Title

Latent-class regression models for simultaneously estimating test accuracy, true prevalence and risk factors for *Brucella abortus*

Authors

A. Campe[1]*; D. Abernethy[3,4]; F. Menzies[4]; M. Greiner[2]

Affiliations

1= Department of Biometry, Epidemiology and Information Processing, University of Veterinary Medicine Hannover and WHO Centre for Research and Training in Veterinary Public Health, Hannover, Germany

2= Federal Institute for Risk Assessment, Dept. Exposure, Germany, and University of Veterinary Medicine, Hannover, Germany

3= Department of Veterinary Tropical Diseases, Faculty of Veterinary Science, University of Pretoria, South Africa

4=Veterinary Epidemiology Unit, Department of Agriculture and Rural Development, Belfast, Northern Ireland.

*= Corresponding Author:

Dr. med. vet. Amely Campe

Veterinary specialist in Epidemiology

Department of Biometry, Epidemiology and Information Processing

University of Veterinary Medicine Hannover, Foundation

Buenteweg 2

D-30559 Hannover

G E R M A N Y

Tel.: +49-511-953-7961

e-mail: amely.campe@tiho-hannover.de

http://www.tiho-hannover.de/bioepi

Short running head

Latent class regression models for *B. abortus*

Abstract

In 2003/2004 a field trial was conducted in Northern Ireland to assess the diagnostic accuracy of six serological tests for bovine brucellosis caused by *Brucella abortus*. Whereas between-test comparisons were used to calculate test performances so far, the present study used a latent class approach to estimate diagnostic test accuracy parameters in the absence of a gold-standard for these six tests simultaneously and to estimate the true prevalence, while accounting for clustering in the study population and risk factors for true prevalence.

Results obtained in this study with regard to prevalence, sensitivity and specificity were largely in accordance with previous findings. Screening tests (SAT and EDTA) appeared to be the most sensitive; however, at low prevalences the EDTA and CFT showed the highest positive predictive values of all investigated tests. The specificities and negative predictive values of all diagnostic tests were found to be very high. Differences of prevalence between three groups of the study population with different risk of exposure could be attributed to the mode of sampling indicating that a more risk-based sampling will result in a higher prevalence than a cross-sectional sampling mode. Age, dairy status and history of abortion were shown to influence the prediction of the latent true infection status.

Keywords

Introduction

A brucellosis eradication scheme commenced in Northern Ireland in 1963 and the disease was almost eradicated before recrudescence occurred in 1997 [1]. This led to a re-evaluation

of the diagnostic tests through a large field trial of approximately 20 000 cattle, which were tested in parallel with six different serological tests, all of which were approved by the EU before 2006 [2]. Bacteriological culture, the normal gold standard, was only available for a small proportion of the animals sampled in order to confirm positive results.

Tests for bovine brucellosis have been studied extensively in other countries as well, in terms of diagnostic accuracy [3] and specific aspects such as conditional dependence [4] or diagnostic equivalence [5]. More recently, tests for brucellosis have also been evaluated using models that do not require a gold standard [6,7].

The latter approach uses the concept of latent class analysis [LCA; 8] and addresses the problem that comparison to an imperfect gold standard invokes bias in the estimation of accuracy parameters. The principle behind this analysis is that the true disease status is latent (i.e. unknown), but can be estimated from a number of measureable items providing information about the true disease status [i.e. diagnostic test results; 9]. The estimation of prevalence and test accuracy has become an issue of major interest in the past 20 years in both veterinary and medical disciplines [10] and LCA has been recommended as the method of choice for validation of diagnostic tests in the absence of a gold-standard [11, 12]. While LCA analysis can be implemented in a frequentist framework and the model parameters can be estimated by classical inference [13, 14] it can also be implemented in a Bayesian framework where prior information about model parameters is utilised. Key assumptions of the approach, (i) different prevalences among the populations included in the model, (ii) invariance of the diagnostic performance of the tests across the populations and (iii) conditional (i.e. given the true disease status) independence of all tests, have been investigated using simulation studies [15]. However, these approaches are often limited in

the level of complexity to account for the number of diagnostic tests, the number of different categories that can be observed for one test result, survey design issues (i.e. clustering of animals within herds, subgroups in the study population), and the analysis of risk factors as covariates. Features such as ability to estimate population-specific diagnostic performance and flexibility in model building implemented in the LCA procedure in SAS [16] may provide new insights into the problem of diagnostic validation without gold standard.This study describes some possibilities and limitations of employing Proc LCA in SAS® [16] as a LCA modelling approach for the estimation of diagnostic test accuracy. Furthermore, the application case will demonstrate the impact of sampling strategies and diagnostic test performance on prevalence estimation under natural exposure conditions.

Data from a surveillance system on bovine brucellosis gathered under natural exposure conditions in Northern Ireland serves as an exemplary dataset. Age, sex, herd type, maternal status and history of abortion are factors available from study data and known to be associated with brucellosis in cattle [17, 18]. Therefore, they are candidate predictors for the latent true infection status.

Material & Methods

Study population

Data were obtained from a field trial conducted in Northern Ireland between 1 January 2003 and 31 October 2004 [2]. Only results of the first test event were used in these analyses. Within the dataset, tested animals were clustered in herds.

Cattle in the study were classified by their putative exposure risk to brucellosis as described elsewhere [2], which reflected the reason why they were tested. A "routine" group

comprised cattle that were tested routinely as part of the Northern Ireland surveillance system, with no anticipated increased risk of brucellosis. A "risk" group was derived from animals tested due to an increased risk of exposure and subjected to more frequent testing. These herds included those contiguous to existing or previously infected herds, or with other forms of direct or indirect contact (e.g. prior movement of cattle, shared ownership or facilities). The third group ("restricted") consisted of cattle in herds where brucellosis had already been identified or the supervising veterinarian had reported strong suspicion of infection [2]. As these test reasons are considered to reflect the exposure status of the cattle, they will be referred to as subgroups in the study population throughout this paper.

Diagnostic tests

The six serological tests applied to the study samples were the complement fixation test (CFT), competitive ELISA (cELISA), serum agglutination test (SAT (31iu)), SAT with EDTA (ethylene diamine tetra-acetic acid, EDTA (31iu)) as an addition to improve specificity, indirect ELISA (iELISA) and Rose Bengal test (RBT). The screening threshold selected for the SAT and EDTA test results was 31 IU as defined by European Union (EU) legislation [19]. The CFT, SAT, EDTA, iELISA and RBT are standard tests in the EU and approved for intra-Community trade testing. The cELISA is a complementary test and not approved for intra-Community trade testing [19, 20]. All test outcomes were recorded as dichotomous.

In the context of LCA, the indicators are the underlying observed variables from which the latent information is derived. In the study presented here, the observed classification is based on the diagnostic test results per animal. Assuming that the indicators are dichotomous (R_j=2; positive, negative), a response pattern Y with $W = \prod_{j=1}^{J} R_j$ combinations of J indicators (j = 1 to 64) can be expected.

LCA modelling

According to the latent class approach, the true infection status in relation to the model parameters (prevalence, diagnostic sensitivity and specificity) is an abstract construction and not subject to a case definition fixed by the investigator (see causal graph in Figure 1). In the context of our study data, the infection status should be interpreted as current or recent exposure to *Brucella abortus*. Proc LCA in SAS® 9.3 was used for the latent-class analysis and its extension of multinomial regression was applied to assess the influence of covariates on the prevalence status (γ-estimate) regardless of the number of (latent) classes of this outcome variable [21, 22, 23]. Thus, a "class" in this context refers to the latent, unobserved expression of the infection status being positive or negative. Furthermore, the unknown diagnostic sensitivity and specificity for all six tests was estimated (ρ-estimates). Prevalence and test accuracy were estimated using maximum likelihood (ML). Structured lack of independence, caused by clustering of animals in herds, was accounted for by choosing the herd identification as the cluster variable. For models with clustering, robust standard errors (SE) were calculated based on Taylor linearization [24].

All models were fitted using the default maximum number of 5000 iterations and the default maximum absolute deviation of 0.000001. In order to assess model identification, 10 sets of random starting values for the γ- and ρ-estimates were generated by specifying a positive integer value in the so-called SEED statement. An identical seed was used in order to make analysis exactly reproducible. Furthermore, it was ascertained if these starting values consistently converge to the same solution. Data-derived flattening priors were invoked in order to enable model identification and to stabilize the estimation of models in case of sparseness issues [16, 25]. For the model selection, only those models were considered
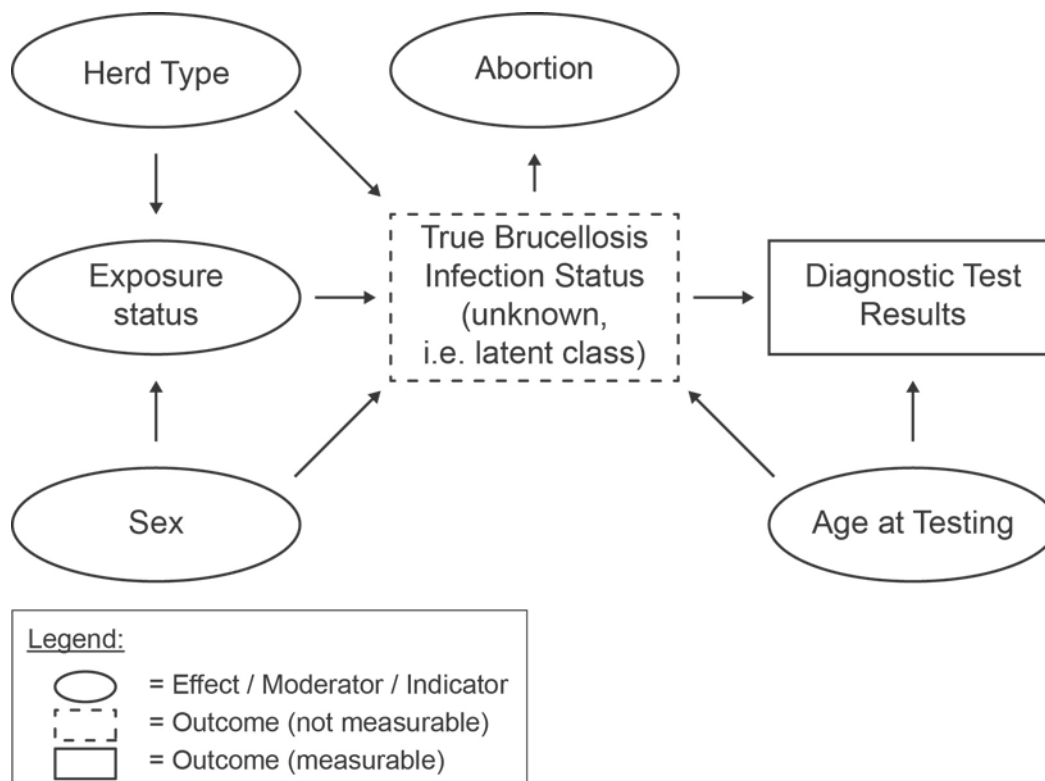
Figure 1: Causal Diagram for conceptualisation of latent class analysis with covariates for bovine *Brucella abortus* infection prevalence estimated using observed diagnostic test outcomes.

where a maximum likelihood solution was identified with 10 randomly selected starting values and all of them were associated with the best fitted model. In those cases, all sets of starting values yielded identical results for the G^2 likelihood ratio test statistic. The seed selected for the best model fit was used for subsequent analyses (i.e. with covariates) in order to provide a standard set of starting values. When evaluating a latent class model the principles of parsimony, interpretability as well as statistical criteria were considered.

Model diagnostics

Despite the fact that the optimal number of latent classes is two (based on the biological principle of an animal either being infected or not infected), sensitivity analyses were conducted for bias or confounding, and model fit was assessed.

The optimal number of latent classes was achieved by comparing alternative models using convergence and goodness-of-fit criteria. Furthermore, G^2-statistics and their p-values (significance level of 0.05) were regarded as meaningful only for models where the degrees of freedom were <60 [16] and where sparseness was not an issue. P-values for the G^2-statistic were derived from a comparison with a chi-square distribution with the corresponding degrees of freedom. As G^2-statistics are only a rough method for comparing model fit, AIC and BIC information criteria were also considered, with the smallest values indicating the best model fit. Using these criteria, models with one to six latent classes were compared.

To assess the impact of conditional dependence the model selection process and the values of the test accuracy estimates were checked and the symmetry of pairwise test results was analysed using the posterior probabilities of the model to assign latent class membership (threshold: ≥70% concordant tuples) and calculating the kappa coefficient (threshold: ≥0.61).

In order to assess whether test accuracy parameters (sensitivity, specificity) were constant across the (sub-)groups of the study population, this so-called "measurement invariance" was investigated by comparing a stratified, not-restricted and a stratified, restricted model. Both models were stratified by the exposure status of the animals. In the not-restricted model the test accuracy parameters were allowed to vary across groups, whereas in the restricted model the test accuracy parameters were constraint to be equal across groups. G^2-statistics model fit criteria and item-response probabilities were used to compare the model alternatives. Measurement invariance was assumed if the criteria favoured the restricted model, because this implies that the imposed restrictions are plausible. To this end, a likelihood-ratio difference test was conducted [16].

Covariates

Latent class analysis covariates were also assessed as to whether and to what degree they predicted the subgroup specific prevalences. The conceptual relationship between the latent variable and the observed indicator(s) as well as predictors is shown as a causal diagram (Fig. 1). The dichotomous candidate predictors, sex (male, female) and abortion (post abortion, no abortion/not applicable), were derived from information about reproductive status. The herd type was classified as dairy or non-dairy; : "non-dairy" herds consisted mainly of beef-cow herds, which are the predominant herd-type in Northern Ireland, and a range of other types, including weaning, rearing and finishing herds. Non-dairy and mixed herds were combined, because the group of mixed herds was relatively small (n=47 herds, 11.19%) and for all of these herds the type is not easily discerned. Dairy herds form a minority of herds in Northern Ireland (approximately 15%); they are associated with increased risk of certain diseases and are readily identified by the breed of female cattle held. The age at the time of

testing was available in years. In order to facilitate a basic analysis of covariates all qualitative exposure variables were dichotomised prior to analysis. Furthermore, the numeric variable age was dichotomized by the mean of its distribution in the overall study population. This was necessary, because there was no linear relationship between age and latent classification [26].

The study population was analysed according to the exposure risk of the herds (Routine/Risk/Restricted), with one model per subgroup. If the pairwise cross-tabulation of covariates and a Cramér's V ≥0.7 indicated considerable correlation between two covariates, only one of these variables was considered for multi-variable modelling to avoid the effects of collinearity.

A logistic regression model was applied with two latent classes to be estimated, where non-infected cattle was the reference latent class.

A non-automated forward selection of the four covariates and their interactions was conducted comparing the model fit of two models by means of a likelihood ratio χ^2 test (significance level: $p<0.05$). Interaction was considered and added to the model by calculating a product variable of the two respective covariates. Exposure variables with a p-value ≤0.05 in the Type III test of the regression analysis were considered to significantly affect the odds of latent class membership. Only results of the final model are reported here.

Results

Based on samples and information acquired for more than one year for 19517 animals from 420 herds, the true prevalence of brucellosis in Northern Ireland was calculated simultaneously with test accuracy estimates for the six applied serological tests. Data of 418

Table 1: Observed absolute number (n) and prevalence (%) of animals testing positive for each serological test. The study population is stratified by exposure groups, where cattle was routinely sampled, sampled due to previous risk or sampled due to known or strong suspicion of infection of the herd.

	Routine (n=5,082)		Risk (n=6,251)		Restricted (n=8,184)	
	n (tested positive)	%	n (tested positive)	%	n (tested positive)	%
CFT	31	0.61	73	1.17	133	1.63
EDTA (31iu)	51	1.00	102	1.63	154	1.88
Competitive ELISA	39	0.77	137	2.19	147	1.80
Indirect ELISA	31	0.61	77	1.23	144	1.76
RBT	34	0.67	86	1.38	150	1.83
SAT (31iu)	102	2.01	225	3.60	305	3.73

further animals was not sufficiently complete for the purpose of this analysis. The observed prevalence per serological test varied between 0.61 and 3.73% (table 1).

The distribution of herds with regard to herd type was 28.81% (n=121), 60.00% (n=252) and 11.19% (n=47) for dairy, non-dairy and mixed herds, respectively. The number of sampled animals per herd was skewed and varied between 1 and 424 (median=22, mean=46.47, interquartile range (IQR=2.0-68.5, sd=62.67). Most of the animals (88%) were tested for the first time between December 2003 and August 2004. The age of the cattle ranged from 3 months to 19 years (median=3.5 years, mean=4.34, IQR=1.92-6.17, sd=2.98). Most of the animals were female (bulls: 1.77%, n=346) and adult (heifers: 27.73%, n=5,412). The percentages of tested animals belonging to the routine, risk and restricted group of herds were 26.04% (n=5,082), 32.03% (n=6,251) and 41.93% (n=8,184), respectively.

Model diagnostics

When assessing the optimal number of latent classes for the pooled sample with all exposure groups, no model fitted the data sufficiently well (G^2 p<0.0001; see Table 2). However, AIC and BIC pointed to a 3-latent-classes solution. As no meaningful categorisation could be assigned to these three classes based on the input data, it was concluded that confounding might be an issue and stratification might improve the model fit. Stratification was conducted for the exposure status, as the most probable confounder, because it reflected on the heterogeneity of sampling schemes applied to gather study data. Stratification indicated that even fewer models converged and identified the ML solution under the given conditions. For these models, the likelihood-ratio G^2 statistic could not be conducted due to the high number of degrees of freedom. Hence, the criteria for relative model fit were considered indicating that a 2-latent-classes solution would be favourable.

Table 2: Optimal number of latent classes for 6 serological tests (with a cut-off of 31 iu for SAT and EDTA; N/W=304.95)

latent classes	ML solution identified with 10 randomly selected starting values (seeds)	percentage of seeds associated with the best fitted model [%]	Number of iterations for conversion	$df_{estimated}$[1]	G^2	p_{G2}	AIC	BIC	l	Entropy[2]
				Pooled models						
1	**yes**	**100**	2	57	10061.84	<0.0001	10073.84	10121.12	-10065.18	1
2	**yes**	**100**	13	50	777.35	<0.0001	803.35	905.78	-5422.94	1
3	**yes**	**100**	106	43	128.30	<0.0001	**168.30**	**325.88**	-5098.41	0.99
4	no	50	2621							
5 and 6	no		NOT CONV							
				stratified by "exposure status" + test accuracy estimates not restricted						
1	**yes**	**100**	2	173	10058.46	-[3]	10094.46	10236.28	-9967.52	1
2	**yes**	**100**	72	152	813.38	-[3]	**891.38**	**1198.66**	-5344.98	1

3	no	30	1450							
4, 5 and 6	no		NOT CONV							

stratified by "exposure status" + test accuracy estimates constraint to be equal across groups (restricted)

1	**yes**	**100**	2	185	10065.18	-[3]	10265.78	10313.06	-10065.18	1
2[4]	**yes**	**100**	13	176	943.35	-[3]	**973.35**	**1091.54**	-5409.97	1
3	no	40	151							
4, 5 and 6	no		NOT CONV							

1 = df<60 allows for trust in G2 statistic
2 = indicates certainty of class assignment to individuals in subsequent analysis
3 = p-values not reported due to the magnitude of degrees of freedom
4 = seed selected for best fitted model: 1444942552
AIC = Akaike Information Criterion
BIC = Bayesian Information Criterion
df = degrees of freedom
G^2 = likelihood ratio test statistic
l = log likelihood
NOT CONV = Model did not converge in 5,000 iterations
N/W=Sparseness; samples size/size of the contingency table
p_{G2} = p-value to the G^2-statistic derived from a chi-square distribution considering the df; if p>=0.05, then: by conventional criteria, this difference is considered not to be statistically significant.

Table 3: Class Membership Probabilities (gamma; within group latent prevalences) and Item Response Probabilities (rho; latent Sensitivities and Specificities) for a two-latent classes model on *Brucella abortus* infections in Northern Ireland cattle - with the positive and negative predictive values (Standard error is shown in parentheses).

	Exposure Group		
	Routine	**Risk**	**Restricted**
Sample Size	5082	6251	8184
Prevalence	0.0104	0.0190	0.0217
	(0.0034)	(0.0033)	(0.0048)
Sensitivity			
CFT	0.5680	0.6017	0.7430
	(0.1258)	(0.0764)	(0.0530)
EDTA (31iu)	0.9341	0.8429	0.8602
	(0.0414)	(0.0334)	(0.0365)
Competitive ELISA	0.5286	0.5100	0.6505
	(0.1366)	(0.0541)	(0.0450)
Indirect ELISA	0.4947	0.5099	0.7146
	(0.1442)	(0.0624)	(0.0567)
RBT	0.5680	0.6822	0.8148
	(0.1357)	(0.0666)	(0.0404)
SAT (31iu)	0.9537	0.9603	0.9759
	(0.0215)	(0.0216)	(0.0109)
Specificity			
CFT	1.0000	0.9998	1.0000
	(0.0000)	(0.0002)	(0.0000)
EDTA (31iu)	1.0000	0.9998	1.0000

	(0.0000)	(0.0002)	(0.0000)
Competitive ELISA	0.9980	0.9876	0.9962
	(0.0008)	(0.0031)	(0.0009)
Indirect ELISA	0.9992	0.9974	0.9980
	(0.0005)	(0.0008)	(0.0008)
RBT	0.9994	0.9993	0.9995
	(0.0003)	(0.0004)	(0.0003)
SAT (31iu)	0.9901	0.9820	0.9837
	(0.0018)	(0.0024)	(0.0024)

Positive predictive value (ppv)			
CFT	1.00	0.98	1.00
EDTA (31iu)	1.00	0.99	1.00
Competitive ELISA	0.74	0.44	0.79
Indirect ELISA	0.87	0.79	0.89
RBT	0.91	0.95	0.97
SAT (31iu)	0.50	0.51	0.57

Negative predictive value (npv)			
CFT	1.00	0.99	0.99
EDTA (31iu)	1.00	1.00	1.00
Competitive ELISA	1.00	0.99	0.99
Indirect ELISA	0.99	0.99	0.99
RBT	1.00	0.99	1.00
SAT (31iu)	1.00	1.00	1.00

As the test accuracy estimates were close to one and the favoured model had only two latent classes, the possible impact of conditional dependence can be considered low. Additionally, in the infected group most pairwise comparisons of test results had less than 70% concordant tuples. In the not infected group ≥97% of the test results were concordant in the pairwise comparison. Kappa coefficients were greater than 0.61 in four out of 15 test comparisons, only. As a difference between the stratified, not-restricted model and the stratified, restricted model, was identified (G^2_Δ=129.97, df=24, p<0.0001) and the criteria of relative model fit favoured the not-restricted model, the assumption of measurement invariance was dropped. Therefore, it can be concluded that sensitivity and specificity vary between the exposure groups in the study population. Hence, the best fitted model would account for the three exposure status sub-samples, separately, and have two latent classes, each.

Latent Prevalence and Diagnostic Test Accuracy

The estimated sensitivities and specificities along with prevalences, positive and negative predictive values are summarised in Tab. 3. The calculated prevalence for the routinely tested animals (0.0104; 95%CI=0.0078-0.0136) was lower than the prevalence of the other two subgroups (0.019, 95%CI=0.0158-0.0227, and 0.0217, 95%CI=0.0187-0.0251, for risk and restricted subgroups, respectively).

For all exposure groups the EDTA (31iu) and the SAT (31iu) test showed the highest sensitivities (0.84-0.93, 0.95-0.98, respectively). The sensitivity of the tests varied among the exposure groups, where best results were achieved for cattle in the restricted subgroup. However, for the test with the highest estimated sensitivity (SAT (31iu)), the probability of test-positive animals being true positives was 50-57% given the estimated prevalence for

each of the study subgroups; whereas the highest positive predictive values could be found for the CFT and the EDTA (31iu) tests (0.98-1.00, 0.99-1.00, respectively). This is due to the fact that the positive predictive value is a function not only of the prevalence and the sensitivity but also of the specificity, which was lower in SAT (31iu) than in the CFT and the EDTA (31iu) tests.

Regardless of the exposure group the CFT, EDTA (31iu) and the RBT test had the highest specificities (0.9998-1.00, 0.9998-1.00, 0.9993-0.9995, respectively). The specificity of tests varied between the exposure groups, where best results were achieved for routinely tested cattle. The negative predictive values did not vary considerably, neither among tests nor among study subgroups (0.99-1.00).

Covariates

As Cramér's V was <0.7 for all pairwise combinations of covariates in all three study subgroups, all covariates were considered appropriate for multi-variable modelling. However, in the contingency tables for the association between sex and abortion as well as between age and abortion empty cells were apparent, which is plausible as bulls or maiden heifers could not have an abortion. After the forward selection process of multi-variable modelling the best fitted final model identified three covariates for a latent infection with brucellosis: Age, dairy status and abortion (see Table 4). Models with interaction terms either showed non-identification of the models or a worse model fit than the models without interaction.

Cattle older than 4.34 years were 1.41 times (in the restricted group) to 2.53 times (on the routine group) more likely to be infected than younger cattle; a finding that was statistically significant only in the risk group (p=0.0426; see Table 4). Being kept on a dairy farm

Table 4: Predictors of latent class infection status tested against the reference latent class "non-infected" identified after forward selection in the final multi-variable model

	Exposure Group		
	Routine	Risk	Restricted
Sample Size	5082	6251	8184
Log likelihood	-843.22	-2085.56	-2345.94
Estimates			
Intercept[1]	-5.3328	-4.6134	-4.0081
	(0.6020)	(0.4836)	(0.5648)
Odds[2]	0.0048	0.0099	0.0182
	(0.0015-0.0157)	(0.0038-0.0256)	(0.0060-0.0550)
Age (<4.34 years [Ref.])	**p=0.0527**	**p=0.0426**	**p=0.2421**
Coefficient[1]	0.9288	0.5731	0.3425
	(0.4544)	(0.2927)	(0.2797)
Odds ratio[2]	2.5316	1.7738	1.4085
	(1.0390-6.1686)	(0.9995-3.1481)	(0.8141-2.4370)

Dairy status			
(non-dairy/Mixed [Ref.])	**p<0.0001**	p=0.1869	**p=0.0052**
Coefficient[1]	-2.0480	-0.5997	-0.9778
	(0.5494)	(0.4676)	(0.4243)
Odds ratio[2]	0.1290	0.5490	0.3762
	(0.0439-0.3786)	(0.2195-1.3727)	(0.1638-0.8640)
Abortion			
(no [Ref.])	**p<0.0001**	**p=0.0220**	**p=0.0008**
Coefficient[1]	3.7204	1.2026	2.1091
	(0.4543)	(0.4702)	(0.5238)
Odds ratio[2]	41.2809	3.3287	8.2408
	(16.944-100.58)	(1.3245-8.3655)	(2.9521-23.004)

N/A = not measurable; p = p-Value of the Type III beta parameter test; 1=standard error in parentheses; 2=95%-confidence interval in parentheses

decreased the probability of being infected significantly in the routinely tested (OR = 0.13) and the restricted group (OR = 0.38). Cattle tested due to an abortion were significantly associated with latent class membership in all exposure groups (OR= 3.3-41.3; see Table 4).

The distribution of infected and not-infected cattle over the combined covariates abortion, dairy status, sex and age was assessed. The most animals considered infected were non-dairy female cattle without abortion, irrespective of age or exposure group (routine: n=42, 77,8%; risk: n=77, 64,7%; restricted: n=136, 75,6%).

Discussion

The objective of this study was to estimate diagnostic accuracy in the absence of a gold-standard while simultaneously calculating the true prevalence given a high number of different diagnostic tests, clustering, stratification and risk factors.

LCA modelling

Diagnostic test performance and true prevalence are best calculated using latent class analysis. It is common knowledge that the degree to which the observed (apparent) prevalence differs from the true prevalence depends on the accuracy (Sensitivity, Specificity) of the diagnostic tests used. However, when calculating the true prevalence, Sensitivity and Specificity estimates can only be derived from previous test evaluation studies and these studies were often conducted in different populations (e.g. age structure, breed), regions or in different laboratories [27]. Therefore, all-in-one solutions like latent class models are preferable, where the true prevalence can be assessed along with test accuracy under the given field conditions [10]. As the estimation of latent disease status becomes more precise

the higher the number of diagnostic tests [28]; six tests, employed under field conditions in Northern Ireland, were incorporated in the study presented here.

Proc LCA is a suitable tool to conduct test accuracy estimations in the future. Although latent class modelling is a well-known method of assessing diagnostic test accuracy and true prevalence [26], it has only seldomly been conducted using Proc LCA in SAS® [16, 29, 30]. However, this tool provides several advantages while yielding the same point and interval estimates and model fit as compared to commonly used software tools such as TAGS [14]. It is not restricted to two latent classes, a maximum number of diagnostic tests or observed two-level diagnostic test results. It also includes an extension to regression modelling in order to investigate covariates and accounts for clustering of animals within herds. In other frameworks, within-herd dependencies are addressed using random effects, while Proc LCA adjusts standard errors in a pseudolikelihood approach [21]. However, one of the most interesting features is the possibility of stratifying the study population and calculating strata specific prevalence and test accuracy estimates, which enabled the exposure status specific estimates presented here. All options can be readily included, without the need for advanced programming skills.

Covariance cannot be directly accounted for in Proc LCA. However, for test evaluation studies it might be desirable that the covariance structure between the diagnostic test results are verifiable. Ignoring the conditional dependence between tests in such studies may yield a simplified and biased overestimation of the diagnostic accuracy parameters and true prevalence [9, 28]. This applies especially in cases where the same biological principle is inherent in all tests as in this case, where all tests are based on detection of antibodies for the diagnosis of *Brucella abortus* infection. Therefore, conditional dependence or

covariance, respectively, was taken into account [31]. However, as we did not conduct a test evaluation study *per se* and, therefore, could not resort to the true disease status in this study [31, 32], we considered external information about covariance [4] to introduce more rather than less bias due to different study conditions [i.e. target population, region; 27]. Furthermore, accounting for covariance parameters of six diagnostic tests would have yielded a high total amount of parameters to be estimated in the model. Therefore, conditional dependence was assessed based on model estimates and model fit. The fact that test accuracy estimates were close to one, the optimal latent class model corresponded with what was biologically expected (two classes) and the pairwise concordance between tests was lower than threshold, suggests that the model estimates are not relevantly biased by conditional dependence.

Stratification of the study population was deemed inevitable from a contextual and statistical point of view. As three different sampling strategies (reflecting different exposure status of cattle) were applied when gathering the study data, a pooled model seemed to be an inappropriate representative for the composition of the study population. When applying stratified modelling the violated measurement invariance [33] indicated a structured sample [16], where the grouping variable is associated with the realised test accuracy and prevalence estimates. . Considering the sampling strategy, it is understandable that the model estimates (Prevalence, Sensitivity, Specificity) vary between the population subgroups. Hence, analyses had to be performed for each of the three (exposure) subgroups, separately [24].

Latent Prevalence and Diagnostic Test Accuracy

Comparison of the Sensitivity and Specificity estimates of this study with previous publications is hindered as we did not use bacterial culture as a gold standard and calculated exposure-group specific test performances. Nevertheless, the results of this study can be considered more reliable and meaningful for brucellosis surveillance than foreign or previous estimations. Estimating test accuracy in the absence of a gold standard by means of LCA is well known [10, 13, 14], less expensive and faster than using culture diagnostic as the gold standard test. Nevertheless, the sensitivity and specificity estimates in this study varied (in some cases greatly) from estimates published in previous reviews of test performance [5, 36]. Furthermore, the analysis presented here was exposure-group specific; whereas calculation of diagnostic accuracy measures across all groups is biased. Nevertheless, the sensitivity and specificity estimates obtained in this study were largely in accordance with the estimates of Abernethy et al. (2012), where the same data were analysed using more traditional methods. In both analyses the SAT (31iu) yielded the best Sensitivity estimates of all tests and both analytical approaches indicated that screening tests (SAT (and EDTA)) compared favourably with the other test systems with regards to Sensitivity. Specificity in both analytical approaches was almost 100% for most tests [see Tab. 7 in Abernethy et al., 2012 and Table 3, here). Comparing both approaches, the one presented here provides prevalence and test accuracy estimates based on all diagnostic tests together, without the need for bacterial culture or a reference panel. This is particularly important where bacteriological culture is not available. For example, in the trial that provided this dataset, only test-positive cattle were slaughtered, and a negative result was based on concordance of multiple tests in a panel. In many other studies, especially where funding is constrained, testing is restricted to few tests and often without slaughter or laboratory follow-up.

Accordingly, the methods described here are less expensive (as slaughter is not a prerequisite), easier to apply and more precise. Furthermore, by considering the low prevalence in Northern Ireland when calculating predictive values, the results of our study (especially those of the sensitivity estimates) appear to be more qualified and are largely in accordance with the predictive values that can be calculated from other authors [6; for RBT and cELISA].As the sampling strategy of a surveillance programme influences the estimated prevalence and test performance, this should be considered in future disease control programmes. In the study presented here (latent) Prevalences ranging from 1.04% to 2.17% agreed with a cohort study on risk associated with brucellosis in Northern Ireland [Prev.:1.28%; 17]. Furthermore, it could be shown that the more risk-based the sampling the higher the prevalence, because then the probability of finding seropositive reactors increases [34]. On the one hand, the impact a more risk-based sampling has on the prevalence is a desired effect and corresponds with requirements of the final eradication stage of brucellosis [35]. On the other hand, our results indicate that the sampling strategy not only impacts on the prevalence but also on test accuracy estimates. This applies not only on *Brucella abortus* surveillance but also to other infectious diseases and should be accounted for more thoroughly in the future when calculating the true prevalence. Therefore, we propose exposure-group specific testing strategies to optimize the diagnosis of infectious diseases within surveillance programmes.

Surveillance programmes should be adapted regarding the choice of diagnostic tests according to the intended purpose of use and the eradication stage ["fitness for purpose", 37]. As bovine brucellosis prevalence declines worldwide, future surveillance programmes need to contemplate the considerable reduction in positive predictive value offered by serological tests, meaning that tests adequate for an overall prevalence reduction might not

be that adequate in later eradication stages (i.e. ELISA test systems). Hence, in a country with a low prevalence in the final eradication stage, it might be advisable to select a test system with a high positive predictive value from an economic point of view. This is to assure that the animals tested positively actually have the disease when being culled. On the other hand, it could have significant financial consequences missing an infected animal, which may outweigh any testing costs. Therefore, an inexpensive screening test with a low probability for animals tested negatively being diseased (i.e. high negative predictive value) might be well complemented by a follow up of animals that test positive using a test with a high positive predictive value. The importance of the negative predictive value from a disease control (and food safety) point of view has been elaborated elsewhere [20]. In this context the EDTA or CFT test fit the requirements as they even have a high specificity under field conditions [38].

Covariates

A logistic regression model was used to predict the latent class membership (i.e. being infected). An advantage of including covariate analysis in a latent class analysis over individual classification or scoring is that this allows uncertainty of the classification to be taken into consideration even in covariate analyses [16].

The study results indicated that sex did not contribute significantly more to the prediction of the latent class membership as compared to the baseline model, which can be related to the biological assumption that the probability of being infected is not affected by sex. Previous findings of differences between sexes in Northern Ireland might reflect age differences as well as different susceptibilities due to performance stress [38]. The age-related increasing

probability of infection is in agreement with the biological principle of an increasing probability of being exposed to the infectious agent the older one becomes [17].

The herd type seems to impact the probability of infection in the way that cattle on dairy farms are less likely to be infected than cattle on other (or mixed) farms. This effect, being most obvious in routinely tested cattle, might be attributable to a higher level of between-herd movement [39] or lower herd sizes in the non-dairy and mixed farms, but should be investigated in the future in more detail.

No distinction was made between the categories "no abortion" and "not applicable" as one was wholly dependent on owner information for pregnancy/abortion status for older heifers or cows for tests other than those triggered by a reported abortion. This might have biased the results. Nevertheless, the impact of having an abortion on the prediction of latent class membership is obvious and can be clearly attributed to abortion being the main clinical presentation of *Br. abortus* infection and the highly infectious nature of an abortion episode [40].

Conclusion

Bovine brucellosis is controlled in Northern Ireland since 1963. After a virtual eradication recrudescence occurred in 1997. Based on surveillance data this paper set out to simultaneously estimate diagnostic test accuracy parameters in the absence of a gold-standard and the true prevalence, while using a high number of different diagnostic tests and accounting for aspects such as clustering, stratification and the influences of covariates.

Considering the above mentioned prerequisites and the fact that diagnostic test performance and true prevalence are calculated most properly using latent class analysis, Proc LCA in SAS is a suitable tool to conduct test accuracy estimations in the future.

The estimated prevalence of *Brucella abortus* in Northern Ireland was low across all exposure groups. Our estimates of risk factors adjusted for misclassification suggest that cattle from non dairy/mixed farms and post abortion cows are rational targets for disease control measures under the given epidemiological circumstances. Regarding the serological surveillance of *Brucella abortus*, the EDTA or CFT test could be chosen as they even have a high specificity under field conditions.

Test performances of the six investigated diagnostic tests varied depending of the sampling strategy of the surveillance system. This applies not only on *Brucella abortus* surveillance but also on other infectious diseases and should be accounted for more thoroughly in the future. Therefore, we propose exposure-group specific testing strategies to optimize the diagnosis of infectious diseases within surveillance programmes.

Furthermore, surveillance programmes should be adapted regarding the choice of diagnostic tests according to the intended purpose of use and the eradication stage. Therefore, countries, with a low prevalence in the final eradication stage, may opt for a test system with high predictive values in order to avoid costs from culling not infected or from missing diseased animals.

Acknowledgments:

in SAS®; and Sonja Hartnack (Veterinary Epidemiology Group, Vetsuisse Faculty University of Zurich, Switzerland) for her support concerning the conceptional planning of analyses.

Declaration of interest

None.

References:

1. **Abernethy DA, *et al.*** Epidemiology of bovine brucellosis in Northern Ireland between 1990 and 2000. *Veterinary Record* 2006; **158**: 717-721.

2. **Abernethy DA, et al.** Field trial of six serological tests for bovine brucellosis. *Veterinary Journal* 2012; **191**: 364-370.

3. **Gall D, Nielsen K.** Serological diagnosis of bovine brucellosis: a review of test performance and cost comparison. *Revue scientifique et technique (International Office of Epizootics)*; 2004a: **23**: 989-1002.

4. **Mainar-Jaime RC, et al**. Specificity dependence between serological tests for diagnosing bovine brucellosis in Brucella-free farms showing false positive serological reactions due to Yersinia enterocolitica O:9. *Canadian Veterinary Journal*; 2005; **46**: 913-916.

5. **Greiner M, Verloo D, de Massis F**. Meta-analytical equivalence studies on diagnostic tests for bovine brucellosis allowing assessment of a test against a group of comparative tests. *Preventive Veterinary Medicine* 2009; **92**: 373-381.

6. **Muma JB, et al.** Evaluation of three serological tests for brucellosis in naturally infected cattle using latent class analysis. *Veterinary Microbiology* 2007; **125**: 187-192.

7. **Sanogo M, et al.** Bayesian estimation of the true prevalence, sensitivity and specificity of the Rose Bengal and indirect ELISA tests in the diagnosis of bovine brucellosis. *Veterinary Journal* 2013; **195**: 114-120.

8. **Lazarsfeld PF, Henry NW.** *Latent structure analysis*. Houghton, Mifflin New York, 1968.

9. **Reboussin BA, Ip EH, Wolfson M.** Locally dependent latent class models with covariates: an application to under-age drinking in the USA. *Journal of the Royal Statistical Society. Series A (Statistics in Society)* 2008; **171**: 877-897.

10. **Enoe C, Georgiadis MP, Johnson WO.** Estimation of sensitivity and specificity of diagnostic tests and disease prevalence when the true disease state is unknown. *Preventive Veterinary Medicine* 2000; **45**: 61-81.

11. **Yang I, Becker MP.** Latent variable modeling of diagnostic accuracy. *Biometrics* 1997; **53:** 948-958.

12. **Pepe MS, Janes H.** Insights into latent class analysis of diagnostic test performance. *Biostatistics* 2007; **8**: 474-484.

13. **Hui SL, Walter SD.** Estimating the error rates of diagnostic-tests. *Biometrics* 1980; **36**: 167-171.

14. **Pouillot R, Gerbier G, Gardner IA.** "TAGS", a program for the evaluation of test accuracy in the absence of a gold standard. *Preventive Veterinary Medicine* 2002; **53**: 67-81.

15. **Toft N, Jørgensen E, Højsgaard S**. Diagnosing diagnostic tests: evaluating the assumptions underlying the estimation of sensitivity and specificity in the absence of a gold standard. *Preventive Veterinary Medicine* 2005; **68**: 19-33.

16. **Collins LM, Lanza ST.** *Latent class and latent transition analysis*: With applications in the social, behavioral, and health sciences. Wiley, 2010.

17. **Stringer LA, et al.** Risk associated with animals moved from herds infected with brucellosis in Northern Ireland. *Preventive Veterinary Medicine* 2008; **84**: 72-84.

18. **Lopes LB, Nicolino R, Haddad JPA.** Brucellosis - risk factors and prevalence: A review. *Open Veterinary Science Journal* 2010; **4**: 72-84.

19. **Anonymous.** Comission Decision amending the Annex C to Council Directive 64/432/EEC and Decision 2004/226/EC as regards diagnostic tests for bovine brucellosis. In: Official Journal of the European Union (Ed.), 2008; L 352/338-345.

20. **Anonymous.** Scientific Opinion on "Performance of brucellosis diagnostic methods for bovines, sheep, and goats". *The EFSA Journal* 2006; **432**: 1-44.

21. **Lanza ST, et al.** PROC LCA: A SAS Procedure for Latent Class Analysis. *Structural Equation Modeling* 2007; **14**: 671-694.

22. **PROC LCA & PROC LTA (Version 1.3.2) [Software].** University Park: The Methodology Center, Penn State. Retrieved from http://methodology.psu.edu, 2015.

23. **SAS Institute Inc.** SAS/SAT User's Guide – versiopn 9.3. SAS Institute Inc. Cary, NC, USA, 2012.

24. **Lanza ST, et al.** PROC LCA & PROC LTA users' guide (Version 1.3.0). University Park: The Methodology Center, Penn State. Retrieved from http://methodology.psu.edu, Accessed 1 September 2013.

25. **Clogg CC, et al.** Multiple imputation of industry and occupation codes in census public-use samples using bayesian logistic regression. *Journal of the American Statistical Association* 1991; **86**:68-78.

26. **Dohoo I, Martin W, Stryhn H.** *Veterinary Epidemiologic Research*. VER Inc. Charlottetown, 2009.

27. **Greiner M, Gardner IA.** Epidemiologic issues in the validation of veterinary diagnostic tests. *Preventive Veterinary Medicine* 2000; **45**: 3-22.

28. **Hadgu A, Dendukuri N, Hilden J.** Evaluation of nucleic acid amplification tests in the absence of a perfect gold-standard test - A review of the statistical and epidemiologic issues. *Epidemiology* 2005; **16**: 604-612.

29. **Koukounari A, et al.** Sensitivities and specificities of diagnostic tests and infection prevalence of Schistosoma haematobium estimated from data on adults in villages northwest of Accra, Ghana. *American Journal of Tropical Medicine and Hygiene* 2009; **80**:435-441.

30. **Busch EL**, Markers of epithelial-mesenchymal transition and colorectal cancer mortality: time-to-event and latent-class analysis (dissertation). Chapel Hill, NC, USA: University of North Carolina, 2015, 154 pp.

31. **Gardner IA, et al.** Conditional dependence between tests affects the diagnosis and surveillance of animal diseases. *Preventive Veterinary Medicine* 2000; **45**: 107-122.

32. **Hanson TE, Johnson WO, Gardner IA.** Log-linear and logistic modeling of dependence among diagnostic tests. *Preventive Veterinary Medicine* 2000; **45**: 123-137.

33. **Millsap RE, Kwok OM.** Evaluating the impact of partial factorial invariance on selection in two populations. *Psychological Methods* 2004; **9**: 93-115.

34. **Hoinville LJ, et al.** Proposed terms and concepts for describing and evaluating animal-health surveillance systems. *Preventive Veterinary Medicine* 2013; **112**: 1-12.

35. **Adone R, Pasquali P.** Epidemiosurveillance of brucellosis. *Revue scientifique et technique (International Office of Epizootics)* 2013; **32**: 199-205.

36. **Gall D, Nielsen K.** Serological diagnosis of bovine brucellosis: a review of test performance and cost comparison. *Revue scientifique et technique (International Office of Epizootics)* 2004b; **23**: 989-1002.

37. **OIE - World Organisation for Animal Health**. Registration of Diagnostic Kits - Legal Basis (http://www.oie.int/our-scientific-expertise/registration-of-diagnostic-kits/background-information/). Accessed 4 November 2015.

38. **Abernethy DA.** The epidemiology and management of bovine brucellosis in Northern Ireland (PhD thesis). Royal Veterinary College, University of London, England, 2008.

39. **Christie TE.** Eradication of brucellosis in Northern Ireland: field problems and experience. *Veterinary Record* 1969; **85**: 268-269.

40. **Aparicio ED.** Epidemiology of brucellosis in domestic animals caused by Brucella melitensis, Brucella suis and Brucella abortus. *Revue scientifique et technique (International Office of Epizootics)* 2013; **32**: 53-60.