Accuracy and inter-rater reliability of lung auscultation by bovine practitioners when compared with ultrasonographic findings

Bart Pardon,[1] Sébastien Buczinski,[2] Piet R Deprez[1]

Abstract

In practice, veterinary surgeons frequently rely on lung auscultation as a confirmation test for pneumonia. To what extent diagnostic accuracy of lung auscultation varies between different practitioners is currently unknown. In this diagnostic test study, 49 Dutch veterinarians each auscultated between 8 and 10 calves, and communicated whether they would decide to treat the animal with antimicrobials or not. They were not allowed to perform any other aspect of the clinical examination. Their decisions were compared with lung ultrasonography findings. The average sensitivity and specificity of lung auscultation were 0.63 (sd=0.2; range=0.2–1.0) and 0.46 (sd=0.3; range=0.0–1.0), respectively. Of the participants, 8.2 per cent were 100 per cent sensitive, 16.3 per cent were 100 per cent specific, and only 4.0 per cent were perfect. The Krippendorff's alpha was 0.18 (95 per cent confidence interval: −0.01 to 0.38), signifying poor reliability between multiple raters. Regardless of the poor diagnostic accuracy in this study, especially the large variation in a confirmation test between different practitioners could potentially cause professional damage as well as misuse of antimicrobials. This study could be seen as a gentle stimulus to regularly evaluate one's diagnostic skills. Both complementary training and the use of more accurate techniques with less inter-rater variation could improve the situation.

Infectious bronchopneumonia (bovine respiratory disease (BRD)) is of major economic importance to the cattle industry and the leading indication for antimicrobial use in calves and beef cattle.[1 2] For economic reasons and because of the increasing pressure to reduce antimicrobial use, timely recognition and diagnostic confirmation of pneumonia at the individual level have become crucial.(EMA/EFSA 2017)[3 4] The increasing herd size in many production systems worldwide directs farmers and veterinarians towards more stepwise combinations of a screening test followed by confirmation tests to diagnose and treat pneumonia.[3] Previous work showed that punctual scoring of clinical signs is not accurate enough to confirm pneumonia.[5] Also inter-rater variability is very large when using scorecards[6] A lot is expected from blood tests and automated detection systems, but to date lung auscultation and thoracic ultrasonography are the most accessible confirmation tests in the field. In the first instance most practitioners will rely on lung auscultation, but its diagnostic accuracy and inter-rater variability are hardly documented. Available studies showed highly variable results and only involved one or two test persons active in research.[5 7] Therefore, the objective of the present study was to determine the diagnostic accuracy and interobserver agreement of lung auscultation to detect pneumonia in calves in a group of bovine practitioners, using lung ultrasonography as a comparative test.

In November 2015 a diagnostic test study was conducted in three Dutch veal farms, including three different test sessions on consecutive days. The study population consisted out of a convenience sample of Dutch bovine practitioners (estimated national number=820 veterinarians) who attended a training

Veterinary Record (2019) doi: 10.1136/vr.105238

[1]Department of Large Animal Internal Medicine, Faculty of Veterinary Medicine, Ghent University, Merelbeke, Belgium
[2]Sciences cliniques, Faculte de Medecine Veterinaire, Saint-Hyacinthe, Quebec, Canada

E-mail for correspondence: bart.pardon@ugent.be

Provenance and peer review Not commissioned; externally peer reviewed.

Received October 24, 2018
Revised January 16, 2019
Accepted June 3, 2019

on thoracic ultrasonography in calves. All veterinarians present (n=60) were eligible, but only 49 were willing to participate. In each farm, one pen was selected by convenience, in which the producer had recently diagnosed BRD. All animals within this pen (8–10 calves) were subject to examination. Animals had been treated with antimicrobials one to five days before, hereby masking other clinical signs associated with pneumonia. Calves aged 5, 13 and 10 weeks in sessions 1, 2 and 3, respectively. All participants auscultated all the calves present in the pen using their own stethoscope and the procedure and pace as they would normally do in practice. They were not allowed to perform other parts of the clinical investigation, and were specifically asked not to observe the animals in detail. Multiple participants were simultaneously present in the pen, but no communication was allowed. All participants reported their findings discretely as a binary variable (normal or abnormal lung sounds, with abnormal signifying the need for antimicrobial treatment) to a recorder which was positioned outside the pen. When all the participants have finished the auscultation, the thorax of all the calves was scanned with a 7.5-MHz linear probe (MyLab One, Esaote, The Netherlands), as described previously.[7] A lung consolidation depth of ≥1 cm was considered as the reference standard test to diagnose bronchopneumonia. First, for each pair of observers, agreement of auscultation with ultrasonography was calculated by means of Cohen's kappa. To estimate inter-rater reliability among all practitioners, Krippendorff's alpha (Kalpha) was calculated,[8] using 5000 bootstrap samples to estimate the 95 per cent confidence interval (95 per cent CI). Kalpha has been reported as the most flexible reliability parameter especially when not all subjects (calves) are classified by all raters.[8] Diagnostic accuracy of lung auscultation was determined for each participant by calculation of sensitivity (Se), specificity (Sp) and percentage correctly classified, using lung ultrasound as the reference standard test. Analyses were performed using a commercial software (SAS V.9.4, SAS Institute, Cary, North Carolina).

A total of 49 Dutch (81.7 per cent, 49/60) practitioners participated in the three different sessions. Of the participants, 71 per cent (35/49) were male and 29 per cent (14/49) were female. The proportion of calves with consolidations was 63.0 per cent (5/8), 78.0 per cent (7/9) and 60.0 per cent (6/10) in sessions 1, 2 and 3, respectively. The mean kappa value for the agreement of auscultation with ultrasonography was 0.38±sd=0.22 (range (R)=0.10–1.0), which is only 'fair' (0.21–0.40) according to Cohen's classification. This agreement was not influenced by session or gender. The average Se and Sp of lung auscultation were 0.63 (sd=0.2, R=0.2–1.0) and 0.46 (sd=0.3, R=0.0–1.0), respectively. Of the participants, 8.2 per cent were 100 per cent sensitive, 16.3 per cent were 100 per cent specific, and only 4.0 per cent were perfect. Figure 1 provides a graphical representation of variation in Se and Sp among the participants. Kalpha was 0.18 (95 per cent CI: −0.01 to 0.38), signifying poor reliability between multiple raters.[9]

The main merit of the present study lies in providing an estimate from a population that actually makes the difference in the field, namely the practitioners. Data are scant on the use in practice of the most commonly available tool for BRD diagnosis. The results showed highly variable, but on average very poor, diagnostic accuracy of lung auscultation in the study group, together with a very poor inter-rater reliability. A limitation is that the sampling method resulted in selection bias, and the results possibly cannot be extrapolated to bovine practitioners in general. Also the age of the studied calves and a difference in aetiological lung pathogens can be potential confounders. However, the authors have conducted this auscultation exercise in different European countries, with similar results. Repeated manipulation by different persons might have influenced respiratory rate and depth, which can influence lung auscultation findings, potentially causing misclassification bias. The most important remark to correctly interpret this study is that the authors only evaluated lung auscultation as a confirmation test, not the whole clinical examination of the veterinarian. Diagnostic accuracy of a complete clinical examination is likely much higher. Although substantial attempts were made to avoid it, the authors cannot exclude that some participants did take some other clinical signs into account, influencing their decision made by lung auscultation. Validation of diagnostic tests is a multistep process that requires test–retest reliability assessment (World Organisation for Animal Health, 2013).[10 11] Knowing how multiple testers classify the same subject (calf) is therefore of critical importance even before discussing test accuracy.[9] Even if the test is not perfectly accurate, it can still be used in practice if reliable, depending on a specific clinical context (eg, for ruling in BRD using a positive specific test or for ruling out BRD using a negative sensitive test). However, if test accuracy depends on the person who performs the test (veterinarian) or any subject (calves) covariate, these factors should be known (1) to improve auscultation teaching and (2) to adequately know calves covariates that impact interpretation of test results.

No gold standard test to diagnose bronchopneumonia currently exists, but lung ultrasound has the highest accuracy for practical use in the field[12] and has good inter-rater agreement.[13] Bayesian methodology could not be used in the current design. The average Se and Sp of lung auscultation were in line with a previous study involving one observer in a Bayesian comparison of lung auscultation and ultrasound (72.9 per cent Se and 53.3 per cent Sp).[8] These results might concern farmers, and especially the low Sp appears worrisome.

Study	TP	FP	FN	TN	Day	Sensitivity (95% CI)	Specificity (95% CI)
Vet6	5	0	0	3	1	1.00 (0.48, 1.00)	1.00 (0.29, 1.00)
Vet7	5	0	0	3	1	1.00 (0.48, 1.00)	1.00 (0.29, 1.00)
Vet1	4	0	1	3	1	0.80 (0.28, 0.99)	1.00 (0.29, 1.00)
Vet11	3	0	2	3	1	0.60 (0.15, 0.95)	1.00 (0.29, 1.00)
Vet2	3	1	2	2	1	0.60 (0.15, 0.95)	0.67 (0.09, 0.99)
Vet3	3	2	2	1	1	0.60 (0.15, 0.95)	0.33 (0.01, 0.91)
Vet9	3	3	2	0	1	0.60 (0.15, 0.95)	0.00 (0.00, 0.71)
Vet4	2	2	3	1	1	0.40 (0.05, 0.85)	0.33 (0.01, 0.91)
Vet5	2	2	3	1	1	0.40 (0.05, 0.85)	0.33 (0.01, 0.91)
Vet10	2	3	3	0	1	0.40 (0.05, 0.85)	0.00 (0.00, 0.71)
Vet8	1	0	4	3	1	0.20 (0.01, 0.72)	1.00 (0.29, 1.00)
Vet14	6	1	1	1	2	0.86 (0.42, 1.00)	0.50 (0.01, 0.99)
Vet21	6	2	1	0	2	0.86 (0.42, 1.00)	0.00 (0.00, 0.84)
Vet16	6	2	1	0	2	0.86 (0.42, 1.00)	0.00 (0.00, 0.84)
Vet18	5	1	2	1	2	0.71 (0.29, 0.96)	0.50 (0.01, 0.99)
Vet15	5	1	2	1	2	0.71 (0.29, 0.96)	0.50 (0.01, 0.99)
Vet22	4	1	3	1	2	0.57 (0.18, 0.90)	0.50 (0.01, 0.99)
Vet24	4	1	3	1	2	0.57 (0.18, 0.90)	0.50 (0.01, 0.99)
Vet27	4	2	3	0	2	0.57 (0.18, 0.90)	0.00 (0.00, 0.84)
Vet29	4	2	3	0	2	0.57 (0.18, 0.90)	0.00 (0.00, 0.84)
Vet25	4	2	3	0	2	0.57 (0.18, 0.90)	0.00 (0.00, 0.84)
Vet13	4	2	3	0	2	0.57 (0.18, 0.90)	0.00 (0.00, 0.84)
Vet12	4	2	3	0	2	0.57 (0.18, 0.90)	0.00 (0.00, 0.84)
Vet17	3	0	4	2	2	0.43 (0.10, 0.82)	1.00 (0.16, 1.00)
Vet19	3	1	4	2	2	0.43 (0.10, 0.82)	0.67 (0.09, 0.99)
Vet20	3	1	4	1	2	0.43 (0.10, 0.82)	0.50 (0.01, 0.99)
Vet23	3	2	4	0	2	0.43 (0.10, 0.82)	0.00 (0.00, 0.84)
Vet26	2	1	5	1	2	0.29 (0.04, 0.71)	0.50 (0.01, 0.99)
Vet28	2	1	5	1	2	0.29 (0.04, 0.71)	0.50 (0.01, 0.99)
Vet35	6	2	0	2	3	1.00 (0.54, 1.00)	1.50 (0.07, 0.93)
Vet48	6	3	0	1	3	1.00 (0.54, 1.00)	0.25 (0.01, 0.81)
Vet39	5	0	1	4	3	0.83 (0.36, 1.00)	1.00 (0.40, 1.00)
Vet31	5	1	1	3	3	1.83 (0.36, 1.00)	1.75 (0.19, 0.99)
Vet33	5	2	1	2	3	0.83 (0.36, 1.00)	0.50 (0.07, 0.93)
Vet49	5	2	1	2	3	1.83 (0.36, 1.00)	0.50 (0.07, 0.93)
Vet40	5	2	1	2	3	1.83 (0.36, 1.00)	0.50 (0.07, 0.93)
Vet36	5	3	1	1	3	1.83 (0.36, 1.00)	0.25 (0.01, 0.81)
Vet45	4	0	2	4	3	1.67 (0.22, 0.96)	1.00 (0.40, 1.00)
Vet42	4	1	2	3	3	0.67 (0.22, 0.96)	0.75 (0.19, 0.99)
Vet30	4	2	2	2	3	0.67 (0.22, 0.96)	0.50 (0.07, 0.93)
Vet47	4	2	2	2	3	0.67 (0.22, 0.96)	0.50 (0.07, 0.93)
Vet38	4	2	2	2	3	0.67 (0.22, 0.96)	0.50 (0.07, 0.93)
Vet46	4	2	2	2	3	0.67 (0.22, 0.96)	0.50 (0.07, 0.93)
Vet41	4	3	2	1	3	0.67 (0.22, 0.96)	0.25 (0.01, 0.81)
Vet32	3	2	3	2	3	0.50 (0.12, 0.88)	0.50 (0.07, 0.93)
Vet37	3	2	3	2	3	0.50 (0.12, 0.88)	0.50 (0.07, 0.93)
Vet34	3	3	3	1	3	0.50 (0.12, 0.88)	0.25 (0.01, 0.81)
Vet43	3	4	3	0	3	0.50 (0.12, 0.88)	0.50 (0.00, 0.60)
Vet44	2	1	4	3	3	0.33 (0.04, 0.78)	0.75 (0.19, 0.99)

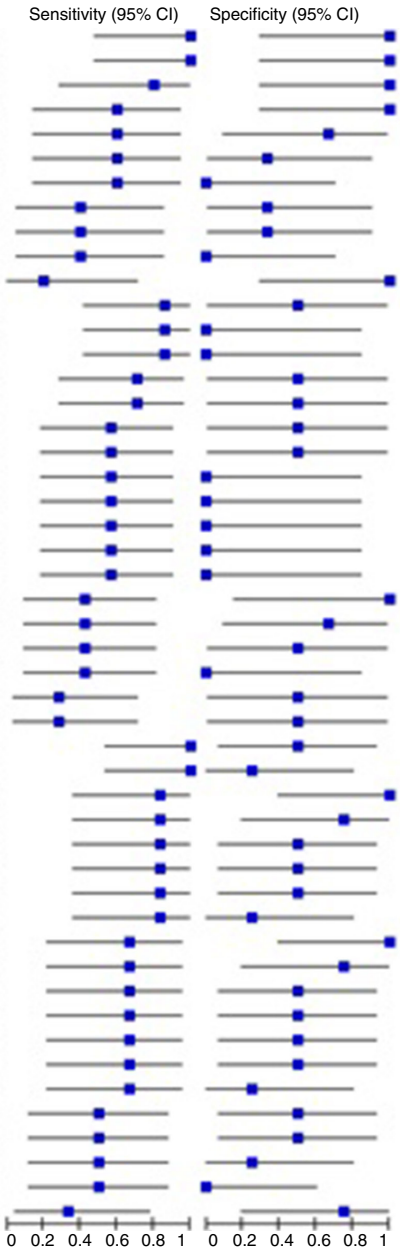

Figure 1 Diagnostic accuracy of lung auscultation by 49 Dutch bovine practitioners in detecting lung consolidation. The veterinarians are classified by day of training (day) and by decreasing sensitivities and specificities on the same day. The figure was obtained using Review Manager (computer program) (V.5.3, Copenhagen: The Nordic Cochrane Centre, The Cochrane Collaboration, 2014). CI, confidence interval; FN (false negatives), *normal vet auscultation/abnormal ultrasound*; FP (false positives),*abnormal vet auscultation/normal ultrasound*; TN (true negatives), *normal vet auscultation/normal ultrasound*; TP (true positives), *abnormal vet auscultation/abnormal ultrasound*.

Airway inflammation and bronchoconstriction can result in adventitious lung sounds long before lung consolidation can be visualised, explaining the large proportion of false positives. The decision of the participants was also possibly biased by informing them that pneumonia cases had occurred in the pen. Anyway, in the scarcely available studies on human beings, physicians hardly scored better (Se=47–69 per cent and Sp=58–75 per cent).[14] Regardless of diagnostic accuracy, especially the low inter-rater reliability should be considered a substantial problem. Such extent of variation in a confirmation test between different practitioners could cause professional damage to the practice as well as potential misuse of antimicrobials.

In conclusion, this study showed on average poor diagnostic accuracy and very low inter-rater reliability for lung auscultation by bovine practitioners. The results could be seen as a gentle stimulus to maintain evaluation of one's diagnostic strengths and weaknesses. Improvements to certain skills can be made by additional training, for example using a high-accuracy reference test such as lung ultrasound. Other solutions can be the use of new technology, such as electronic stethoscopes and neural networks,[15 16] or the systematic use of lung ultrasound as a field test to confirm bronchopneumonia. However, the latter test is potentially also subject to inter-rater variability, especially in inexperienced operators.

Acknowledgements The authors wish to thank Zoetis The Netherlands for practical organisation of the training, and the practitioners for their kind participation and interest in the subject. S Pardon is acknowledged for the inspiration.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests The present study was done within the framework of a Zoetis BRD master class on lung ultrasonography in calves. The authors declare that the present results have not been influenced by this pharmaceutical company.

Ethics approval The study protocol was approved by the ethical committee of the Faculty of Veterinary Medicine, Ghent University.

References

1 Pardon B, Catry B, Dewulf J, *et al*. Prospective study on quantitative and qualitative antimicrobial and anti-inflammatory drug use in white veal calves. *J Antimicrob Chemother* 2012;67:1027–38.
2 Delabouglise A, James A, Valarcher JF, *et al*. Linking disease epidemiology and livestock productivity: The case of bovine respiratory disease in France. *PLoS One* 2017;12:e0189090.
3 Wolfger B, Timsit E, White BJ, *et al*. A systematic review of bovine respiratory disease diagnosis focused on diagnostic confirmation, early detection, and prediction of unfavorable outcomes in feedlot cattle. *Vet Clin North Am Food Anim Pract* 2015;31:351–65.
4 EMA/EFSA. EMA and EFSA Joint Scientific Opinion on measures to reduce the need to use antimicrobial agents in animal husbandry in the European Union, and the resulting impacts on food safety (RONAFA). EFSA Journal 15, 4666. 2017.
5 Buczinski S, Forté G, Francoz D, *et al*. Comparison of thoracic auscultation, clinical score, and ultrasonography as indicators of bovine respiratory disease in preweaned dairy calves. *J Vet Intern Med* 2014;28:234–42.
6 Buczinski S, Ménard J, Timsit E. Incremental Value (Bayesian Framework) of Thoracic Ultrasonography over Thoracic Auscultation for Diagnosis of Bronchopneumonia in Preweaned Dairy Calves. *J Vet Intern Med* 2016;30:1396–401.
7 Buczinski S, Faure C, Jolivet S, *et al*. Evaluation of inter-observer agreement when using a clinical respiratory scoring system in pre-weaned dairy calves. *N Z Vet J* 2016;64:243–7.
8 Zapf A, Castell S, Morawietz L, *et al*. Measuring inter-rater reliability for nominal data - which coefficients and confidence intervals are appropriate? *BMC Med Res Methodol* 2016;16:93.
9 Gwet KL. Handbook of inter-rater reliability: the definitive guide to measuring the extent of agreement among raters. 4th edn: Advanced Analytics LLC, 2014.
10 World Organisation for Animal Health. Principles and methods for validation of diagnostic assays for infectious diseases. In: Manual of diagnostic tests and vaccines for terrestrial animals, 2013.
11 OIE Terrestrial Manual. Principles of validation of diagnostic assays for infectious diseases. 2018 http://www.oie.int/fileadmin/Home/eng/Health_standards/tahm/1.01.06_VALIDATION.pdf.
12 Ollivett TL, Buczinski S. On-Farm use of ultrasonography for bovine respiratory disease. *Vet Clin North Am Food Anim Pract* 2016;32:19–35.
13 Buczinski S, Buathier C, Bélanger AM, *et al*. Inter-rater agreement and reliability of thoracic ultrasonographic findings in feedlot calves, with or without naturally occurring bronchopneumonia. *J Vet Intern Med* 2018;32:1787–92.
14 Wipf JE, Lipsky BA, Hirschmann JV, *et al*. Diagnosing pneumonia by physical examination: relevant or relic? *Arch Intern Med* 1999;159:1082–7.
15 Mang AV, Buczinski S, Booker CW, *et al*. Evaluation of a computer-aided lung auscultation system for diagnosis of bovine respiratory disease in feedlot cattle. *J Vet Intern Med* 2015;29:1112–6.
16 Datta S, Choudhury AD, Deshpande P, *et al*. Automated Lung Sound Analysis for Detecting Pulmonary Abnormalities. 2017 39th Annual International Conference of the Ieee Engineering in Medicine and Biology Society (Embc), 4594-4598. 2017.