Preventive Veterinary Medicine xxx (2010) xxx–xxx

Contents lists available at ScienceDirect

Preventive Veterinary Medicine

journal homepage: www.elsevier.com/locate/prevetmed

Effect of classifying disease states in genetic association studies for paratuberculosis

J.B. Osterstock[a,*], S. Sinha[b], C.M. Seabury[c], N.D. Cohen[d]

[a] Texas AgriLife Research, Amarillo, TX, USA
[b] Department of Statistics, Texas A&M University, College Station, TX, USA
[c] Department of Veterinary Pathobiology, Texas A&M University, College Station, TX, USA
[d] Department of Large Animal Clinical Sciences, Texas A&M University, College Station, TX, USA

ARTICLE INFO

Article history:
Received 22 July 2009
Received in revised form 26 February 2010
Accepted 9 March 2010

Keywords:
Genetic association study
Cattle-microbiological diseases
Monte Carlo simulation
Bias
Paratuberculosis

ABSTRACT

Genetic association studies are a means to elucidate underlying genetic regulation of host–pathogen interaction, immune response, and the fate of infection. Diseases such as paratuberculosis in cattle lack definitive diagnostic criteria, thereby complicating the definition of infection status as an outcome for genetic association studies. A study was performed to evaluate the potential bias in estimates of effect and differences in statistical power associated with parallel test interpretation, latent probability of infection adjusted for imperfect test sensitivity and specificity, and multinomial outcomes in cohorts of cattle simulated using Monte Carlo sampling methods. Test results were simulated for microbial culture of feces for *Mycobacterium avium* subsp. *paratuberculosis* (MAP) and serum ELISA for anti-MAP antibody using estimates of test sensitivity and specificity. A range of disease allele frequencies and levels of association were considered. Case–control study populations were drawn from the simulated cohorts and the association between the disease allele and infection status was evaluated using logistic regression for binary outcomes and polytomous regression for multinomial outcomes. For the majority of the classification and analytical methods evaluated, estimates of effect were biased toward the null. Frequentist approaches to analysis of the latent probability of infection and multinomial classifications based upon results of culture of feces for MAP demonstrated the smallest degree of bias. Power to detect associations was generally low for all models, but improved with larger effects and higher allele frequencies. Imperfect specificity of serum ELISA was a major factor in the degree of bias observed and statistical power. The results of this study indicate that the method of classifying infection status must be considered carefully in genetic association studies for paratuberculosis and other diseases with similar challenges in defining infection status, and study designs should be modified to accommodate relative advantages and disadvantages of available methods.

1. Introduction

With the rapid increase in availability of genomic data and development of high-throughput technologies to identify genetic polymorphisms, association studies are being used with increasing frequency to examine the effects of genetic variation on manifestation of infectious diseases in human and animal populations. These studies contribute to our understanding of the elements of host–pathogen interaction, genetic regulation of innate disease resistance, and biomarkers for use in genetic selection programs for livestock. However, the ever increasing availability of variable genetic markers for study has also increased the complexity of epidemiologic and statistical approaches in such studies.

* Corresponding author at: Texas AgriLife Research, 6500 Amarillo Blvd. West, Amarillo, TX 79106, USA. Tel.: +1 806 677 5600; fax: +1 806 677 5644.
E-mail address: jbosterstock@ag.tamu.edu (J.B. Osterstock).

0167-5877/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.prevetmed.2010.03.005

One of the emerging challenges addressed herein is the definition of phenotypes for diseases that may have multiple clinical manifestations or lack definitive diagnostic criteria. The method of phenotypic classification and subsequent analyses may have profound effects on type I and type II error rate, magnitude of effect estimates relative to the true association, and optimal study design.

Paratuberculosis is a chronic inflammatory bowel disease of ruminants caused by infection with *Mycobacterium avium* subsp. *paratuberculosis* (MAP). The disease is most prevalent in dairy cattle and estimates of economic losses attributed to paratuberculosis in U.S. dairy herds due to decreased milk production, diagnostic tests and control programs, and reduced value of culled animals due to weight loss during advanced clinical stages range from $200 million to $1.5 billion annually (Stabel, 1998; Ott et al., 1999). Animals are typically infected early in life, but clinical manifestations of disease are often delayed by several years (Larsen et al., 1975; Nielsen and Ersboll, 2006). Additionally, diagnostic tests are insensitive and often fail to detect infected animals early in the course of disease. Currently, effective treatments and vaccines are unavailable. Therefore, control of this disease is limited to intensive control via surveillance using available tests, culling of infected animals, herd management protocols designed to reduce opportunities for young-stock to be exposed to feces, colostrum, or milk containing MAP, and the implementation of biosecurity protocols designed to reduce the probability of transmission within infected herds as well as movement of infected animals between herds.

Evidence exists that genetic variation in cattle is associated with differential susceptibility to paratuberculosis, with heritability estimates ranging from 6% to 15% in Holsteins (Koets et al., 2000; Mortensen et al., 2004; Gonda et al., 2006). To date, several studies have focused on examining specific genetic polymorphisms in candidate genes including caspase recruitment domain 15 (CARD15) (Pinedo et al., 2009), interferon-gamma (IFNγ) (Reddacliff et al., 2005), and Toll-like receptors 2 and 4 (TLR) (Mucha et al., 2009). Likewise, a genome wide association analysis has also emerged using the Illumina Bovine SNP50 assay (Settles et al., 2009). Nevertheless, results to date are inconsistent and putative disease associations have not been fully established. Potential causes for inconsistency of disease associations include differences and inaccuracy in defining the disease phenotype, differences in the genotypic distributions among sample populations, potential genotyping errors, and general approaches to data analysis.

There are several potential methods for using diagnostic test data to define paratuberculosis disease phenotypes. Tests may be interpreted in parallel where a positive result from any test is sufficient to classify the subject as infected whereas a negative test is required for all tests to classify the subject as uninfected. This is a common strategy for improving the cumulative diagnostic sensitivity of the testing strategy, although decreases in cumulative diagnostic specificity will be encountered. Alternatively, test results may be classified using a latent variable to describe the probability of the disease given the diagnostic test outcomes. This parameter is estimated using the prevalence of infected animals in the population and estimates of sensitivity and specificity for the tests. This approach has advantages in that the imperfect sensitivity and specificity of available tests is explicitly considered in deriving the latent variable estimate. However, this variable is susceptible to inaccuracies in available estimates of sensitivity and specificity, and assumptions of conditional dependence must be evaluated (Gardner et al., 2000). Analysis of latent parameters can be performed using frequentist approaches that utilize point estimates for sensitivity, specificity, and prevalence or Bayesian frameworks where these factors are defined as prior distributions and parameter estimates are derived from iterative sampling procedures. Finally, diagnostic test results may be used to generate a multinomial outcome. For example, all possible combinations of test results could be used to generate a multinomial outcome or correlated binary and continuous test outcomes may be simultaneously modeled to generate summary estimates of effect (Liu et al., 2009). The primary objective of this study was to examine the differences in the distribution of estimates of the effect of genetic exposure on the disease phenotype relative to the true value for these effects as well as the statistical power to detect associations observed using parallel test interpretation, latent disease classification, and multinomial terms in a simulated genetic association study for paratuberculosis in cattle.

2. Materials and methods

2.1. Simulation of case–control datasets

A simulation study was conducted to evaluate the differences in the observed estimates of the association parameter between a genetic exposure and the disease phenotype and its statistical power using different methods of identifying diseased subjects over a range of scenarios using available software (@Risk 5.1, Palisade Corp., Ithaca, NY). For each scenario, we simulated $R = 100$ case–control datasets. For each dataset we considered a cohort of N subjects, and defined their genetic exposure variable G by simulating a binomial random variable with $n = 2$ and success probability p (i.e., $\sim Bin(2,p)$) where p represented the frequency of the disease allele in the population. Possible genotypes included subjects that were homozygous ($G = 2$) or heterozygous ($G = 1$) for the disease allele and subjects that lacked the disease allele ($G = 0$). Values of p evaluated included 0.05, 0.1, and 0.2. The number of subjects (N) was adjusted by scenario to produce a population with at least 500 subjects eligible for selection as cases. True disease status (Y) was simulated for each subject using a Bernoulli distribution with success probability defined using a logistic function

$$P(Y = 1|G) = \frac{e^{\beta_0+\beta_1 G}}{1 + e^{\beta_0+\beta_1 G}} \quad (1)$$

where Y represents the true infection status (0 for uninfected and 1 for infected) and G represents the observed genotype. Here, β_0 is the intercept parameter and β_1 represents the log-odds ratio parameter for the effect of G on disease penetrance.

The infection status (Y) was determined by establishing 3 potential values for β_1 corresponding to odds ratios

Please cite this article in press as: Osterstock, J.B., et al., Effect of classifying disease states in genetic association studies for paratuberculosis. PREVET (2010), doi:10.1016/j.prevetmed.2010.03.005

(OR, defined as e^{β_1}) of 1.2, 1.5, and 2. The value of β_0 was controlled such that the prevalence of infected cattle was fixed at 5% for all possible values of β_1 given the respective genotypic distributions which is consistent with previously reported prevalence estimates for paratuberculosis in dairy cattle (Collins et al., 1994; NAHMS, 1997; van Schaik et al., 2003). Importantly, we assumed that all animals were uniformly exposed to MAP to remove variability attributed to exposure distribution as a confounding factor. The disease phenotypes for the simulated subjects were based on bacterial culture of feces for MAP and serum antibody measured by ELISA. The probability of having a positive test result was defined by

$$P(Test+|Y=1) \sim Bin(1, Se); \quad (2)$$

or

$$P(Test+|Y=0) \sim Bin(1, (1-Sp)); \quad (3)$$

where *Se* and *Sp* were the sensitivity and specificity of the individual tests, respectively. Point estimates and 95% confidence intervals for sensitivity and specificity were obtained from a review of available literature regarding tests for paratuberculosis in cattle, and specifically, for identification of cattle infected with MAP (Eamens et al., 2000; Dargatz et al., 2001; Collins et al., 2005; McKenna et al., 2005; Nielsen and Toft, 2008). The respective sensitivities and specificities as well as their distributions were defined as follows:

Se_{ELISA} = 0.15 (95% interval 0.09, 0.22), ~Beta(19.4, 109);
Sp_{ELISA} = 0.97 (95% interval 0.94, 0.99), ~Beta(162, 5);
$Se_{culture}$ = 0.25 (95% interval 0.16, 0.38), ~Beta(17, 48);
and
$Sp_{culture}$ = 1.

The effect of passive shedding of MAP by uninfected animals was not considered and therefore culture of feces for MAP was assumed to have perfect specificity. Importantly, development of the sampling frame for the case–control study population was performed using fixed values of sensitivity and specificity for each test. The distributions around these point estimates were only used in subsequent Bayesian analyses.

2.2. Classification of infection status

Based on simulated diagnostic test data, 3 potential classifications of infection status were considered. First, parallel test interpretation was used to define a dichotomous variable for infection status ($Y_{parallel}$) where a subject was eligible for consideration as a case if either test result (culture or ELISA) was positive and a control if both tests were negative. The second classification scheme used a latent variable for the probability of infection adjusting for imperfect sensitivity and specificity. This latent variable (Y_{latent}) was defined as

$$P(Y_{latent}=1|culture=1,\ ELISA=1) = P(Y_{latent}=1|culture=1,\ ELISA=0) = 1; \quad (4)$$

or

$$P(Y_{latent}=1|culture=0,\ ELISA=1) = \frac{Se_{ELISA} * Prev}{Se_{ELISA} * Prev + (1-Sp_{ELISA})(1-Prev)}; \quad (5)$$

or

$$P(Y_{latent}=1|culture=0,\ ELISA=0) = 1 - \frac{Sp_{cumulative} * (1-Prev)}{Sp_{cumulative} * (1-Prev) + (1-Se_{cumulative}) * Prev}; \quad (6)$$

where the cumulative sensitivity and specificity were defined by

$$Sp_{cumulative} = Sp_{ELISA} * Sp_{culture}; \quad (7)$$

and

$$Se_{cumulative} = 1 - (1-Se_{ELISA})(1-Se_{culture}) - Cov+ \quad (8)$$

and *Cov+* was the test-positive covariance parameter included to account for the dependence between positive serum ELISA and bacterial culture of feces results. The positive test covariance parameter was estimated to be 0.065 (95% interval 0.043, 0.092; ~Beta(25.8, 371)) as previously reported (Scott et al., 2007). Covariance of negative test results to adjust cumulative specificity estimates was assumed to be insignificant (Scott et al., 2007).

The final approach utilized to characterize infection status based on diagnostic test results was a categorical variable with a multinomial distribution. Subjects that were negative on both serum ELISA and culture of feces for MAP were assigned to the referent category (0). Subjects with MAP isolated from feces, but negative serum ELISA results were assigned to category 1. Subjects with positive serum ELISA, but negative fecal culture results were assigned to category 2. Subjects that were positive on both tests were assigned to category 3.

Each simulation was performed using Monte Carlo sampling with a new random seed used for each simulation. The iterations of each simulation were used as a sampling frame for a case–control study design. For all simulations, 500 cases were selected based on the described criteria. An equal number of controls were randomly selected from remaining subjects. The number of cases and controls included in this study were based on estimation of the minimum sample size necessary to detect an association with an OR of 1.2 for a disease allele frequency of 0.1 with 95% confidence and 80% statistical power. Each simulation and subsequent analyses were repeated 100 times for all possible combinations of p and β_1.

2.3. Estimating effect of disease allele on infection status

The association between genotype for the proposed disease allele and disease status was estimated using generalized linear models with a logit link function and binomial distribution (Eq. (9)) for parallel and frequentist latent classifications with available software (Stata 10 MP, Stata Corp., College Station, TX)

$$\text{logit}(\pi_i) = \log\frac{\pi_i}{1-\pi_i} = \beta_0 + \beta_1 G + \varepsilon, \quad Y_i \sim Bin(n_i, \pi_i) \quad (9)$$

Please cite this article in press as: Osterstock, J.B., et al., Effect of classifying disease states in genetic association studies for paratuberculosis. PREVET (2010), doi:10.1016/j.prevetmed.2010.03.005

For simulations using parallel test interpretation, cases were defined as subjects with at least 1 positive test result (culture of feces for MAP or ELISA). For simulations using frequentist latent test interpretations, cases were defined as subjects with a latent probability of infection >0.5.

Bayesian analysis of the latent probability of infection was performed in the statistical software R using a logit link function and binomially distributed error term for misclassification of the disease status. Parameter estimates were derived from Markov chain Monte Carlo (MCMC) sampling technique along with a Metropolis-Hastings algorithm. The number of iterations necessary to achieve convergence (5000) was set as the burn-in and an additional 15,000 iterations were used to derive posterior estimates of effect.

The multinomial outcome was analyzed using polytomous logistic regression (Eq. (10)) with available software (Stata 10 MP)

$$\text{logit}(\pi_i) = \log\frac{\pi_i}{\pi_0} = \beta_{i0} + \beta_{i1}G + \varepsilon, \quad Y_i \sim Bin(n_i, \pi_i) \qquad (10)$$

For multinomial outcomes, estimates of effect ($\hat{\beta}_1$) were derived for each category referent to the baseline category (i.e., negative results for both tests). All subjects in categories 1, 2, or 3 were eligible for selection as cases.

Summary parameters were calculated from obtained $\hat{\beta}_1$ including mean $\hat{\beta}_1$, mean width of the 95% confidence interval surrounding $\hat{\beta}_1$, and the proportion of 95% confidence intervals that contained the true value of β_1 used to develop the respective groups of simulations. Summary parameters were compared between methods for classification of disease status and among the simulated values of p used to describe the genotypic distributions. The statistical power associated with each classification method was estimated by determining the proportion of simulations for which the Z-score (Eq. (11)) was $\geq$1.96 corresponding to a 95% confidence level ($\alpha = 0.05$)

$$z = \frac{\widehat{\beta_1}}{SE} \qquad (11)$$

where SE is the standard error for the parameter estimate.

2.4. Sensitivity analysis

Additional scenarios were evaluated to assess the sensitivity of the effect estimates and power calculations on the assumptions regarding diagnostic test sensitivity and specificity. Simulations were performed as described above for 3 additional scenarios: (1) assuming sensitivity of fecal culture and serum ELISA remained fixed, but specificity of serum ELISA increased from 97% to 100%; (2) assuming test characteristics for fecal culture remained the same, but sensitivity of serum ELISA increased from 15% to 25% with a specificity of 97%; and (3) assuming test characteristics for fecal culture remained the same, but sensitivity of serum ELISA increased to 25% and specificity of the serum ELISA increased to 100%. In Bayesian analyses, the distribution of the more sensitive ELISA was assumed to be the same as the distribution for culture of feces for MAP (~Beta(17, 48)). For these scenarios, p and e^{β_1} were fixed at 0.1 and 1.5, respectively, as preliminary data (not shown) illustrated that observed changes in effect estimates and power were consistent across tested ranges of p and e^{β_1}. Results were summarized in tables and graphs to examine the effect of changes in sensitivity and specificity on mean $\hat{\beta}_1$, mean 95% confidence interval width, proportion of 95% confidence intervals that contained the true value of β_1, and calculated statistical power with all methods of classifying disease phenotype and analysis.

3. Results

The cohort size (N) necessary to achieve the proposed number of cases varied among scenarios. Specifically, latent classification methods typically required approximately 3.5 times the number of subjects to provide 500 subjects eligible for selection as cases compared to parallel and multinomial classification methods. The exception was scenarios where serum ELISA specificity increased to 100%, where N was 1.2 and 1.5 times greater in latent models associated with ELISA sensitivities of 15% and 25%, respectively.

Differences were observed between $\hat{\beta}_1$ for all methods of classification of infection status and analysis (Table 1). For the majority of scenarios and infection status classification approaches evaluated, the mean of $\hat{\beta}_1$ was biased toward the null (Table 1). The exception was Bayesian latent analyses which demonstrated substantial bias away from the null in all scenarios. Frequentist latent classifications and the *Culture Only* and *Culture and ELISA* classifications of multinomial models consistently had the smallest differences between the mean estimate of $\hat{\beta}_1$ and its true value. The Bayesian latent method consistently had the largest difference between the mean estimate of $\hat{\beta}_1$ and its true value. The magnitude of the bias was not impacted by the disease allele frequency over the range of values evaluated in parallel, frequentist latent, and multinomial models, but did increase as allele frequency increased in Bayesian latent models. The extent of bias decreased substantially in parallel, frequentist latent, and multinomial models for scenarios in which the specificity of the serum ELISA was increased to 100% (Fig. 1), and changes in sensitivity did not result in substantial changes in the magnitude of observed bias for these models. Bayesian latent models showed increases in bias with increased specificity, with modest increases in bias as sensitivity increased.

The proportion of 95% confidence intervals for parameter estimates that contained the true value was generally high (Table 2) and all Bayesian models contained the true value of e^{β_1} within 95% credible intervals. However, increases in frequency of the disease allele were associated with a decrease in this proportion for other models. These findings are consistent with observed changes in width of the 95% confidence interval which showed a trend toward narrower intervals as allele frequency and magnitude of effect increased (Table 3). Increasing the serum ELISA specificity to 100% was associated with decreases in width of the 95% confidence interval for all models except for the *ELISA Only* classification of the multinomial model. However, increases in specificity resulted in larger proportions of 95% confidence intervals containing the true value of e^{β_1} despite observed reduction in the mean interval width due to the concomitant decreases in observed bias in $\hat{\beta}_1$.

Please cite this article in press as: Osterstock, J.B., et al., Effect of classifying disease states in genetic association studies for paratuberculosis. PREVET (2010), doi:10.1016/j.prevetmed.2010.03.005

Table 1
Mean distance between effect estimate and the true value of β_1 (calculated as true-estimate) in the log-odds scale for parallel, latent, Bayesian, and multinomial models over a range of values for magnitude of the true value of e^{β_1} (OR), disease allele frequency (p), and sensitivity (Se) and specificity (Sp) of serum ELISA in simulated genetic association studies.

OR	p	ELISASe	ELISASp	Parallel	Latent		Multinomial		
					Frequentist	Bayesian	Culture only	ELISA only	Culture and ELISA
1.2	0.05	15%	97%	0.139	−0.012	−0.215	0.039	0.184	0.282
1.2	0.10	15%	97%	0.082	0.039	−0.133	0.002	0.118	0.069
1.2	0.20	15%	97%	0.137	−0.007	−0.192	0.037	0.176	0.076
1.5	0.05	15%	97%	0.221	0.055	−0.397	0.059	0.319	−0.066
1.5	0.10	15%	97%	0.223	0.018	−0.371	0.017	0.310	0.067
1.5	0.10	15%	100%	0.035	0.024	−0.922	0.034	0.053	0.033
1.5	0.10	25%	97%	0.213	0.038	−0.444	0.013	0.294	0.049
1.5	0.10	25%	100%	−0.027	−0.012	−0.944	0.005	−0.050	−0.016
1.5	0.20	15%	97%	0.281	0.043	−0.426	0.041	0.376	0.091
2.0	0.05	15%	97%	0.411	0.008	−0.540	0.052	0.589	0.178
2.0	0.10	15%	97%	0.414	0.078	−0.611	0.025	0.585	0.137
2.0	0.20	15%	97%	0.408	0.060	−0.571	0.017	0.572	0.006

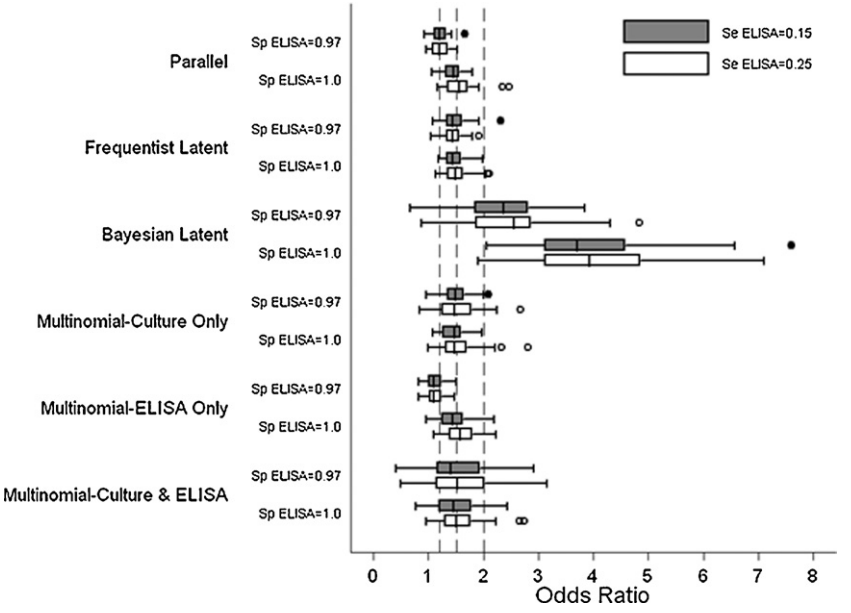

Fig. 1. Boxplot demonstrating the effect of imperfect specificity (Sp) of the serum ELISA on extent of bias in observed effect estimates relative to the true value of e^{β_1} (dashed line) by method employed for classifying infection status and analysis. Line within boxplot represents median effect estimate derived from 100 simulated case–control genetic association studies. Left and right boundaries of boxes represent 25th and 75th percentiles and whiskers represent minimum and maximum values within 1.5 times the interquartile range. Allele frequency was fixed at 0.1.

Table 2
Proportion of 95% confidence intervals (credible intervals for Bayesian methods) containing the true value for parallel, latent, Bayesian, and multinomial models over a range of values for magnitude of the true value of e^{β_1} (OR), disease allele frequency (p), and sensitivity (Se) and specificity (Sp) of serum ELISA in simulated genetic association studies.

OR	p	ELISASe	ELISASp	Parallel	Latent		Multinomial		
					Frequentist	Bayesian	Culture only	ELISA only	Culture and ELISA
1.2	0.05	15%	97%	86%	98%	100%	98%	82%	94%
1.2	0.10	15%	97%	88%	100%	100%	96%	86%	96%
1.2	0.20	15%	97%	76%	100%	100%	92%	72%	92%
1.5	0.05	15%	97%	84%	96%	100%	96%	72%	96%
1.5	0.10	15%	97%	68%	92%	100%	100%	48%	100%
1.5	0.10	15%	100%	96%	98%	100%	98%	92%	98%
1.5	0.10	25%	97%	68%	96%	100%	96%	50%	98%
1.5	0.10	25%	100%	96%	84%	100%	90%	96%	96%
1.5	0.20	15%	97%	28%	90%	100%	96%	16%	100%
2.0	0.05	15%	97%	42%	96%	100%	98%	18%	98%
2.0	0.10	15%	97%	22%	96%	100%	96%	10%	98%
2.0	0.20	15%	97%	6%	94%	100%	94%	0%	98%

Please cite this article in press as: Osterstock, J.B., et al., Effect of classifying disease states in genetic association studies for paratuberculosis. PREVET (2010), doi:10.1016/j.prevetmed.2010.03.005

Table 3
Mean width of 95% confidence intervals (credible intervals for Bayesian methods) for effect estimates in log-odds scale for parallel, latent, Bayesian, and multinomial models over a range of values for magnitude of the true value of e^{β_1} (OR), disease allele frequency (p), and sensitivity (Se) and specificity (Sp) of serum ELISA in simulated genetic association studies.

OR	p	ELISASe	ELISASp	Parallel	Latent		Multinomial		
					Frequentist	Bayesian	Culture only	ELISA only	Culture and ELISA
1.2	0.05	15%	97%	0.789	0.777	3.193	1.245	0.865	3.067
1.2	0.10	15%	97%	0.575	0.565	3.126	0.899	0.627	2.094
1.2	0.20	15%	97%	0.429	0.430	3.105	0.680	0.469	1.528
1.5	0.05	15%	97%	0.770	0.752	2.989	1.162	0.852	2.481
1.5	0.10	15%	97%	0.563	0.557	2.927	0.851	0.619	1.881
1.5	0.10	15%	100%	0.531	0.517	2.812	0.598	0.723	1.115
1.5	0.10	25%	97%	0.561	0.556	2.753	0.912	0.612	1.520
1.5	0.10	25%	100%	0.534	0.506	2.678	0.662	0.653	0.978
1.5	0.20	15%	97%	0.423	0.421	2.854	0.648	0.465	1.465
2.0	0.05	15%	97%	0.756	0.734	2.740	1.062	0.847	2.435
2.0	0.10	15%	97%	0.555	0.539	2.699	0.795	0.619	1.722
2.0	0.20	15%	97%	0.417	0.420	2.639	0.626	0.461	1.364

Increases in sensitivity of the serum ELISA did not substantially alter the proportion of 95% confidence intervals containing the true value or the width of these intervals.

The power to detect a significant effect of genotype on infection status at the 95% confidence level also varied among the scenarios tested (Table 4). Increases in statistical power were typically observed with increases in disease allele frequency and magnitude of effect. In addition, increasing specificity of the serum ELISA to 100% dramatically increased statistical power. For frequentist latent classifications, the observed increase in power associated with increases in serum ELISA specificity were comparatively modest, but improved nonetheless. Similarly, increases in power associated with increasing the sensitivity of the serum ELISA from 15% to 25% were generally modest for all models evaluated. The exception was a decrease in power for the *Culture Only* group in the multinomial model which suffered a small decrease in power with increasing sensitivity. Overall, the majority of the classifications evaluated failed to achieve the threshold of 80% power routinely used in sample size estimation, including sample size calculations employed here. The only scenarios where this threshold were met in the present study were parallel interpretations when specificity was set to 100%, frequentist latent analysis when the effect was consistent with an OR of 2, frequentist latent analysis with an OR of 1.5 if tests included an ELISA with improved sensitivity or specificity, and the *Culture Only* classification within the multinomial model with an OR of 2 and allele frequency $\geq$0.1.

4. Discussion

Several studies have been performed evaluating the associations between selected candidate genes and susceptibility to paratuberculosis in cattle. Significant associations have been reported for a SNP in CARD15 (Pinedo et al., 2009) and polymorphisms in TLR1, 2, and 4 (Mucha et al., 2009). Methods for classification of phenotype vary between these studies including parallel interpretation of diagnostic test results (Pinedo et al., 2009) and presence of characteristic clinical signs without diagnostic confirmation (Mucha et al., 2009). The results of this study help to identify the potential effects of different approaches to classifying disease phenotype based on diagnostic test data on statistical power to detect an association and biases in effect estimates.

The methods employed to classify infection status and to analyze obtained results may vary substantially between studies due to differences in disease expression of interest and preferences regarding statistical methodology. In this study, we focused on paratuberculosis infection as the outcome and employed a variety of statistical approaches to our analyses. For diseases like paratuberculosis that demonstrate prolonged latent stages of infection and varied clinical manifestations of disease, some researchers may choose to focus on other aspects of disease such as fecal shedding of MAP or age at onset of clinical disease. Alternative disease outcomes will require a different set of assumptions in model development, particularly with regard to the sensitivity and specificity of diagnostic tests and prevalence of the condition within the population. The effects of changes in sensitivity and particularly specificity on extent of bias in effect estimates, proportion of 95% confidence intervals containing the true value of β_1, and statistical power observed in this study demonstrate the importance of these assumptions. In addition, assumptions must also be made with respect to defining these terms. The frequentist latent classification scheme in this study utilized fixed estimates for sensitivity and specificity whereas the Bayesian approaches defined distributions around these estimates, thus allowing these values to vary during the iterative sampling procedures. The results obtained from frequentist latent and Bayesian classification methods varied substantially in all outcomes assessed here, with frequentist latent classifications performing better than Bayesian estimation with respect to the extent of bias in effect estimates and statistical power. This is likely due to the fact that the same point estimates for *Se* and *Sp* were used in generation of the sampling frame and subsequent analysis. The bias observed in the Bayesian latent models was the greatest among all methods evaluated. The source of this bias is in the specification of the likelihood function in the Bayesian framework. Analysis of the effect of the disease allele on disease status using the entire cohort (population) yielded unbiased esti-

Please cite this article in press as: Osterstock, J.B., et al., Effect of classifying disease states in genetic association studies for paratuberculosis. PREVET (2010), doi:10.1016/j.prevetmed.2010.03.005

Table 4
Estimated power to detect an association between presence of a disease allele and infection status for parallel, latent, Bayesian, and multinomial models over a range of values for magnitude of the true value of e^{β_1} (OR), disease allele frequency (p), and sensitivity (Se) and specificity (Sp) of serum ELISA in simulated genetic association studies. Power was estimated as the proportion of simulations for which the effect estimate was significant at the 95% confidence level.

OR	p	ELISASe	ELISASp	Parallel	Latent		Multinomial		
					Frequentist	Bayesian	Culture only	ELISA only	Culture and ELISA
1.2	0.05	15%	97%	0%	12%	2%	14%	2%	0%
1.2	0.10	15%	97%	8%	18%	0%	14%	8%	10%
1.2	0.20	15%	97%	2%	42%	4%	18%	0%	8%
1.5	0.05	15%	97%	8%	46%	2%	30%	6%	20%
1.5	0.10	15%	97%	20%	76%	4%	48%	6%	14%
1.5	0.10	15%	100%	82%	84%	52%	70%	50%	28%
1.5	0.10	25%	97%	26%	80%	16%	42%	10%	24%
1.5	0.10	25%	100%	86%	90%	70%	64%	78%	38%
1.5	0.20	15%	97%	22%	92%	30%	58%	4%	12%
2.0	0.05	15%	97%	24%	98%	6%	64%	10%	20%
2.0	0.10	15%	97%	46%	100%	36%	90%	16%	28%
2.0	0.20	15%	97%	74%	100%	66%	98%	22%	54%

mates when using Bayesian latent class models (results not shown). However, when limited to the case–control datasets drawn from the target population, the likelihood function no longer produces unbiased estimates. More specifically after incorporating specificity and sensitivity, the disease model becomes no longer a logistic model. As a result, the equivalence of retrospective and prospective sampling (Prentice and Pyke, 1979) does not hold regarding the estimation of the log-odds ratio parameter.

Alternative approaches could be used to reduce the extent of this bias including semiparametric methods (Scott and Wild, 1997) and approaches that rely on estimates of the distribution of the disease allele obtained in practice from external sources or previous reports, but are not necessarily applicable within the context of this simulation study.

For some diseases, including paratuberculosis, it is unlikely that alternative diagnostic tests could be developed that would result in a substantial change in test sensitivity or specificity. Strategic sampling could improve the performance of these tests such as limiting testing to older animals or those with characteristic clinical signs. However, this would necessitate redefining the intent of the study as targeting associations with manifestation of clinical signs consistent with infection rather than infection itself if progression of disease is associated with genetic variants that are inherited independent of variants associated with susceptibility to infection. Alternatively, covariates such as age could be included in the model to account for age-dependent differences in disease progression and test performance (Nielsen and Toft, 2006). Age was not considered in this study as a potential modifier of manifestation of disease, and therefore sensitivity and specificity, due the limited amount of information available to accurately define this parameter.

It is apparent from the results obtained in this study that imperfect specificity contributes significantly to observed bias in estimates of effect, proportion of 95% confidence intervals containing the true value of β_1, and statistical power. Of all classification methods evaluated here, the frequentist latent classification scheme appeared to be the most robust with respect to imperfect specificity. The effect of imperfect specificity is particularly important for diseases with low prevalence, estimated to be 5% in the present study, because of the large number of uninfected subjects that are susceptible to false-positive test results with the imperfect test. In fact, under the assumptions of sensitivity within this study, the proportion of subjects classified as cases that are infected is less than the proportion of subjects classified as cases that are uninfected. An important factor affecting the specificity of serum ELISA for paratuberculosis is potential exposure to non-MAP *Mycobacterium* spp. in the environment which may demonstrate substantial variability under differing management conditions and in different geographical regions (Osterstock et al., 2007; Roussel et al., 2007). Therefore, researchers designing genetic association studies for paratuberculosis should pay particular attention to the potential for regional or local variation in specificity of serum ELISA for paratuberculosis, and for the effects of imperfect specificity on study design, statistical power, and potential bias in estimates of effect.

The effects of imperfect specificity on the extent of bias in effect estimates is also demonstrated in the differences among groups within the multinomial classification methods. The *Culture Only* and *Culture and ELISA* groups both demonstrated smaller differences from the true value than the *ELISA Only* category. This can be directly attributed to the fact that misclassification of uninfected subjects as cases would not have occurred within these groups due to the assumptions regarding specificity of culture of feces for MAP. However, it should also be noted that in the multinomial model, bias should be aggregated over all classifications due to the fact that 3 parameters, one for each category, are used to estimate the same quantity ($\hat{\beta}_1$).

Differences in specificity also account for the observed differences in bias of effect estimates between frequentist and Bayesian latent classifications. The frequentist approach dichotomized outcomes using a threshold of 0.5 for the latent probability of infection. Given the sensitivity and specificity assumptions for serum ELISA and culture of feces for MAP, the only subjects that would have been eligible for consideration as cases under the frequentist classification scheme would have been those with a pos-

Please cite this article in press as: Osterstock, J.B., et al., Effect of classifying disease states in genetic association studies for paratuberculosis. PREVET (2010), doi:10.1016/j.prevetmed.2010.03.005

itive culture result. Therefore, misclassification of cases would not have occurred under the latent scheme whereas sampling in the Bayesian method could have resulted in misclassification of cases in the absence of a similarly applied threshold.

Some seemingly paradoxical changes were observed for some of the classification methods under differing scenarios. For example, increasing specificity decreased the width of 95% confidence intervals with the exception of the *ELISA Only* group in the multinomial models and increases in sensitivity improved power to detect an association except for the *Culture Only* group in the multinomial model. In the multinomial model, it is important to recognize that changes in test accuracy alter the distribution of cases within the 3 groups representing subjects with 1 or more positive test results. By increasing serum ELISA specificity to 100%, thereby eliminating cases sampled from uninfected subjects with false-positive serum ELISA results, the proportion of cases within the *Culture Only* group increases due to the higher sensitivity of bacterial culture of feces for MAP relative to serum ELISA. Therefore, the reciprocally smaller number of cases within the ELISA only group results in an increase in standard error and subsequently, an increase in width of the 95% confidence interval with increases in specificity. Similarly, the relative contribution of the *Culture Only* group to all cases in the multinomial model decreases when sensitivity of the ELISA increases, therefore negatively impacting statistical power.

Several important elements of genetic association studies were not addressed in the present study. The impact (or effect) of carrying 2 copies of the disease allele was assumed to yield a higher probability of disease equal to twice the magnitude of effect associated with carriage of a single allele consistent with an additive model of inheritance. Alternative modes of inheritance (e.g. dominant or recessive) were not examined. The additive mode was selected because well defined genetic associations and corresponding modes of inheritance are largely absent for bovine paratuberculosis, with additive modes evaluated within similar recent analyses reporting tentative genetic associations with paratuberculosis (Pinedo et al., 2009; Settles et al., 2009). It is likely that variation within several different bovine genes is associated with differential susceptibility to paratuberculosis and that the precise mode of inheritance may vary among them (Pinedo et al., 2009; Settles et al., 2009). Further study under alternative modes of inheritance is warranted to enable broader inference on the effects of method for classifying infection status on properties of statistical tests used to evaluate potential associations. In addition, disease penetrance associated with the disease allele was assumed to be 1 and the consequences of incomplete or partial penetrance were not evaluated. However, incomplete penetrance would be expected to yield results similar to those for scenarios in which a smaller OR was used to describe the magnitude of association between the disease allele and infection status for a given allele frequency.

The potential difference in genetic regulation of disease resistance associated with specific disease alleles and traits or diagnostic tests used to classify infection status was not evaluated in this study. For instance, genetic polymorphisms associated with cell-mediated immunity may be expected to have less influence on the variability in serum ELISA status than shedding of MAP in feces, and thus fecal culture outcomes and/or the results of cell-mediated immunity assays such as the interferon-gamma assay. This would be an important consideration for studies targeting polymorphisms in genes with known functions relative to MAP infection or immune response. The application of multinomial outcomes in these studies may be more appropriate to discriminate the associations for each type of test.

The results of this study provide important insights into the design of genetic association studies, including the number of subjects that need to be sampled to identify a suitable number of cases. Several approaches for defining the sampling frame could have been used to perform the simulations in this study. We elected to define a fixed number of cases and controls for all disease classification approaches derived on the basis of that specific classification method. Alternatively, the size of the sampling frame from which the cases and controls were derived could have been fixed. In the present study, the frequentist latent and Bayesian classification schemes required substantially more subjects within the sampling frame to achieve the targeted number of cases. This is due in large part to the impact of correction for imperfect test specificity in a population with low disease prevalence. For parallel and multinomial classifications, a substantial number of uninfected subjects with false-positive serum ELISA tests were eligible for inclusion as cases. Thus, the relative benefits of improved power and reduced bias that may be achieved with frequentist latent models by reducing the potential influence of cases derived from false-positive test results must be considered in light of the sampling inefficiency introduced by using latent classification methods. Selection of controls could also be altered to improve some of the statistical properties of the tests evaluated here. We elected to use 1-to-1 matching of cases-to-controls. A reasonable alternative would have been to select multiple controls per case which could be utilized as a means to improve statistical power.

5. Conclusion

In conclusion, the results of this study demonstrate some of the differences in observed estimates of effect and statistical power associated with different methods for classifying infection status for diseases that lack definitive diagnostic criteria, particularly in subclinical stages, such as paratuberculosis. This information is important for the design and interpretation of genetic association studies as it helps to guide sample size estimation and identify potential biases that may exist in association parameters. For instance, for a fixed or small sampling frame, frequentist latent classification of infection status may not be suitable because of the limited power observed when serum ELISA sensitivity was assumed to be low, despite the expectation that the estimates of effect would more closely approximate true values. Additional studies using known diagnostic test results for sample populations and observed allele frequencies would be beneficial to further

Please cite this article in press as: Osterstock, J.B., et al., Effect of classifying disease states in genetic association studies for paratuberculosis. PREVET (2010), doi:10.1016/j.prevetmed.2010.03.005

characterize the statistical properties of models based on these and other approaches to phenotypic classification.

References

Collins, M.T., Sockett, D.C., Goodger, W.J., Conrad, T.A., Thomas, C.B., Carr, D.J., 1994. Herd prevalence and eographic distribution of, and risk factors for, bovine paratuberculosis in Wisconsin. J. Am. Vet. Med. Assoc. 204, 636–641.

Collins, M.T., Wells, S.J., Petrini, K.R., Collins, J.E., Schultz, R.D., Whitlock, R.H., 2005. Evaluation of five antibody detection tests for diagnosis of bovine paratuberculosis. Clin. Diagn. Lab Immunol. 12, 685–692.

Dargatz, D.A., Byrum, B.A., Barber, L.K., Sweeney, R.W., Whitlock, R.H., Shulaw, W.P., Jacobson, R.H., Stabel, J.R., 2001. Evaluation of a commercial ELISA for diagnosis of paratuberculosis in cattle. J. Am. Vet. Med. Assoc. 218, 1163–1166.

Eamens, G.J., Whittington, R.J., Marsh, I.B., Turner, M.J., Saunders, V., Kemsley, P.D., Rayward, D., 2000. Comparative sensitivity of various faecal culture methods and ELISA in dairy cattle herds with endemic Johne's disease. Vet. Microbiol. 77, 357–367.

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.

Gonda, M.G., Chang, Y.M., Shook, G.E., Collins, M.T., Kirkpatrick, B.W., 2006. Genetic variation of *Mycobacterium avium* ssp. *paratuberculosis* infection in US Holsteins. J. Dairy Sci. 89, 1804–1812.

Koets, A.P., Adugna, G., Janss, L.L., van Weering, H.J., Kalis, C.H., Wentink, G.H., Rutten, V.P., Schukken, Y.H., 2000. Genetic variation of susceptibility to *Mycobacterium avium* subsp. *paratuberculosis* infection in dairy cattle. J. Dairy Sci. 83, 2702–2708.

Larsen, A.B., Merkal, R.S., Cutlip, R.C., 1975. Age of cattle as related to resistance to infection with *Mycobacterium paratuberculosis*. Am. J. Vet. Res. 36, 255–257.

Liu, J., Pei, Y., Papasian, C.J., Deng, H.W., 2009. Bivariate association analyses for the mixture of continuous and binary traits with the use of extended generalized estimating equations. Genet. Epidemiol. 33, 217–227.

McKenna, S.L., Keefe, G.P., Barkema, H.W., Sockett, D.C., 2005. Evaluation of three ELISAs for *Mycobacterium avium* subsp. *paratuberculosis* using tissue and fecal culture as comparison standards. Vet. Microbiol. 110, 105–111.

Mortensen, H., Nielsen, S.S., Berg, P., 2004. Genetic variation and heritability of the antibody response to *Mycobacterium avium* subspecies *paratuberculosis* in Danish Holstein cows. J. Dairy Sci. 87, 2108–2113.

Mucha, R., Bhide, M.R., Chakurkar, E.B., Novak, M., Mikula Sr., I., 2009. Toll-like receptors TLR1, TLR2 and TLR4 gene mutations and natural resistance to *Mycobacterium avium* subsp. *paratuberculosis* infection in cattle. Vet. Immunol. Immunopathol. 128, 381–388.

NAHMS, 1997. Johne's disease on U.S. dairy operations. USDA:APHIS:VS, CEAH, National Animal Health Monitoring System, Fort Collins, CO #N245.1097.

Nielsen, S.S., Ersboll, A.K., 2006. Age at occurrence of *Mycobacterium avium* subspecies *paratuberculosis* in naturally infected dairy cows. J. Dairy Sci. 89, 4557–4566.

Nielsen, S.S., Toft, N., 2006. Age-specific characteristics of ELISA and fecal culture for purpose-specific testing for paratuberculosis. J. Dairy Sci. 89, 569–579.

Nielsen, S.S., Toft, N., 2008. Ante mortem diagnosis of paratuberculosis: a review of accuracies of ELISA, interferon-gamma assay and faecal culture techniques. Vet. Microbiol. 129, 217–235.

Osterstock, J.B., Fosgate, G.T., Norby, B., Manning, E.J., Collins, M.T., Roussel, A.J., 2007. Contribution of environmental mycobacteria to false-positive serum ELISA results for paratuberculosis. J. Am. Vet. Med. Assoc. 230, 896–901.

Ott, S.L., Wells, S.J., Wagner, B.A., 1999. Herd-level economic losses associated with Johne's disease on US dairy operations. Prev. Vet. Med. 40, 179–192.

Pinedo, P.J., Buergelt, C.D., Donovan, G.A., Melendez, P., Morel, L., Wu, R., Langaee, T.Y., Rae, D.O., 2009. Association between CARD15/NOD2 gene polymorphisms and paratuberculosis infection in cattle. Vet. Microbiol. 134, 346–352.

Prentice, R.L., Pyke, R., 1979. Logistic disease incidence models and case–control studies. Biometrika 66, 403–412.

Reddacliff, L.A., Beh, K., McGregor, H., Whittington, R.J., 2005. A preliminary study of possible genetic influences on the susceptibility of sheep to Johne's disease. Aust. Vet. J. 83, 435–441.

Roussel, A.J., Fosgate, G.T., Manning, E.J., Collins, M.T., 2007. Association of fecal shedding of mycobacteria with high ELISA-determined seroprevalence for paratuberculosis in beef herds. J. Am. Vet. Med. Assoc. 230, 890–895.

Scott, H.M., Fosgate, G.T., Libal, M.C., Sneed, L.W., Erol, E., Angulo, A.B., Jordan, E.R., 2007. Field testing of an enhanced direct-fecal polymerase chain reaction procedure, bacterial culture of feces, and a serum enzyme-linked immunosorbent assay for detecting *Mycobacterium avium* subsp. *paratuberculosis* infection in adult dairy cattle. Am. J. Vet. Res. 68, 236–245.

Scott, A.J., Wild, C.J., 1997. Fitting regression models to case–control data by maximum likelihood. Biometrika 84, 57–71.

Settles, M., Zanella, R., McKay, S.D., Schnabel, R.D., Taylor, J.F., Whitlock, R., Schukken, Y., Van Kessel, J.S., Smith, J.M., Neibergs, H., 2009. A whole genome association analysis identifies loci associated with *Mycobacterium avium* subsp. *paratuberculosis* infection status in US Holstein cattle. Anim. Genet. 40, 655–662.

Stabel, J.R., 1998. Johne's disease: a hidden threat. J. Dairy Sci. 81, 283–288.

van Schaik, G., Schukken, Y.H., Crainiceanu, C., Muskens, J., VanLeeuwen, J.A., 2003. Prevalence estimates for paratuberculosis adjusted for test variability using Bayesian analysis. Prev. Vet. Med. 60, 281–295.

Please cite this article in press as: Osterstock, J.B., et al., Effect of classifying disease states in genetic association studies for paratuberculosis. PREVET (2010), doi:10.1016/j.prevetmed.2010.03.005