Preventive Veterinary Medicine 150 (2018) 162–167

Contents lists available at ScienceDirect

Preventive Veterinary Medicine

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

Diagnosing intramammary infection: Controlling misclassification bias in longitudinal udder health studies

Denis Haine[a,d,*], Ian Dohoo[b,d], Daniel Scholl[c], Simon Dufour[a,d]

[a] Faculté de médecine vétérinaire, Université de Montréal, 3200 rue Sicotte, St-Hyacinthe, Québec, Canada J2S 2M2
[b] Atlantic Veterinary College, Centre for Veterinary Epidemiological Research, University of Prince Edward Island, 550 University Avenue, Charlottetown, Prince Edward Island, Canada C1A 4P3
[c] College of Agriculture and Biological Sciences, South Dakota State University, Box 2207, South Dakota Brookings, SD 57007, USA
[d] Canadian Bovine Mastitis and Milk Quality Research Network, St-Hyacinthe, Québec, Canada J2S 2M2

ARTICLE INFO

Keywords:
Misclassification
Dairy cow
mastitis
Bias
Epidemiologic methods

ABSTRACT

Using imperfect tests may lead to biased estimates of disease frequency and of associations between risk factors and disease. For instance in longitudinal udder health studies, both quarters at risk and incident intramammary infections (IMI) can be wrongly identified, resulting in selection and misclassification bias, respectively. Diagnostic accuracy can possibly be improved by using duplicate or triplicate samples for identifying quarters at risk and, subsequently, incident IMI.

The objectives of this study were to evaluate the relative impact of selection and misclassification biases resulting from IMI misclassification on measures of disease frequency (incidence) and of association with hypothetical exposures. The effect of improving the sampling strategy by collecting duplicate or triplicate samples at first or second sampling was also assessed.

Data sets from a hypothetical cohort study were simulated and analyzed based on a separate scenario for two common mastitis pathogens representing two distinct prevailing patterns. *Staphylococcus aureus*, a relatively uncommon pathogen with a low incidence, is identified with excellent sensitivity and almost perfect specificity. Coagulase negative staphylococci (CNS) are more prevalent, with a high incidence, and with milk bacteriological culture having fair Se but excellent Sp. The generated data sets for each scenario were emulating a longitudinal cohort study with two milk samples collected one month apart from each quarter of a random sample of 30 cows/herd, from 100 herds, with a herd-level exposure having a known strength of association. Incidence of IMI and measure of association with exposure (odds ratio; OR) were estimated using Markov Chain Monte Carlo (MCMC) for each data set and using different sampling strategies (single, duplicate, triplicate samples with series or parallel interpretation) for identifying quarters at risk and incident IMI.

For *S. aureus* biases were small with an observed incidence of 0.29 versus a true incidence of 0.25 IMI/100 quarter-month. In the CNS scenario, diagnostic errors in the two samples led to important selection (40 IMI/100 quarter-month) and misclassification (23 IMI/100 quarter-month) biases for estimation of IMI incidence, respectively. These biases were in opposite direction and therefore the incidence measure obtained using single sampling on both the first and second test (29 IMI/100 quarter-month) was exactly the true value.

In the *S. aureus* scenario the OR for association with exposure showed little bias (observed OR of 3.1 versus true OR of 3.2). The CNS scenario revealed the presence of a large misclassification bias moving the association towards the null value (OR of 1.7 versus true OR of 2.6). Little improvement could be brought using different sampling strategies aiming at improving Se and/or Sp on first and/or second sampling or using a two out of three interpretation for IMI definition.

Increasing number of samples or tests can prevent bias in some situations but efforts can be spared by holding to a single sampling approach in others. When designing longitudinal studies, evaluating potential biases and best sampling strategy is as critical as the choice of test.

* Corresponding author at: Faculté de médecine vétérinaire, Université de Montréal, 3200 rue Sicotte, St-Hyacinthe, Québec, Canada J2S 2M2.
E-mail address: denis.haine@umontreal.ca (D. Haine).

https://doi.org/10.1016/j.prevetmed.2017.11.010
Received 31 January 2017; Received in revised form 31 October 2017; Accepted 9 November 2017

1. Introduction

A cohort study is the standard method to estimate the incidence of diseases and identify their natural history, by analyzing the association between a baseline exposure and risk of disease over the follow-up period. A disease-free population is identified, i.e. subjects with the outcome at baseline are excluded from the follow-up, while new, incident cases of exposure are identified. Therefore, it is assumed that prevalent and non-prevalent cases can be differentiated with no error so that only susceptible individuals are included in the follow-up. Incident cases are likewise supposed to be correctly identified.

However, using imperfect tests may lead to biased estimates of disease frequency and of association with exposure. For instance in longitudinal udder health studies in which bacteriological culture is commonly used for diagnosis of intramammary infections (IMI), both quarters at risk of becoming infected and, later on, incident IMI can be wrongly identified. If the wrong classification at baseline and at follow-up are both misclassification biases, in the former the bias resulting from IMI misclassification could be considered a selection bias, as the wrong (diseased) subjects are included in the cohort (Rothman et al., 2012) while in the latter, it would be commonly defined as misclassification bias (Delgado-Rodriguez and Llorca, 2004). Different methods can be used to limit or address these biases, such as study design, improving the diagnostic procedures for identifying quarters at risk and incident IMI, or addressing the biases analytically (McInturff et al., 2004; Dufour et al., 2012a). Improvement of the diagnostic procedures is commonly achieved in udder health studies by using duplicate or triplicate samples in order to improve the sensitivity (Se) and/or specificity (Sp) of the bacteriological cultures. With series interpretation, animals are declared positive only if they test positive to both tests, resulting in an increased Sp but decreased Se. With parallel interpretation, a positive result in either test is sufficient to declare the animal positive, increasing Se but decreasing Sp. Regarding test characteristics for IMI diagnostic, Dohoo et al. (2011a) reported that series interpretation of duplicate samples provided the highest Sp but lowest Se, whereas parallel interpretation of duplicate samples resulted in the highest Se but lowest Sp. If triplicate samples provided the best combination of Se and Sp compared with a single sample, the gain in Sp was modest and there were little or no gain in Se.

Little is known about the relative impact in cohort studies of the selection bias resulting from misidentification of quarters at risk of becoming infected compared to the more traditional misclassification bias. Furthermore, the impact on measures of association of any reduction of this selection bias using more accurate diagnostic procedures is unknown. With the often limited resources available for milk samples analyses, a better understanding of the relative impact of these biases would allow a more appropriate distribution of these resources in a manner that would optimize the balance between cost and precision of a study.

The objectives of this study were to evaluate the relative impact of selection and misclassification biases resulting from IMI misclassification on measures of disease frequency (incidence) and of association with a hypothetical exposure. The effect of improving the sampling strategy was also assessed.

2. Materials and methods

Data sets from a hypothetical cohort study were simulated and analyzed based on a separate scenario for two common mastitis pathogens studied in udder health researches and representing two distinct prevailing patterns, *Staphylococcus aureus* and coagulase negative staphylococci (CNS). *S. aureus*, a relatively uncommon pathogen (prevalence $< 5\%$) with a low incidence (1 NIMI/100 quarter-month), can be identified with excellent Se ($\sim 90\%$) and almost perfect Sp ($> 99\%$, at 100 CFU/ml) by bacteriological culture (Zadoks et al., 2001; Dohoo et al., 2011b; Dufour et al., 2012b). Coagulase negative staphylococci are more prevalent (10–30%), with a high incidence (~ 30 NIMI/100 quarter-month), and with milk bacteriological culture having a fair Se ($\sim 60\%$) but an excellent Sp (95%, at 200 CFU/ml; Dohoo et al., 2011b; Dufour et al., 2012a).

The generated data sets for each scenario were emulating a longitudinal cohort study with two milk samples collected one month apart from each quarter of a random sample of 30 cows/herd, from 100 herds, with a herd-level exposure having a known strength of association. The true IMI status (S_1) on first milk sample collection was used to identify quarters at risk of IMI at the beginning of the cohort, while the second (S_2) was used to identify the true outcome (acquisition of a new IMI). A hypothetical exposure at the herd-level with known strength of association (OR $\sim$ 3.0) was generated at baseline (S_1). To make the scenario more realistic, exposure was equally associated with odds of a prevalent IMI on the first milk sample as with odds of IMI acquisition on the second sample (as observed in Dufour et al., 2012a). Exposure was randomly associated with the odds of eliminating an existing IMI. Correlation of these two specific types of IMI by cow and by herd were obtained from Dufour et al. (2012a,b) to produce realistic datasets. As demonstrated in Dufour et al. (2012a,b), IMI incidence has a much greater effect on IMI prevalence than the elimination rate. The parameters to generate these data sets are given in Table 1. For each scenario, 100 data sets were generated.

Sensitivity and Sp to diagnose *S. aureus* and CNS were represented as Beta distributions with the following shape parameters: (46.8, 6.09) and (45, 30.3) for Se; and (1, i.e. uniform distribution) and (4.26, 1.17) for Sp, for *S. aureus* and CNS, respectively. Analyses for each of the two scenarios were conducted separately. On each datasets new S_1' and S_2' variables were generated by applying the scenario misclassification parameters to the S_1 and S_2 samples. Incidence and measures of association with the hypothetical exposure were computed using first the S_1' and S_2' variables (total bias), then S_1' and S_2 (selection bias only), and finally the S_1 and S_2' variables (misclassification bias only). If the selection and misclassification biases were deemed important, the effect of improving Se and Sp on the first and/or second sampling(s) was assessed by applying different sampling strategies having the objective to improve Se and/or Sp. This is commonly achieved in udder health studies by carrying on duplicate or triplicate samplings with parallel or

Table 1
Parameters used to generate the simulated data sets (CNS: coagulase negative staphylococci).

Parameters	*S. aureus*	CNS
Exposure distribution (0–1) of the binary herd-level predictor	0.5	0.5
Exposure distribution (0–1) of the binary cow-level predictor	0.5	0.5
Exposure distribution (0–1) of the binary quarter-level predictor	0.5	0.5
Herd-level variance for prevalence of intramammary infection (IMI)	0.14	0.363
Cow-level variance for prevalence of IMI	2.25	0.294
Intercept for IMI prevalence; aiming at a prevalence of 2.5%	−6.7	
Intercept for IMI prevalence; aiming at a prevalence of 40%		−2.15
odds ratio (OR) of association between herd-level variable and IMI prevalence	3	3
Herd-level variance for incidence of IMI	0.838	0.27
Cow-level variance for incidence of IMI	2.926	0.256
Intercept for IMI incidence, aiming at an incidence of 1 IMI/100 quarter-month	−8.3	
Intercept for IMI incidence, aiming at an incidence of 30 IMI/100 quarter-month		−2.4
OR of association between herd-level variable and IMI incidence	3	3
Herd-level variance for elimination of IMI	0.15	0.112
Cow-level variance for elimination of IMI	2.246	0.7
Intercept for IMI persistency, aiming at 61 IMI/100 quarter-month	−0.6	
Intercept for IMI persistency, aiming at 21 IMI/100 quarter-month		1.6

Table 2
Gain or loss in sensitivity (Se) and specificity (Sp) using different hypothetical sampling strategies and interpretations (values added/subtracted to/from the original distributions for Se and Sp).

	Duplicate samples, series interpretation		Duplicate samples, parallel interpretation		Triplicate samples, out of 3 interpretation	
	Se	Sp	Se	Sp	Se	Sp
S. aureus	−0.10	0	+0.10	0	0	0
CNS	−0.25	+0.05	+0.15	−0.05	0	+0.10

Table 3
Existence of bias for measure of intramammary infection incidence (IMI/100 quarter-month; median and 95% credible interval; 1 month interval between baseline and follow-up).

	S. aureus	CNS
True incidence	0.25 [0.09–0.53]	28.94 [25.48–32.72]
Single sample, total bias	0.29 [0.12–0.63]	28.57 [24.38–32.71]
Single sample, selection bias only	0.33 [0.13–0.66]	40.69 [36.43–45.07]
Single sample, misclassification bias only	0.22 [0.08–0.49]	22.59 [18.68–26.99]

series interpretations (Table 2):

- duplicate samples—parallel interpretation on S_1 and S_2;
- duplicate samples—parallel interpretation on S_1 and single sample on S_2;
- duplicate samples—single sample on S_1, parallel interpretation on S_2;
- duplicate samples—series interpretation on S_1 and S_2;
- duplicate samples—series interpretation on S_1, single sample on S_2;
- duplicate samples—single sample on S_1, series interpretation on S_2;
- duplicate samples—parallel interpretation on S_1 and series interpretation on S_2;
- duplicate samples—series interpretation on S_1 and parallel interpretation on S_2; and
- triplicate samples with a two out of three interpretation (on S_1 and S_2).

Changes in Se and Sp resulting from these various sampling strategies investigated were based on Dohoo et al. (2011b) and educated knowledge (Table 2). The gains or losses in Se/Sp were added/subtracted to/from the original distributions for Se and Sp.

Poisson and logistic 3-level (quarters within cows within herds) regression models were used to estimate the cluster-specific incidence of IMI and measure of association with exposure (odds ratio; OR), respectively, using Markov Chain Monte Carlo (MCMC) implemented using the Stan modelling language (Carpenter et al., 2017) through the rstan (Stan Development Team, 2016) interface to R (Core Team, 2015). Herd- and cow-specific random effects and standard deviations prior distributions for both model were specified as a normal distribution (`Normal(0,1)`) and a Student-*t* distribution (`studentT(3,0,10)`), respectively. Each MCMC sample used four sampling chains with 100 burn-in samples followed by 400 monitored samples. Models were run under the Amazon EC2 cloud-computing

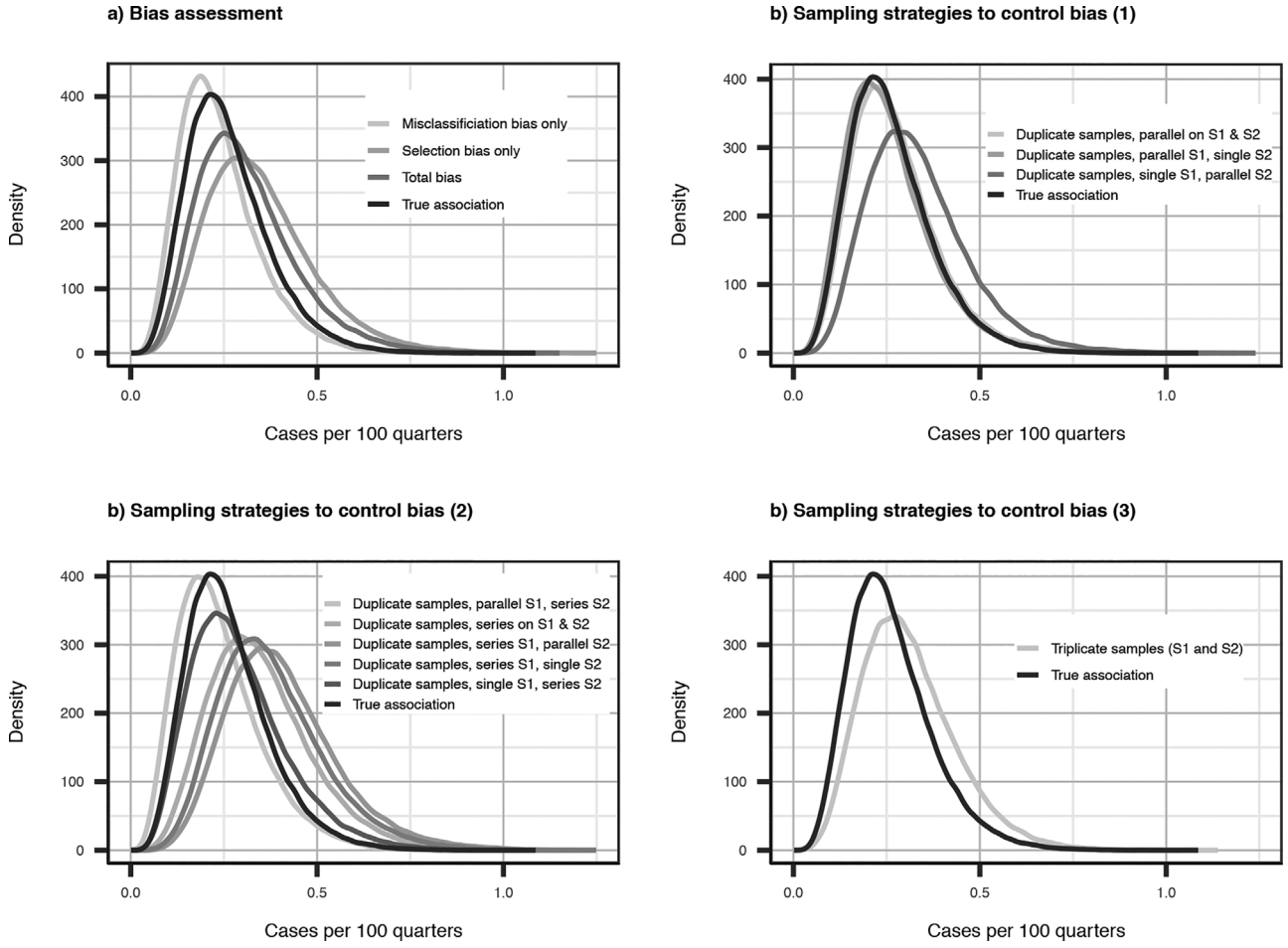

Fig. 1. Bias assessment (a) and effect of various sampling strategies (b) on measure of intramammary infection incidence—*Staphylococcus aureus* (1 month interval between baseline and follow-up).

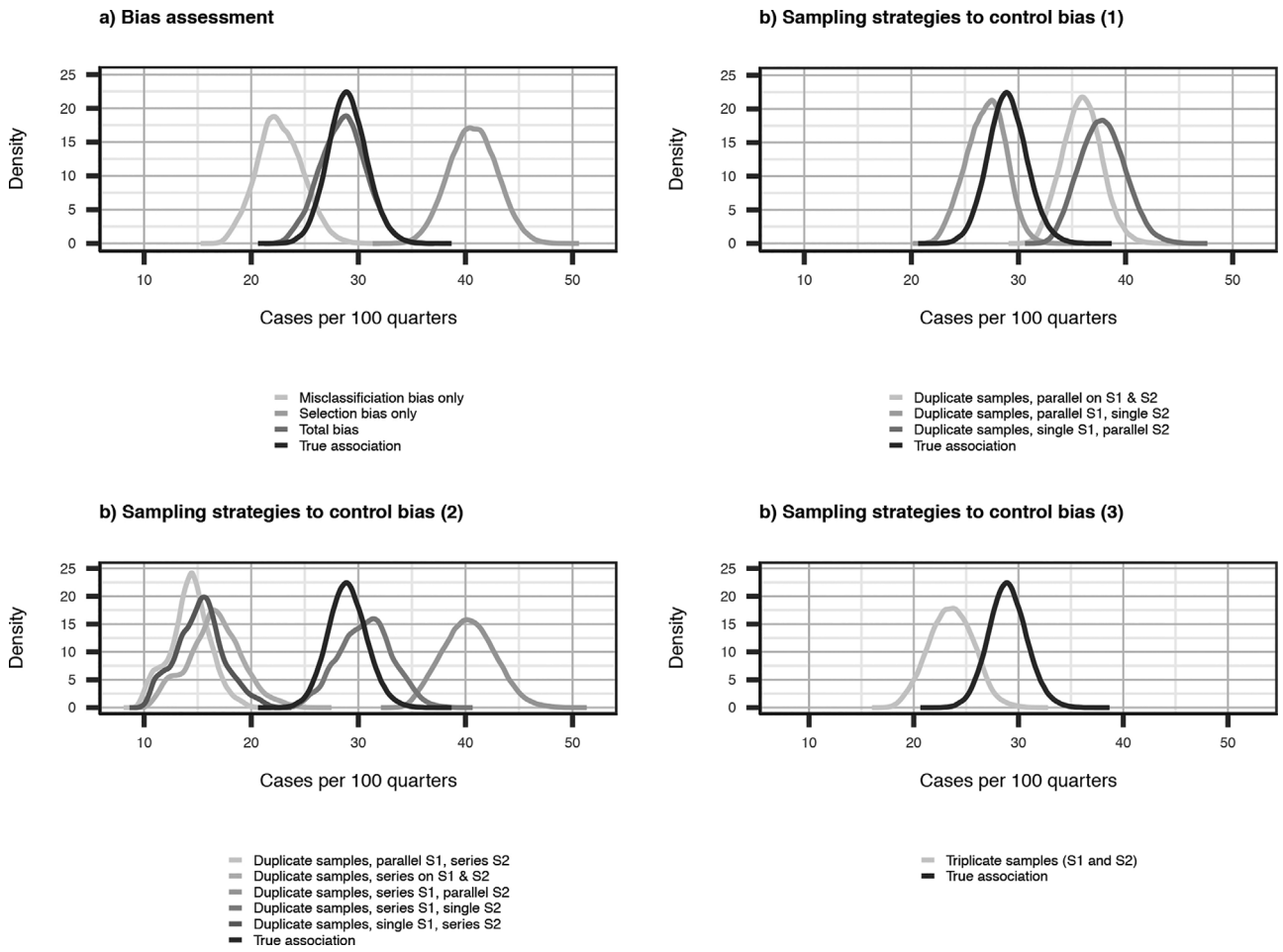

Fig. 2. Bias assessment (a) and effect of various sampling strategies (b) on measure of intramammary infection incidence—CNS (1 month interval between baseline and follow-up).

Table 4
Effect of sampling strategy on estimated measure of intramammary infection incidence (IMI/100 quarter/month; median and 95% credible interval. S_1: first sample; S_2: second sample; 1 month interval between the two).

	S. aureus	CNS
True incidence	0.25 [0.09–0.53]	28.94 [25.48–32.72]
Duplicate samples, parallel interpretation on S_1 and S_2	0.26 [0.10–0.56]	35.93 [32.50–39.62]
Duplicate samples, parallel on S_1 and single on S_2	0.24 [0.09–0.53]	27.05 [23.35–30.38]
Duplicate samples, parallel on S_2, single on S_1	0.32 [0.13–0.66]	37.81 [33.97–42.13]
Duplicate samples, series on S_1 and S_2	0.33 [0.13–0.66]	16.61 [11.62–21.81]
Duplicate samples, series on S_1, single on S_2	0.36 [0.16–0.71]	30.72 [25.78–35.43]
Duplicate samples, series on S_2, single on S_1	0.27 [0.10–0.60]	15.25 [10.94–19.96]
Duplicate samples, parallel on S_1 and series on S_2	0.22 [0.07–0.51]	14.34 [10.36–18.05]
Duplicate samples, series on S_1 and parallel on S_2	0.39 [0.17–0.76]	40.27 [35.75–45.25]
Triplicate samples (on S_1 and S_2)	0.29 [0.12–0.60]	23.58 [19.51–27.73]

environment (one node with a quad-core Intel(R) Xeon(R) CPU E5-2670 v2 and Ubuntu Server 14.04 LTS 64-bit operating system). Data sets generation and estimation procedures were compiled into an R package available at https://github.com/dhaine/misclass.

Table 5
Existence of bias for measure of association between exposure and probability of incident intramammary infection (odds ratio; median and 95% credible interval; 1 month interval between baseline and follow-up).

	S. aureus	CNS
True association	3.18 [1.31–8.55]	2.58 [1.89–3.55]
Single sample, total bias	3.11 [1.40–7.75]	1.72 [1.44–2.10]
Single sample, selection bias only	3.16 [1.44–7.61]	2.66 [2.10–3.41]
Single sample, misclassification bias only	3.14 [1.26–8.73]	1.70 [1.34–2.14]

3. Results

3.1. Incidence

For *S. aureus* biases were small with an observed incidence of 0.29 versus a true incidence of 0.25 IMI/100 quarter-month (Table 3, Figs. 1a and 2a). In the CNS scenario, diagnostic errors in the two samples led to important selection (40 IMI/100 quarter-month) and misclassification (23 IMI/100 quarter-month) biases for estimation of IMI incidence, respectively. These biases were in opposite direction and therefore the incidence measure obtained using single sampling on both the first and second test (29 IMI/100 quarter-month) was exactly the true value and no specific sampling strategies had to be considered when estimating CNS IMI incidence (Table 4, Figs. 1b and 2b).

3.2. Association

In the *S. aureus* scenario the OR for association with exposure

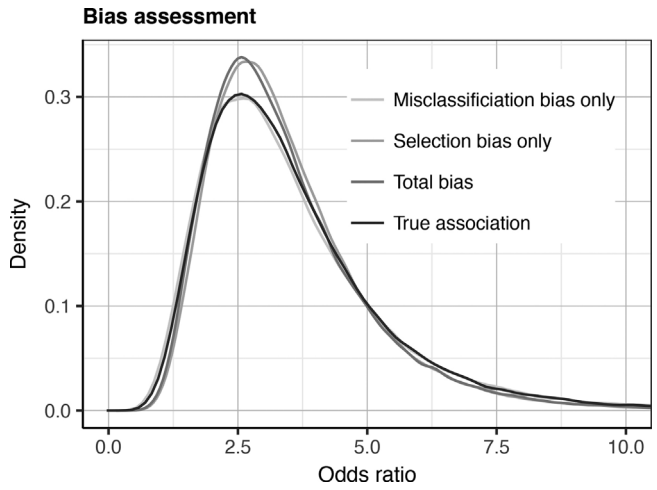

Fig. 3. Bias assessment for measure of intramammary infection association—*Staphylococcus aureus* (1 month interval between baseline and follow-up).

showed little bias (observed OR of 3.1 versus true OR of 3.2) (Table 5 and Fig. 3). Therefore no particular sampling strategy was considered. The CNS scenario, however, revealed the presence of a large misclassification bias moving the association towards the null value (OR of 1.7 versus true OR of 2.6) (Fig. 4a). Little improvement could be brought using different sampling strategies aiming at improving Se and/or Sp on first and/or second sampling and using a two out of three interpretation for IMI definition (Table 6 and Fig. 4b). This last strategy only moved the measure of association to an OR of 2.0.

Table 6
Effect of sampling strategy on measure of association between exposure and probability of incident intramammary infection (odds ratio; median and 95% credible interval. S_1: first sample; S_2: second sample; 1 month interval between the two).

	CNS
True association	2.58 [1.89–3.55]
Duplicate samples, parallel interpretation on S_1 and S_2	1.75 [1.46–2.19]
Duplicate samples, parallel on S_1 and single on S_2	1.73 [1.42–2.15]
Duplicate samples, parallel on S_2, single on S_1	1.76 [1.47–2.16]
Duplicate samples, series on S_1 and S_2	1.61 [1.30–1.96]
Duplicate samples, series on S_1, single on S_2	1.69 [1.42–2.05]
Duplicate samples, series on S_2, single on S_1	1.68 [1.34–2.08]
Duplicate samples, parallel on S_1 and series on S_2	1.68 [1.32–2.12]
Duplicate samples, series on S_1 and parallel on S_2	1.74 [1.46–2.15]
Triplicate samples (on S_1 and S_2)	2.04 [1.66–2.54]

4. Discussion

With the scenarios studied, our results indicated that selection and misclassification biases of a low prevalent and incident disease, diagnosed with high Se and most importantly with close to perfect Sp, are minimal and do not require specific sampling strategies to improve unit at risk nor case identification. On the other hand, when investigating a highly prevalent and incident disease, diagnosed with an average Se and high Sp, a bias toward the null would be observed for measure of association with exposure and this bias could not be controlled by modulating the sampling strategy.

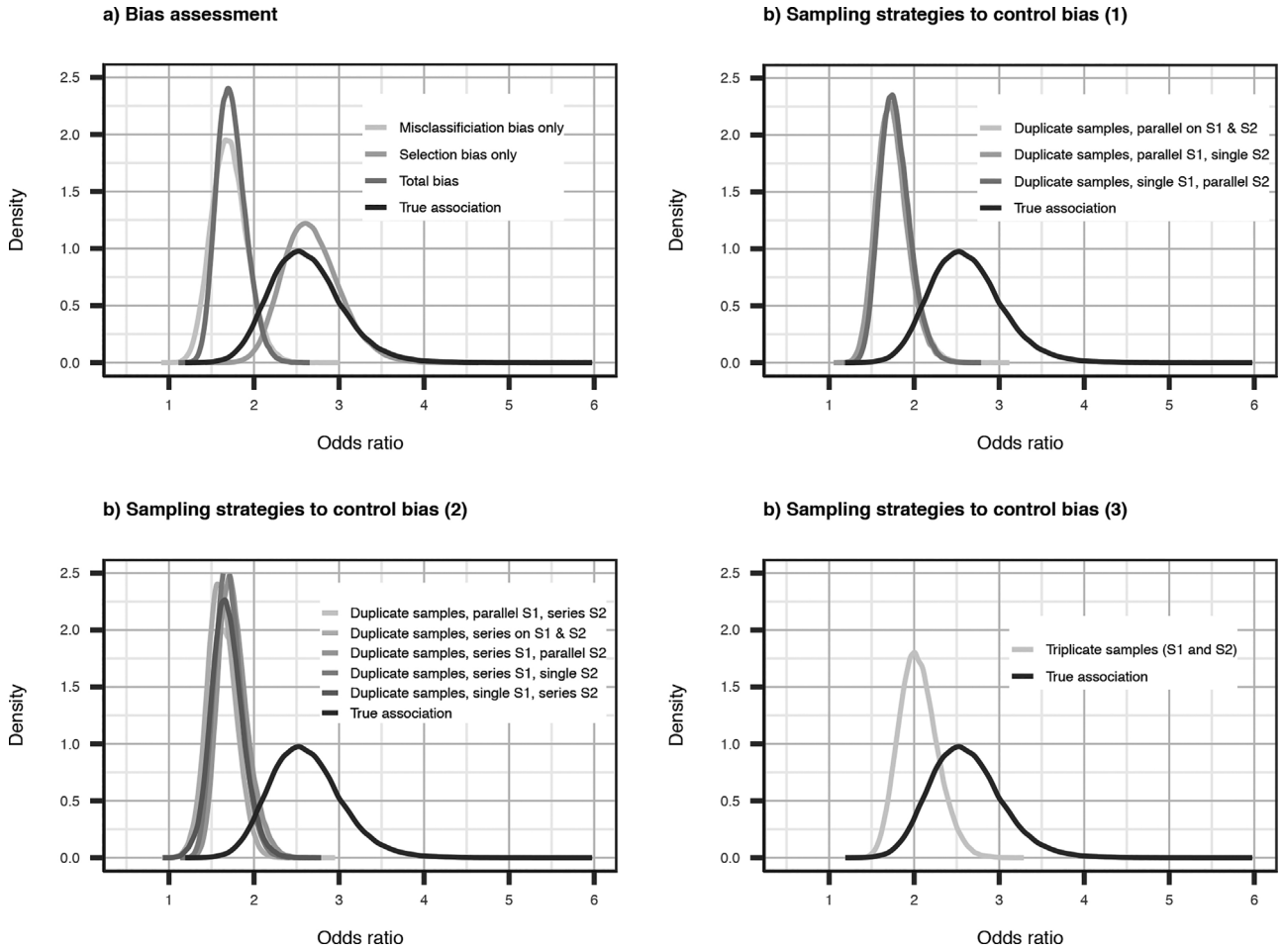

Fig. 4. Bias assessment (a) and effect of various sampling strategies (b) on measure of intramammary infection association—CNS (1 month interval between baseline and follow-up).

Many studies have looked into the effect of misclassification into statistical inferences, including biased prevalence and incidence rate estimates (Rogan and Gladen, 1978; Quade et al., 1980) and biased relative risk estimates (Barron, 1977; Greenland, 1980). Nondifferential misclassification of disease leads in general to bias toward null in the estimated associations as well as reduced statistical efficiency (Bross, 1954; Barron, 1977; Copeland et al., 1977). This bias depends mainly on the Sp of the test used for rare diseases (Copeland et al., 1977). If Sp of the test is perfect, then bias is absent (Poole, 1985). The effect of misclassification of disease at baseline to calculate incidence has less frequently been considered. In longitudinal studies, the nondifferential misclassification of disease at baseline, especially imperfect sensitivity, can lead to over- or under-estimation of the observed cumulative incidence risk ratios (Pekkanen et al., 2006). This bias can be significant for disease with a low true incidence, a high true prevalence, a substantial disease duration (i.e. as long as the interval between first and second test), and a poor test Se. To minimize bias, disease subjects at baseline should be excluded from the cohort based on a highly sensitive test (Pekkanen and Sunyer, 2008). Case identification during the follow-up period should use a highly specific test having a high positive predictive value (Brenner and Gefeller, 1993). While these recommendations are still valid and should be followed, we have shown that a more prevalent and incident disease diagnosed with an imperfect sensitivity and/or specificity will give biased measure of association despite attempts to improve its diagnosis.

The direction and magnitude of biases affecting incidence and association measures are hardly predictable. We provide an R package (https://github.com/dhaine/misclass) that allow the udder health researcher to estimate the potential biases that would be present in his/her study according to pathogen, population, and test characteristics. Improvement in Se and/or Sp of the tests can also be assessed with the package to the help the design and planning of the study. Therefore, researchers can evaluate the impact of adding additional samplings with various tests interpretations or of using a different test with different accuracy. As we have shown here with CNS, testing more or using more expensive tests may be of little or no use in terms of reducing bias. It is then necessary to correct the bias into the analytic stage, for instance by incorporating the Se and Sp of the test in the modelling strategy (Magder and Hugues, 1997). The uncertainty in the estimates can be included in a Bayesian analysis in the form of prior distributions (McInturff et al., 2004). A latent class model (Hui and Walter, 1980) would therefore return the posterior inference on regression parameters and the Se and Sp of both tests.

In this study we are reporting cluster-specific estimates and not the marginal, population-average ones, which would be smaller and less spread out, without suggesting any bias. However, for population-average parameters to be more than 10% lower than the cluster-specific parameters, cluster variance would have to be greater than 0.68 (Dohoo et al., 2009), which would only be the case for *S. aureus*. As the purpose was on the comparisons between the various sampling strategies and not on inference of results about a population, cluster-specific estimates were used to save computing time.

5. Conclusion

Increasing number of samples or tests can prevent bias in some situations, but efforts can be spared by holding to a single sampling approach in others. When designing longitudinal studies, evaluating potential biases and best sampling strategy is as critical as the choice of test. An R package was developed for such appraisal. Correcting remaining biases using analytical methods can complement choice of a good sampling strategy.

Conflict of interest

None.

Acknowledgements

We thank Pr. Henrik Stryhn for helpful comments during the preparation of this manuscript and for statistical guidance in simulating data.

This research was financed by the senior author (SD) Natural Sciences and Engineering Research Council of Canada Discovery Grant.

References

Barron, B.A., 1977. The effects of misclassification on the estimation of relative risk. Biometrics 33 (2), 414–418. http://dx.doi.org/10.1002/0471667196.ess0146.pub2.

Brenner, H., Gefeller, O., 1993. Use of the positive predictive value to correct for disease misclassification in epidemiological studies. Am. J. Epidemiol. 138, 1007–1015.

Bross, I., 1954. Misclassification in 2 × 2 tables. Biometrics 10 (4), 478–486.

Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M.A., Li, P., Riddell, A., 2017. Stan: a probabilistic programming language. J. Stat. Softw. 76, 1–32.

Copeland, K.T., Checkoway, H., McMichael, A.J., Holbrook, R.H., 1977. Bias due to misclassification in the estimation of relative risk. Am. J. Epidemiol. 105 (5), 488–495.

Delgado-Rodriguez, M., Llorca, J., 2004. Bias. J. Epidemiol. Community Health 58 (8), 635–641. http://dx.doi.org/10.1136/jech.2003.008466.

Dohoo, I., Andersen, S., Dingwell, R., Hand, K., Kelton, D., Leslie, K., Schukken, Y., Godden, S., 2011a. Diagnosing intramammary infections: comparison of multiple versus single quarter milk samples for the identification of intramammary infections in lactating dairy cows. J. Dairy Sci. 94 (11), 5515–5522. http://linkinghub.elsevier.com/retrieve/pii/S0022030211005789.

Dohoo, I.R., Smith, J., Andersen, S., Kelton, D.F., Godden, S., 2011b. Diagnosing intramammary infections: evaluation of definitions based on a single milk sample. J. Dairy Sci. 94 (1), 250–261. http://www.sciencedirect.com/science/article/pii/S0022030210006922.

Dohoo, I., Martin, W., Stryhn, H., 2009. Veterinary Epidemiologic Research, 2nd ed. VER Inc.

Dufour, S., Dohoo, I.R., Barkema, H.W., DesCôteaux, L., DeVries, T.J., Reyher, K.K., Roy, J.-P., Scholl, D.T., 2012a. Epidemiology of coagulase-negative staphylococci intramammary infection in dairy cattle and the effect of bacteriological culture misclassification. J. Dairy Sci. 95 (6), 3110–3124. http://www.sciencedirect.com/science/article/pii/S002203021200286X.

Dufour, S., Dohoo, I.R., Barkema, H.W., DesCôteaux, L., DeVries, T.J., Reyher, K.K., Roy, J.-P., Scholl, D.T., 2012b. Manageable risk factors associated with the lactational incidence, elimination, and prevalence of *Staphylococcus aureus* intramammary infections in dairy cows. J. Dairy Sci. 95, 1283–1300.

Greenland, S., 1980. The effect of misclassification in the presence of covariates. Am. J. Epidemiol. 112 (4), 564–569.

Hui, S.L., Walter, S.D., 1980. Estimating the error rates of diagnostic tests. Biometrics 36, 167–171.

Magder, L.S., Hugues, J.P., 1997. Logistic regression when the outcome is measured with uncertainty. Am. J. Epidemiol. 146, 195–203.

McInturff, P., Johnson, W.O., Cowling, D., Gardner, I.A., 2004. Modelling risk when binary outcomes are subject to error. Stat. Med. 23 (7), 1095–1109. http://dx.doi.org/10.1002/sim.1656/abstract.

Pekkanen, J., Sunyer, J., 2008. Problems in using incidence to analyze risk factors in follow-up studies. Eur. J. Epidemiol. 23 (9), 581–584.

Pekkanen, J., Sunyer, J., Chinn, S., 2006. Nondifferential disease misclassification may bias incidence risk ratios away from the null. J. Clin. Epidemiol. 59, 281–289.

Poole, C., 1985. Exceptions to the rule about nondifferential misclassification. Am. J. Epidemiol. 122 (3), 508.

Quade, D., Lachenbruch, P.A., Whaley, F.S., McClish, D.K., Haley, R.W., 1980. Effects of misclassifications on statistical inferences in epidemiology. Am. J. Epidemiol. 111 (5), 503–515.

Core Team, R., 2015. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org.

Rogan, W.J., Gladen, B., 1978. Estimating prevalence from the results of a screening test. Am. J. Epidemiol. 107 (1), 71–76.

Rothman, K.J., Lash, T.L., Greenland, S., 2012. Modern Epidemiology. Lippincott Williams & Wilkins.

Stan Development Team, 2016. RStan: The R Interface to Stan. http://mc-stan.org.

Zadoks, R.N., Allore, H.G., Barkema, H.W., Sampimon, O.C., Wellenberg, G.J., Gröhn, Y.T., Schukken, Y.H., 2001. Cow- and quarter-level risk factors for *Streptococcus uberis* and *Staphylococcus aureus* mastitis. J. Dairy Sci. 84 (12), 2649–2663. http://www.ncbi.nlm.nih.gov/pubmed/11814021.