe) and specificity (HSp) of the serial testing program are more relevant to the regulatory veterinarians managing the eradication effort.

Factors influencing estimates of HSe and HSp include animal-level Se and Sp, the size of the sample (n), the true prevalence of the disease within positive herds (TP), the number of test-positive animals used to define a positive herd (cutoff)(Martin et al., 1992), and among herd variation in the Se, Sp and TP.(Donald et al., 1994) The herd-level predictive value of a positive classification/designation (HPVP) and of a negative classification/designation (HPVN) are furthermore dependent on the prevalence of herds with the disease (HTP). Additional information regarding herd-level testing and factors affecting herd-tests are reviewed by Christensen and Gardner (2000).

Martin *et al.* (1992) described a relatively simple way of estimating Hse, HSp, HPVP, and HPVN. However, because the binomial distribution is used for estimating the number of test positive animals in each sample, this method requires the assumption that the sample size used for testing is $\leq 20\%$ of the herd size.(Christensen and Gardner, 2000; Jordan, 1996) Nonetheless, when testing for bTb in Michigan, all animals over the age of 12 months in each herd were tested. This problem may be overcome by estimating the number of test positive animals using the hypergeometric distribution.(Cameron and Baldock, 1998). Jordan (1996) used this approach in a “shareware” program (HERDACC version 3) that is used to estimate HSe and HSp, using point estimates for Se, Sp, and TP, sampling from binomial or hypergeometric distributions. The results are presented for five different prevalence estimates and up to 10 different cut-off values

used to define a “positive” herd. Though this software program simplifies the calculations, it does not account for the uncertainty and variability in the input variables and the clustering of disease positive animals e.g. within a herd. Clustering of disease positive animals within a herd is plausible, and is the foundation for the utility of this testing scheme to detect infected herds. The bTb eradication programs of the early 1900’s recognized that the CFT and CCT used in serial interpretation detected infected herds (not necessarily infected animals) and that this worked in practice, because the infectiousness of the disease was such that it was exceedingly rare that only one animal in a herd would be infected without the existence of multiple infected herdmates, i.e., clustering. The shareware program, ‘HerdTest’ (Jordan and McEwen, 1998), permits the user to include uncertainty about Se, Sp, and TP in the model, and furthermore allows for a more complex sampling protocols. The objective of this study was to use the HerdTest program to estimate HSe, HSp, HPVP, and HPVN for several different bTb testing scenarios in Michigan.

MATERIALS AND METHODS

Diagnostic tests and testing procedure

The statewide testing procedure, and implementation of the CFT and CCT, mycobacterial culture (CULT), and polymerase chain reaction (PCR) has been described elsewhere. (Norby et al., 2003a) In short, the CFT was performed by public or private practicing veterinarians, using an intradermal injection of 0.1 ml of USDA bovine PPD (1mg/ml) (USDA and APHIS, 1999) in the caudal fold on either side of the base of the

tail. The injection site was palpated approximately 72 hours later, and the test was considered positive (“suspect”) if there was any sign of inflammation (Bruning-Fann, personal communication, 2002) If the CFT was “suspect”, a USDA or state veterinarian administered the CCT within 7 days of when the CFT was read. For the CCT, 0.1 ml of USDA bovine PPD and 0.1 ml of USDA avian PPD (biologically balanced, 1mg/ml) were injected intradermally on the neck approximately 12.5 cm apart. The thickness of the skin was measured before injection and approximately 72 hours after injections. The difference (‘after’ minus ‘before’) in skin-thickness was plotted on a scattergram (USDA, VS, Form 6-22D) used to interpret the CCT results as “negative”, “suspect”, or “reactor”. Necropsy and histological exam was conducted on CCT positive animals at the Diagnostic Center for Population and Animal Health at Michigan State University. Histological exam and CULT was performed on fresh samples of select tissues from all cattle that were necropsied. Samples that were suspect for *M. bovis* on histological exam were tested with PCR at the National Animal Disease Center and CULT was performed at the National Veterinary Services Laboratories, Ames, Iowa. For six animals, tissues were also cultured at the Michigan Department of Community Health.(Norby et al., 2003a) The necropsy, culture, and PCR procedure has been described elsewhere.(Miller et al., 1997; Norby et al., 2003a; Schmitt et al., 1997) In the CULT procedure, a very specific genetic probe was used to classify mycobacterial species into isolates belonging to the *Mycobacterium* complex, which include *M. bovis*, *tuberculosis*, *africanum*, and *microtii*.(Schmitt et al., 1997) Since it was unlikely that cattle in Michigan were infected with mycobacteria belonging to the *Mycobacterium* complex other than *M. bovis*, it is assumed that all isolates that were positive on culture were *M. bovis*.

Analyses

The program 'HerdTest' as described by Jordan and McEwen (1998) was used to estimate values for HSe, HSp, HPVP, HPVN for diagnostic tests used in the bTb eradication program in Michigan. The input variables for this model included: Se, Sp, TP, and HTP. Other inputs included a choice between binomial or hypergeometric sampling, how many test-positive animals were needed to denote a positive herd, and a selection of transmission mode (free-living-contagion or obligate-parasite). Input variables may be either point estimates or frequency distributions. HSe, HSp, HPVP, and HPVN were estimated for several different test scenarios: 1) serial interpretation of the CFT and CCT(Norby et al., 2003b), 2) serial interpretation of the CFTCCT_{SER} and mycobacterial culture (CULT) and PCR used in parallel (Se_{MI} and Sp_{MI}), 3) as for scenario 2, but Sp fixed 1.0, and 4) Se increased to 0.9 (4a) or 0.95 (4b) and Sp=1.0. All four scenarios were performed using the distribution of herd sizes in 10 counties (figure 5-1) in the northern part of the lower peninsula in Michigan, as reported by the National Agricultural Statistics Service, USDA (Table 5-1). In addition, the herd-level parameters were estimated for herd sizes fixed at 10, 25, 50, 100, 250, 500, and 1000 head of cattle, using Se and Sp from scenarios 2 and 3. The results for HSe, HSp, HPVP, and HPVN are reported as median, first and third quartiles, and the range.

Figure 5-1. Geographical location of the 10-county-area in northern Michigan.



Input distribution for Se, Sp, TP, and HTP

The input parameters for HerdTest were entered as frequency distributions to account for the uncertainty and variability in parameter estimates.(Jordan and McEwen, 1998) The Se and Sp for $CFTCCT_{SER}$ (scenario 1) were estimated by Norby *et al.* (2003a, b). However, there are no published estimates on the Se and Sp of all the tests (CFT, CCCT, CULT, and histologic exam and PCR) used in series in Michigan (Scenario 2 and 3). Hence, distributions for $CFTCCT_{SER}$, CULT and PCR used in serial interpretation were estimated in collaboration with veterinary epidemiologists at APHIS,

USDA (Lansing, Michigan). These distributions were created by selecting the most likely values (median) and a range that was likely to cover 95% of all possible values. For example, it was estimated that the most likely value for Se_{MI} was 0.60 and the 5th and 95th percentiles were 0.50 and 0.70, respectively. These estimates were based on empirical evidence and the Se estimate of CULT as reported by Norby *et al.* (2003b). The Sp was assumed to be very close to one and for scenario 2, it was suggested to follow a uniform distribution ranging between 0.995 and 1.0. For scenario 3, Sp was fixed at 1.0. The true within-herd prevalence was based on estimates from Norby *et al.* (2003b) and the median was judged to be 0.060 with 5th and 95th percentiles of 0.005 and 0.20, respectively. The input estimate of HTP was based on the number of herds that were detected each year in the bTb 'infected zone', which consist of Alpena, Montmorency, Alcona, and Oscoda counties. Approximate 8 new bTb herd diagnoses are reported each year from this area, and the median and percentiles for HTP were hence estimated to be 0.02, 0.005, and 0.05, respectively.

RISKview 4.0.4 (Palisades Corporation, Newfield, NY) was used to fit beta distributions that followed the specifications for TP, HTP and sensitivities (Table 5-1). For example, beta distribution for TP was Beta(1.7,22). The nature of the beta distributions and RISKview made it impossible to fit beta distributions that exactly followed the specifications for median and 5th and 95th percentiles for TP, HTP, and sensitivities. However, both median, 5th and 95th percentiles as estimated by RISKview were within 0.01 of the preferred values. The resulting distribution with median and 5th and 95th percentiles are shown in Table 5-1.

Herd size distributions and sampling technique

The herd level parameters were estimated for different fixed herd sizes, as describe earlier. Furthermore herd level parameters were estimated for a local area containing 10 counties in the upper part of the lower peninsula (figure 5-1). The 10 county area was based on the current division into a 'disease zone', 'surveillance zone', and 'disease free zone' in the area where bTb is present in Michigan. The 10 county area consist of the 'disease zone' and the surrounding counties (the 'surveillance zone'). It is was considered unlikely that bTb is present outside this 10 county area, and using an empirical herd distribution for the entire Michigan and a very low herd-level prevalence could negatively bias the results of the herd-level performance of the tests in the area where bTb is present. Empirical distributions of herds in all of Michigan and the 10 county area were reported for 1997 (Table 5-2) in 'Michigan Agricultural Statistics' (Michigan Agricultural Statistics Service, National Agricultural Statistics Service (NASS), USDA, <http://www.usda.gov/nass/mi/>). The last category of herd sizes was '≥500. It was assumed that all herds in this category were evenly distributed between 500 and 1,000 animals.

Table 5-1. Input variables for sensitivity (Se) Specificity (Sp), and within herd prevalence (TP) and herd prevalence (HTP) used for the shareware program 'HerdTest'. [Jordan and McEwen, 1998].

Input variables	Distribution	5th percentile	Median	95 th percentile
^a Se CFTCCT _{SER}	Beta(7.1,2.5)	0.438	0.757	0.948
^a Sp CFTCCT _{SER}	Beta(98.6,1.7)	0.950	0.986	0.998
^{b,c} Se _{MI}	Beta(47,31)	0.493	0.603	0.708
^b Sp _{MI}	Uni(0.995,1)	0.995	0.998	1.0
^c Sp ₁	Fixed at 1.0	1.0	1.0	1.0
^d Se ₉₀	Beta(90,10)	0.834	0.903	0.951
^e Se ₉₅	Beta(95,5)	0.900	0.953	0.983
TP	Beta(1.7,22)	0.007	0.060	0.202
HTP	Beta(3,130)	0.0047	0.020	0.054

^aSe and Sp as estimated when using only the caudal fold and comparative cervical tuberculin tests in series (CFTCCT_{SER}), scenario 1.

^bSe and Sp as estimated for the serial use of the CFTCCT and mycobacterial culture and/or histopathological exam and PCR (Se_{MI}, Sp_{MI}), scenario 2.

^cSe and Sp used in scenario 3.

^dSe set equal to 0.90, Sp set equal to 1.0, scenario 4a.

^eSe set equal to 0.95, Sp set equal to 1.0, scenario 4b.

Table 5-2. Distributions of herds in 10 counties and all of Michigan in 1997.

Herd size	No. of herds in 10 county area (%)	No. of herds in Michigan (%)
1-9 ^a	197 (19.4%)	3,913 (25.3%)
10-19 ^a	168 (16.5%)	2,929 (18.9%)
20-49 ^a	315 (31.0%)	3,803 (24.6%)
50-99 ^a	163 (16.0%)	2,148 (13.9%)
100-199 ^a	107 (10.5%)	1,497 (9.7%)
200-499 ^a	58 (5.7%)	955 (6.2%)
≥ 500 ^{a,b}	8 (0.8%)	223 (1.4%)
Total	1,016 (99.9%)^c	15,468 (100%)

^aCategories of herds sizes. It is assumed that the herd sizes within each category are evenly distributed over the entire interval. The data are reported in agricultural statistics for Michigan in 1997, (<http://www.usda.gov/nass/mi/>). ^bIt was assumed that herd sizes for the seventh category (≥ 500) was limited between 500 and 1,000. ^cThe discrepancy to 100% is due to round-off errors.

The hypergeometric distribution was selected as the sampling technique, because all animals in the herds were sampled.(Jordan and McEwen, 1998) The free-living-contagion mode where herds were allowed to have no infected animals was chosen, because susceptible cattle may be infected with *M. bovis* from other sources than bTb infected cattle within the same herd.(Kaneene et al., 2002)

Spearman's correlation coefficient (r_s) was used to investigate how input variables affected the output variables.(Jordan and McEwen, 1998)

RESULTS

The results of the simulated herd-level parameters are shown in Tables 5-3 and 5-4 for all four scenarios. For example, scenario 2 showed a median HSe of 0.75, quartiles of 0.369, and 1.0 and a range of 0 - 1. HSe, HSp, HPVP, and HPVN for scenario 2 and 3 at different fixed herd sizes are shown in figure's 5-2 and 5-3, respectively. Increasing Se to 0.9 and 0.95 (scenario 4a and 4b) had only minor effects on HSe (0.783 and 0.833, respectively) and HPVN (0.996 and 0.997, respectively), and no effect on HSp and HPVN. When Sp was not fixed at 1.0 (scenario 1 and 2), the HPVP was very low (0.042 and 0.143, respectively).

Table 5-3. Herd-level sensitivity (HSe) and specificity (HSp) for four different scenarios with individual animal sensitivity (Se) and Specificity (Sp) used in 10 counties in the northern part of the lower peninsula in Michigan.

Scenario ^a	HSe			HSp		
	Median	Q1, Q3	Range	Median	Q1, Q3	Range
1	0.840	0.423,1	0.0, 1.0	0.693	0.372, 0.826	0.0, 1.0
2	0.750	0.369, 1.0	0.0, 1.0	0.940	0.838, 0.980	0.023, 1.0
3	0.712	0.302, 1.0	0.0, 1.0	1.0	1.0, 1.0	0.981, 1.0
4a	0.783	0.375, 1.0	0.0, 1.0	1.0	1.0, 1.0	1.0, 1.0
4b	0.833	0.412, 1.0	0.0, 1.0	1.0	1.0, 1.0	1.0, 1.0

^aMedian and 5th and 95th for the input variables (Se, Sp, TP, and HTP) are shown in Table 5-1.

Spearman's rank correlation coefficients were calculated for all scenarios to show the extent of dependence between input parameters and herd-test performance. They are shown specifically for scenario 3 in Table 5-4. The correlation between input variables

and herd-level test performance were similar for the other scenarios (results not shown).

When Sp was 1.0, there was no effect of herd size on the herd-level parameters, and correlations between HSe and HPVP and the input variables were non-existent.

Table 5-4. Herd-level predictive value positive (HPVP) and negative (HPVN) for four different scenarios with individual animal sensitivity (Se) and Specificity (Sp) used in 10 counties in the northern part of the lower peninsula in Michigan.

Scenario ^a	HPVP			HPVN		
	Median	Q1, Q3	Range	Median	Q1, Q3	Range
1	0.042	0.020, 0.078	0.0, 1.0	0.995	0.986, 1.0	0.922, 1.0
2	0.143	0.066, 0.296	0.0, 1.0	0.996	0.988, 1.0	0.935, 1
3	1.0	1.0, 1.0	1.0, 1.0	0.995	0.987, 1.0	0.942, 1.0
4a	1.0	1.0, 1.0	1.0, 1.0	0.996	0.988, 1.0	0.939, 1.0
4b	1.0	1.0, 1.0	1.0, 1.0	0.997	0.989, 1.0	0.935, 1.0

^aMedian and 5th and 95th for the input variables (Se, Sp, TP, and HTP) are shown in Table 5-1.

Table 5-5. Spearman's rank correlation between herd-level input variables and herd-level output variables for Scenario 3. *p*-values are shown in brackets ($H_0: r_s=0$).

	Herd size	Sensitivity	Specificity	^e P(HD+)	^f P(D+)
^a HSe	+0.83 (0.00)	+0.02 (0.63)	-0.10 (0.002)	-0.04 (0.21)	+0.35 (0.00)
^b HSp	-0.83 (0.00)	+0.03 (0.37)	+0.51 (0.00)	-0.06 (0.05)	+0.03 (0.00)
^c HPVP	-0.38 (0.00)	+0.00 (1.0)	+0.43 (0.00)	+0.55 (0.00)	+0.20 (0.00)
^d HPVN	+0.69 (0.00)	+0.04 (0.23)	+0.03 (0.28)	-0.32 (0.00)	+0.41 (0.00)

^aHerd-level sensitivity.

^bHerd-level specificity.

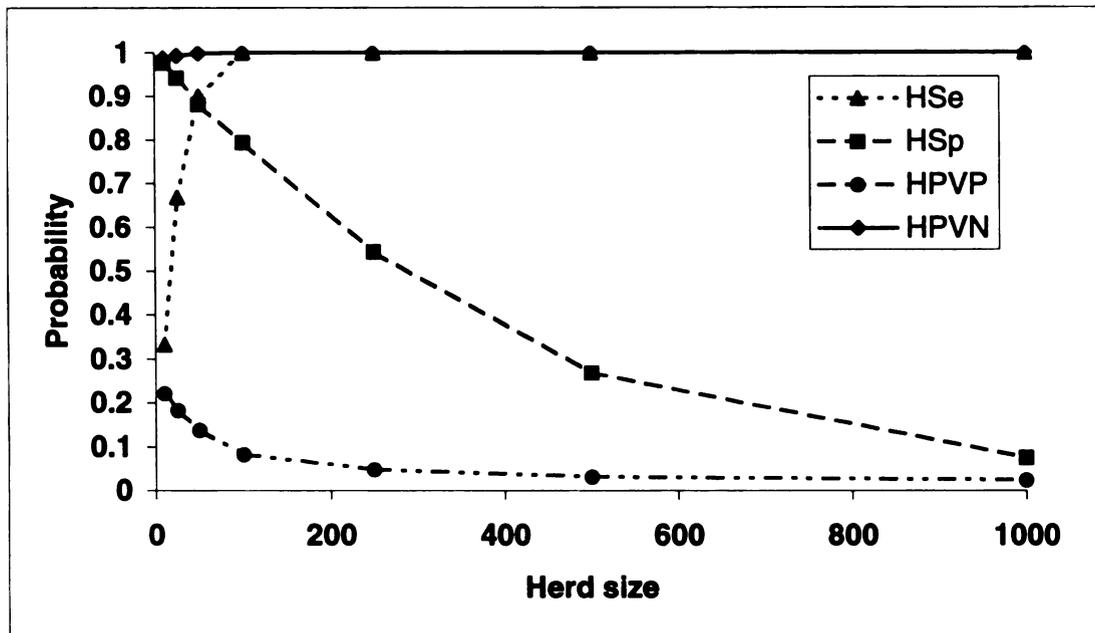
^cHerd-level predictive value positive.

^dHerd-level predictive value negative.

^eProbability that a randomly selected herd is disease positive.

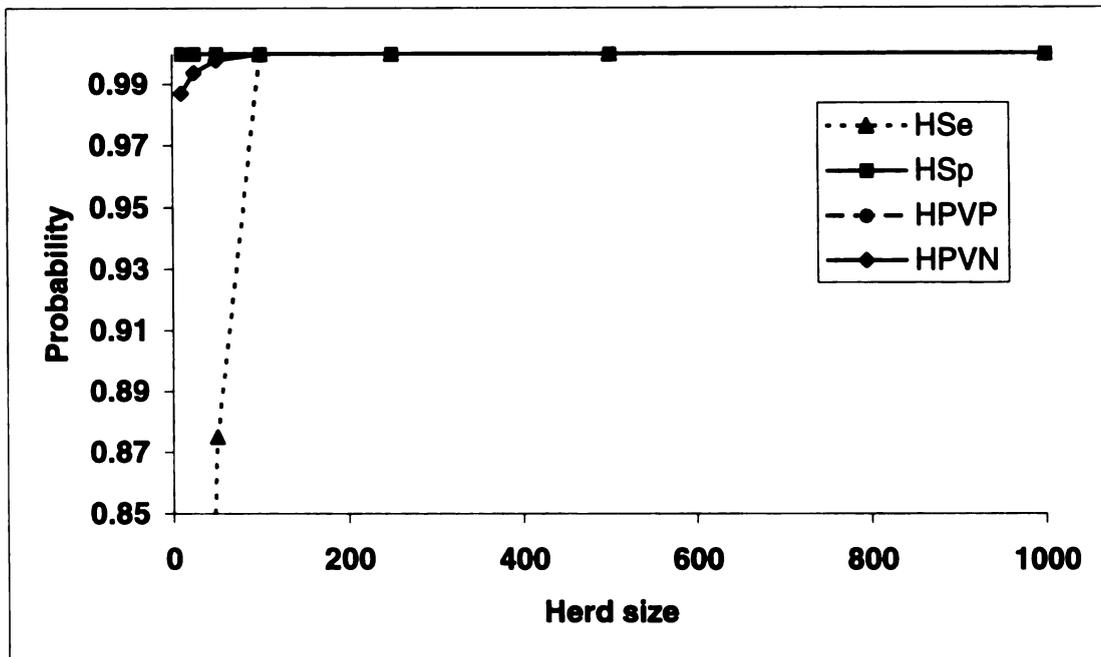
^fProbability that a randomly selected animal is disease positive.

Figure 5-2. Scenario 2 - Medians for herd level sensitivity (HSe), specificity (HSp), predictive value positive (HPVP), and negative (HPVN) for different fixed herd sizes using input values from scenario 2.



The following input variables were used for individual animal sensitivity (Se) and specificity (Sp): Median and 5th and 95th percentiles for Se and Sp were 0.603 (0.493,0.708), Sp=0.998 (0.989,1.0), respectively. Medians and 5th and 95th percentiles for within herd prevalence (P) and among herd prevalence (HP) were 0.060 (0.007,0.202) and 0.020 (0.005,0.054), respectively. The cut point for disease positive animals was 1.

Figure 5-3. Scenario 3 - Medians for herd level sensitivity (HSe), specificity (HSp), predictive value positive (HPVP), and negative (HPVN) for different fixed herd sizes in scenario 3.



The following input variables were used for individual animal sensitivity (Se) and specificity (Sp): Median and 5th and 95th percentiles for Se and Sp were 0.603 (0.493,0.708), 1.0 (1.0,1.0), respectively. Medians and 5th and 95th percentiles for within herd prevalence (P) and among herd prevalence (HP) were 0.060 (0.007,0.202) and 0.020 (0.005,0.054), respectively. The cut point for disease positive animals was 1. The data point for HSe for herd sizes 10 and 25 were 0.333 and 0.667. HSe and HPVP were 1.0 for all used herd sizes.

DISCUSSION

In this study we estimated the herd-level sensitivity, specificity, and predictive values of bTb diagnostic tests as they are applied to cattle herds in Michigan. Results from four different testing scenarios were investigated with different input values for estimating HSe, HSp, HPVP, and HPVN. Scenario 3 (Se = 0.60, Sp = 1.0, TP = 0.06 and HTP = 0.005) was considered to be most applicable to the situation in Michigan, because all the tests are used in serial interpretation and because false positive results on CULT

and PCR are not likely.(Norby et al., 2003a) The other scenarios were investigated to examine the effect of only using the $CFTCCT_{SER}$ in the eradication program as well as investigating the effect of increasing the Se of the testing scheme.

For scenario 3, HSp and $HPVP$ are both 1.0 with range (1.0-1.0), i.e. if a herd is diagnosed positive for bTb, it is certain that the herd is truly infected (i.e., no false positive herds). However, it is worth noting that unless $Sp=1.0$, then $HPVP$ is very small (scenarios 1 and 2).

The $HPVN$ is very high for all scenarios, with the median from 0.995 to 0.997. Hence a negative herd test will almost always be correct (i.e. few false positive herds).

The effect of herd size on HSe , HSp , HPP , and $HPVN$ was investigated by keeping values for herd size fixed (figure 5-3 and 5-4) at 10, 25, 50, 100, 250, 500, and 1000, while all other input values were kept fixed as described for scenario 2 and 3. Using inputs for Se , Sp , TP , and HTP from scenario 2 (figure 5-2), shows that the median for $HPVP$ drops quickly as herd size increased and slowly gravitates toward 0 when the herd size is larger than 250. The HSp also decreases with increasing herd size, however much slower than $HPVP$. Herd-level sensitivity increased with increasing herd size and the median reached 1.0 at a herd size of approximately 100. However, it is important to remember that both HSp , $HPVP$, and HSe in particular had wide ranges. When $Sp=1$ (scenario 3) (figure 5-3), changes in herd sizes only affected HSe and $HPVN$, and these changes were virtually identical to the changes described above for scenario 2.

Using the empirical herd distribution from the 10 counties, the HSe was only relatively high because the majority of herd (843 of 1,016) had less than 100 cattle. If the herd distribution for all of Michigan was used, the results were only slightly higher for

HSe (0.667) and slightly lower for HPVN (0.931). This was done assuming that TP and HTP are as high in the rest of Michigan as it is in the 10 county area. This is not a reasonable assumption, and a simulation using a uniform distribution (Uniform(0.00001,0.0001)) showed that Hsp, HPVP, and HPVN remained the same as in the previous scenario, but HSe only took on values of 0 and 1 (median=1, Q1=0, and Q3=1, simulation not shown).

Under the most likely scenario in Michigan (scenario 3), it was obvious that HSe was low, in particular in smaller herds (figure 5-3). It was therefore investigated if an increase in Se would increase HSe. If Se was increased to 0.90 (Beta(90,10)) or 0.95 (Beta(95,5)) there was none or only a slight increase in HSe, 0.783, and 0.833, respectively. Hence using a test scenario with higher Se does not improve HSe substantially (unless Se=1.0), because the average herd size is small. This is further supported by the Spearman's rank correlations between Se and herd-level parameters, where the correlations are very small (Table 5-5) and not statistically significant. However, herd size (=sample size) has a strong influence on all herd-level parameters (Table 5-5). The correlation coefficients also show that Hse increase with increasing P(D+), whereas only HPVP increase if P(HD+) increases. Increasing Se will have no or very little effect on HPVN, HSp, and HPVP. Hence, the eradication program will only gain minor improvement by increasing the Se. Specificity has most influence on HSp and HPVP.

We used 1997 census data of herd distributions in Michigan as reported by Michigan Agricultural Statistics. Because only animals older than 12 months of age are tested for bTb, the true number of animals tested will be lower than the used herd

distribution. However, since the testing has been ongoing for several years now, it is likely that the distribution of herds used in this study is a fairly accurate estimate of the situation in Michigan.

We chose to use the free-living contagion mode for simulations in this study, allowing for cattle to get infected from other sources beside bTb-infected herdsmates. In Michigan, it was shown that there is a statistical significant association with a herd being positive and the prevalence of bTB infected white-tailed deer(Kaneene et al., 2002), hence white-tailed deer may serve as a reservoir for *M. bovis* and infect cattle directly or contaminate pastures and feed for the cattle. Nevertheless, a simulation was conducted to investigate the effect of using the obligate parasite mode. Using the obligate-parasite mode had no effect on HSp and HPVP. However, HSe was increased (median=0.933) and the range was decreased to 0.25-1. There was only a slight increase of 0.03 in HPVN and the range was virtually unchanged. This means that if our assumption about free-living-contagion mode as compared to obligate contagion mode is wrong, only slight errors in HSe and HPVN would have occurred, but estimates of HSe would be increased.

The 'HerdTest' Model

We chose the model constructed by Jordan and McEwen (1998) (implemented in the software program 'HerdTest') to estimate HSe, HSp, HPVP, and HPVN for the series of tests used to diagnose cattle herds infected with bTb in Michigan. This model has the advantage that uncertainty and variation in Se, Sp, TP, and HTP may be accounted for using frequency distribution for these input variables. Another advantage of this program is the estimation of Spearman's rank correlation coefficient. The r_s can

then be used to assess how changes in Se and Sp may influence the herd-level parameters. In the bTb situation in Michigan, it is clear that an increase in Se will have an advantageous effect on HSe, but only for smaller herds.

All the bTb tests used in Michigan are used in series, and hence the Se of these tests can never be better than the Se of the first test (CFT) used in the series of tests. The Se of the skin test(s) could be improved by removing the CCT from the series of tests, however, this also increases the number of cattle with false positive test results, and hence increases the number and cost of animals sent to necropsy for further work up. Another solution would be to use an additional test with high Se in parallel with the skin tests. This would increase the overall Se, but most likely decrease Sp. If another test(s) is used in parallel with the skin tests, it is important that this new test and the skin tests are conditionally independent to obtain the optimal effect of using tests in parallel.(Gaboric et al., 1996; Gardner et al., 2000) The gamma-interferon (γ -INF) has received much attention as an alternative to (or for use in combination with) the skin tests used in the United States.(Whipple et al., 1995) However, because both the skin tests and the γ -INF are measuring a cellular immune response(Wood et al., 1991) they may be conditionally dependent and the gain of using these tests in parallel may be obsolete. If the skin tests (or the γ -INF) could be used in parallel with another test like for example an ELISA detecting antibodies, then there may be a gain in Se sensitivity and hence HSe. Gaboric *et al.* (1996) showed that the sensitivity of the CFT and ELISA used in parallel (Se=95%) was higher than for the CFT (Se=84.9%) or ELISA (Se=65.6) used individually. However, as stated earlier, this would signify a decrease in the Sp of the test(s) and most likely affect HSp and HPVP in a negative sense.

Comments and recommendation to the eradication program in Michigan

It is apparent that the skin tests have been used very successfully in the United States to eradicate bTb from most states,(Anonymous, 1992) and this simulation study also shows that both HSp, and HSe are high in larger herds. The problem lies in detection of disease positive animals in small herds if only one (or few) animals are infected. This also reflects a characteristic of an eradication program where the number of diseased herds and diseased animals within an infected herd tend to decrease as the disease prevalence decreases.(Christensen and Gardner, 2000).

The tests used to eradicate bTb in Michigan work well on the herd-level. If the tests are used in a test-and-removal scheme, i.e. removing only animals that were positive on the bTb tests from the herd, it is likely that *M. bovis* infected animals will remain in a herd after testing.(Norby et al., 2003a) Such animals are a source for new infections in the herd, prolonging the eradication. Hence, it will take much longer for Michigan to obtain bTb free status. It is pertinent that special attention is paid to small herds with less than approximately 100 cattle, because the HSe and HPVN are low in small herds, and it might be worthwhile to continue the testing of smaller herds more frequently or for a longer period of time.

It is likely that a few herds may turn up bTb positive after the disease is considered eradicated. This was for example seen in Denmark where three herds have been diagnosed with bTb after the routine monitoring with tuberculin tests every three years was discontinued in 1980.(Flensburg, 1995) The three cases were in one instance traced to the owner who had open kidney tuberculosis caused by *M. bovis*, the second

case was most likely caused by the owners father, and in the third herd, no evidence of bTb could be found. If there truly is a non-bovine reservoir in Michigan as assumed in our simulations, this may cause the eradication effort to prolong considerably.

REFERENCES

- Anonymous 1992. Assessment of risk factors for *Mycobacterium bovis* in the United States (Fort Collins, Colorado USA, USDA, APHIS, VS).
- Cameron, A.R., Baldock, F.C., 1998, A new probability formula for surveys to substantiate freedom from disease. *Prev. Vet. Med.* 34, 1-17.
- Christensen, J., Gardner, I.A., 2000, Herd-level interpretation of test results for epidemiologic studies of animal diseases. *Prev. Vet. Med.* 45, 83-106.
- Corso, B., Wagner, B., Norden, D. 1996. Assessing the risks associated with *M. bovis* in Michigan free-ranging deer. (Fort Collins, Co, Centers for Epidemiology and Animal Health, USDA, APHIS), pp. 1-2.
- Donald, A.W., Gardner, I.A., 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.* 19, 167-187.
- Flensburg, J., 1995, Regional and country status reports, Denmark, In: Thoen, C.O., Steele, J.H. (Eds.) *Mycobacterium bovis* in animals and humans. Iowa State University Press, Ames, pp. 206-212.
- Gaboric, C.M., Salman, M.D., Ellis, R.P., Triantis, J., 1996, Evaluation of five-antigen ELISA for diagnosis of tuberculosis in cattle and Cervidae. *JAVMA* 209, 962-966.
- Gardner, I.A., Stryhn, H., Lind, P., Collins, M.T., 2000, Conditional dependence between tests affects the diagnosis and surveillance of animal diseases. *Prev. Vet. Med.* 45, 107-122.
- Jordan, d., 1996, Aggregate testing for the evaluation of Johne's disease health status. *Aust. Vet. J.* 73, 16-19.
- Jordan, D., McEwen, S.A., 1998, Herd-level test performance based on uncertain estimates of individual test performance, individual true prevalence and herd true prevalence. *Prev. Vet. Med.* 36, 187-209.
- Kaneene, J.B., Bruning-Fann, C.S., Granger, L.M., Miller, R., Porter-Spalding, B.A., 2002, Environmental and farm management factors associated with tuberculosis on cattle farms in northeastern Michigan. *JAVMA* 221, 837-842.
- Martin, S.W., Shoukri, M., Thorburn, M.A., 1992, Evaluating the health status of herds based on tests applied to individuals. *Prev. Vet. Med.* 14, 33-43.

Miller, J., Jenny, A., Rhyan, J., Saari, D., Suarez, D., 1997, Detection of *Mycobacterium bovis* in formalin-fixed, paraffin-embedded tissues of cattle and elk by PCR amplification of an IS6110 sequence specific for *Mycobacterium tuberculosis* complex organisms. JVDI 9, 244-249.

Norby, B., Bartlett, P.C., Fitzgerald, S.D., Granger, L., Bruning-Fann, C., Whipple, D.L., Payeur, J.B., 2003a, The sensitivity of gross necropsy, caudal fold and comparative cervical tests for the diagnosis bovine tuberculosis. JVDI In Press.

Norby, B., Tempelman, R.J., Hanson, T., Kaneene, J.B., Bartlett, P.C., 2003b, Estimation of sensitivity and specificity of bovine tuberculosis skin tests in Michigan when a perfect reference is not available, Dissertation, Michigan State University, MI

Schmitt, S., Fitzgerald, S., Cooley, T., Bruning-Fann, C., Berry, D., Carlson, T., Minnis, R., Payer, J., Sikarskie, J, 1997, Bovine tuberculosis in free-ranging white-tail deer in Michigan. J Wildlife Dis 33, 749-758.

Stuth, J., Fay, L.D. 1975. Laboratory record for necropsy #76-50 (East Lansing, Michigan: Wildlife Pathology Laboratory).

USDA, APHIS 1999. Bovine tuberculosis eradication. Uniform methods and rules, effective January 22, 1999, (United States Department of Agriculture.), pp. 1-34.

Whipple, D.L., Bolin, C.A., Davis, A.J., Jarnagin, J.L., Johnson, D.C., Payer, J.B., Saari, D.A., Wilson, A.J., Wolf, M.M., 1995, Comparison of the sensitivity of the caudal fold skin test and a commercial gamma-interferon assay for diagnosis of bovine tuberculosis. Am J Vet Res 56, 415-419.

Wood, P.R., Corner, L.A., Rothel, J.S., Baldock, C., Jones, S.L., Cousins, D.B., McCormick, B.S., Francis, B.R., Creeper, J., Tweddle, N.E., 1991, Field comparison of the interferon-gamma assay and the intradermal tuberculin test for the diagnosis of bovine tuberculosis. Austr Vet J 68, 286-290.

CHAPTER SIX

HERD AND SPATIAL FACTORS AFFECTING THE PROPORTION OF FALSE POSITIVE RESULTS ON THE CAUDAL FOLD TUBERCULIN TEST IN MICHIGAN CATTLE.

ABSTRACT

A survey of beef and dairy herds in Michigan tested with the caudal fold tuberculin test was carried out to quantify herd, and ecologic factors associated with the proportion of false positive tuberculin skin tests. Furthermore, a spatial analysis was used to identify clusters of herds with unusually high proportion of false positive results on the caudal fold tuberculin skin test. A total of 4,989 herds were included in the analysis of herd factors and a subset of 3,301 herds that had reliable information on geographical location were used to identify spatial clusters and analyze ecologic factors associated with the proportion of false positive results on the caudal fold tuberculin test. The overall proportion of false positive test results for beef and dairy was 4.82%, and 3.41% and 5.58% for beef and dairy herds individually. Negative binomial regression with the testing veterinarian included as a random effect was used to assess statistically significant factors associated with the proportion of false positive results on the caudal fold tuberculin skin test. The proportion of false positive test results was higher in dairy cattle than in beef cattle and considerably higher in areas of Michigan where bovine tuberculosis is present. The adjusted proportion of false positive test results was 2.25 and 4.38 cases per 100 animals tested for beef and dairy herds, respectively. Thirty-two clusters with a high proportion of false positive results on the caudal fold test were detected. The primary cluster had a radius of 19.5 km and was located in the center of one of the counties in the bovine tuberculosis infected area in Michigan. The study

identified factors that affect the proportion of false positive results on the caudal fold tuberculin tests that could be used by regulatory veterinarians to assess the bovine tuberculosis status of tested herds. Furthermore, regulatory veterinarians could use results from this study to identify geographic areas and/or veterinarian that consistently have high or low false positive test results.

INTRODUCTION

The bovine tuberculosis (bTb) skin tests have been used successfully for control and eradication of bTb in the United States since the eradication program was begun in 1917.(Anonymous, 1992) In 1992, there were only 10 states that had not achieved bTb-free status.(Frey, 1995) Nevertheless, bTb has become endemic in the cattle population in certain areas in the United States, including the northern part of the Michigan's lower peninsula.(Kaneene et al., 2002; Norby et al., 2003a) Although the methods for production of the tuberculin used in the skin tests have changed over time, tuberculin skin tests are still being used in series to detect individual animals and herds that are infected with bTb.(Angus, 1978) The caudal fold tuberculin test (CFT) is used to screen herds, and animals suspect (positive) on the CFT are retested with the comparative cervical test (CCT).(Norby et al., 2003a; USDA and APHIS, 1999)

When the prevalence of bTb infected animals is relatively high, e.g., 10%, the ratio between true positives and false positives (FP) test results is also relatively high. However, as the prevalence decreases, this ratio also decreases.(Frey, 1995) In other words, the predictive value of a positive CFT result decreases as the prevalence decreases.

The sensitivity (Se) and specificity (Sp) of the CFT has been estimated as used in Michigan,(Norby et al., 2003a; Norby et al., 2003b) and in several other studies in the United States(Gaboric et al., 1996; Whipple et al., 1995) and around the world.(Francis and Seiler, 1978; Lepper et al., 1977; Leslie et al., 1975a; Leslie et al., 1975b; Pollock et al., 2000; Wood et al., 1991; Wood et al., 1992) From these studies, it is evident that Sp

is not a constant parameter, hence neither is the proportion of false positive (PFP) as $PFP = 1 - Sp$. Variation in Sp can be due to: 1) differences in the amount and concentration of the tuberculin used for the CFT, 2) different (gold) standards used for estimation of Se and Sp of the CFT, 3) regional variation in cross-reacting mycobacteria that cause FP results on the CFT.(Cooney et al., 1997; Corner and Pearson, 1978a, 1978b), and 4) variation in the interpretation of the CFT by the testing veterinarians.

The objective of the current study was to use available data from bTb testing performed in Michigan to describe possible associations between herd and ecological/geographical factors and the proportion of FP results on the CFT (PFP), and to describe PFP clustering within Michigan.

MATERIALS AND METHODS

Diagnostic tests and testing procedure.

The statewide testing procedure, and implementation of the CFT and CCT, mycobacterial culture (CULT), and polymerase chain reaction (PCR) has been described elsewhere.(Norby et al., 2003a) In short, the CFT was performed by public or private practicing veterinarians, using an intradermal injection of 0.1 ml of USDA bovine PPD (1mg/ml)(USDA and APHIS, 1999) in the caudal fold on either side of the base of the tail. The injection site was palpated approximately 72 hours later, and the test was considered positive (“suspect”) if there was any sign of inflammation (Bruning-Fann, personal communication, 2002). If the CFT was “suspect”, a USDA or state veterinarian administered the CCT within 7 days of when the CFT was read. For the CCT, 0.1 ml of

USDA bovine PPD and 0.1 ml of USDA avian PPD (biologically balanced, 1mg/ml) were injected intradermally on the neck approximately 12.5 cm apart. The thickness of the skin was measured before injection and approximately 72 hours after injections. The difference ('after' minus 'before') in skin-thickness was plotted on a scattergram (USDA, VS, Form 6-22D) used to interpret the CCT results as "negative", "suspect", or "reactor".

Necropsy and histological exam were conducted on CCT positive animals at the Diagnostic Center for Population and Animal Health at Michigan State University. Culture and PCR were performed on fresh samples or paraffin embedded samples, respectively, at the National Veterinary Services Laboratories, and National Animal Disease Center, Ames, Iowa. For six animals, tissues were also cultured at the Michigan Department of Community Health.(Norby et al., 2003a) The necropsy, culture, and PCR procedure has been described elsewhere.(Miller et al., 1997; Schmitt et al., 1997)

The Michigan Department of Agriculture (MDA) maintains records of all cattle and herds in Michigan tested with the bTb tuberculin skin tests. The records include a herd ID, postal address, the date that testing was performed, the number of cattle that were tested, the number of 'suspect' (positive) and negative on the CFT, an ID-number of the veterinarian performing the testing, and whether or not the whole herd was tested ('whole herd test').

Between January 1st, 1996 and November 30th, 2001, 8,260 bTb negative cattle herds (4,223 beef and 4,037 dairy) underwent a 'whole herd test' with the CFT. Some herds were tested more than once during this period. These herds were all judged to have been negative for bTb because all cattle were negative on the CCT or cows positive on CCT were later shown to be negative by necropsy, CULT and PCR.(Norby et al., 2003a)

For a herd to be included in this study, the testing private veterinary practitioner had to have tested at least 3 herds.

Considering all herds with at least one CFT suspect animal, the crude rate of FP was 12% in beef herds and 7% in dairy herds. The minimum herd size for inclusion in our study was determined as the herd size that would have at least one FP CFT result with 95% confidence (C) when the proportion of false positive results (PFP) were 12% for beef and 7% for dairy herds.

$$CO \geq \frac{\log(1-C)}{\log(1-PFP)} \text{ (Martin et al., 1992)}$$

Therefore, the herd size in this study was limited to beef herds with > 12 cattle and dairy herds with > 22 cattle.

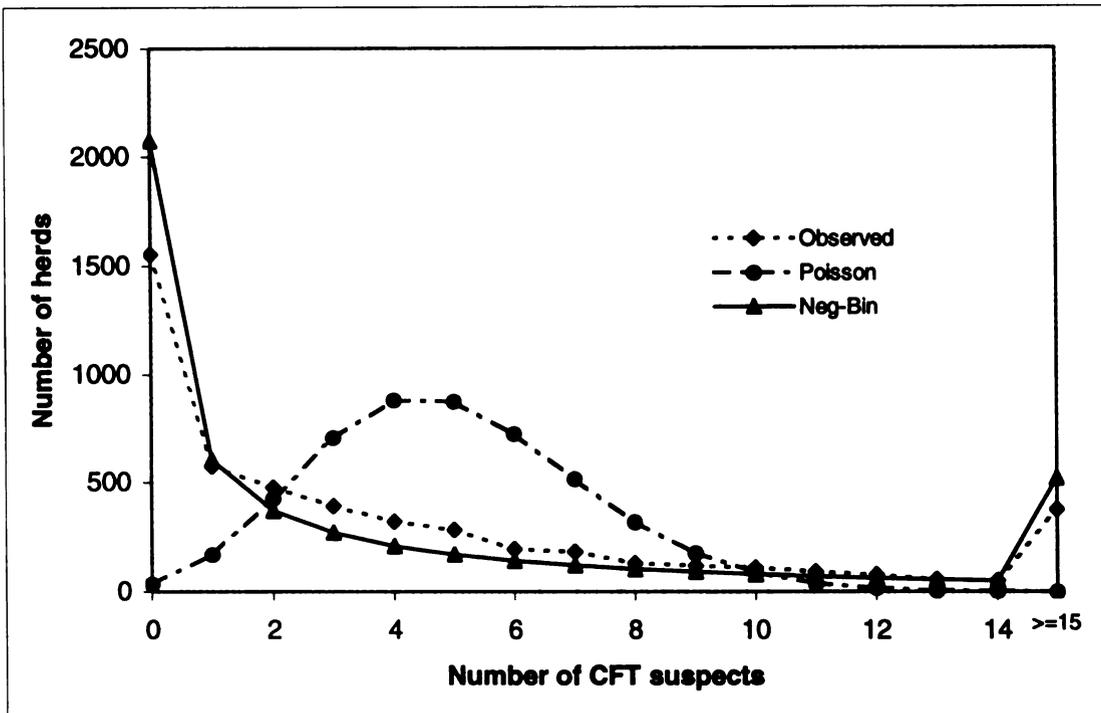
Statistical analysis

The data were analyzed using SAS statistical software package (SAS Institute, Cary, NC). Distributions of the observed data, the expected Poisson distribution ($P(Y=y) = e^{-\mu} \mu^y / y!$) (Rice, 1995) and the expected negative binomial distribution

$$P(Y = y) = \left(1 + \frac{m}{k}\right)^{-k} \left(\frac{\Gamma(k+m)}{y! \Gamma(k)}\right) \left(\frac{m}{m+k}\right)^y, m = \text{mean}, k = \left(\frac{\bar{X}^2}{\hat{\sigma}^2 - \bar{x}}\right), \text{ (Rice,}$$

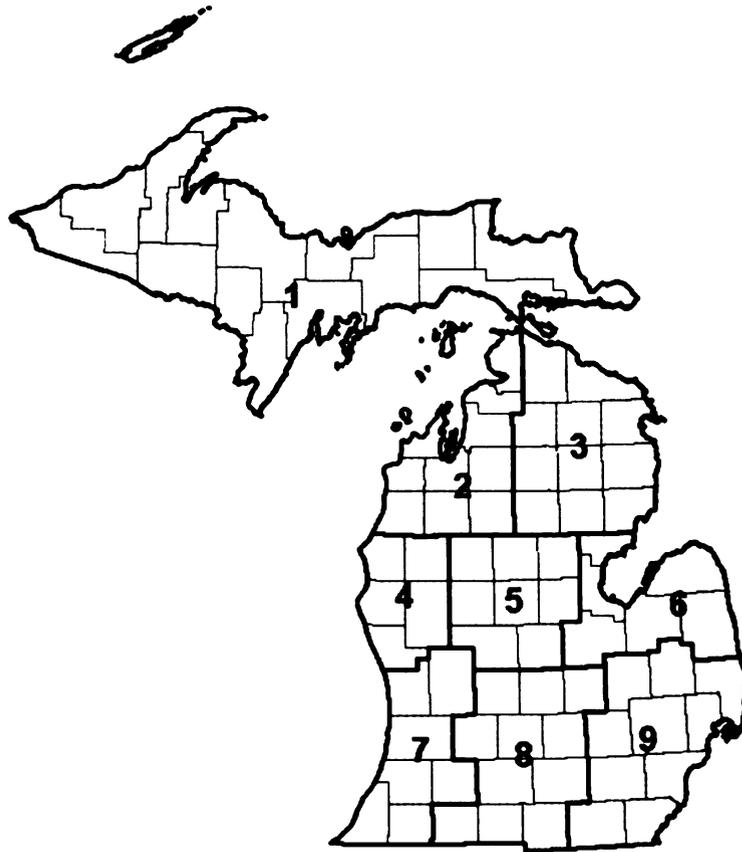
1995) are shown in figure 6-1. The data were checked for errors and descriptive statistics and frequency distributions were calculated.

Figure 6-1. Observed and expected frequencies of false positive test results per farm. Neg-Bin: Negative Binomial.



Associations between PFP and herd factors were tested using a negative binomial regression model(Wassink et al., 2003) in which the testing veterinarian was included as a random effect. Such a model allowed us to distinguish between sources of overdispersion due to residual variation and clustered random effects.(Tempelman and Gianola, 1996) The dependent variable of interest was the proportion of FP results (PFP) on the CFT for any given 'whole herd test' meeting the above stated selection criteria. The outcome variable for the negative binomial regression was the number of FP results on the CFT at a given 'whole herd' test, where the natural log (ln) of number of animals tested used as offsets.(Stokes et al., 2000) Thus we were fitting a loglinear mixed model to the proportion of CFT suspects to the independent variables.

Figure 6-2. Nine agricultural regions in Michigan.



The PFP was calculated as the number of animals suspect on the CFT/total number of animals tested, assuming that all the CFT suspect results were false positive. Explanatory variables, including herd type, agricultural region (figure 6-2), and the season of testing were each independently compared with PFP using a univariate negative binomial regression model. Explanatory variables with a likelihood ratio chi-squared test less than $p < 0.25$ in the univariate analysis were evaluated by the multivariable negative binomial regression model. Backward stepwise elimination procedure was used to obtain the final multivariable model. Inclusion of explanatory variables in the final model were based on the likelihood ratio chi-squared test with $p < 0.05$. In the multivariable negative

binomial regression, the identity of the testing veterinarians (VET) was included as a random variable. The goodness of fit measure, pseudo R^2 , was calculated as follows: $(\log\text{-likelihood null model} - \log\text{-likelihood full model})/\log\text{-likelihood full model}$, where the null model only includes the intercept.(Anonymous, 1997) Using a generalized estimating equations (GEE) method (Liang and Seger, 1986), the log-link mixed model used was as follows:

$$\text{Number of CFT suspects in a herd}_i = \alpha + \ln(\text{herd size})_i + \beta_j \text{var}_j + u_k + e_{ijk}.$$

α = intercept,

$\ln(\text{herd size})_i$ = natural logarithm of the number of cattle tested in a herd (offset variable).

β_j = regression coefficients for var_j ,

var_j = factor j for herd i .

u_k = random effects term,

e_{ijk} = residual random error.

Association between PFP and the number of herds tested by a veterinarian was estimated using Pearson product moment correlation ($H_0: \rho=0$).

Spatial analysis

The exact spatial coordinates (latitude and longitude) of the tested herds were not available, so the geographic location of herds was obtained by mapping (geocoding) the address of the herd to spatial data on roads in Michigan using ArcINfO (ESRI, Redlands, CA). The spatial reference data on roads in Michigan were obtained from the 2000

Census (<http://www.michigan.gov/cgi/>). Herds that had an address match $\geq 80\%$ to the reference address were considered an acceptable match.

Spatial clustering of high PFP was assessed using the spatial scan statistic, (Kulldorff and Nargawalla, 1995), because this method can assess the location of clusters. Ward and Carpenter (2000) reviewed the theoretical basis and general characteristics of this statistical method of detecting disease clusters and its use in veterinary epidemiology. In short, the spatial scan statistic creates multiple geographic circles (windows) centered around possible clusters in the study region. The radius of the window is determined by the investigator and may vary between 0 and an upper limit, which is recommended not to exceed 50% of the study area. (Kulldorff et al., 2003) Each geographic window is assessed by testing whether the probability of being a case within the circle is equal to the probability of being a case outside the circle. The null hypothesis is that these two probabilities are equal, which corresponds to complete spatial randomness. (Kulldorff and Nargawalla, 1995) The maximum likelihood test is used as the test statistic. The p -value for the maximum likelihood test is obtained through Monte Carlo simulation. The Poisson model was used to model our data and allowed control for the effects of herd type on spatial clusters of herds with a high proportion of CFT suspects. In this study, the scan statistic was used to detect geographical areas with unusually high proportions of FP results on the CFT. All spatial analyses were performed in SaTScan Ver. 3.1.2, a shareware program (<http://www.satscan.org/>), using a scan window of 2%.

Ecological associations

Spatial information on topography and wetland location in Michigan was obtained from <http://www.michigan.gov/cgi/>, and association with PFP was assessed through univariate negative binomial regression as described above for herd factors. A multivariable analysis was performed controlling for herd factors that were significant in the previous multivariable analysis. The testing veterinarian was again included as a random effect.

RESULTS

Descriptive statistics

Table 6-1 shows the number of observations, mean, standard deviation (Std), 0th (minimum), 25th, 50th (median), 75th and 100th (maximum) percentiles of the proportion of CFT false positive (PFP) in all herds, beef and dairy herds, and each of 9 agricultural regions. Table 6-2 shows the similar statistics for herd size stratified by herd type and agricultural region. The PFP clearly differs between beef and dairy herds and among agricultural regions. For example, the mean PFP in beef dairy herds was 0.0341 and 0.0558 in dairy herds. The distribution of PFP by herd is shown in figure 6-3 ($n = 4,989$), and it is obvious that the variability of PFP is larger among small herds as compared to larger herds.

The mean, std, 0th, 25th, 50th, 75th, and 100th percentiles of the number of herds tested by individual veterinarians were 56.0, 39.6, 4, 31, 47, 71, and 217, respectively. The distribution of the mean PFP by the number of herds tested by individual

veterinarians are shown in figure 6-4. No obvious pattern was observed between PFP and the number of herds a veterinarian had tested ($p=0.1060$). The number of herds tested in 1996, 1997, 1998, 1999, 2000, and 2001 were 7, 4, 44, 31, 1689, and 3214, respectively.

Table 6-1. Proportion of false positive (FP) on caudal fold tuberculin test (CFT) in Michigan, by herd type and agricultural region.

Variable	^a Obs.	Mean	Std	Percentiles					
				^b 0 %	25 %	^c 50 %	75 %	^d 100 %	
Herd type									
Beef	1735	0.0341	0.0583	0.00	0.00	0.00	0.0500	0.4500	0.4500
Dairy	3254	0.0558	0.0580	0.00	0.0155	0.0408	0.0770	0.4490	0.4490
Beef and Dairy	4989	0.0482	0.0590	0.00	0.00	0.0319	0.0700	0.4500	0.4500
^cAgric. regions									
1	438	0.0395	0.0520	0.00	0.00	0.0225	0.0625	0.3080	0.3080
2	299	0.0815	0.0801	0.00	0.0185	0.0645	0.1159	0.4333	0.4333
3	596	0.0664	0.0717	0.00	0.00	0.0476	0.1003	0.4375	0.4375
4	292	0.0437	0.0587	0.00	0.00	0.0253	0.0635	0.3667	0.3667
5	798	0.0620	0.0683	0.00	0.0086	0.0443	0.0909	0.4500	0.4500
6	624	0.0365	0.0425	0.00	0.00	0.0252	0.0542	0.2667	0.2667
7	549	0.0398	0.0534	0.00	0.00	0.0235	0.0588	0.3846	0.3846
8	1036	0.0373	0.0453	0.00	0.00	0.0247	0.0581	0.3488	0.3488
9	357	0.0387	0.0442	0.00	0.00	0.0263	0.0597	0.2571	0.2571

^aNumber of observations. ^bMinimum, ^cMedian, ^dMaximum.

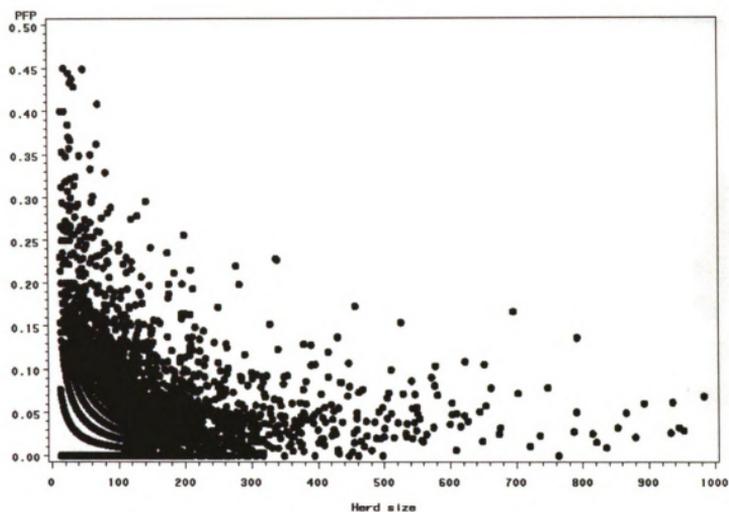
^eAgricultural regions.

Table 6-2. Herd size by herd type and agricultural region.

Variable	Mean	Std	Percentiles						
			^b 0 %	25 %	^c 50 %	75 %	^d 100 %		
Herd type									
Beef	47.55	50.74	13	19	29	53	390		
Dairy	131.63	149.68	23	50	89	158	2738		
Beef and Dairy	102.39	130.81	13	32	62	126	728		
^aAgric. Regions									
1	78.28	65.99	13	30	59	103	533		
2	96.91	95.08	13	29	61	127	526		
3	75.24	73.96	13	26	48	97.5	546		
4	108.87	132.31	14	35	64	119	1191		
5	90.69	136.17	13	30	55	107	2738		
6	112.69	133.11	13	45	79	135.5	1817		
7	133.23	152.21	13	40	80	180	1733		
8	113.22	159.5	13	31	60	128	1780		
9	105.82	130.00	13	30	70	136	1418		

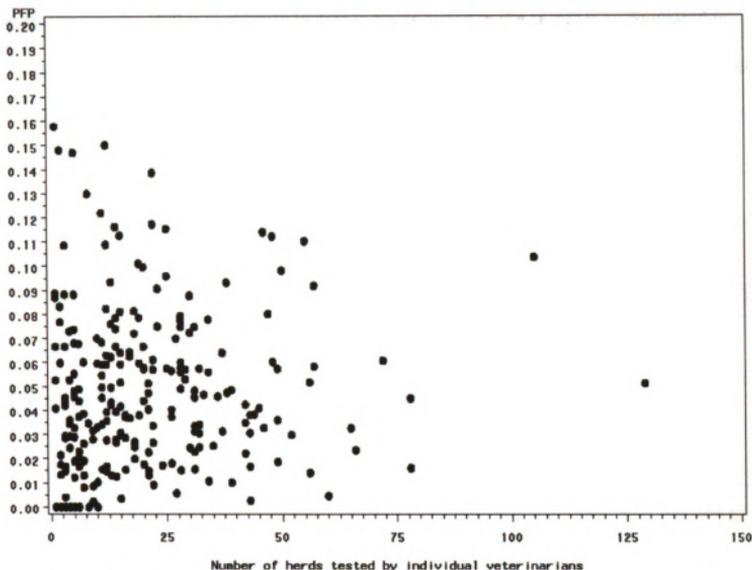
^aAgricultural regions. ^bMinimum, ^cMedian, ^dMaximum.

Figure 6-3. Distribution of proportion of CFT false positive (PFP) by herd size.



• is PFP in individual herds. The herd size was cut off at 1,000 to better show the dispersion at lower herd sizes, and because only 13 herds had more than 1,000 animals.

Figure 6-4. Distribution of the mean proportion of CFT false positive test results (PFP) by veterinarian.



• is the mean PFP by individual veterinarian testing 3 or more herds in Michigan. The Number of herds tested by an individual veterinarian was cut off at 150 to better show the dispersion at lower number of herds tested.

Negative binomial regression model

Then mean and variance of the number of CFT suspects for all herds were 4.97 and 84.64, respectively, indicating that the data were overdispersed, and negative binomial regression model was appropriate to model the data.

By univariate analyses (PFP vs one factor at a time), PFP was significantly lower in beef herds as compared to dairy herds (3.36 versus 5.51 cases per hundred tested animals, $p < 0.0001$) (Table 6-4). Cattle herds in agricultural region 2, 3, and 5, had

significantly higher PFP than what was found in agricultural region 9 (8.27, 6.83, and 6.24 versus 4.02 cases per hundred tested animals, respectively, $p < 0.0001$) (Table 6-3). Furthermore, the PFP was slightly higher in herds tested in the summer as compared to herds tested in the winter (Season) (5.15 versus 4.66, $p = 0.09$) (Table 6-3). Several other variables expressing the location of the herds and the time of year that herds were also tested, but they were not included in the final model to avoid multicollinearity.

In the final model, herd type and agricultural region were associated with the proportion of false positive results on the CFT (Table 6-4). The regression coefficients are adjusted for the effect of the random effect and offset variable.(Stokes et al., 2000) The regression function was hence:

$$\text{PFP} = \exp(-3.1275 + \beta_1(\text{herd type}) + \beta_2(\text{agricultural region}))$$

Three biologically plausible two-way interaction terms between herd type, agricultural region, and season of testing were tested with the final model, but none were found to be statistically significant. The adjusted OR for beef cattle was 0.41 as compared to dairy herds. Herds in agricultural regions 1, 2, 3, 5, and 6 had higher odds of PFP as compared to herds in region 9. The pseudo R^2 was 0.009.

Table 6-3. Univariate analysis of factors associated with the proportion of false positive (PFP) on caudal fold tuberculin test (CFT) in Michigan.

Parameter	^c Obs.	^d Cases	^e OR	^f 95% CI	^g LR	^h df	<i>p</i> -value
Herd type							
Beef	1735	3.36	0.61	0.57, 0.66			
Dairy	3254	5.51	1.00		170.28	1	<0.0001
^aAgric. region							
1	438	4.07	1.01	0.86, 1.19			
2	299	8.27	2.06	1.74, 2.43			
3	596	6.83	1.70	1.47, 1.97			
4	292	4.36	1.08	0.91, 1.29			
5	798	6.24	1.55	1.35, 1.79			
6	624	3.65	0.91	0.78, 1.05			
7	549	3.94	0.98	0.84, 1.14			
8	1036	3.98	0.99	0.86, 1.13			
9	357	4.02	1.00		291.77	8	<0.0001
^bQuarter							
1 st	1337	4.97	1.06	0.97, 1.17		3	
2 nd	1097	5.15	1.10	1.00, 1.22		3	
3 rd	975	4.66	1.00	0.90, 1.11		3	
4 th	1580	4.66	1.00		6.4	3	0.0937

^aAgricultural region. ^bCalendar quarters. ^cNumber of observations. ^dNumber of CFT suspects per 100 animals tested. ^eOdds Ratio. ^f95% confidence interval of the OR.

^gLikelihood ratio statistic. ^hDegrees of freedom.

Table 6-4. Multivariable analysis of factors associated with number of CFT suspect results.

Parameter	^b Estimate	^c Cases	^d OR	^e 95% CI	^f LR	^g df	<i>p</i> -value
Intercept	-3.1275						
Herd type							
Beef	-0.6656	2.25	0.41	0.44, 0.51			
Dairy	0	4.38	1.00		49.79	1	<0.0001
^aAgric. region							
1	0.3225	6.05	1.18	1.01, 1.38			
2	0.4419	6.82	2.27	1.92, 2.67			
3	0.3240	6.06	2.01	1.74, 2.33			
4	0.3356	6.13	1.05	0.89, 1.25			
5	0.3922	6.49	1.66	1.45, 1.91			
6	-0.0421	4.20	1.19	1.03, 1.38			
7	0.2566	5.66	1.00	0.87, 1.16			
8	0.1501	5.09	1.02	0.89, 1.17			
9	0	4.38	1.00		20.27	8	0.0094

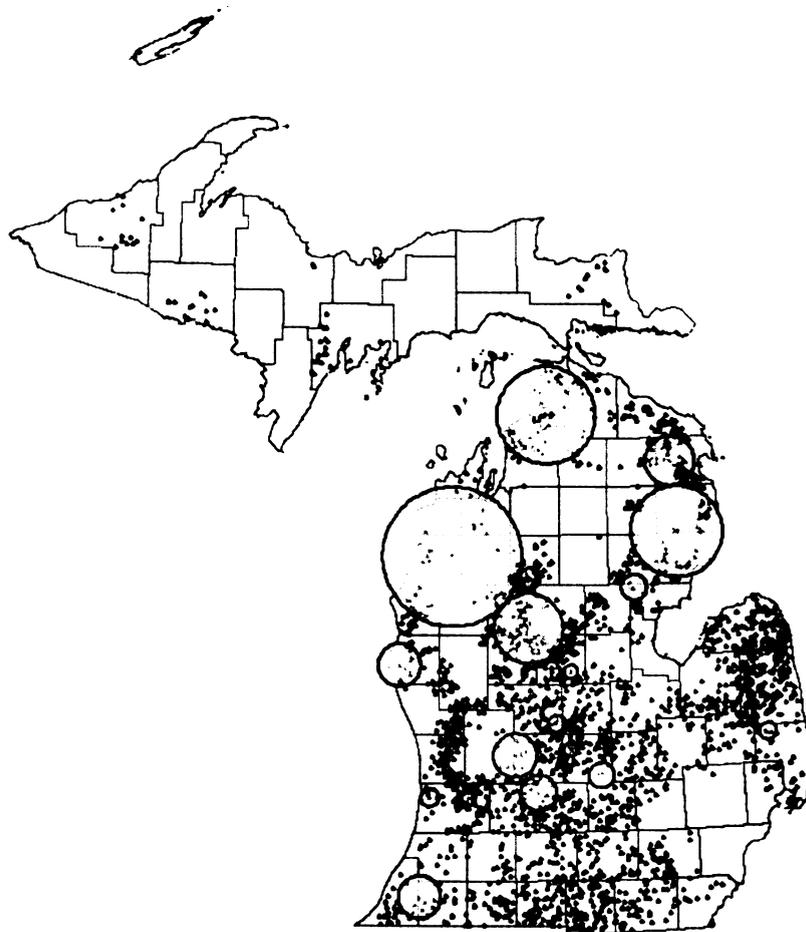
^aAgricultural region. ^bParameter estimate ^cNumber of CFT suspects per 100 animals tested. ^dOdds Ratio. ^e95% confidence interval of the OR. ^fLikelihood ratio statistic. ^gDegrees of freedom.

Spatial analysis

Spatial coordinates (x,y coordinates, Michigan GeoRef) were obtained for 1039 beef and 2,262 dairy herds, respectively. These herds (*n* = 3,301) were used in the spatial analysis. A spatial distribution of these herds in Michigan is shown in figure 6-5.

Significant ($p < 0.05$) clustering of false positive animals on the CFT was detected using the spatial scan statistic. The most likely cluster was found in the center of Alpena county and had a radius of 19.5 km. A total of 759 CFT FP cattle were observed in this area, and the expected number was 331.1. An additional 32 secondary clusters ($p < 0.05$) were located throughout Michigan (figure 6-5).

Figure 6-5. Spatial distribution of herds tested for bovine tuberculosis with the caudal fold tuberculin test in Michigan, including spatial clusters of herds with a high proportion of CFT false positive test results.



- Individual herds.
- Spatial clusters.

Ecological associations

The same herds as used for the spatial analysis was used to investigate associations between PFP and the elevation (in meters) of respective herds, and of herds that were located within 0.25 and 0.50 miles of defined wetlands. Results of the univariate analysis are shown in Table 6-5. Furthermore, the elevation of the farm was significantly positively associated with PFP ($p < 0.0001$). Neither of the variables of distance to wetland or elevation were significantly associated with PFP when the testing veterinarian was added as a random effect together with herd type and agricultural region.

Table 6-5. Univariate analysis of ecological factors associated with the proportion of false positive (PFP) on caudal fold tuberculin test (CFT) in Michigan.

Parameter	^c Obs.	^d Cases	^e OR	^f 95% CI	^g LR	^h df	<i>p</i> -value
^aWetland within 0.25 miles							
1	129	4.98	1.24	1.01, 1.51			
2	3172	4.02	1.00		4.54	1	0.0332
^aWetland within 0.50 miles							
1	317	4.66	1.16	1.02, 1.32			
2	2984	4.02	1.0		5.07	1	0.0243

^aHerds were located within 0.25 miles of designated wetland area. ^bHerds were located within 0.50 miles of designated wetland. ^dNumber of CFT suspects per 100 animals tested. ^eOdds Ratio. ^f95% confidence interval of the OR. ^gLikelihood ratio statistic. ^hDegrees of freedom.

DISCUSSION

In populations with a very low disease prevalence (< 1 %), the specificity of a test maybe be estimated by the proportion of false positive animals. The overall individual animal bTb prevalence in Michigan is well below 1 %. The overall PFP for both dairy and beef herds together was 0.0482, which means that the specificity of the CFT is approximately 0.952. This is slightly higher than the specificity of the CFT (0.939) estimated in cattle located in the bTb-infected area in northern Michigan.(Norby et al., 2003b) Although small, this difference may be due to the specificity being overestimated when evaluated in a population of disease free animals.(Greiner and Gardner, 2000) For example, the quantity of cross-reacting mycobacteria may vary between geographical areas.

In this study, it was assumed that none of the herds we studied had any cattle truly infected with bTb. All CFT suspect cattle had either been found negative on CCT, or were found positive on CCT and then subsequently negative on necropsy, histopathology, culture and PCR.(Norby et al., 2003a) The herd-level predictive value negative (HPVN) is high when applied to a population of cattle consistent with low prevalence of bTb positive herds as seen in Michigan.(Norby et al., 2003c) Nonetheless, it is possible that a few herds that were truly infected with bTb were included in this study. However, as bTb is only present in the northern lower peninsula and the prevalence is very low among herds, the number of bTb positive herds potentially included in this study would not be more than 1 or 2, and thus our results would not be greatly affected.

The study was limited to herds where the testing veterinarian had tested at least 3 herds for bTb, and beef and dairy herds with at least 13 and 23 animals, respectively. This was done to reduce the amount of variation that inexperienced veterinarians might introduce in the PFP. Small herd sizes were eliminated because a single FP animal in a small herd would have an unjustifiably high influence on the mean PFP.

There was a large variation in the PFP among herds and overdispersion was clearly present (Table 6-1), indicating that a negative binomial regression was preferred over a Poisson regression. Studies on incidence density rate of foot rot in sheep showing similar data distributions have shown that such data were best fit by a negative binomial distribution.(Wassink et al., 2003) Because a large number of herds had 0% FP results a zero-inflated negative binomial model(Anonymous, 1997) could be warranted for this analysis. However, such a model did not improve the fit of the model.

Univariate analyses showed that PFP was associated with herd type, agricultural region, and the time of year the herd was tested, however, the season of testing was not significant in the final model. Hence, the final model included herd type, agricultural region, and the testing veterinarian as a random effect. Beef herd had a significantly lower PFP than dairy herds. This was somewhat surprising, because beef herds in general are more likely to be on pasture where they may ingest mycobacteria that cross-react with the CFT. However, because dairy cattle are more used to being handled and hence calmer when tested, it may be easier to read the CFT results, and more subtle inflammatory reaction in the skin were easier to detect. Test results may be correlated between herds when performed by the same veterinarian, and it cannot be assumed that

such measurements are independent. Hence, the veterinarian performing the herd test was included as a random effect in the final generalized estimating equation model.

The PFP was higher in the northern agricultural regions of the lower peninsula in Michigan as compared to the southern areas, and the primary cluster of FP animals were also detected in this area by the spatial analysis (see below). Bovine tuberculosis has only been detected in the northern part of the lower peninsula, and it is probable that veterinarians in this area may be more likely to judge a small reaction in the skin as a CFT suspect, because an animal is more likely to be infected in this area. It is human nature that people tend to see what they expect to see, and may also be more careful or willing to report a “suspect” CFT in an area where bTb is known to occur. It could also be explained by the fact that regulatory veterinarians perform more of the testing in this area, and that they were more rigorous in their interpretation of the CFT.

Our results may be useful for the bTb control effort by adjusting the observed proportion of FP on the CFT by herd type and geographical location, and hence improve the interpretation of a CFT, i.e. determine if the herd needs further testing. Also, our results could help determine if veterinarians with an unusually low (or high) PFP test record need to be contacted for retraining in the use of the CFT. In the current bTb testing scheme in Michigan, all animals with a suspect CFT have to be retested with the CCT within 7 days of reading the CFT (USDA and APHIS, 1999) and animals positive on the CCT will be culled and submitted for mycobacterial culture and/or PCR. The results of this study make it possible to statistically test whether the observed numbers of CFT suspects are significantly different from what would be expected according to herd type and geographical location of the herd. Such an analysis may be performed using standard

statistical tests for comparison of binomial proportion.(Rosner, 2000) However, this approach should be validated on available test data from Michigan. Similarly, veterinarians with unusually low (or high) PFP on the CFT can be identified and called in for retraining.

We used the spatial scan statistic(Kulldorff and Nargawalla, 1995) to investigate if there were spatial clusters of PFP on the CFT, because it permits identification of the location of high (or low) clusters of PFP. An advantage of the scan statistic and SaTScan in particular is that the ID's of herds in each cluster are produced. This makes it easy for regulatory veterinarians to investigate herds with particularly high PFP. The primary cluster was located in the bTb-infected region in Michigan, but secondary clusters were also detected outside this area (figure 6-5). This suggests that there might be other factors than being a herd in the bTb area that affect the PFP. Such factors may include: differences in mycobacteria found in the environment that cause cross-reaction on the CFT or veterinarians that consistently have high PFP on the CFT or the testing veterinarian. A spatial analysis may also be used to detect areas with unusually low PFP. A scanning window of 2% was used in this analysis, which resulted in some of the secondary cluster representing only one herd. We chose a 2% scanning window to be able to detect fairly small clusters (specific areas) with unusually high PFP. However, this procedure may also be used to detect areas with unusually low PFP.

We were not able to investigate the effect of age and other individual animal factors that may affect the PFP of the tested herds because these data were unavailable. Further studies must be conducted for this purpose. However, the associations between PFP and two ecological factors, elevation and the presence of wetland close to the farm

were investigated. On univariate analysis, there were significant positive associations between PFP and the elevation of the farm and whether a farm was located in the proximity of wetland (Table 6-5). In both instances an increase in elevation and the presence of wetland close to the farm, increase the expected PFP for cattle herds. This may be explained by a quantitative and/or qualitative difference in non-typical mycobacterial flora in the environment, because exposure to certain non-typical mycobacteria may cause non-specific reactions on the tuberculin skin tests.(Cooney et al., 1997; Corner and Pearson, 1978a, 1978b) Johnson-Ifearulundu and Kaneene (1999) showed that the prevalence of Johne's disease among and within herds was positively associated with acidic soil, increased soil iron content, and soil type. This suggests that there may be an association between certain environmental factors and the presence of mycobacteria in the environment. In this study it was not possible to investigate the effect of soil type, iron content and pH, because these data were only available for less than ¼ of the counties in Michigan.

In this study we showed that the PFP varies between beef and dairy herds and by geographical region when adjusting for the veterinarian performing the CFT. These results can be used by regulatory veterinarians to adjust the observed PFP by herd type and geographical region and to detect geographical locations in Michigan where particularly high or low PFP are present.

We recommend that geographical locations are obtained for all tested farms to facilitate future spatial analyses.

REFERENCES

- Angus, R., 1978, Production of reference PPD tuberculins for veterinary use in the United States. *J Biol Stand* 6, 221-227.
- Anonymous 1992. Assessment of risk factors for *Mycobacterium bovis* in the United States (Fort Collins, Colorado USA, USDA, APHIS, VS).
- Anonymous, 1997, STATA statistical software release 8 reference manual. STATA Press, College Station, TX.
- Cooney, R., Kazda, J., Quinn, J., Cook, B., Müller, K., Monaghan, M., 1997, Environmental mycobacteria in Ireland as a source of non-specific sensitization to tuberculins. *Irish Vet. J.* 50, 370-373.
- Corner, L.A., Pearson, C.W., 1978a, Pathogenicity in cattle of atypical mycobacteria isolated from feral pigs and cattle and the correlation of lesions with tuberculin sensitivity. *Austr. Vet. J.* 54, 280-286.
- Corner, L.A., Pearson, C.W., 1978b, Response of cattle to inoculation with atypical mycobacteria of bovine origin. *Austr. Vet. J.* 54, 379-382.
- Francis, J., Seiler, R.J., 1978, The sensitivity and specificity of various tuberculin tests using bovine PPD and other tuberculins. *Vet Rec* 103, 420-425.
- Frey, G.H., 1995, Bovine tuberculosis eradication: The program in the United States, In: Thoen, C.O., Steele, J.H. (Eds.) *Mycobacterium bovis* infection in animals and humans. Iowa State University Press, Ames, IA, pp. 119-129.
- Gaboric, C.M., Salman, M.D., Ellis, R.P., Triantis, J., 1996, Evaluation of five-antigen ELISA for diagnosis of tuberculosis in cattle and Cervidae. *JAVMA* 209, 962-966.
- Greiner, M., Gardner, I.A., 2000, Epidemiologic issues in the validation of veterinary diagnostic tests. *Prev. Vet. Med.* 45, 3-22.
- Johnson-Ifearegulu, Y., Kaneene, J.B., 1999, Distribution and environmental risk factors for paratuberculosis in dairy cattle herds in Michigan. *AJVR* 60, 589-596.
- Kaneene, J.B., Bruning-Fann, C.S., Granger, L.M., Miller, R., Porter-Spalding, B.A., 2002, Environmental and farm management factors associated with tuberculosis on cattle farms in northeastern Michigan. *JAVMA* 221, 837-842.
- Kulldorff, M., Nargawalla, N., 1995, Spatial disease clusters: detection and inference. *Stat. Med.* 14, 799-810.

Kulldorff, M., Rand, K., Gherman, G., Williams, G., DeFrancesco, D. 2003. SatScan version 3.1.2. Software for the spatial and space-time scan statistics (National Cancer Institute, Bethesda, MD).

Lepper, A.W.D., Newton-Tabrett, D.A., Carpenter, M.T., Williams, O.J., 1977, The use of bovine PPD tuberculin in the single caudal fold test to detect tuberculosis in beef cattle. *Austr Vet J* 53, 208-213.

Leslie, I.W., Hebert, C.N., Barnett, D.N., 1975a, Comparison of the specificity of human and bovine tuberculin PPD for testing cattle. 2 - Southeastern England. *Vet Rec* 96, 335-338.

Leslie, I.W., Hebert, C.N., Burn, K.J., MacClancy, B.N., Donnelly, W.J.C., 1975b, Comparison of the specificity of human and bovine tuberculin PPD for testing cattle. 1 - Republic of Ireland. *Vet Rec* 96, 332-334.

Liang, K.Y., Zeger, S.L., 1986, Longitudinal data analysis using generalized linear models. *Biometrika*, 73, 13-22.

Martin, S.W., Shoukri, M., Thorburn, M.A., 1992, Evaluating the health status of herds based on tests applied to individuals. *Prev. Vet. Med.* 14, 33-43.

Miller, J., Jenny, A., Rhyan, J., Saari, D., Suarez, D., 1997, Detection of *Mycobacterium bovis* in formalin-fixed, paraffin-embedded tissues of cattle and elk by PCR amplification of an IS6110 sequence specific for *Mycobacterium tuberculosis* complex organisms. *JVDI* 9, 244-249.

Norby, B., Bartlett, P.C., Fitzgerald, S.D., Granger, L., Bruning-Fann, C., Whipple, D.L., Payeur, J.B., 2003a, The sensitivity of gross necropsy, caudal fold and comparative cervical tests for the diagnosis bovine tuberculosis. *JVDI* In Press.

Norby, B., Bartlett, P.C., Grooms, D., Kaneene, J.B., Bruning-Fann, C., 2003c, Herd-level sensitivity, specificity, and predictive values of bovine tuberculosis skin tests in Michigan, Dissertation, Michigan State University, MI

Norby, B., Tempelman, R.J., Hanson, T., Kaneene, J.B., Bartlett, P.C., 2003b, Estimation of sensitivity and specificity of bovine tuberculosis skin tests in Michigan when a perfect reference is not available, Dissertation, Michigan State University, MI.

Pollock, J.M., Girvin, R.M., Lightbody, K.A., Clements, R.A., Neill, S.D., Buddle, B.M., Andersen, P., 2000, Assessment of defined antigens for the diagnosis of bovine tuberculosis in skin test-reactor cattle. *Vet Rec* 146, 659-665.

Rice, J.A., 1995, *Mathematical statistics and data analysis*, Second Edition. Duxbury Press, Belmont, CA.

Rosner, B., 2000, *Fundamentals of biostatistics.*, Fifth Edition. Duxbury, Pacific Grove, CA, 792 p.

Schmitt, S., Fitzgerald, S., Cooley, T., Bruning-Fann, C., Berry, D., Carlson, T., Minnis, R., Payer, J., Sikarskie, J, 1997, Bovine tuberculosis in free-ranging white-tail deer in Michigan. *J Wildlife Dis* 33, 749-758.

Stokes, M.E., Davis, C.S., Koch, G.G., 2000, *Categorical data analysis using the SAS system*, Second Edition. SAS Institute Inc., Cary, NC.

Tempelman, R.J., Gianola, D., 1996, A mixed effect model for overdispersed count data in animal breeding. *Biometrics* 52, 265-279.

USDA, APHIS 1999. Bovine tuberculosis eradication. Uniform methods and rules, effective January 22, 1999, (United States Department of Agriculture.), pp. 1-34.

Ward, M.P., Carpenter, T.E., 2000, Techniques for analysis of disease clustering in space and in time in veterinary epidemiology. *Prev. Vet. Med.* 45, 257-284.

Wassink, G.J., Grogono-Thomas, R., Moore, L.J., Green, L.E., 2003, Risk factors associated with the prevalence of foot rot in sheep from 1999 to 2000. *Vet. Rec.* 152, 351-358.

Whipple, D.L., Bolin, C.A., Davis, A.J., Jarnagin, J.L., Johnson, D.C., Payer, J.B., Saari, D.A., Wilson, A.J., Wolf, M.M., 1995, Comparison of the sensitivity of the caudal fold skin test and a commercial gamma-interferon assay for diagnosis of bovine tuberculosis. *Am J Vet Res* 56, 415-419.

Wood, P.R., Corner, L.A., Rothel, J.S., Baldock, C., Jones, S.L., Cousins, D.B., McCormick, B.S., Francis, B.R., Creeper, J., Tweddle, N.E., 1991, Field comparison of the interferon-gamma assay and the intradermal tuberculin test for the diagnosis of bovine tuberculosis. *Austr Vet J* 68, 286-290.

Wood, P.R., Corner, L.A., Rothel, J.S., Ripper, J.L., Fifis, T., McCormick, B.S., Francis, B., Melville, L., Small, K., De Witte, K., Tolson, J., Ryan, T.J., de Lisle, G.W., Cox, J.C., Jones, S.L., 1992, A field evaluation of serological and cellular diagnostic tests for bovine tuberculosis. *Vet Micro* 31, 71-79.

Zeger, S. L., Liang, K. Y., 1986, Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*, 42, 121-130.

SUMMARY

While likely not the last word on bovine tuberculosis testing in Michigan, this study offer some insight to the performance of the tuberculin skin tests in individual animals and the herd-level performance of these tests, as well as factors that may affect these tests. Estimates of sensitivities and specificities of the caudal fold test were in general close to estimates reported in other studies in the United States. Only very little information is available on the performance of the caudal fold and comparative cervical tests when used in serial interpretation. However, our results seem to agree with limited results from the United States.

Besides estimating the sensitivity of the two tuberculin skin tests, we were also able to estimate the sensitivity of gross necropsy. Gross necropsy is often used alone or in combination with other tests as the ‘gold standard’ for estimation of sensitivity (and specificity) of the skin tests, hence the sensitivity of gross necropsy cannot be estimated. However, in this study, we were able to estimate the sensitivity of the gross post-mortem necropsy exam, which perhaps can be viewed as the “upper detection limit” of any practical slaughter-based system for bTb.

Four different ‘two-population-two-tests’ latent-class models were used to estimate the sensitivity and specificity of the tuberculin skin tests. The models allowed us to account for possible dependence between the tests conditional on true disease status. All the models performed well, especially taking into account the relatively small sample of animals and low disease prevalence. Both the sensitivities and specificities in

particular are high. So the potential for these tests to perform well in an eradication program is high.

The next logical step in evaluation of the tuberculin skin tests, was to investigate how well they work in a population of cattle herds infected with bovine tuberculosis. Historically, eradication of bovine tuberculosis from many countries around the world demonstrates that the skin tests perform very well on the herd-level. Because many factors affect the herd-level sensitivity, specificity, and predictive values, it is not straightforward to predict how individual animal sensitivity, specificity and prevalence in particular, affect the herd-level parameters. Assuming that only one tuberculosis positive animal is enough to declare the herd tuberculosis positive and the array of tests used to classify a single animal as positive for bovine tuberculosis has a specificity of 1.0, then the herd-level performance of the tuberculin skin tests in Michigan is very good. When the prevalence is low, individual animal sensitivity has little effect on herd-level sensitivity, which seems to be more correlated with the disease prevalence and the size of the tested herd. More importantly, the predictive value negative of a tested herd is very high, only decreasing slightly when the herd size decreases. These results show that to fulfill the goal of eradication of bovine tuberculosis particular attention should be paid to smaller herds.

It has long been known that some cattle that are not infected with *Mycobacterium bovis* still react to the tuberculin injected intradermally when tested, i.e. they have false positive test results. This adds to the cost of the eradication program because tuberculosis negative animals that test positive on the caudal fold test must be retested with the comparative cervical test. Furthermore, animals with false positive results on

the comparative cervical test are euthanized, necropsied, and selected samples are submitted for tissue culture, histologic exam and possibly PCR. If factors that predispose animals to a false negative result could be identified the cost of the eradication program could be lowered. Furthermore, they could be used to adjust the proportion of false positive animals, and a decision made whether to retest the herd or not. Furthermore, veterinarians that consistently have high (or low) proportion of caudal fold positive animals could be called in for retraining.

CONCLUSIONS

In this study I sought to answer five research objectives regarding the performance of tests used to diagnose bovine tuberculosis (bTb). The first objective of this study was to determine the sensitivity of the caudal fold tuberculin test (CFT) and the CFT and the comparative cervical tuberculin tests used in serial interpretation as well as gross necropsy using a gold standard approach. The sensitivity (Se) of the caudal fold tuberculin tests (CFT) was as expected to be higher than when the CFT and the comparative cervical test (CCT) were used in serial interpretation (CFTCCT_{SER}). Our Se estimate of the CFT is slightly higher than that reported by other studies in the United States.(Gaboric et al., 1996; Whipple et al., 1995) This might be explained by differences in disease prevalence or status in our and their cattle populations. Our estimates of Se for the CFTCCT_{SER} are very close to similar estimates reported by Roswurm and Konyha. (1973). It is well accepted that the CFT and CFTCCT_{SER} are more suited to the detection of infected herds rather than the detection of infected cattle, i.e. some bTb-infected animals will escape detection when the skin tests are used to assess the infection status of individual animals. This finding supports a disease control policy of total herd depopulation over a test and slaughter approach, because findings in this study suggest that the test and slaughter approach would fail to detect all infected animals in approximately 29% of the herds. The sensitivity of necropsy was almost as high as CFTCCT_{SER}. Furthermore, the sensitivity of gross necropsy increased to 100% when 2 or more lesions were present. Such animals with multiple lesion sites would expectedly shed greater numbers of infective *Mycobacterium bovis* and thus would be a

more important source of disease transmission. The gross post-mortem necropsy exam can perhaps be viewed as the “upper detection limit” of any practical slaughter-based system for bTb.

The second objective was to determine the sensitivity and specificity of the caudal fold tuberculin test (CFT) and the CFT and comparative cervical tuberculin tests used in serial interpretation using a method that does not require the assumption of a gold standard and which can account for correlation between tests. Sensitivity and Sp of the two tuberculin skin tests were estimated using a two-test-two-population approach using four different simulation-based Bayesian models allowing for correction for possible correlation between the two tests. There was a small significant negative correlation between the CFTCCT_{SER} and bacteriologic culture. This means that animals that are negative on one tests are less likely to be negative on the other test than would be expected if the tests were independent conditional on disease status. A negative correlation is difficult to explain, but the lack of some biologic phenomenon may affect both tests. Our estimates of the Se of the CFT were very similar to other U.S. estimates. (range:0.804-0.849).(Anonymous, 1992; Gaboric et al., 1996; Whipple et al., 1995) The estimates of Sp of the CFT are similar to estimates reported by the USDA(Anonymous, 1992), but slightly higher than estimated by Gaboric *et al.* (1996). Our results for the CFTCCT_{SER} were close to other reported estimates.(Anonymous, 1992; Roswurm and Konyha, 1973) Results obtained completing this objective will be a good estimate of how the tuberculin skin tests perform in Michigan. However, it is important to keep in mind that as the eradication program progresses, the bTb prevalence within herds is likely to change and hence sensitivity and specificity estimates may do the same.

Completing the third objective presents herd-level sensitivity (HSe) and specificity (HSp) estimates of the of the caudal fold tuberculin test (CFT) and the CFT and comparative cervical tuberculin test when used in the bovine tuberculosis infected area in Michigan. From a review of the literature, this seems to be the first attempt to address herd-level performance of the tuberculin skin tests, so it is not possible to compare our results to other work. However, our results reflect the behavior of changes in Se and Sp and herd size as has been reviewed by Christensen and Gardner (2000). The Sp is most likely equal to 100%, because all the tests are used in series in the eradication program as reflected in scenarios 3 and 4. This will automatically result in a HSp of 100%. Equally the herd-level predictive value positive (HPVP) will also be 100%. However, from a regulatory/eradication standpoint, the most interesting parameter is probably herd-level predictive value negative (HPVN), because some herds may have a negative herd-level test and still be infected with bTb. The HSe is very sensitive to the herd size (sample size), and the smaller a herd is, the lower the HSe is. Decreasing herd sizes also means a decrease in HPVN. Hence it may be beneficial to the eradication effort to retest smaller herds more frequently than larger herds. It is worth to note, that if Se and Sp changes over the course of the eradication program, then HSe, HSp, HPVP, and HPVN may change also.

The fourth objective was to determine herd, geographic, and ecologic factors that may affect the outcome of the CFT. Not unexpectedly, there was a large variation in the proportion of false positive results (FPF) on the CFT among herds and overdispersion was clearly present, indicating that a negative binomial regression was preferred over a

Poisson regression. Studies on incidence density rate of foot rot in sheep showing similar data distributions have shown that such data were best fit by a negative binomial distribution.(Wassink et al., 2003) The study showed that there was a much larger variation in the predicted proportion of false positive results on the CFT (PFP) among smaller herds than among large herds. This is to be expected because the result of just one or a few false positive results will give a proportionately higher PFP than in a large herd. The multiple negative binomial regression showed that the predicted proportion of false positive results on the CFT is approximately half as large in beef cattle as compared to dairy herds. This was somewhat surprising, because beef herds in general are more likely to be on pasture where they may ingest mycobacteria that cross react with the CFT. However, because dairy cattle are more used to be handled and hence calmer when tested, it may be easier to read the CFT results, and more subtle inflammatory reaction in the skin were easier to detect. The PFP was higher in the northern agricultural regions of the lower peninsula in Michigan as compared to the southern areas, and the primary cluster of FP animals were also detected in this area by the spatial analysis. Bovine tuberculosis has only been detected in the northern part of the lower peninsula, and it is probable that veterinarians in this area may be more likely to judge a small reaction in the skin as a CFT suspect, because an animal is more likely to be infected in this area. It is human nature that people tend to see what they expect to see, and may also be more careful or willing to report a “suspect” CFT in an area where bTb is known to occur. It could also be explained by the fact that regulatory veterinarians perform more of the testing in this area, and that they were more rigorous in their interpretation of the CFT.

Finally, completing the fifth objective stressed that there maybe 'unidentified' factors that cause PFP to cluster. This may be explained by a quantitative and/or qualitative difference in non-typical mycobacterial flora in the environment, because exposure to certain non-typical mycobacteria may cause non-specific reaction on the tuberculin skin tests.(Cooney et al., 1997; Corner and Pearson, 1978a, 1978b; Wassink et al., 2003) Johnson-Ifearulundu and Kaneene (1999) showed that the prevalence of Johne's disease among and within herds was positively associated with acidic soil, increased soil iron content, and soil type. This suggests that there may be an association between certain environmental factors and the presence of mycobacteria in the environment. In this study it was not possible to investigate the effect of soil type, iron content and pH, because these data were only available for less than ¼ of the counties in Michigan. Further studies will have to shed light on this subject.

Overall, this study presents regulatory veterinarians with some tools that may be of help in evaluating the current bTb eradication program in Michigan or if adjustments need to be made to the current program.

REFERENCES

Anonymous 1992. Assessment of risk factors for *Mycobacterium bovis* in the United States (Fort Collins, Colorado USA, USDA, APHIS, VS).

Christensen, J., Gardner, I.A., 2000, Herd-level interpretation of test results for epidemiologic studies of animal diseases. *Prev. Vet. Med.* 45, 83-106.

Cooney, R., Kazda, J., Quinn, J., Cook, B., Müller, K., Monaghan, M., 1997, Environmental mycobacteria in Ireland as a source of non-specific sensitization to tuberculins. *Irish Vet. J.* 50, 370-373.

Corner, L.A., Pearson, C.W., 1978a, Pathogenicity in cattle of atypical mycobacteria isolated from feral pigs and cattle and the correlation of lesions with tuberculin sensitivity. *Austr. Vet. J.* 54, 280-286.

Corner, L.A., Pearson, C.W., 1978b, Response of cattle to inoculation with atypical mycobacteria of bovine origin. *Austr. Vet. J.* 54, 379-382.

Gaboric, C.M., Salman, M.D., Ellis, R.P., Triantis, J., 1996, Evaluation of five-antigen ELISA for diagnosis of tuberculosis in cattle and Cervidae. *JAVMA* 209, 962-966.

Johnson-Ifearegulu, Y., Kaneene, J.B., 1999, Distribution and environmental risk factors for paratuberculosis in dairy cattle herds in Michigan. *AJVR* 60, 589-596.

Roswurm, J.D., Konyha, 1973, The comparative cervical tuberculin test as an aid to diagnosing bovine tuberculosis. *Proc. Annu. Meet. U.S. Anim. Health Assoc.* 77, 368-389.

Wassink, G.J., Grogono-Thomas, R., Moore, L.J., Green, L.E., 2003, Risk factors associated with the prevalence of foot rot in sheep from 1999 to 2000. *Vet. Rec.* 152, 351-358.

Whipple, D.L., Bolin, C.A., Davis, A.J., Jarnagin, J.L., Johnson, D.C., Payer, J.B., Saari, D.A., Wilson, A.J., Wolf, M.M., 1995, Comparison of the sensitivity of the caudal fold skin test and a commercial gamma-interferon assay for diagnosis of bovine tuberculosis. *Am J Vet Res* 56, 415-419.

APPENDIX B.

The author takes no responsibility for errors and incorrect results obtained with this code.

Conditional independence model:

```
*****
* This code is based on a paper by WO Johnson.          *
* Wesley O. Johnson et al., Am J of Epidemiology      *
* 2001, 153: 921-924, "Screening without a gold      *
* standard: The Hui-Walter paradigm revisited.      *
*                                                    *
*                                                    *
* x-ijk refers to the observed number of in the cell *
* of the i-th and j-th tests (i=j=1,2) and the k-th *
* population (here k=1,2) (2 populations)          *
* *****;
proc iml;
/* Seed values for the binomial distributions */
seed1=1;

/* Number of burn-in samples */
burnin = 2000;

/* Total number of samples (cycles) */
ncycle = 12000;

/* Population 1 */
x111=23; x121=7; x211=4; x221=162;
/* Population 2*/
x112=10; x122=5; x212=1; x222=282;

/* Starting values for prevalence, sensitivity, and specificity */
P1=0.1; P2=0.5; Se1=0.9; Se2=0.5; Sp1=0.9; Sp2=0.5;

/* Prior beta-distributions {beta(a,b)} for prevalence, sensitivity,
and specificity */
/* Population 1 and 2 */
/* P1 ~ beta(1.27,9.65) and P2 ~ beta(1.73,2.71) */
/* Sensitivity for tests 1 and 2 */
/* Se1 ~ beta(2.82,2.49) and Se2 ~ beta(8.29,1.81) */
/* Specificity for tests 1 and 2*/
/* Sp1 ~ beta(15.7,2.49) and Sp2 ~ beta(10.69,2.71) */

/* Coding of the alpha and beta parameters for the prior distributions
*/
/* a=alpha and b=beta parameters for the beta distributions */
/* Alpha and beta parameters for prior distributions of prevalence */
```

```

aP1=13.63;
bP1=57.94;
aP2=1;
bP2=1;
/* Alpha and beta parameters for prior distributions of sensitivity */
aSe1=1;
bSe1=1;
aSe2=1;
bSe2=1;
/* Alpha and beta parameters for prior distributions of Specificity */
aSp1=1;
bSp1=1;
aSp2=1;
bSp2=1;

/* Creating subroutine named gibbs*/
Start gibbs;

/* Marginal frequencies for observed counts */
/* Sample sizes for the two populations */
n1 = (x111 + x121 + x211 + x221);
n2 = (x112 + x122 + x212 + x222);

/* Observed marginal counts, d indicates summation over a row or
column */
x1dd = (x111 + x121 + x112 + x122);
x2dd = (x211 + x221 + x212 + x222);
xd1d = (x111 + x112 + x211 + x212);
xd2d = (x121 + x122 + x221 + x222);

*Start Gibbs;
/* Imputation step */
/* Sampling of z-ijk's., which are bin(x-ijk, p-ijk) */
if x111^=0 then
do;
p111 = (p1*Se1*Se2/(p1*Se1*Se2+(1-p1)*(1-Sp1)*(1-Sp2)));
z111 = ranbin(seed1, x111, p111);
end;
else z111 = 0;

if x121^=0 then
do;
p121 = (p1*se1*(1-Se2)/(p1*Se1*(1-Se2)+(1-p1)*(1-Sp1)*Sp2));
z121 = ranbin(seed1, x121, p121);
end;
else z121 = 0;

if x211^=0 then
do;
p211 = (p1*(1-Se1)*Se2/(p1*(1-Se1)*Se2+(1-p1)*Sp1*(1-Sp2)));
z211 = ranbin(seed1, x211, p211);
end;
else z211 = 0;

if x221^=0 then

```

```

        do;
        p221 = (p1*(1-Se1)*(1-Se2)/(p1*(1-Se1)*(1-Se2)+(1-
p1)*Sp1*Sp2));
        z221 = ranbin(seed1, x221, p221);
        end;
        else z221 = 0;

        if x112 then
        do;
        p112 = (p2*Se1*Se2/(p2*Se1*Se2+(1-p2)*(1-Sp1)*(1-Sp2)));
        z112 = ranbin(seed1, x112, p112);
        end;
        else z112 = 0;

        if x122^=0 then
        do;
        p122 = (p2*se1*(1-Se2)/(p2*Se1*(1-Se2)+(1-p2)*(1-Sp1)*Sp2));
        z122 = ranbin(seed1, x122, p122);
        end;
        else z122 = 0;

        if x212^=0 then
        do;
        p212 = (p2*(1-Se1)*Se2/(p2*(1-Se1)*Se2+(1-p2)*Sp1*(1-Sp2)));
        z212 = ranbin(seed1, x212, p212);
        end;
        else z212 = 0;

        if x222^=0 then
        do;
        p222 = (p2*(1-Se1)*(1-Se2)/(p2*(1-Se1)*(1-Se2)+(1-
p2)*Sp1*Sp2));
        z222 = ranbin(seed1, x222, p222);
        end;
        else z222 = 0;

/* Latent marginal counts */
        zdd1 = (z111 + z121 + z211 + z221);
        zdd2 = (z112 + z122 + z212 + z222);

        z1dd = (z111 + z121 + z112 + z122);
        z2dd = (z211 + z221 + z212 + z222);

        zd1d = (z111 + z211 + z112 + z212);
        zd2d = (z121 + z221 + z122 + z222);

/* Posterior step */
/* Sampling marginal probability estimates */
/* Posterior beta for prevalence 1 (P1) */
        P1a = rangam(seed1, (aP1+zdd1));
        P1b = rangam(seed1, (bP1+n1-zdd1));
        P1 = P1a/(P1a+P1b);

/* Posterior beta for prevalence 2 (P2) */
        P2a = rangam(seed1, (aP2+zdd2));
        P2b = rangam(seed1, (bP2+n2-zdd2));
        P2 = P2a/(P2a+P2b);

```

```

/* Posterior beta for Sensitivity 1 (Se1) */
  Sela = rangam(seed1, (aSe1+z1dd));
  Selb = rangam(seed1, (bSe1+z2dd));
  Se1 = Sela/(Sela+Selb);

/* Posterior beta for Sensitivity 2 (Se2) */
  Se2a = rangam(seed1, (aSe2+zd1d));
  Se2b = rangam(seed1, (bSe2+zd2d));
  Se2 = Se2a/(Se2a+Se2b);

/* Posterior beta for Specificity 1 (Sp1) */
  Spla = rangam(seed1, (aSp1+x2dd-z2dd));
  Sp1b = rangam(seed1, (bSp1+x1dd-z1dd));
  Sp1 = Spla/(Spla+Sp1b);

/* Posterior beta for Specificity 1 (Sp2) */
  Sp2a = rangam(seed1, (aSp2+xd2d-zd2d));
  Sp2b = rangam(seed1, (bSp2+xd1d-zd1d));
  Sp2 = Sp2a/(Sp2a+Sp2b); *print Sp2beta;
finish gibbs;

/* Burn in*/

/* Running gibbs sampler within burn in*/
P1mpost = j(ncycle,1,0); /* Vector for marginal posterior for P1 */
P2mpost = j(ncycle,1,0);
Selmpost = j(ncycle,1,0);
Se2mpost = j(ncycle,1,0);
Sp1mpost = j(ncycle,1,0);
Sp2mpost = j(ncycle,1,0);
cycle = j(ncycle,1,0);
num=0; /* 'Counter' for number of saved samples */

do iter=1 to burnin;
  run gibbs;
  do;
    num=num+1;
    P1mpost[num]=P1;
    P2mpost[num]=P2;
    Selmpost[num]=Se1;
    Se2mpost[num]=Se2;
    Sp1mpost[num]=Sp1;
    Sp2mpost[num]=Sp2;
    cycle[num]=num;
  end;
end;

  create burninsmpl var {cycle P1mpost P2mpost Selmpost
                        Se2mpost Sp1mpost Sp2mpost};
append;

/* Running gibbs sampler after burn in*/
P1mpost = j(ncycle,1,0); /* Vector for marginal posterior for P1 */
P2mpost = j(ncycle,1,0);

```

```

Se1mpost = j(ncycle,1,0);
Se2mpost = j(ncycle,1,0);
Sp1mpost = j(ncycle,1,0);
Sp2mpost = j(ncycle,1,0);
cycle = j(ncycle,1,0);
num=0; /* 'Counter' for number of saved samples */

do iter = 1 to ncycle;
  run gibbs;
  do;
    num = num+1;
    P1mpost[num]=P1;
    P2mpost[num]=P2;
    Se1mpost[num]=Se1;
    Se2mpost[num]=Se2;
    Sp1mpost[num]=Sp1;
    Sp2mpost[num]=Sp2;
    cycle[num]=num; /* collects the number of cycles (=ncycle)
*/
  end;
end;

/* Creating prior distributions */
start prior;
/* Prior beta for prevalence 1 (P1) */
Plapr = rangam(seed1, aP1); /*Creating alpha parameter for beta
dist*/
P1bpr = rangam(seed1, bP1); *print P1b;
P1betapr = Plapr/(Plapr+P1bpr); *print P1beta;

/* Prior beta for prevalence 2 (P2) */
P2apr = rangam(seed1, aP2);
P2bpr = rangam(seed1, bP2);
P2betapr = P2apr/(P2apr+P2bpr);

/* Prior beta for Sensitivity 1 (Se1) */
Selapr = rangam(seed1, aSe1);
Selbpr = rangam(seed1, bSe1);
Selbetapr = Selapr/(Selapr+Selbpr);

/* Prior beta for Sensitivity 2 (Se2) */
Se2apr = rangam(seed1, aSe2);
Se2bpr = rangam(seed1, bSe2);
Se2betapr = Se2apr/(Se2apr+Se2bpr);

/* Prior beta for Specificity 1 (Sp1) */
Splapr = rangam(seed1, aSp1);
Sp1bpr = rangam(seed1, bSp1);
Sp1betapr = Splapr/(Splapr+Sp1bpr);

/* Prior beta for Specificity 2 (Sp2) */
Sp2apr = rangam(seed1, aSp2);
Sp2bpr = rangam(seed1, bSp2);
Sp2betapr = Sp2apr/(Sp2apr+Sp2bpr); *print Sp2beta;
finish prior;

```

```

P1pr = j(ncycle,1,0); /* Vector for marginal posterior for P1 */
P2pr = j(ncycle,1,0);
Selpr = j(ncycle,1,0);
Se2pr = j(ncycle,1,0);
Sp1pr = j(ncycle,1,0);
Sp2pr = j(ncycle,1,0);
cyclepr = j(ncycle,1,0);

numpr=0;

do iter = 1 to ncycle;
  run prior;
  do;
    numpr = numpr+1;
    P1pr[numpr]=P1betapr;
    P2pr[numpr]=P2betapr;
    Selpr[numpr]=Selbetapr;
    Se2pr[numpr]=Se2betapr;
    Sp1pr[numpr]=Sp1betapr;
    Sp2pr[numpr]=Sp2betapr;
    cyclepr[numpr]=numpr; /* collects the number of cycles
(=ncycle) */
  end;
end;
  create margdist var {cycle P1pr P2pr Selpr Se2pr Sp1pr Sp2pr
P1mpost P2mpost Selmpost Se2mpost Sp1mpost
Sp2mpost};
  append;
quit;
run;

/* Macro for graphing the prior distributions and */
/* printing summary statistics */
title1 'Prior distribution for ...';
title2 'Cycles: Burn in = 200 and total = 1200';
%macro table(dsp);
  proc univariate noprint data=margdist; var &dsp;
    histogram / vscale=count beta (noprint theta=0 sigma=1 color=red
    percents=1 2.5 5 10 50 90 95 97.5 99);
    output out=&dsp mean=mean std=std pctlpre=P_
    pctlpts=0, 1, 2.5, 5, 10, 50, 90, 95, 97.5, 99, 100;
    data &dsp; set &dsp; var="&dsp";
  %mend;
%table(P1pr) %table(P2pr); %table(Selpr);
%table(Se2pr); %table(Sp1pr); %table(Sp2pr);
data margdistmacrohp; set P1pr P2pr Selpr Sp1pr Se2pr Sp2pr;
title 'Summary statistics for prior distributions for P1, P2, Sel, Se2,
Sp1, Sp2';
proc print; id var; var P_0 P_1 P_2_5 P_5 P_50 P_95 P_97_5 P_99 P_100;
format P_1 P_2_5 P_5 P_50 P_95 P_97_5 P_99 P_100;
run;

title1 'Trace Plot burnin';
proc gplot data=burninsmpl;

```

```

plot P1mpost*cycle / overlay;
plot P2mpost*cycle / overlay;
plot Selmpost*cycle / overlay;
plot Se2mpost*cycle / overlay;
plot Sp1mpost*cycle / overlay;
plot Sp2mpost*cycle / overlay;
symbol1
  value=circle
  cv=black;
run; quit;

title1 'Trace Plot';
proc gplot data=margdist;
  plot P1mpost*cycle / overlay;
  plot P2mpost*cycle / overlay;
  plot Selmpost*cycle / overlay;
  plot Se2mpost*cycle / overlay;
  plot Sp1mpost*cycle / overlay;
  plot Sp2mpost*cycle / overlay;
  symbol1
    value=circle
    cv=black;
run; quit;

data trace;
  join burninsmpl margdist;
run;

/* Macro for graphing the posterior distributions and */
/* printing summary statistics */
title1 '';
title2 '';
%macro table(ds);
  proc univariate noprint data=margdist; var &ds;
    histogram / vscale=count beta (noprint theta=0 sigma=1 color=red
      percents=1 2.5 5 10 50 90 95 97.5 99);
    output out=&ds mean=mean std=std mode=mode pctlpre=P_
      pctlpts=0, 1, 2.5, 5, 10, 50, 90, 95, 97.5, 99, 100;
    data &ds; set &ds; var="&ds";
  %mend;
%table(P1mpost) %table(P2mpost); %table(Selmpost);
%table(Se2mpost); %table(Sp1mpost); %table(Sp2mpost);
data margdistmacroh; set P1mpost P2mpost Selmpost Sp1mpost Se2mpost
Sp2mpost;
title 'Summary statistics for posterior distributions for P1, P2, Sel,
Se2, Sp1, Sp2';
proc print; id var; var P_0 P_1 P_2_5 P_5 P_50 mode P_95 P_97_5 P_99
P_100;
format P_0 P_1 P_2_5 P_5 P_50 mode P_95 P_97_5 P_99 P_100;
run;

```

Conditional dependence model (M2).

```
proc iml;
  /* Seed values for the binomial distributions */
  seed=1;

  /* Number of burn-in samples */

  /* Total number of samples (cycles) */
  total = 12000;
  /* Number of burn-in samples */
  burnin = 2000; *burnin sample;
  skip = 1; *sample every 100 cycles;

  /* Starting values for prevalence, sensitivity, and specificity etc*/
  Selpre=.3; Se2pre=.4; Sp1pre=0.5; Sp2pre=0.4;
  p111pre=.4; p101pre=.4; p011pre=.4; p001pre=.4;
  p110pre=.4; p100pre=.4; p010pre=.4; p000pre=.4;
  P1=.2; P2=.2; Se1=.6; Se2=.5; Sp1=.6; Sp2=.5; p111=.1; p000=.1;

  /* files to save different parameters */
  alphaSsav=.;
  alphaCsav=.;
  P1par=P1;
  P2par=P2;
  Selpar=Sel;
  Se2par=Se2;
  Sp1par=Sp1;
  Sp2par=Sp2;
  p111par=p111;
  p000par=p000;
  uSsav=0;
  uCsav=0;
  dcp=dc;
  dspar=ds;

  /* Observed cell values, x */
  /* Population 1 */
  x111=23; x101=7; x011=4; x001=162;
  /* Population 2*/
  x112=10; x102=5; x012=1; x002=282;

  /* Coding of the alpha and beta parameters for the prior distributions
  */
  /* a=alpha and b=beta parameters for the beta distributions */
  /* Alpha and beta parameters for prior distributions of prevalence */
  aP1=1;
  bP1=1;
  aP2=1; *5.98;
  bP2=1;

  /* Alpha and beta parameters for prior distributions of sensitivity */
  aS1=1;
  bS1=1;
  aS2=1;
  bS2=1;
```

```

/* Alpha and beta parameters for prior distributions of Specificity */
aC1=1;
bC1=1;
aC2=360.4; *1; *21.11; *20.6; *360.4;
bC2=1.36; *1; *2.06; *1.2; *1.36;
/* prior parameters for Dirichlet distribution */
aDRL=1;
bDRL=1;
cDRL=1;
dDRL=1;

/* Gibbs and MH algorithm */
do ncycle = 1 to total; /*Start MH and Gibbs mixing */

/* Sampling full conditional distributions */
  if x111^=0 then
    do;
      pa111 = (p1*p111/(p1*p111+(1-p1)*(1-Sp1-Sp2+p000)));
      a111 = ranbin(seed, x111, pa111);
    end;
  else a111 = 0;

  if x101^=0 then
    do;
      pa101 = (p1*(Se1-p111)/(p1*(Se1-p111)+(1-p1)*(Sp2-p000)));
      a101 = ranbin(seed, x101, pa101);
    end;
  else a101 = 0;

  if x011^=0 then
    do;
      pa011 = (p1*(Se2-p111)/(p1*(Se2-p111)+(1-p1)*(Sp1-p000)));
      a011 = ranbin(seed, x011, pa011);
    end;
  else a011 = 0;

  if x001^=0 then
    do;
      pa001 = (p1*(1-Se1-Se2+p111)/(p1*(1-Se1-Se2+p111)+(1-
p1)*p000));
      a001 = ranbin(seed, x001, pa001);
    end;
  else a001 = 0;

  if x112^=0 then
    do;
      pa112 = (p2*p111/(p2*p111+(1-p2)*(1-Sp1-Sp2+p000)));
      a112 = ranbin(seed, x112, pa112);
    end;
  else a112 = 0;

  if x102^=0 then
    do;
      pa102 = (p2*(Se1-p111)/(p2*(Se1-p111)+(1-p2)*(Sp2-p000)));
      a102 = ranbin(seed, x102, pa102);
    end;
  else a102 = 0;

```

```

    if x012^=0 then
      do;
        pa012 = (p2*(Se2-p111)/(p2*(Se2-p111)+(1-p2)*(Sp1-p000)));
        a012 = ranbin(seed, x012, pa012);
      end;
    else a012 = 0;

    if x002^=0 then
      do;
        pa002 = (p2*(1-Se1-Se2+p111)/(p2*(1-Se1-Se2+p111)+(1-
p2)*p000));
        a002 = ranbin(seed, x002, pa002);
      end;
    else a002 = 0;

/* Calculating latent cell counts */
A1=a111+a101+a011+a001;
b111=x111-a111;
b101=x101-a101;
b011=x011-a011;
b001=x001-a001;
B1=b001+b011+b101+b111;
N1=A1+B1;

A2=a112+a102+a012+a002;
b112=x112-a112;
b102=x102-a102;
b012=x012-a012;
b002=x002-a002;
B2=b002+b012+b102+b112;
N2=A2+B2;

a11d=a111+a112;
a10d=a101+a102;
a01d=a011+a012;
a00d=a001+a002;

b11d=b111+b112;
b10d=b101+b102;
b01d=b011+b012;
b00d=b001+b002;

A=A1+A2;
B=B1+B2;
N=A+B;

/* Sampling full conditional for prevalence */
P1a = rangam(seed, (aP1+A1));
P1b = rangam(seed, (bP1+B1));
P1 = P1a/(P1a+P1b);

P2a = rangam(seed, (aP2+A2));
P2b = rangam(seed, (bP2+B2));
P2 = P2a/(P2a+P2b);

```

```

do iter=1 to 1; /*Start MH*/
  aGAM = rangam(seed, (a11d+aDRL));
  bGAM = rangam(seed, (a10d+bDRL));
  cGAM = rangam(seed, (a01d+cDRL));
  dGAM = rangam(seed, (a00d+dDRL));

/* Proposed probabilities for P11, P10, P01, P00 */
  p111 = aGAM/(aGAM+bGAM+cGAM+dGAM);
  p101 = bGAM/(aGAM+bGAM+cGAM+dGAM);
  p011 = cGAM/(aGAM+bGAM+cGAM+dGAM);
  p001 = dGAM/(aGAM+bGAM+cGAM+dGAM);
  Se1 = p111+p101;
  se2 = p111+p011;
/* end;*/

/* Acceptance rate (P-move) for Sensitivity */
  logAlphaS = log(min(Se1pre, Se2pre)-Se1pre*se2pre)-
log(min(Se1, Se2)-Se1*Se2)+
      (aS1-1)*(log(Se1)-log(Se1pre))+(bS1-1)*(log(1-Se1)-
log(1-Se1pre))+
      (aS2-1)*(log(Se2)-log(Se2pre))+(bS2-1)*(log(1-Se2)-
log(1-Se2pre));
  AlphaS1=exp(logAlphaS);

  AlphaS = min(alphaS1,1);
  if (alphaS < 1) then
    do;
      uS = ranuni(seed);
      if (uS < alphaS) then
        do;
          Se1pre=Se1;
          Se2pre=Se2;
          p111pre=p111;
          p101pre=p101;
          p011pre=p011;
        end;
      else
        Se1 = Se1pre;
        Se2 = Se2pre;
        p111=p111pre;
        p101=p101pre;
        p011=p011pre;
      end;
    end;
  else
    do;
      Se1pre = Se1;
      Se2pre = Se2;
      p111pre=p111;
      p101pre=p101;
      p011pre=p011;
    end;

/* Acceptance rate (P-move) for Specificity */
  eGAM = rangam(seed, (b11d+aDRL));
  fGAM = rangam(seed, (b10d+bDRL));
  gGAM = rangam(seed, (b01d+cDRL));

```

```

hGAM = rangam(seed, (b00d+dDRL));

p110 = eGAM/(eGAM+fGAM+gGAM+hGAM);
p100 = fGAM/(eGAM+fGAM+gGAM+hGAM);
p010 = gGAM/(eGAM+fGAM+gGAM+hGAM);
p000 = hGAM/(eGAM+fGAM+gGAM+hGAM);
Sp1 = p000 + p010;
Sp2 = p000 + p100;

logAlphaC = log(min(Sp1pre, Sp2pre) - Sp1pre*Sp2pre) -
log(min(Sp1, Sp2) - Sp1*Sp2) +
(aC1-1)*(log(Sp1) - log(Sp1pre)) + (bC1-1)*(log(1-Sp1) -
log(1-Sp1pre)) +
(aC2-1)*(log(Sp2) - log(Sp2pre)) + (bC2-1)*(log(1-Sp2) -
log(1-Sp2pre));
AlphaC1=exp(logAlphaC);

alphaC = min(AlphaC1, 1);
if (alphaC < 1) then
do;
uC = ranuni(seed);
if (uC < alphaC) then
do;
Sp1pre=Sp1;
Sp2pre=Sp2;
p100pre=p100;
p010pre=p010;
p000pre=p000;
end;
else
Sp1 = Sp1pre;
Sp2 = Sp2pre;
p100=p100pre;
p010=p010pre;
p000=p000pre;
end;
else
do;
Sp1pre = Sp1;
Sp2pre = Sp2;
p100pre=p100;
p010pre=p010;
p000pre=p000;
end;

dc=(p000-Sp1*Sp2)/(sqrt(Sp1*(1-Sp1)*Sp2*(1-sp2)));
ds=(p111-Se1*Se2)/(sqrt(Se1*(1-Se1)*Se2*(1-Se2)));

end; /*end MH */

if(mod(ncycle,skip)=0 & ncycle > burnin) then /* sampling according to
ncycle */
/* burnin, and skip */
do;
cycle=cycle//ncycle;
alphaSsav=alphaSsav//alphaS;

```

```

alphaCsav=alphaCsav//alphaC;
P1par=P1par//P1;
P2par=P2par//P2;
Selpar=Selpar//Se1;
Sp1par=Sp1par//Sp1;
Se2par=Se2par//Se2;
Sp2par=Sp2par//Sp2;
  p111par=p111par//p111;
  p000par=p000par//p000;
  dspar=dspar//ds;
  dcpardcpar=dcpar//dc;
uSsav=uSsav//uS;
uCsav=uCsav//uC;
end;
end; /*End MH gibs mixing */

create output1 var {cycle alphaSsav alphaCsav usav P1par P2par
Selpar Se2par Sp1par Sp2par p111par p000par dspar dcpar};
append;

quit;
run;

/* Macro used to calculate median and 95% CI's for parameter estimates
*/
title1 '';
title2 '';
%macro table(ds);
  proc univariate noprint data=output1; var &ds;
    histogram / vscale=count beta (noprint theta=0 sigma=1 color=red
      percents=1 2.5 5 10 50 90 95 97.5 99);
    output out=&ds mean=mean std=std mode=mode pctlpre=P_
      pctlpts=0, 1, 2.5, 5, 10, 50, 90, 95, 97.5, 99, 100;
    data &ds; set &ds; var="&ds";
  %mend;
%table(P1par); %table(P2par); %table(Selpar); %table(Se2par);
%table(Sp1par); %table(Sp2par);
%table(p111par); %table(p000par); %table(dspar); %table(dcpar);
data margdistmacroh; set P1par P2par Selpar Sp1par Se2par Sp2par
p111par p000par dspar dcpar;
title 'Summary statistics for posterior distributions for P1, Se1, Se2,
Sp1, Sp2, p111 p000 dspar dcpar';
  proc print; id var; var P_1 P_2_5 P_5 P_50 /*mean*/ P_95 P_97_5 P_99;
  format P_1 P_2_5 P_5 P_50 /*mean*/ P_95 P_97_5 P_99;
run;

/*Creating trace plot to assure sufficient burnin */
title1 'Trace Plot ...';
proc gplot data=output1;
  plot P1par*cycle / overlay;
  plot P2par*cycle / overlay;
  plot Selpar*cycle / overlay;
  plot Se2par*cycle / overlay;
  plot Sp1par*cycle / overlay;
  plot Sp2par*cycle / overlay;
  symbol1

```

```
value=circle  
cv=black;  
run; quit;
```

Conditional dependence model (M3):

```
proc iml;
  /* Seed values for the binomial distributions */
  seed=1;

  /* Total number of samples (cycles) */
  total = 12000;
  /* Number of burn-in samples */
  burnin = 2000; *burnin sample;
  skip = 1; *sample every 100 cycles;

  /* Starting values for prevalence, sensitivity, and specificity etc*/
  Selpre=.3; Se2pre=.4; Splpre=0.5; Sp2pre=0.4;
  p11pre=.4; p10pre=.4; p01pre=.4; p00pre=.4;
  p110pre=.4; p100pre=.4; p010pre=.4; p000pre=.4;
  P1=.2; P2=.2; Se1=.6; Se2=.5; Sp1=.6; Sp2=.5; p111=.1; p000=.1;
  p001=.1; p101=.1; p100=.1;
  ds=1;

  /* files to save different parameters */
  alphaSsav=.;
  alphaCsav=.;
  P1par=P1;
  P2par=P2;
  Selpar=Sel;
  Se2par=Se2;
  Splpar=Sp1;
  p111par=p111;
  uSsav=0;
  uCsav=0;
  dspar=ds;

  /* Observed cell values, x */
  /* Population 1 */
  x111=23; x101=7; x011=4; x001=162;
  /* Population 2*/
  x112=10; x102=5; x012=1; x002=282;

  /* Coding of the alpha and beta parameters for the prior distributions
  */
  /* a=alpha and b=beta parameters for the beta distributions */
  /* Alpha and beta parameters for prior distributions of prevalence */
  aP1=13.63;
  bP1=57.94;
  aP2=5.98;
  bP2=66.68;

  /* Alpha and beta parameters for prior distributions of sensitivity */
  aS1=1;
  bS1=1;
  aS2=1;
  bS2=1;

  /* Alpha and beta parameters for prior distributions of Specificity */
  aC1=1;
  bC1=1;
```

```

/* prior parameters for Dirichlet distribution */
aDRL=1;
bDRL=1;
cDRL=1;
dDRL=1;

/* Gibbs and MH algorithm */
do ncycle = 1 to total; /*Start MH and Gibbs mixing */

    a111=x111;

    if x101^=0 then
        do;
            pa101 = (p1*p101/(p1*p101+(1-p1)*(1-Sp1)));
            a101 = ranbin(seed, x101, pa101);
        end;
    else a101 = 0;

    a011=x011;
    if x001^=0 then
        do;
            pa001 = (p1*p001/(p1*p001+(1-p1)*Sp1));
            a001 = ranbin(seed, x001, pa001);
        end;
    else a001 = 0;

    a112=x112;
    if x102^=0 then
        do;
            pa102 = (p2*p101/(p2*p101+(1-p2)*(1-Sp1)));
            a102 = ranbin(seed, x102, pa102);
        end;
    else a102 = 0;

    a012=x012;
    if x002^=0 then
        do;
            pa002 = (p2*p001/(p2*p001+(1-p2)*Sp1));
            a002 = ranbin(seed, x002, pa002);
        end;
    else a002 = 0;

/* Calculating latent cell counts */
A1=a111+a101+a011+a001;
b111=0;
b101=x101-a101;
b011=0;
b001=x001-a001;
B1=b001+b101;
N1=A1+B1;

A2=a112+a102+a012+a002;
b112=0;
b102=x102-a102;
b012=0;
b002=x002-a002;
B2=b002+b102;

```

```

N2=A2+B2;

a11d=a111+a112;
a10d=a101+a102;
a01d=a011+a012;
a00d=a001+a002;

/* b11d=0;*/
b10d=b101+b102;
/* b01d=0;*/
b00d=b001+b002;

A=A1+A2;
B=B1+B2;
N=A+B;

/* Sampling full conditional for prevalence */
P1a = rangam(seed, (aP1+A1));
P1b = rangam(seed, (bP1+B1));
P1 = P1a/(P1a+P1b);

P2a = rangam(seed, (aP2+A2));
P2b = rangam(seed, (bP2+B2));
P2 = P2a/(P2a+P2b);

Spla = rangam(seed, (aC1+b00d));
Splb = rangam(seed, (bC1+b10d));
Sp1 = Spla/(Spla+Splb); *print spl;

do iter=1 to 1; /*Start MH*/
/* do until (p111>Se1*Se2);*/
aGAM = rangam(seed, (a11d+aDRL));
bGAM = rangam(seed, (a10d+bDRL));
cGAM = rangam(seed, (a01d+cDRL));
dGAM = rangam(seed, (a00d+dDRL));

/* Proposed probabilities for P11, P10, P01, P00 */
p111 = aGAM/(aGAM+bGAM+cGAM+dGAM);
p101 = bGAM/(aGAM+bGAM+cGAM+dGAM);
p011 = cGAM/(aGAM+bGAM+cGAM+dGAM);
p001 = dGAM/(aGAM+bGAM+cGAM+dGAM);
Se1 = p111+p101;
se2 = p111+p011;
se12 = Se1*Se2; *not needed!;
/* end;*/

/* Acceptance rate (P-move) for Sensitivity */
logAlphaS = log(min(Se1pre, Se2pre) - Se1pre*se2pre) -
log(min(Se1, Se2) - Se1*Se2) +
(aS1-1)*(log(Se1) - log(Se1pre)) + (bS1-1)*(log(1-Se1) -
log(1-Se1pre)) +
(aS2-1)*(log(Se2) - log(Se2pre)) + (bS2-1)*(log(1-Se2) -
log(1-Se2pre));
AlphaS1=exp(logAlphaS);
AlphaS = min(alphaS1, 1);
if (alphaS < 1) then

```

```

do;
  uS = ranuni(seed);
  if (uS < alphaS) then
    do;
      Selpre=Sel;
      Se2pre=Se2;
      p111pre=p111;
      p101pre=p101;
      p011pre=p011;
    end;
  else
    Sel = Selpre;
    Se2 = Se2pre;
    p111=p111pre;
    p101=p101pre;
    p011=p011pre;
  end;
else
do;
  Selpre = Sel;
  Se2pre = Se2;
  p111pre=p111;
  p101pre=p101;
  p011pre=p011;
end;

ds=(p111-Sel*Se2)/(sqrt(Sel*(1-Sel)*Se2*(1-Se2)));

end; /*end MH */

if(mod(ncycle,skip)=0 & ncycle > burnin) then /* sampling according to
ncycle */
/* burnin, and skip */
do;
  cycle=cycle//ncycle;
  alphaSsav=alphaSsav//alphaS;
  alphaCsav=alphaCsav//alphaC;
  P1par=P1par//P1;
  P2par=P2par//P2;
  Selpar=Selpar//Sel;
  Sp1par=Sp1par//Sp1;
  Se2par=Se2par//Se2;
  Sp2par=Sp2par//Sp2;
  p111par=p111par//p111;
  p000par=p000par//p000;
  dspar=dspar//ds;
  uSsav=uSsav//uS;
  uCsav=uCsav//uC;
end;
end; /*End MH gibbs mixing */

create output1 var {cycle alphaSsav alphaCsav usav P1par P2par
Selpar Se2par Sp1par p111par p000par dspar};
append;

quit;

```

```

run;

/* Macro used to calculate median and 95% CI's for parameter estimates
*/
title1 '';
title2 '';
%macro table(ds);
  proc univariate noprint data=output1; var &ds;
    histogram / vscale=count beta (noprint theta=0 sigma=1 color=red
      percents=1 2.5 5 10 50 90 95 97.5 99);
    output out=&ds mean=mean std=std mode=mode pctlpre=P_
      pctlpts=0, 1, 2.5, 5, 10, 50, 90, 95, 97.5, 99, 100;
    data &ds; set &ds; var="&ds";
  %mend;
%table(P1par); %table(P2par); %table(Selpar); %table(Se2par);
%table(Sp1par); %table(Sp2par);
%table(p111par); %table(p000par); %table(dspar); %table(dcpa);
data margdistmacroh; set P1par P2par Selpar Sp1par Se2par p111par
p000par dspar; * dcpa;
title 'Summary statistics for posterior distributions for P1, Sel, Se2,
Sp1, Sp2, p111 p000 dspar dcpa';
proc print; id var; var P_1 P_2_5 P_5 P_50 /*mean*/ P_95 P_97_5 P_99;
format P_1 P_2_5 P_5 P_50 /*mean*/ P_95 P_97_5 P_99;
run;

/*Creating trace plot to assure sufficient burnin */
title1 'Trace Plot ...';
proc gplot data=output1;
  plot P1par*cycle / overlay;
  plot P2par*cycle / overlay;
  plot Selpar*cycle / overlay;
  plot Se2par*cycle / overlay;
  plot Sp1par*cycle / overlay;
/* plot Sp2par*cycle / overlay;*/
  symbol1
    value=circle
    cv=black;
run; quit;

```

Conditional dependence model (M4)[Georgiadis et al., 2003]:

- * this program is based on a paper by MP Georgiadis, WO Johnson
IA Gardner and R Singh. Appl. Statist. (2003) 52, pp. 1-14
- * Test 1: MAT (Prior info is selected for this parameter)
Test 2: ELISA (No prior info [Parameters to be 'estimated']);
- * Notation
P1: prevalence in population 1
P2: prevalence in population 2
Se11, Se12, Se21, Se22 conditional probabilities (eta in paper)
Sp11, Sp12, Sp21, Sp22 Conditional probabilities (theta in paper)
Se1=Se11+Se12: sensitivity of test 1 (eta in paper)
Sp1=Sp21+Sp22: specificity of test 1 (theta in paper)
D: diseased (D in paper)
H: healthy (D with bar over in paper)
l: lambda
g: gamma
w: xxxxx. Vector of parameters: w=(P1, P2, Se1, Sp1, lD, gD, lH, gH)
lD=Se11/Se1
gD=Sp21/(1-Se1)
lH=Sp22/Sp1
gH=Sp12/(1-Sp1)
r: rho
rD: correlation parameter for diseased animals (D)
rH: correlation parameter for healthy animals (H)

See paper p. 5 for the reparameterizations using lambda and gamma;

```
proc iml;
/* files to save different parameters */
P1par=P1;
P2par=P2;
Se1par=Se1;
Se2par=Se2;
Sp1par=Sp1;
Sp2par=Sp2;
lDpar=lD;
lHpar=lH;
gDpar=gD;
gHpar=gH;
rDpar=rD;
rHpar=rH;
dspar=ds;
dcpar=dc;
/* Seed values for the binomial distributions */
seed1=1;

/* Starting values for prevalence, sensitivity, and specificity */
P1=0.1; P2=0.5; Se1=0.1; Sp1=0.1;
lD=0.1; lH=0.1; gD=0.1; gH=0.1;
```

```

/* Observed cell values, x */
/* Population 1 */
x111=27; x121=2; x211=7; x221=63;
/* Population 2*/
x112=32; x122=5; x212=11; x222=243;

/* Priors for prevalence, Se1, Sp1, lambda, and gamma*/
/* Coding of the alpha and beta parameters for the prior distributions
*/
/* a=alpha and b=beta parameters for the beta distributions */
/* Alpha and beta parameters for prior distributions of prevalence */
aP1=19.16; *13.63;
bP1=27.79; *57.94;
aP2=3.11; *5.98;
bP2=36.56; *66.68;
/* Alpha and beta parameters for prior distributions of sensitivity
*/
aSe1=1;
bSe1=1;
/* Alpha and beta parameters for prior distributions of Specificity */
aSp1=360.4;
bSp1=1.36;
/* Alpha and beta parameters for prior distributions of lambda */
a1D=1;
b1D=1;
a1H=1;
b1H=1;
/* Alpha and beta parameters for prior distributions of gamma */
agD=1;
bgD=1;
agH=1;
bgH=1;

/* Total number of samples (cycles) */
total = 12000;
burnin = 2000; *burnin sample;
skip = 1; *sample every 100 cycles;
/* Sample every 100 cycles */

/* Starting Gibbs sampler */
do ncycle=1 to total;

/* Marginal frequencies for observed counts*/
/* Observed marginal counts, d indicates summation over a row or
column */
xdd1=(x111+x121+x211+x221);
xdd2=(x112+x122+x212+x222);
x1dd=(x111+x121+x112+x122);
x2dd=(x211+x212+x221+x222);
x21d=(x211+x212);
x22d=(x221+x222);
x11d=(x111+x112);
x12d=(x121+x122);

```

```

*Start Gibbs;
/* Imputation step */
/* Sampling of z-ijk's., which are bin(x-ijk, p-ijk) */
if x111^=0 then
  do;
    p111 = (p1*Se1*1D/(p1*Se1*1D+(1-p1)*(1-Sp1)*(1-gH)));
    z111 = ranbin(seed1, x111, p111);
  end;
else z111 = 0; *print z111sample;

  if x121^=0 then
    do;
      p121 = (p1*se1*(1-1D)/(p1*Se1*(1-1D)+(1-p1)*(1-Sp1)*gH));
      z121 = ranbin(seed1, x121, p121);
    end;
  else z121 = 0;

  if x211^=0 then
    do;
      p211 = (p1*(1-Se1)*gD/(p1*(1-Se1)*gD+(1-p1)*Sp1*(1-lH)));
      z211 = ranbin(seed1, x211, p211);
    end;
  else z211 = 0;

  if x221^=0 then
    do;
      p221 = (p1*(1-Se1)*(1-gD)/(p1*(1-Se1)*(1-gD)+(1-p1)*Sp1*1H));
      z221 = ranbin(seed1, x221, p221);
    end;
  else z221 = 0;

  if x112 then
    do;
      p112 = (p2*Se1*1D/(p2*Se1*1D+(1-p2)*(1-Sp1)*(1-gH)));
      z112 = ranbin(seed1, x112, p112);
    end;
  else z112 = 0;

  if x122^=0 then
    do;
      p122 = (p2*se1*(1-1D)/(p2*Se1*(1-1D)+(1-p2)*(1-Sp1)*gH));
      z122 = ranbin(seed1, x122, p122);
    end;
  else z122 = 0;

  if x212^=0 then
    do;
      p212 = (p2*(1-Se1)*gD/(p2*(1-Se1)*gD+(1-p2)*Sp1*(1-lH)));
      z212 = ranbin(seed1, x212, p212);
    end;
  else z212 = 0;

  if x222^=0 then
    do;
      p222 = (p2*(1-Se1)*(1-gD)/(p2*(1-Se1)*(1-gD)+(1-p2)*Sp1*1H));

```

```

        z222 = ranbin(seed1, x222, p222);
        end;
        else z222 = 0;

/* Latent marginal counts */
        zdd1 = (z111 + z121 + z211 + z221);
        zdd2 = (z112 + z122 + z212 + z222);

        z1dd = (z111 + z121 + z112 + z122);
        z2dd = (z211 + z221 + z212 + z222);

        z11d = (z111+z112);
        z12d = (z121+z122);

        z21d = (z211+z212);
        z22d = (z221+z222);

        z11d = (z111+z112);
        z12d = (z121+z122);

/* Posterior step */
/* Sampling marginal probability estimates */
/* Posterior beta for prevalence 1 (P1) */
        P1a = rangam(seed1, (aP1+zdd1));
        P1b = rangam(seed1, (bP1+xdd1-zdd1));
        P1 = P1a/(P1a+P1b);

/* Posterior beta for prevalence 2 (P2) */
        P2a = rangam(seed1, (aP2+zdd2));
        P2b = rangam(seed1, (bP2+xdd2-zdd2));
        P2 = P2a/(P2a+P2b);

/* Posterior beta for Sensitivity 1 (Se1) */
        Sela = rangam(seed1, (aSe1+z1dd));
        Selb = rangam(seed1, (bSe1+z2dd));
        Se1 = Sela/(Sela+Selb);

/* Posterior beta for Specificity 1 (Sp1) */
        Sp1a = rangam(seed1, (aSp1+x2dd-z2dd));
        Sp1b = rangam(seed1, (bSp1+x1dd-z1dd));
        Sp1 = Sp1a/(Sp1a+Sp1b);

/* Posterior beta for lD */
        lDa = rangam(seed1, (alD+z11d));
        lDb = rangam(seed1, (blD+z12d));
        lD = lDa/(lDa+lDb);

/* Posterior beta for lH*/
        lHa = rangam(seed1, (alH+x22d-z22d));
        lHb = rangam(seed1, (blH+x21d-z21d));
        lH = lHa/(lHa+lHb);

/* Posterior beta for gD */
        gDa = rangam(seed1, (agD+z21d));
        gDb = rangam(seed1, (bgD+z22d));
        gD = gDa/(gDa+gDb);

```

```

/* Posterior beta for gH */
  gHa = rangam(seed1, (agH+x12d-z12d));
  gHb = rangam(seed1, (bgH+x11d-z11d));
  gH = gHa/(gHa+gHb);

/* Posterior for Se2 */
  Se2 = lD*Se1+gD*(1-Se1);

/* Posterior for Sp2 */
  Sp2 = gH*(1-Sp1)+lH*Sp1;

/* Posterior for correlation for diseased subjects */
  Se11=Se1*lD;
  rD = (Se11-Se1*Se2)/sqrt(Se1*(1-Se1)*Se2*(1-Se2));
  ds=(Se11-Se1*Se2)/(sqrt(Se1*(1-Se1)*Se2*(1-Se2)));

/* Posterior for correlation for healthy subjects */
  Sp22=Sp1*lH;
  rH = (Sp22-Sp1*Sp2)/sqrt(Sp1*(1-Sp1)*Sp2*(1-Sp2));
  dc=(sp22-Sp1*Sp2)/(sqrt(Sp1*(1-Sp1)*Sp2*(1-sp2)));

if(mod(ncycle,skip)=0 & ncycle > burnin) then

  do;
    cycle=cycle//ncycle;
    P1par=P1par//P1;
    P2par=P2par//P2;
    Selpar=Selpar//Se1;
    Se2par=Se2par//Se2;
    Sp1par=Sp1par//Sp1;
    Sp2par=Sp2par//Sp2;
    lDpar=lDpar//lD;
    lHpar=lHpar//lH;
    gDpar=gDpar//gD;
    gHpar=gHpar//gH;
    rDpar=rDpar//rD;
    rHpar=rHpar//rH;
    dspar=dspar//ds;
    dcpa=dcpar//dc;
  end;

end;
  create Marios var {cycle P1par P2par Selpar Sp1par Se2par Sp2par
                    lDpar lHpar gDpar gHpar rDpar rHpar dspar dcpar};
  append;

quit;
run;

/* Macro used to calculate median and 95% CI's for parameter estimates
*/
title1 'Posterior Marginal Distribution for ...';
title2 'Cycles: Burn in = 5000, total = 1000000, Skip = 100';

```

```

%macro table(ds);
  proc univariate noprint data=Marios; var &ds;
    histogram / vscale=count beta (noprint theta=0 sigma=1 color=red
      percents=1 2.5 5 10 50 90 95 97.5 99);
    output out=&ds mean=mean std=std mode=mode pctlpre=P_
      pctlpts=0, 1, 2.5, 5, 10, 50, 90, 95, 97.5, 99, 100;
  data &ds; set &ds; var="&ds";
%mend;
%table(P1par); %table(P2par); %table(Selpar); %table(Sp1par);
%table(Se2par); %table(Sp2par); %table(lDpar); %table(lHpar);
%table(gDpar); %table(gHpar); %table(rDpar); %table(rHpar);
%table(dspar); %table(dcpar);
data mariosmacro; set P1par P2par Selpar Sp1par Se2par Sp2par
  lDpar lHpar gDpar gHpar rDpar rHpar dspar
  dcpar;
title 'Summary statistics for posterior distributions for P1, Sel, Se2,
Sp1, Sp2';
proc print; id var; var P_1 P_2_5 P_5 P_50 /*mean*/ P_95 P_97_5 P_99;
  format P_1 P_2_5 P_5 P_50 /*mean*/ P_95 P_97_5 P_99;
run;

/*Creating trace plot to assure sufficient burnin */
title1 'Trace Plot ...';
proc gplot data=marios;
  plot P1par*cycle / overlay;
  plot P2par*cycle / overlay;
  plot Selpar*cycle / overlay;
  plot Sp1par*cycle / overlay;
  plot Se2par*cycle / overlay;
  plot Sp2par*cycle / overlay;
  plot lDpar *cycle / overlay;
  plot lHpar *cycle / overlay;
  plot gDpar *cycle / overlay;
  plot gHpar *cycle / overlay;
  plot rDpar *cycle / overlay;
  plot rHpar *cycle / overlay;
  plot dspar *cycle / overlay;
  plot dcpar *cycle / overlay;
  symbol1
    value=circle
    cv=black;
run; quit;

```

BIBLIOGRAPHY

Adams, L.G., 2001, In vivo and in vitro diagnosis of *Mycobacterium bovis* infection. Rev. sci. tech. int. Epiz. 20, 304-324.

Affronti, L.F., 1988, Mycobacterial antigens: reagents for tuberculin skin testing and serodiagnosis of tuberculosis. Plenum Press, New York, 1-37 pp.

Angus, R., 1978, Production of reference PPD tuberculins for veterinary use in the United States. J Biol Stand 6, 221-227.

Angus, R.D., Payeur, J.B., Elksen, L.A., 1989, Preliminary evaluation of an intradermal tuberculin test utilizing cell wall protein extracted from *M. bovis*. Proceedings of the U.S. Animal Health Association 93, 111-124.

Anonymous 1992. Assessment of risk factors for *Mycobacterium bovis* in the United States (Fort Collins, Colorado USA, USDA, APHIS, VS).

Anonymous, 1997, STATA statistical software release 8 reference manual. STATA Press, College Station, TX.

Black, M.A., and B.A. Craig. 2002, Estimating disease prevalence in the absence of a gold standard. Statist. Med. 21, 2653-2669.

Bruning-Fann, C.S., Schmitt, S.M., Fitzgerald, S.D., Fierke, J.S., Friedrich, P.D., Kaneene, J.B., Clarke, K.A., Butler, K.L., Payeur, J.B., Whipple, D.L., Cooley, T.M., Miller, J.M., Muzo, D.P., 2001, Bovine tuberculosis in free-ranging carnivores from Michigan. J. of Wildlife Dis. 37, 58-64.

Cameron, A.R., Baldock, F.C., 1998, A new probability formula for surveys to substantiate freedom from disease. Prev. Vet. Med. 34, 1-17.

Casella, G., Berger, R.L., 2001, Statistical Inference. Second ed: Duxbury Press.
Chib, S., and E. Greenberg. 1995, Understanding the Metropolis-Hastings algorithm. Am. Statistician 49, 327-335.

Christensen, J., Gardner, I.A., 2000, Herd-level interpretation of test results for epidemiologic studies of animal diseases. Prev. Vet. Med. 45, 83-106.

Cooney, R., Kazda, J., Quinn, J., Cook, B., Müller, K., Monaghan, M., 1997, Environmental mycobacteria in Ireland as a source of non-specific sensitization to tuberculins. Irish Vet. J. 50, 370-373.

- Corner, L.A., Pearson, C.W., 1978a, Pathogenicity in cattle of atypical mycobacteria isolated from feral pigs and cattle and the correlation of lesions with tuberculin sensitivity. *Austr. Vet. J.* 54, 280-286.
- Corner, L.A., Pearson, C.W., 1978b, Response of cattle to inoculation with atypical mycobacteria of bovine origin. *Austr. Vet. J.* 54, 379-382.
- Corso, B., Wagner, B., Norden, D. 1996. Assessing the risks associated with *M. bovis* in Michigan free-ranging deer. (Fort Collins, CO, Centers for Epidemiology and Animal Health, USDA, APHIS), pp. 1-2.
- Dempster, A., N. Laird, and D. Rubin, 1977. Maximum likelihood from incomplete data via the EM algorithm. *J. Royal Statist. Soc. Ser. B* 39, 1-38.
- Dendukuri, N., Joseph, L., 2001, Bayesian approach to modeling the conditional dependence between multiple diagnostic tests. *Biometrics* 57, 158-167.
- Donald, A., 1993, Prevalence estimation using diagnostic tests when there are multiple, correlated disease states in the same animal or farm. *Prev. Vet. Med.* 15, 125-145.
- Donald, A.W., Gardner, I.A., 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.* 19, 167-187.
- Eisenach, K., Cave, M., Bates, J., Crawford, J., 1990, Polymerase chain reaction amplification of a repetitive DNA sequence specific for *Mycobacterium tuberculosis*. *J Infect Dis* 161, 977-981.
- Enøe, C., Andersen, S., Sørensen, V., Willeberg, P., 2001, Estimation of sensitivity, specificity and predictive values of two serologic tests for the detection of antibodies against *Actinobacillus pleuropneumoniae* serotype 2 in the absence of a reference test (gold standard). *Prev. Vet. Med.* 51, 227-243.
- Enøe, C., M.P. Geordiadis, and W.O. Johnson. 2000, Estimation of sensitivity and specificity of diagnostic tests and disease prevalence when the true disease status is unknown. *Prev. Vet. Med.* 45, 61-81.
- Fitzgerald, S., Kaneene, J., Butler, K., Clarke, K., Fierke, J., Schmitt, M., Bruning-Fann, C., Mitchell, R., Berry, D., Payer, J., 2000, Comparison of postmortem techniques for the detection of *Mycobacterium bovis* in white-tailed deer (*Odocoileus virginianus*). *J Vet Diagn Invest* 12, 322-327.
- Flensburg, J., 1995, Regional and country status reports, Denmark, In: Thoen, C.O., Steele, J.H. (Eds.) *Mycobacterium bovis* in animals and humans. Iowa State University Press, Ames, pp. 206-212.

- Francis, J., 1947, Bovine tuberculosis, Including a contrast with human tuberculosis. Staple Press, London, 220 p.
- Francis, J., Seller, R.J., 1978, The sensitivity and specificity of various tuberculin tests using bovine PPD and other tuberculins. *Vet. Rec.* 103, 420-425.
- Frey, G.H., 1995, Bovine tuberculosis eradication: The program in the United States, In: Thoen, C.O., Steele, J.H. (Eds.) *Mycobacterium bovis* infection in animals and humans. Iowa State University Press, Ames, IA, pp. 119-129.
- Gaboric, C.M., M.D. Salman, R.P. Ellis, and J. Triantis. 1996, Evaluation of five-antigen ELISA for diagnosis of tuberculosis in cattle and Cervidae. *JAVMA* 209, 962-966.
- Gardner, I.A., Stryhn, H., Lind, P., Collins, M.T., 2000, Conditional dependence between tests affects the diagnosis and surveillance of animal diseases. *Prev. Vet. Med.* 45, 107-122.
- Gart, J., Buck, A., 1966, Comparison of a screening test and a reference test in epidemiologic studies. *Am J Epidemiol* 83, 593-602.
- Gelfand, A.E., and A.F.M. Smith. 1990, Sampling-based approaches to calculating marginal densities. *J. Am. Statist. Assoc.* 85, 398-409.
- Gelman, A., J.B. Carlin, H.S. Stern, and D.B. ZRubin. 2000, Bayesian data analysis. Edited by C. Chatfield and J. V. Zidek. First CRC reprint ed, Chapman & Hall, Texts in statistical science series. Boca Raton: Chapman & Hall/CRC.
- Georgiadis, M.P., Gardner, I.A., Hedrick, R.P., 1998, Field evaluation of sensitivity and specificity of a polymerase chain reaction (PCR) for detection of *N. salmonis* in rainbow trout. *J. Aquat. Anim. Health* 10, 372-380.
- Georgiadis, M.P., Johnson, W.O., Gardner, I.A., Singh, R., 2003, Correlation-adjusted estimation of sensitivity of two diagnostic tests. *Appl. Statist* 52, 1-14.
- Greiner, M., Gardner, I.A., 2000, Epidemiologic issues in the validation of veterinary diagnostic tests. *Prev. Vet. Med.* 45, 3-22.
- Hui, S.L., Walter, S.D., 1980, Estimating the error rates of diagnostic tests. *Biometrics* 36, 167-171.
- Johnson, W.O., Gastwirth, J.L., Pearson, M., 2001, Screening without a "gold standard": The Hui-Walter Paradigm revisited. *Am. J. Epidemiol.* 153, 921-924.
- Johnson-Ifearulundu, Y., Kaneene, J.B., 1999, Distribution and environmental risk factors for paratuberculosis in dairy cattle herds in Michigan. *AJVR* 60, 589-596.

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

Jordan, D., McEwen, S.A., 1998, Herd-level test performance based on uncertain estimates of individual test performance, individual true prevalence and herd true prevalence. Prev. Vet. Med. 36, 187-209.

Joseph, L., Gyorkos, T., Coupal, L., 1995, Bayesian estimation of disease prevalence and the parameters of diagnostic tests in the absence of a gold standard. Am. J. Epidemiol. 153, 921-924.

Kaneene, J.B., Bruning-Fann, C.S., Granger, L.M., Miller, R., Porter-Spalding, B.A., 2002, Environmental and farm management factors associated with tuberculosis on cattle farms in northeastern Michigan. JAVMA 221, 837-842.

Kleeberg, H.H., 1960, The tuberculin test in cattle. JSAVMA 31, 213-225.

Kulldorff, M., Nargawalla, N., 1995, Spatial disease clusters: detection and inference. Stat. Med. 14, 799-810.

Kulldorff, M., Rand, K., Gherman, G., Williams, G., DeFrancesco, D. 2003. SatScan version 3.1.2. Software for the spatial and space-time scan statistics (National Cancer Institute, Bethesda, MD).

Lepper, A.W.D., Newton-Tabrett, D.A., Carpenter, M.T., Williams, O.J., 1977, The use of bovine PPD tuberculin in the single caudal fold test to detect tuberculosis in beef cattle. Austr Vet J 53, 208-213.

Lepper, A.W.D., Pearson, C.W., Corner, L.A., 1977, Anergy to tuberculin in beef cattle. Aust. Vet. J. 53, 214-216.

Leslie, I. W., C. N. Hebert, K. J. Burn, B. N. MacClancy, and W. J. C. Donnelly. 1975, Comparison of the specificity of human and bovine tuberculin PPD for testing cattle. 1 - Republic of Ireland. Vet Rec 96, 332-334.

Leslie, I.W., Hebert, C.N., 1975, Comparison of the specificity of human and bovine tuberculin PPD for testing cattle. 3 - National trial in Great Britain. Vet Rec 96, 338-341.

Leslie, I.W., Hebert, C.N., Barnett, D.N., 1975a, Comparison of the specificity of human and bovine tuberculin PPD for testing cattle. 2 - Southeastern England. Vet Rec 96, 335-338.

Leslie, I.W., Hebert, C.N., Burn, K.J., MacClancy, B.N., Donnelly, W.J.C., 1975b, Comparison of the specificity of human and bovine tuberculin PPD for testing cattle. 1 - Republic of Ireland. Vet Rec 96, 332-334.

Liang, K.Y., Zeger, S.L., 1986, Longitudinal data analysis using generalized linear models. *Biometrika*, 73, 13-22.

Martin, S.W., Shoukri, M., Thorburn, M.A., 1992, Evaluating the health status of herds based on tests applied to individuals. *Preventive Veterinary Medicine* 14, 33-43.

Mendoza-Blanco, J.R., X.M. Tu, and S. Iyengar. 1996, Bayesian inference on prevalence using a missing-data approach with simulation-based techniques: Applications to HIV screening. *Statist. Med.* 15, 2161-2167.

Miller, J., Jenny, A., Rhyan, J., Saari, D., Suarez, D., 1997, Detection of *Mycobacterium bovis* in formalin-fixed, paraffin-embedded tissues of cattle and elk by PCR amplification of an IS6110 sequence specific for *Mycobacterium tuberculosis* complex organisms. *J Vet Diagn Invest* 9, 244-249.

Monaghan, M.L., Doherty, M.L., Collins, J.D., Kazda, J.F., Quinn, P.J., 1994, The tuberculin test. *Vet Micro* 40, 111-124.

Ng, K.H., Watson, J.D., Prestidge, R., Buddle, B.M., 1995, Cytokine mRNA expressed in tuberculin skin test biopsies from BCG -vaccinated and *Mycobacterium bovis* inoculated cattle. *Immunol. Cell. Biol.* 73, 362-368.

Norby, B., Bartlett, P.C., Fitzgerald, S.D., Granger, L., Bruning-Fann, C., Whipple, D.L., Payeur, J.B., 2003a, The sensitivity of gross necropsy, caudal fold and comparative cervical tests for the diagnosis bovine tuberculosis. *JVDI* In Press.

Norby, B., Bartlett, P.C., Grooms, D., Kaneene, J.B., Bruning-Fann, C., 2003, Herd-level sensitivity, specificity, and predictive values of bovine tuberculosis skin tests in Michigan, Dissertation, Michigan State University, MI.

Norby, B., Tempelman, R.J., Hanson, T., Kaneene, J.B., Bartlett, P.C., 2003, Estimation of sensitivity and specificity of bovine tuberculosis skin tests in Michigan when a perfect reference is not available, Dissertation, Michigan State University, MI.

O'Reilly, L., 1995, *Tuberculin Skin Tests: Sensitivity and Specificity.*, First Edition. Iowa State Press, Ames, 355 p.

Pollock, J.M., Girvin, R.M., Lightbody, K.A., Clements, R.A., Neill, S.D., Buddle, B.M., Andersen, P., 2000, Assessment of defined antigens for the diagnosis of bovine tuberculosis in skin test-reactor cattle. *Vet Rec* 146, 659-665.

Prophet, E., Mills, B., Arrington, J., Sobin, L., eds, 1992, *Laboratory methods in histotechnology.* American Registry of Pathology, Washington, DC.

Qu, Y., and A. Hadgu. 1998, A model for evaluating the sensitivity and specificity for correlated diagnostic tests in efficacy studies with an imperfect reference test. *J. Am. Statist. Ass.* 93, 920-928.

Qu, Y., M. Tan, and M.H. Kutner. 1996, Random effects models in latent class analysis for evaluating accuracy of diagnostic tests. *Biometrics* 52, 797-810.

Radunz, B.L., Lepper, A.W.D., 1985, Suppression of skin reactivity to bovine tuberculin in repeat tests. *Aust. Vet. J.* 62, 191-194.

Reisner, E., Gatson, A., Woods, G., 1994, Use of gen-probe accupobes to identify *Mycobacterium avium* complex, *Mycobacterium tuberculosis* complex, *Mycobacterium kansasii* and *Mycobacterium gordonae* directly from BACTEC TB broth cultures. *J Clin Microbiol* 32, 2995-2998.

Rice, J.A., 1995, *Mathematical statistics and data analysis*, Second Edition. Duxbury Press, Belmont, CA.

Rosner, B., 2000, *Fundamentals of biostatistics.*, Fifth Edition. Duxbury, Pacific Grove, CA, 792 p.

Roswurm, J.D., Konyha, 1973, The comparative cervical tuberculin test as an aid to diagnosing bovine tuberculosis. *Proc. Annu. Meet. U.S. Anim. Health Assoc.* 77, 368-389.

Roswurm, J.D., Konyha, 1973, The comparative cervical tuberculin test as an aid to diagnosing bovine tuberculosis. *Proc. Annu. Meet. U.S. Anim. Health Assoc.* 77, 368-389.

Schmitt, S., Fitzgerald, S., Cooley, T., Bruning-Fann, C., Berry, D., Carlson, T., Minnis, R., Payer, J., Sikarskie, J, 1997, Bovine tuberculosis in free-ranging white-tail deer in Michigan. *J Wildlife Dis* 33, 749-758.

Singer, R.S., Boyce, W.M., Gardner, I.A., Johnson, W.O., Fisher, A.S., 1998, Evaluation of bluetongue virus diagnostic tests in free-ranging bighorn sheep. *Prev. Vet. Med.* 35, 265-282.

Staquet, M., Rozenzweig, M., Lee, M., Muggai, F.M., 1981, Methodology for assessment of new dichotomous diagnostic tests. *J. Chronic Dis.* 34, 599-610.

Stokes, M.E., Davis, C.S., Koch, G.G., 2000, *Categorical data analysis using the SAS system*, Second Edition. SAS Institute Inc., Cary, NC.

Stuth, J., Fay, L.D. 1975. Laboratory record for necropsy #76-50 (East Lansing, Michigan: Wildlife Pathology Laboratory).

- Tanner, M.A. 1996, Tools for statistical inference: Methods for the exploration of posterior distributions and likelihood functions. 3rd edition. New York: Springer.
- Tempelman, R.J., Gianola, D., 1996, A mixed effect model for overdispersed count data in animal breeding. *Biometrics* 52, 265-279.
- Torrance-Rynard, V.L., and S.D. Walter. 1997, Effects of dependent errors in the assessment of diagnostic test performance. *Statist. Med.* 16, 2157-2175
- USDA, A. 1999. Bovine Tuberculosis Eradication: Uniform Methods and Rules, Effective January 22, 1999, (United States Department of Agriculture.), pp. 1-34.
- Vacek, P.M. 1985, The effect of conditional dependence on the evaluation of diagnostic tests. *Biometrics* 41, 959-968.
- Walter, S.D., Irwig, L.M., 1988, Estimation of error rates, disease prevalence and relative risk from misclassified data: a review. *J. Clin. Epidemiol* 41, 923-937.
- Ward, M.P., Carpenter, T.E., 2000, Techniques for analysis of disease clustering in space and in time in veterinary epidemiology. *Prev. Vet. Med.* 45, 257-284.
- Wassink, G.J., Grogono-Thomas, R., Moore, L.J., Green, L.E., 2003, Risk factors associated with the prevalence of foot rot in sheep from 1999 to 2000. *Vet. Rec.* 152, 351-358.
- Weng, T.S. 1996, Evaluation of a new diagnostic test against a reference test less than perfect in accuracy. *Commun. Statist. Simul. Comput.* 35, 533-555.
- Whipple, D.L., Bolin, C.A., Davis, A.J., Jarnagin, J.L., Johnson, D.C., Payer, J.B., Saari, D.A., Wilson, A.J., Wolf, M.M., 1995, Comparison of the sensitivity of the caudal fold skin test and a commercial gamma-interferon assay for diagnosis of bovine tuberculosis. *AJVR* 56, 415-419.
- Willeberg, P., J.M. Wedam, I.A. Gardner, J.C. Holmes, J.A. Mousing, J. Kyrval, C. Enøe, and S. Andersen. 1997, A comparative study of visual and traditional post-mortem inspection of slaughter pigs: Estimation of sensitivity, specificity and differences in non-detection rates. *Epidémiologie et Santé Animale* 31/32:04.20.1-04.20.3.
- Wood, P.R., Corner, L.A., Rothel, J.S., Baldock, C., Jones, S.L., Cousins, D.B., McCormick, B.S., Francis, B.R., Creeper, J., Tweddle, N.E., 1991, Field comparison of the interferon-gamma assay and the intradermal tuberculin test for the diagnosis of bovine tuberculosis. *Austr. Vet. J.* 68, 286-290.
- Wood, P.R., Corner, L.A., Rothel, J.S., Ripper, J.L., Fifis, T., McCormick, B.S., Francis, B., Melville, L., Small, K., De Witte, K., Tolson, J., Ryan, T.J., de Lisle, G.W., Cox, J.C.,

Jones, S.L., 1992, A field evaluation of serological and cellular diagnostic tests for bovine tuberculosis. *Vet Micro* 31, 71-79.

Yang, I., and M. P. Becker. 1997, Latent variable modeling of diagnostic accuracy. *Biometrics* 53, 948-958.

Zeger, S. L., Liang, K.Y., 1986, Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*, 42, 121-130.

MICHIGAN STATE UNIVERSITY LIBRARIES



3 1293 02504 9648