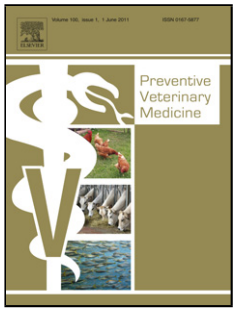

Title: A Bayesian latent class model to estimate the accuracy of pregnancy diagnosis by transrectal ultrasonography and laboratory detection of pregnancy-associated glycoproteins in dairy cows

Authors: G.T. Fosgate, B. Motimele, A. Ganswindt, P.C. Irons

PII: S0167-5877(17)30074-0
DOI: http://dx.doi.org/doi:10.1016/j.prevetmed.2017.07.004
Reference: PREVET 4278

To appear in: *PREVET*

Received date: 2-2-2017
Revised date: 27-6-2017
Accepted date: 4-7-2017

Please cite this article as: Fosgate, G.T., Motimele, B., Ganswindt, A., Irons, P.C., A Bayesian latent class model to estimate the accuracy of pregnancy diagnosis by transrectal ultrasonography and laboratory detection of pregnancy-associated glycoproteins in dairy cows.Preventive Veterinary Medicine http://dx.doi.org/10.1016/j.prevetmed.2017.07.004

A Bayesian latent class model to estimate the accuracy of pregnancy diagnosis by transrectal ultrasonography and laboratory detection of pregnancy-associated glycoproteins in dairy cows

G. T. Fosgate [a,*], B. Motimele[a,b], A. Ganswindt[c], P.C. Irons[a,d]

[a]University of Pretoria, Faculty of Veterinary Science, Department of Production Animal Studies, Onderstepoort, 0110, South Africa

[b]Agricultural Research Council-Animal Production Institute, Irene 0062, South Africa

[c]University of Pretoria, Faculty of Veterinary Science, Endocrine Research Laboratory, Department of Anatomy and Physiology, Onderstepoort 0110, South Africa

[d]Murdoch University, College of Veterinary Medicine, School of Veterinary and Life Sciences, Western Australia, 6150, Australia

*Corresponding author at: University of Pretoria, Faculty of Veterinary Science, Department of Production Animal Studies, Private Bag X04, Onderstepoort, 0110, South Africa

Tel: +27 12 529 8257

Fax: +27 12 529 8315

E-mail address: geoffrey.fosgate@up.ac.za (G.T. Fosgate).

Abstract

Accurate diagnosis of pregnancy is an essential component of an effective reproductive management plan for dairy cattle. Indirect methods of pregnancy detection can be performed soon after breeding and offer an advantage over traditional direct methods in not requiring an experienced veterinarian and having potential for automation. The objective of this study was to estimate the sensitivity and specificity of pregnancy-associated glycoprotein (PAG) detection ELISA and transrectal ultrasound (TRUS) in dairy cows of South Africa using a Bayesian latent class approach. Commercial dairy cattle from the five important dairy regions in South Africa were enrolled in a short-term prospective cohort study. Cattle were examined at 28-35 days after artificial insemination (AI) and then followed up 14 days later. At both sampling times, TRUS was performed to detect pregnancy and commercially available PAG detection ELISAs were performed on collected serum and milk. A total of 1236 cows were sampled and 1006 had complete test information for use in the Bayesian latent class model. The estimated sensitivity (95% probability interval) and specificity for PAG detection serum ELISA were 99.4% (98.5, 99.9) and 97.4% (94.7, 99.2), respectively. The estimated sensitivity and specificity for PAG detection milk ELISA were 99.2% (98.2, 99.8) and 93.4% (89.7, 96.1), respectively. Sensitivity of veterinarian performed TRUS at 28-35 days post-AI varied between 77.8% and 90.5% and specificity varied between 94.7% and 99.8%. In summary, indirect detection of pregnancy using PAG ELISA is an accurate method for use in dairy cattle. The method is descriptively more sensitive than veterinarian-performed TRUS and therefore could be an economically viable addition to a reproductive management plan.

Keywords: Dairy cattle; Pregnancy; Sensitivity; Specificity; Bayesian

1. Introduction

Global milk production was estimated to be 466 metric tons in 2013 (Lagrange et al., 2015). Population growth in combination with continued urbanization is expected to continue to drive strong demand for increasing production. The average dairy herd size has increased over the past decades but the number of herds has decreased or remained relatively stable. Consequently, dairy cows are being managed in a smaller number of larger herds (Barkema et al., 2015). Increasing farm size is driven by "economies of scale" and supported by improved technology (von Keyserlingk et al., 2013). Increases in global production are largely due to increasing production per cow, rather than increasing cow numbers. Optimal reproductive programs maximize individual cow production by ensuring regular calvings during the cow's lifetime thus limiting the time spent in the tail of the lactation curve, when daily production is lower.

Sound reproductive management depends on establishing and maintaining pregnancy as soon as possible after the end of the voluntary waiting period after calving, or after the onset of the breeding season in the case of seasonal systems. This is achieved by ensuring normal cyclicity, breeding the cows at the optimal time in the estrus cycle and accurate pregnancy determination as soon as possible after breeding. Follow-up confirmation for the maintenance of pregnancy is also required. The prompt detection of open cows is desirable to enable treatments to be applied that initiate estrus and allow repeated breeding attempts when a pregnancy is not present, thus reducing days open. The economic advantage of earlier detection of open cows is manifested through shorter inter-calving intervals (Oltenacu et al., 1990). The optimal situation would be a highly accurate test able to detect pregnancy before the first expected estrus following breeding, but no such tests have been developed.

Whereas the recurrence of estrus at the expected interval of 18-24 days is an obvious early indication of non-pregnancy, the weak expression of estrus behavior and the impracticality of maintaining consistent estrus detection in large dairy herds make this method unreliable. Diagnostic tests for pregnancy have therefore become indispensable to support ongoing increases in reproductive efficiency.

Pregnancy detection can be performed using both direct and indirect methods. Direct methods include transrectal palpation and transrectal ultrasonography (TRUS). Indirect methods utilize chemical markers of pregnancy including milk or plasma progestagen concentrations, or pregnancy-associated protein and pregnancy-associated glycoprotein (PAG) measurements (Fricke et al., 2016). Bovine PAGs are produced by the binucleate trophoblastic cells of the placenta early after implantation (Wooding et al., 2005). PAGs can be detected in the maternal circulation from 22 days of pregnancy (Zoli et al., 1992) until 2-3 months after parturition (Humblot, 2001; Whitlock and Maxwell, 2008).

The major advantages of indirect methods of pregnancy diagnosis are that a skilled, and thus comparatively costly, examiner is not required on the farm and they can be performed sooner after breeding compared to direct detection methods. Direct palpation of the conceptus early in gestation has also been associated with pregnancy loss (Franco et al., 1987; Thurmond and Picanso, 1993). The growing trend of larger herd sizes in conjunction with a trend towards fewer veterinarians entering rural veterinary practice (Villarroel et al., 2010) further support the development of indirect methods that will not require a highly-skilled operator to be present at the time of testing. The potential for the use of indirect pregnancy detection methods in automated systems incorporated into the milking parlor provides additional motivation for their development. PAG detection ELISAs are amenable to the design of

automated systems since they have been validated for use in serum (Romano and Larson, 2010; Karen et al., 2015) and milk (Leblanc, 2013).

Accurate pregnancy diagnosis is critical to adoption in the field. High sensitivity of the diagnostic method is very important when considering the economic benefit of pregnancy detection using these modalities (Giordano et al., 2013). The economic losses associated with inducing estrus in pregnant cows (false-negative on the assay) are significant (Fricke et al., 2016). The specificity of the indirect method relative to traditional direct methods is also an important consideration in the selection of the best diagnostic method as improved specificity reduces the losses associated with the late detection of non-pregnancy and the resulting reduced opportunity cost. The cost of the diagnostic tests is also an important consideration.

Previously performed diagnostic evaluations estimated the sensitivity and specificity of PAG detection relative to direct pregnancy detection as the reference standard. This analytic approach is incapable of determining whether or not PAG detection is more accurate than the traditional pregnancy detection methods, and is not possible during early gestation when the true pregnancy status of the animals cannot be accurately determined. The objective of this study was to estimate the sensitivity and specificity of PAG detection ELISA and transrectal ultrasound (TRUS) in dairy cows of South Africa using a Bayesian latent class approach. The hypothesis was that the laboratory detection of PAGs would be a more sensitive but less specific method for the detection of pregnancies compared to veterinarian-performed TRUS.

2. Materials and methods

2.1 Study design

A short-duration prospective cohort study (July – September 2014) was implemented on commercial dairy farms within the five important dairy regions of South Africa. Selected regions were Kwa-Zulu Natal Province, Mpumalanga Province, the west coast region of the Western Cape Province, the western coastal region of the Eastern Cape Province, and the interior of the Eastern Cape. One veterinary practice from each location was conveniently selected for participation in the study. Veterinary practitioners received training concerning study procedures and objectives at a workshop that preceded study commencement. These practitioners performed all veterinary activities on a single client herd that was conveniently selected after receiving informed consent from the herd owner. The sample size for each study location was estimated based on the desire to have 97 pregnant cows available for estimation of sensitivity and 97 additional cows available for estimation of specificity. This was based on assumed sensitivity and specificity of 95% and the desire to estimate these values +/- 0.05 at the 95% level of confidence using exact binomial methodology (Fosgate, 2005, 2009). The pregnancy proportion in sampled cows was unknown at the time of enrolment and assumed to be 50% for the sample size calculation. The sample size per study location was increased by 20% to account for losses during follow-up and errors including mislabeling of collected specimen tubes. The sample size was therefore estimated as 233 cows per location for a total 1165 cows (Supplemental Material). The study was reviewed and approved by the Animal Ethics Committee at the Faculty of Veterinary Science, University of Pretoria (Protocol No. V043-14).

2.3 Study animals and determination of reproductive status

Cows in commercial dairy herds that were presented to one of the participating veterinarians for pregnancy diagnosis were utilized for this study. Participating veterinarians were requested to enroll all eligible cows from the selected study farm until the necessary sample size was obtained. Cows were eligible to be enrolled in the study at 28-35 days after breeding by artificial insemination (AI) if no estrus activity had been detected post-breeding. Veterinarians excluded cows when the stage of pregnancy appeared inconsistent with the reported breeding to reduce the possibility of enrolling cows that were not at the correct stage of pregnancy (28-35 days post-AI). Cows that were diagnosed with any uterine or ovarian pathology by the veterinarian at the time of enrollment were excluded if treatment was required since treatment would likely cause fetal loss in pregnant cows. Veterinary diagnosis of pregnancy was performed using transrectal ultrasound (TRUS) and pregnancy was diagnosed by visualization of the embryonic/fetal membranes or by direct observation of the fetus. Participating veterinarians did not receive specific training in TRUS as part of the study and the previous experience level ranged from 1 to 8 years. Veterinarians performed TRUS blinded to the results of previous veterinary examinations and were unmasked after making their pregnancy diagnosis. Additional data that were collected at the time of examination included breed, lactation number, days in milk (DIM), and days post-AI.

2.4 Sample collection

At the time of transrectal examination, veterinarians collected whole blood from the coccygeal vein into plain evacuated 10 ml vacutainer tubes (Becton, Dickinson and Company, Franklin Lakes, New Jersey USA) and raw milk from at least two teats into 11 ml tubes containing a commercial milk preservative (Broad Spectrum Microtabs II, Advanced Instruments, Inc., Norwood, Massachusetts, USA). Samples were held at ambient

temperature and shipped by overnight courier to the Endocrine Research Laboratory within the Faculty of Veterinary Science, University of Pretoria. The first sampling performed at 28-35 days post-AI was labelled 'A' and the same cows were sampled again two weeks later (42-49 days post-AI) and this collection time was labelled 'B'. Upon arrival to the laboratory, blood samples were centrifuged at 1,800 g for 15 min and serum was separated and stored at -20˚C until analysis. Milk samples were stored at 4˚C until analysis. Samples were labelled with unique serial identification numbers upon arrival to the laboratory. A single laboratory technician performed all testing for an individual assay and independent laboratory technicians performed the different tests. Technicians were blinded to cow identification, TRUS pregnancy status of cows, and the results of other assays. Samples were only exposed to a single freeze-thaw cycle at room temperature prior to laboratory analysis.

2.5 Enzyme Immunoassays

Cow sera were evaluated for PAGs using a quantitative serum ELISA (Bovine Pregnancy Test Kit, IDEXX Laboratories, Inc., Westbrook, Maine, USA) and the manufacturer's qualitative modification of this test for use without an ELISA plate reader (Visual Pregnancy Kit, IDEXX Laboratories, Inc., Westbrook, Maine, USA). The level of PAGs in the milk was assessed using a quantitative milk ELISA (Milk Pregnancy Test Kit, IDEXX Laboratories, Inc., Westbrook, Maine, USA). Assays were performed according to manufacturer's instructions using the dedicated software (xCheckPlus, IDEXX Laboratories, Inc.). Each 96-well ELISA plate included positive and negative controls added in duplicate. A 450 nm wavelength filter was used to read the color change of the serum and milk pregnancy test kits. The quantitative result was determined as the sample optical density (OD) minus the mean OD of the plate negative control. A cutoff of >0.3 was used to classify serum ELISA results as positive and negative (pregnant versus open). An OD difference of >0.1 was similarly

used to classify milk ELISA results as pregnant versus open. This classification meant that results that the manufacturer typically recommends for "recheck" were classified as "pregnant". This approach for dichotomization of results was chosen after consultation with the manufacturer's technical staff. The qualitative visual ELISA was performed independently of the serum ELISA using dedicated kits provided by the manufacturer. This test follows a similar protocol but the positive/negative determination is performed by the visual inspection of the ELISA plate rather than using an ELISA plate reader. Any perceptible color change in the visual ELISA was classified as a positive ("pregnant") result. All testing was performed at the Endocrine Research Laboratory within the Department of Anatomy and Physiology, Faculty of Veterinary Science, University of Pretoria.

2.6 Statistical analysis

Data were described by scatter plots and boxplots using the ggplot2 package (Wickham, 2009) within R (R Core Team, 2017). Kappa and prevalence-adjusted bias-adjusted kappa (PABAK) were calculated to determine the agreement between tests on the dichotomous (positive/negative) scale (Byrt et al., 1993). Spearman's rho was used to estimate the correlation between quantitative test results. Agreement analyses were performed in commercially available software (IBM SPSS Statistics Version 23, International Business Machines Corp., Armonk, New York, USA). Kappa values of ≤ 0.20, 0.21-0.40, 0.41-0.60, 0.61-0.80, and 0.81-1.00 were classified as poor, fair, moderate, good, and very good agreement, respectively (Altman, 1991). Statistical significance was set as $P < 0.05$.

The sensitivity and specificity of the veterinarian TRUS examinations (28-35 days post-AI) and two PAG detection assays (quantitative serum and milk ELISAs) were estimated within a Bayesian latent class analytical framework. Results from the visual (qualitative) ELISA were

not included as the test is essentially the same as the quantitative serum ELISA but without the quantification step using an ELISA reader. The model also estimated the pregnancy proportion of the cows sampled during the study. The model was based on the Hui-Walter paradigm (Hui and Walter, 1980) but modified for a Bayesian analysis (Enoe et al., 2000). The model included adjustment for conditional dependence between tests (Vacek, 1985) and similar diagnostic test models have been described in more detail elsewhere (Fosgate et al., 2002; Fosgate et al., 2010). The base approach was a three-test model (TRUS A, serum ELISA, milk ELISA) that was stratified by the five herds (one each per veterinarian). TRUS sensitivity and specificity were estimated within each strata while the accuracy of serum and milk ELISA were estimated across all strata. Pairwise conditional dependence terms were added between the serum and milk ELISA results but veterinarian TRUS was assumed conditionally independent of the PAG detection ELISAs.

It was desired to descriptively compare the accuracy of TRUS and PAG detection at 28-35 days post-AI using the data collected within this study and therefore the same mildly informative prior probabilities were used for all tests. It would have been impossible to disentangle the effects of prior probabilities from the likelihood of the data had the analysis incorporated different priors. Sensitivity was modelled using beta (9,1) distributions and specificity was modelled using beta (9.5,0.5) distributions. The TRUS pregnancy proportion at 42-49 days post-AI (TRUS 'B') was used to elicit prior probabilities for the pregnancy proportions at the 28-35 days post-AI sampling (TRUS 'A'). This was performed by setting the first parameter of the beta distribution as the number of TRUS 'B' pregnant cows / 10 and the second parameter as TRUS 'B' open cows / 10. For example, Herd 'A' had a TRUS 'B' pregnancy proportion of 150/206 and therefore the prior probability for the pregnancy proportion at the 'A' sampling was elicited as beta(15,5.6). Markov chain Monte Carlo

(MCMC) techniques were implemented in available statistical software (WinBUGS Version 1.4, MRC Biostatistics Unit, Cambridge, UK). Iterate values of the MCMC process were assessed for autocorrelation and only every 10th iterate was retained to reduce the impact of this correlation. Two simulation chains with different initial values were performed and convergence was assessed by evaluating plots of model parameter iterates and by calculating the Gelman-Rubin statistic (Toft et al., 2005). The first 200,000 iterations were discarded as the burn-in and inferences were made based on the subsequent 20,000 (post-thinning). Median values were used as point estimates and 95% probability intervals (PI) were calculated as the 2.5th to 97.5th percentiles of the posterior distributions. More information concerning the analytical approach including the WinBUGS code has been provided as supplemental material.

A sensitivity analysis was performed by replacing the mildly informative prior probabilities with non-informative, flat priors (beta 1,1), for all model parameters. The sensitivity model was assessed and implemented in the same manner as described for the initial model.

3. Results

3.1 Descriptive

One thousand, two hundred and thirty-six cows were sampled during the study with 226, 257, 241, 224, and 288 from each of the five veterinary practices (a single commercial dairy farm requested for each practice). One thousand and thirty-two cows had sufficient data that could be used for descriptive analyses (Figure 1). Cows were excluded when specimen delivery to the laboratory was delayed and when cows had a breeding date outside the acceptable range for the study (28-35 days post-AI for 'A' sampling). Sampled cows included 868 Holsteins, 105 Jersey, and 59 Holstein-Jersey crosses. The median (interquartile range; IQR) days in

milk was 144 days (109, 211) and 191 cows had missing information. The median (IQR) lactation number was 2 (1, 3) with 198 cows without this information reported. Twenty cows from one of the veterinary practices were excluded from the latent class model because they were collected from a herd different than the primary study herd. Therefore, 1006 cows had complete test results that could be utilized in the Bayesian analysis (Table 1). Cows with discordant test results at the 28-35 day post-AI sampling were analyzed by PAG detection ELISA on the subsequent sampling at 42-49 days post-AI. However, these data were not suitable for the estimation of diagnostic accuracy and therefore are not presented here.

The agreement between tests ranged from moderate (kappa = 0.58) to very good (kappa = 0.97; Table 2) with similar results after adjusting for bias and prevalence. The lowest agreement was between TRUS B and the three PAG detection assays. The highest agreement was between the quantitative serum and visual ELISAs. Quantitative test results from the serum and milk ELISAs had a strong non-linear correlation (Figure 2). Quantitative results from the serum ELISA tended to be low in cows that were determined to be open by the veterinarian at both sampling times but higher when at least one of the examinations suggested the presence of a conceptus (Figure 3). Some of these distributions included notable outliers with values appearing to be more consistent within other categories (e.g. large PAG values for some open cows suggesting the presence of a pregnancy). Quantitative results from the milk ELISA demonstrated a similar overall pattern but with less separation among the groups.

3.2 Diagnostic accuracy

The final Bayesian model estimated a sensitivity (95% PI) and specificity (95% PI) of the quantitative serum ELISA as 99.4% (98.5%, 99.9%) and 97.4% (94.7%, 99.2%), respectively (Table 3). The sensitivity and specificity of the milk ELISA were 99.2% (98.2%, 99.8%) and

93.4% (89.7%, 96.1%), respectively. The sensitivity of TRUS performed at 28-35 days post-AI varied between 77.8% and 90.5% for the five veterinarians and specificity varied between 94.7% and 99.8%.

The final Bayesian model included adjustment for sensitivity and specificity covariance modeled between the quantitative serum and milk PAG detection ELISAs. These dependencies (95% PI) were estimated as $9.1x10^{-4}$ ($-5.7x10^{-6}$, $4.0x10^{-3}$) and $-3.9x10^{-5}$ ($-2.6x10^{-3}$, $4.4\ x10^{-3}$) for sensitivity and specificity, respectively. The results of the sensitivity analysis were descriptively very similar to the final model used for making inferences (Figure 4).

4. Discussion

Diagnosis of pregnancy in dairy cattle using PAG detection is an accurate tool that could be incorporated into an effective reproductive management system. There was very good agreement between serum and milk PAG detection assays indicating a potential to develop automated systems for use in the milking parlor. However, the levels observed within the milk tended to be lower than the corresponding levels in serum. There is an inverse relationship between milk production and PAG concentrations (Lopez-Gatius et al., 2007; Ricci et al., 2015) and the possible requirement of adjusting the positive threshold based on production level requires further investigation. The lower concentration in milk and potential sources of variability independent of pregnancy status are the likely sources of the imperfect agreement between serum and milk results and will likely reduce the overall accuracy of any automated system.

There was good agreement between PAG assays and TRUS performed at 28-35 days with a moderate decrease in agreement when compared to TRUS performed at 42-49 days of pregnancy. The 42-49 day range is the usual timing for performing TRUS by the participating practitioners, which is considered accurate from day 35 (Romano et al., 2006). It was anticipated that TRUS would be more accurate at the second sampling (42-49 days) and the lower agreement with the PAG detection assays was unexpected. The source of the lower agreement could have been due to early pregnancy loss occurring between the time of the sampling for PAG testing and the second TRUS examination. The relatively long half-life of PAGs in the maternal circulation can cause detectable amounts of PAGs to persist for several days after death of the conceptus (Zoli et al., 1992). The percentage of early pregnancy loss (> 28 days) has been estimated to range between 3 and 43% in cattle (as reviewed by Pohler et al., 2016) and this likely affected some cows enrolled in the study. TRUS would immediately recognize these losses but the circulating levels of PAGs might remain above the positive threshold leading to false positive results, even in samples drawn on the same day as the TRUS examination. The majority of cows diagnosed as pregnant on TRUS A but subsequently open on TRUS B had high serum PAG levels similar to cows that were identified as pregnant by TRUS on both 'A' and 'B' examinations. These high values could indicate early pregnancy loss and misclassification as pregnant on the serum PAG ELISA at the 'B' examination. It is also possible that some of these cows were truly pregnant and thus a false negative on TRUS 'B'; however, TRUS is expected to have high sensitivity at this time of gestation and therefore few false negatives are expected. The relationship was similar for the concentrations of PAGs in milk but the lower values overall made the relationship less apparent.

The sensitivities estimated for the PAG serum ELISA were near perfect and the specificity was also quite high, especially for the quantitative serum ELISA. The high sensitivity was expected but the near perfect specificity estimate for the quantitative serum ELISA was surprising due to the possibility of pregnancy loss and the relatively long half-life of PAGs in the maternal circulation. A specificity of 97.4% suggests that only 26 out of 1000 non-pregnant cows would yield a false-positive result. This compares favorably to the simple average specificity of the TRUS performed by the five veterinarians at the same stage of gestation (mean TRUS specificity = 97.2%) However, the estimated sensitivity is substantially higher than what was estimated for TRUS at 28-35 days post-AI (mean TRUS sensitivity = 83.5%). The current study was unable to estimate the sensitivity and specificity of TRUS when performed at 42-49 days of pregnancy, which is the more typical timing in South Africa. Direct detection of pregnancy at this stage is expected to have near perfect specificity and adequate sensitivity and it is a limitation of the study that these expectations could not be verified. Data concerning the birth of calves could also be used to evaluate the specificity of the tests evaluated with this study. The fact that the data collection was performed on commercial dairies presented obstacles related to data access and although requested as part of this study, compliance was an issue and these data were unavailable for analysis.

The accuracy of the commercially available PAG ELISAs employed in this study has been estimated previously. The sensitivity of this quantitative commercial serum ELISA has been estimated to be between 98.4% and 100% for cows >30 days of pregnancy. Within the same studies, specificity was estimated to range between 88.6% and 100% (Ricci et al., 2015; Commun et al., 2016). The sensitivity of the commercial milk ELISA has been estimated to range between 98% and 100% for cows >30 days of pregnancy. Specificity of this milk

assay was estimated to range between 83% and 100% in these same studies (Leblanc, 2013; Lawson et al., 2014; Ricci et al., 2015; Commun et al., 2016). The specificity of the milk ELISA was estimated to be lower than the corresponding estimates for the serum ELISA (when both performed in the same study), which is similar to what was found in the present study. All previous studies estimated sensitivity and specificity relative to TRUS as the reference standard. The results reported in this study are similar to the previous studies despite the use of the latent class analytical approach. A more recent paper also employing Bayesian latent class methods (Dufour et al., 2017), reported a sensitivity of 99% and a specificity of 95% for the same milk ELISA employed in the present study. Results were quite similar despite the different study locations and variations in the sampling and modeling strategies. This similarity despite study differences provides strong evidence that PAG detection has high accuracy for pregnancy diagnosis in dairy cattle.

The results presented here were generated from an observational field study and inferences should be tempered by a thorough evaluation of the potential sources of bias. The herds in the study were not randomly selected and represent a convenience sample selected by participating veterinarians based on having an adequate number of cows and records of suitable quality to perform the study. These herds likely represent some of the most efficient dairy operations in South Africa and this is a potential source of selection bias. It would be unusual for diagnostic accuracy to vary among farms based on management procedures but the potential impact of this selection bias is unknown.

There are also potential limitations based on the chosen analytical approach. The typical method for estimating the sensitivity and specificity of indirect methods to detect pregnancy is through the use of TRUS as a reference standard. The current study was performed with

the explicit desire to estimate the accuracy of TRUS in addition to the PAG detection assays. Therefore, the usual approach was not possible and a Bayesian latent class analysis was chosen. The assumptions of these models must be assessed in an effort to evaluate the validity of presented results (Toft et al., 2005). These assumptions include different pregnancy proportions across all modeled populations, equal sensitivity and specificity across all modeled populations, conditional independence between tests, and appropriate prior probability specification. The results of the Bayesian model suggested that the pregnancy proportion varied among herds and it would be unusual for two herds to have exactly the same pregnancy proportions due to differences in heat detection, nutrition, breeding strategies, and other important management factors. The sensitivity and specificity of TRUS was independently estimated within each herd but it was necessary to assume that the accuracy of PAG detection did not vary by herd. The accuracy of PAG detection can vary by parity, breed, and DIM (Ricci et al., 2015) so it is theoretically possible that the overall results could be confounded by variations in these factors among study herds. However, the very high overall estimates of sensitivity and specificity suggests that this bias is likely quite small. The model included conditional dependence terms to relax the assumption of conditional independence and the sensitivity analysis showed that the prior probabilities did not have an undue influence on model results. The evaluation of convergence also suggested that the model was consistent with the data but inferences might have changed if a different modeling approach was employed.

The participating veterinarians had variable experience in the use of TRUS and the fact that veterinarians were not selected at random suggests that accuracy estimates of TRUS are likely affected by this selection bias. Furthermore, veterinarians were requested to enroll all suitable cows that were presented for pregnancy diagnosis but we have no data to verify

whether this design requirement was followed. It is theoretically possible that veterinarians excluded cows for reasons that were not recorded and this is another source of selection bias that should be considered

These results represent findings of a latent class approach to estimating the accuracy of both direct and indirect methods of pregnancy diagnosis in dairy cattle. The precision of our reported estimates for the PAG detection assays were within our desired level of precision (+/- 5%) but TRUS accuracy was not estimated at a similar level of precision. This likely occurred due to the stratification by veterinarian, results being markedly different than sample size assumptions, and extensive variability in the pregnancy proportions on each farm. Despite the discussed limitations, latent class models likely provide more reliable estimates of sensitivity and specificity for commercial PAG detection assays than what has been previously estimated relative to TRUS, which is an imperfect reference standard.

5. Conclusions

Indirect detection of pregnancy in dairy cattle using PAG ELISA is an accurate method for use in dairy cattle. The method is descriptively more sensitive than veterinarian-performed TRUS at 28-35 days post-AI and therefore could be an economically viable addition to a reproductive management plan. However, the potential for false-positive results after pregnancy loss due to a relatively long half-life in the maternal circulation must be considered. Veterinarians will continue to be an integral component of the design, implementation and monitoring of reproductive management plans, because PAG detection does not detect reproductive pathology or other causes of reproductive failures due to cow health. There is the potential for milk PAG detection to be included in automated systems within the milking parlor but further work is required due to the lower levels detected in milk.

Conflict of interest

The research was funded by a cooperative agreement with IDEXX Laboratories. The funding agency did not influence the design, conduct, analysis, or reporting of the study.

Acknowledgements

This work was funded by cooperative agreement with IDEXX laboratories. We also acknowledge the assistance of the veterinary practitioners that were involved in study design and data collection: Dr. Stephen Hughes (Howick Veterinary Clinic), Drs. Louis Hoek and Alf Lategan (Cape Cross Veterinary Services), Dr. Gordon Adam (Tygerberg Animal Hospital, Darling), Dr. Brendan Farrell (Bosberg Veterinary Clinic), and Dr. Frikkie Nel (Standerton Dierehospitaal). We are extremely grateful for the laboratory analyses performed by Annicky Modirwa, Stephanie Ganswindt, and Hettie Rossouw. The study would not have been possible without the gracious funding of IDEXX and the hard work of Brad Roe as the IDEXX Business Project Manager.

References

Altman, D.G., 1991. Practical statistics for medical research. Chapman and Hall London; New York.

Barkema, H.W., von Keyserlingk, M.A., Kastelic, J.P., Lam, T.J., Luby, C., Roy, J.P., LeBlanc, S.J., Keefe, G.P., Kelton, D.F., 2015. Invited review: Changes in the dairy industry affecting dairy cattle health and welfare. J. Dairy Sci. 98, 7426-7445.

Byrt, T., Bishop, J., Carlin, J.B., 1993. Bias, prevalence and kappa. J. Clin. Epidemiol. 46, 423-429.

Commun, L., Velek, K., Barbry, J.B., Pun, S., Rice, A., Mestek, A., Egli, C., Leterme, S., 2016. Detection of pregnancy-associated glycoproteins in milk and blood as a test for early pregnancy in dairy cows. J. Vet. Diagn. Invest. 28, 207-213.

Dufour, S., Durocher, J., Dubuc, J., Dendukuri, N., Hassan, S., Buczinski, S., 2017. Bayesian estimation of sensitivity and specificity of a milk pregnancy-associated glycoprotein-based

ELISA and of transrectal ultrasonographic exam for diagnosis of pregnancy at 28-45 days following breeding in dairy cows. Prev. Vet. Med. 140, 122-133.

Enoe, C., Georgiadis, M.P., Johnson, W.O., 2000. Estimation of sensitivity and specificity of diagnostic tests and disease prevalence when the true disease state is unknown. Prev. Vet. Med. 45, 61-81.

Fosgate, G.T., 2005. Modified exact sample size for a binomial proportion with special emphasis on diagnostic test parameter estimation. Stat. Med. 24, 2857-2866.

Fosgate, G.T., 2009. Practical sample size calculations for surveillance and diagnostic investigations. J. Vet. Diagn. Invest. 21, 3-14.

Fosgate, G.T., Adesiyun, A.A., Hird, D.W., Johnson, W.O., Hietala, S.K., Schurig, G.G., Ryan, J., 2002. Comparison of serologic tests for detection of Brucella infections in cattle and water buffalo (Bubalus bubalis). Am. J. Vet. Res. 63, 1598-1605.

Fosgate, G.T., Urdaz-Rodriguez, J.H., Dunbar, M.D., Rae, D.O., Donovan, G.A., Melendez, P., Dobek, G.L., Alleman, A.R., 2010. Diagnostic accuracy of methods for detecting Anaplasma marginale infection in lactating dairy cattle of Puerto Rico. J. Vet. Diagn. Invest. 22, 192-199.

Franco, O.J., Drost, M., Thatcher, M.J., Shille, V.M., Thatcher, W.W., 1987. Fetal survival in the cow after pregnancy diagnosis by palpation per rectum. Theriogenology 27, 631-644.

Fricke, P.M., Ricci, A., Giordano, J.O., Carvalho, P.D., 2016. Methods for and Implementation of Pregnancy Diagnosis in Dairy Cows. Vet. Clin. North Am. Food Anim. Pract. 32, 165-180.

Giordano, J.O., Fricke, P.M., Cabrera, V.E., 2013. Economics of resynchronization strategies including chemical tests to identify nonpregnant cows. J. Dairy Sci. 96, 949-961.

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

Humblot, P., 2001. Use of pregnancy specific proteins and progesterone assays to monitor pregnancy and determine the timing, frequencies and sources of embryonic mortality in ruminants. Theriogenology 56, 1417-1433.

Karen, A., Sousa, N.M., Beckers, J.F., Bajcsy, A.C., Tibold, J., Madl, I., Szenci, O., 2015. Comparison of a commercial bovine pregnancy-associated glycoprotein ELISA test and a pregnancy-associated glycoprotein radiomimmunoassay test for early pregnancy diagnosis in dairy cattle. Anim. Reprod. Sci. 159, 31-37.

Lagrange, V., Whitsett, D., Burris, C., 2015. Global market for dairy proteins. J. Food Sci. 80 Suppl. 1, A16-22.

Lawson, B.C., Shahzad, A.H., Dolecheck, K.A., Martel, E.L., Velek, K.A., Ray, D.L., Lawrence, J.C., Silvia, W.J., 2014. A pregnancy detection assay using milk samples: evaluation and considerations. J. Dairy Sci. 97, 6316-6325.

Leblanc, S.J., 2013. Short communication: field evaluation of a pregnancy confirmation test using milk samples in dairy cows. J. Dairy Sci. 96, 2345-2348.

Lopez-Gatius, F., Garbayo, J.M., Santolaria, P., Yaniz, J., Ayad, A., de Sousa, N.M., Beckers, J.F., 2007. Milk production correlates negatively with plasma levels of pregnancy-associated glycoprotein (PAG) during the early fetal period in high producing dairy cows with live fetuses. Domest. Anim. Endocrinol. 32, 29-42.

Oltenacu, P.A., Ferguson, J.D., Lednor, A.J., 1990. Economic evaluation of pregnancy diagnosis in dairy cattle: a decision analysis approach. J. Dairy Sci. 73, 2826-2831.

Pohler, K.G., Peres, R.F., Green, J.A., Graff, H., Martins, T., Vasconcelos, J.L., Smith, M.F., 2016. Use of bovine pregnancy-associated glycoproteins to predict late embryonic mortality in postpartum Nelore beef cows. Theriogenology 85, 1652-1659.

R Core Team, 2017. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Ricci, A., Carvalho, P.D., Amundson, M.C., Fourdraine, R.H., Vincenti, L., Fricke, P.M., 2015. Factors associated with pregnancy-associated glycoprotein (PAG) levels in plasma and milk of Holstein cows during early pregnancy and their effect on the accuracy of pregnancy diagnosis. J. Dairy Sci. 98, 2502-2514.

Romano, J.E., Larson, J.E., 2010. Accuracy of pregnancy specific protein-B test for early pregnancy diagnosis in dairy cattle. Theriogenology 74, 932-939.

Romano, J.E., Thompson, J.A., Forrest, D.W., Westhusin, M.E., Tomaszweski, M.A., Kraemer, D.C., 2006. Early pregnancy diagnosis by transrectal ultrasonography in dairy cattle. Theriogenology 66, 1034-1041.

Thurmond, M.C., Picanso, J.P., 1993. Fetal loss associated with palpation per rectum to diagnose pregnancy in cows. J. Am. Vet. Med. Assoc. 203, 432-435.

Toft, N., Jorgensen, E., Hojsgaard, S., 2005. Diagnosing diagnostic tests: evaluating the assumptions underlying the estimation of sensitivity and specificity in the absence of a gold standard. Prev. Vet. Med. 68, 19-33.

Vacek, P.M., 1985. The effect of conditional dependence on the evaluation of diagnostic tests. Biometrics 41, 959-968.

Villarroel, A., McDonald, S.R., Walker, W.L., Kaiser, L., Dewell, R.D., Dewell, G.A., 2010. A survey of reasons why veterinarians enter rural veterinary practice in the United States. J. Am. Vet. Med. Assoc. 236, 849-857.

von Keyserlingk, M.A., Martin, N.P., Kebreab, E., Knowlton, K.F., Grant, R.J., Stephenson, M., Sniffen, C.J., Harner, J.P., 3rd, Wright, A.D., Smith, S.I., 2013. Invited review: Sustainability of the US dairy industry. J. Dairy Sci. 96, 5405-5425.

Whitlock, B.K., Maxwell, H.S., 2008. Pregnancy-associated glycoproteins and pregnancy wastage in cattle. Theriogenology 70, 550-559.

Wickham, H., 2009. Ggplot2 : elegant graphics for data analysis.

Wooding, F.B., Roberts, R.M., Green, J.A., 2005. Light and electron microscope immunocytochemical studies of the distribution of pregnancy associated glycoproteins (PAGs) throughout pregnancy in the cow: possible functional implications. Placenta 26, 807-827.

Zoli, A.P., Guilbault, L.A., Delahaut, P., Ortiz, W.B., Beckers, J.F., 1992. Radioimmunoassay of a bovine pregnancy-associated glycoprotein in serum: its application for pregnancy diagnosis. Biol. Reprod. 46, 83-92.

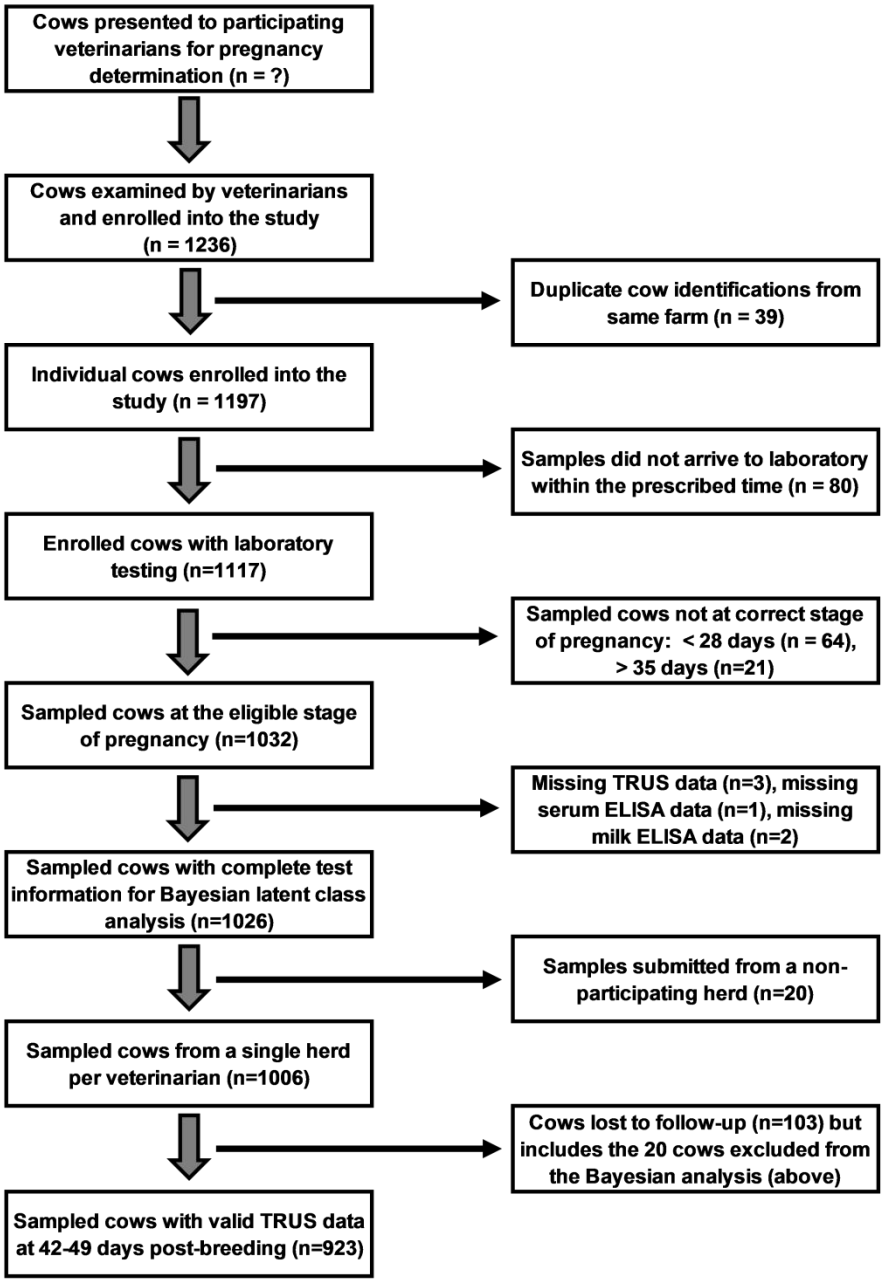

Figure 1. Flow diagram for the cows sampled from commercial dairy farms in South Africa for the evaluation of PAG detection ELISA (serum and milk) and veterinarian-performed transrectal unltrasound examination (TRUS).

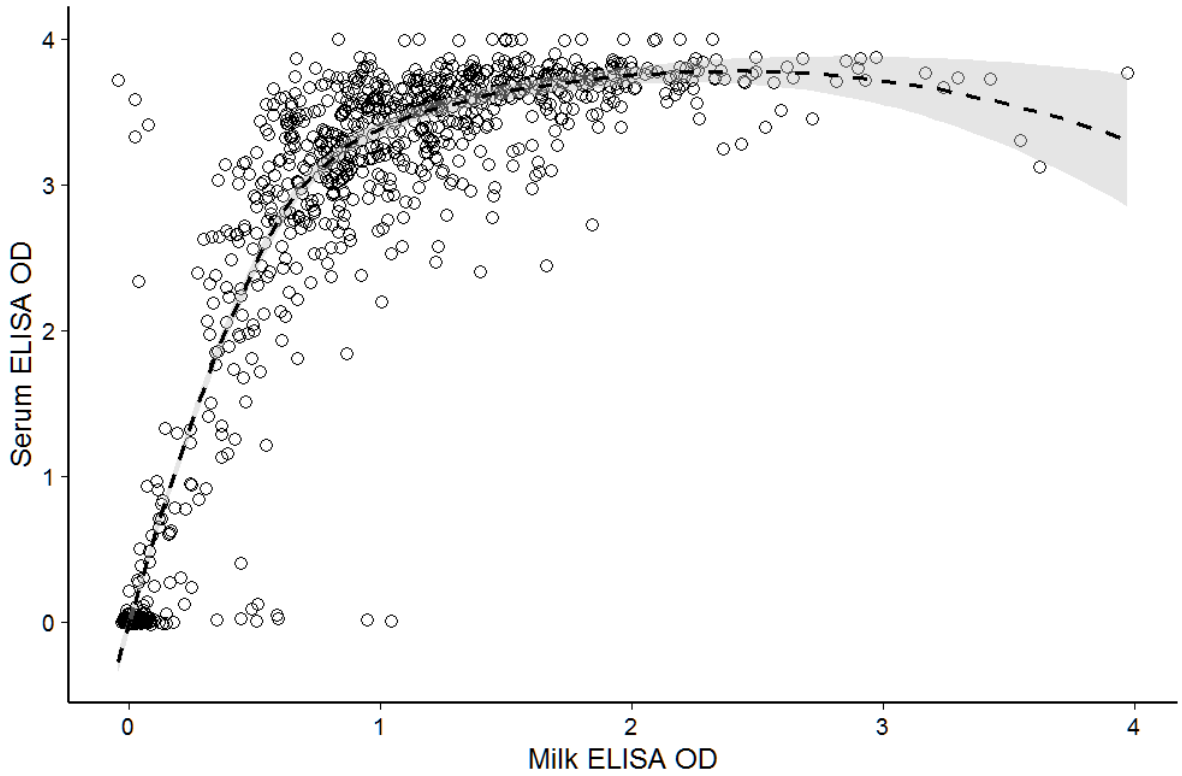

Figure 2. Correlation between the quantitative serum and milk pregnancy-associated glycoprotein detection ELISA results in 1029 dairy cows sampled 28-35 days post artificial insemination from 5 agricultural regions in South Africa (Spearman's rho = 0.858, P < 0.001).

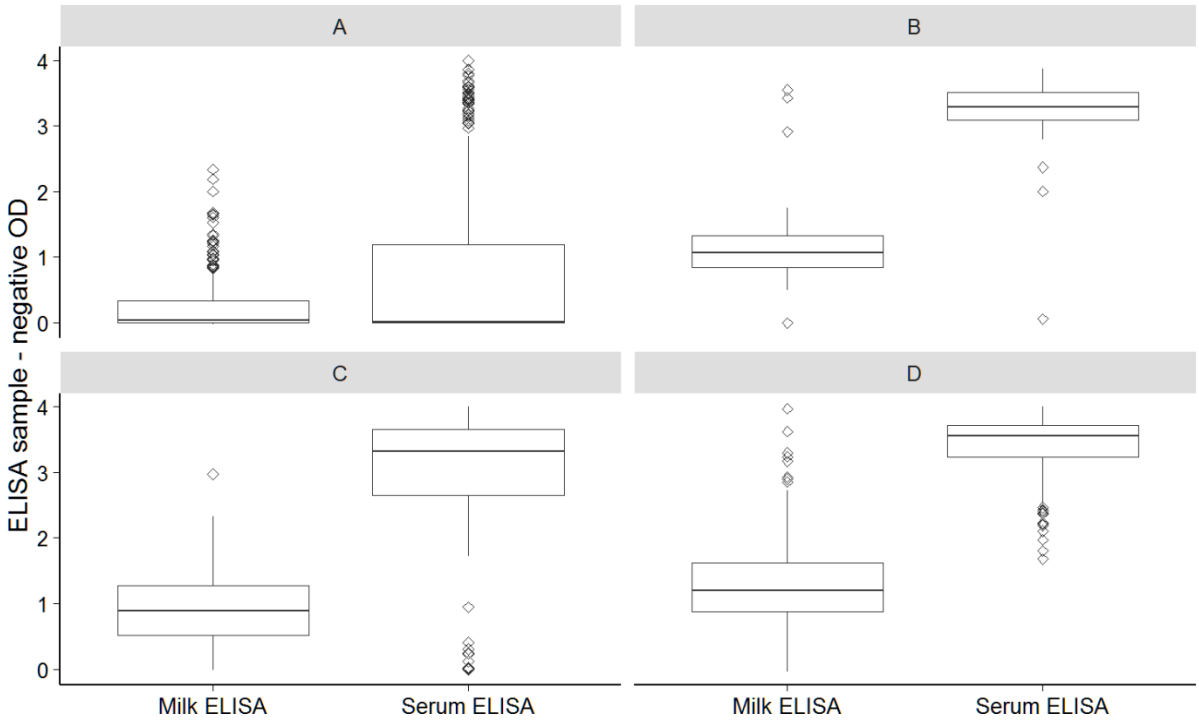

Figure 3. Descriptive results for the quantitative serum and milk ELISA optical densities (OD) categorized by transrectal ultrasound examination (TRUS) findings performed at 28-35 and 42-49 days post-breeding in 923 cows sampled from five agricultural regions in South Africa. Categories are listed as TRUS findings at 28-35 days post-breeding followed by the second TRUS examination two weeks later (42-49 days post-breeding): Open – Open (A); Open – Pregnant (B); Pregnant – Open (C); Pregnant – Pregnant (D). Diamonds represent outlier values.

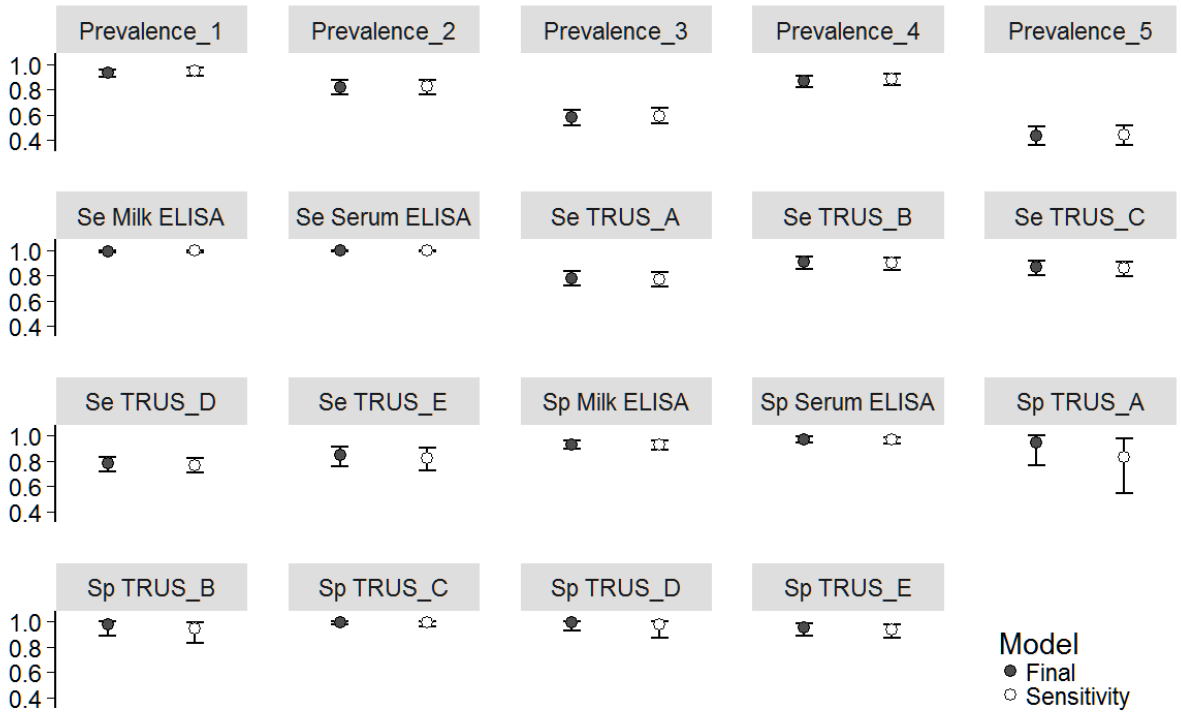

Figure 4. Medians (circles) and 95% probability intervals (caps) for the final Bayesian latent class and sensitivity analysis models estimating the sensitivity (Se) and specificity (Sp) of transrectal ultrasound examination (TRUS) at 28-35 days post-breeding for five veterinarians (TRUS_A – TRUS_E), pregnancy-associated glycoprotein detection ELISA (serum and milk), and pregnancy proportions (Prevalence_1 – Prevalence_5) in 1006 dairy cows from five commercial dairy herds in South Africa. Sensitivity analysis incorporated non-informative priors (beta 1,1) for all model parameters.

Tables

Table 1. Cross-classified test results for pregnancy detection using transrectal ultrasound examination (TRUS) and pregnancy-associated glycoprotein (PAG) detection at 28-35 days post-breeding in 1006 dairy cows from five commercial dairy herds in South Africa.

	PAG detection		TRUS	
Veterinarian	Serum ELISA	Milk ELISA	Pregnant	Open
A	O	O	1	8
	O	P	2	2
	P	O	2	2
	P	P	160	48
B	O	O	1	25
	O	P	0	2
	P	O	0	2
	P	P	120	13
C	O	O	0	91
	O	P	0	4
	P	O	0	4
	P	P	121	20
D	O	O	0	23
	O	P	0	1
	P	O	0	1
	P	P	152	44
E	O	O	4	78
	O	P	1	6
	P	O	2	1
	P	P	57	12

O = open. P = pregnant.

Table 2. Agreement among methods for pregnancy detection using transrectal ultrasound examination (TRUS) at two times post-breeding and pregnancy-associated glycoprotein (PAG) detection in 1029 dairy cows from five agricultural regions in South Africa.

Test 1	Test 2	n	Agreement (%)	Kappa (95% CI)	PABAK
TRUS at 28-35 days	TRUS at 42-49 days	926	88	0.76 (0.72, 0.80)	0.77
	Visual PAG ELISA	1025	85	0.65 (0.61, 0.70)	0.70
	Serum PAG ELISA	1028	85	0.66 (0.62, 0.71)	0.70
	Milk PAG ELISA	1027	84	0.64 (0.59, 0.69)	0.68
TRUS at 42-49 days	Visual PAG ELISA	925	80	0.58 (0.53, 0.63)	0.61
	Serum PAG ELISA	928	82	0.60 (0.55, 0.65)	0.63
	Milk PAG ELISA	927	81	0.58 (0.53, 0.63)	0.61
Visual PAG ELISA	Serum PAG ELISA	1028	99	0.97 (0.951, 0.987)	0.98
	Milk PAG ELISA	1026	97	0.92 (0.89, 0.94)	0.94
Serum PAG ELISA	Milk PAG ELISA	1029	97	0.92 (0.90, 0.95)	0.94

CI = confidence interval. PABAK = prevalence-adjusted and bias-adjusted kappa.

Table 3. Prior and posterior distributions for pregnancy proportions and the accuracy of pregnancy detection for transrectal ultrasound examination (TRUS) and pregnancy-associated glycoprotein (PAG) detection assays based on a Bayesian latent class analysis of data from 1006 dairy cows within South Africa.

		Prior	Posterior
Population or test	**Measure**	**Median (95% PI)**	**Median (95% PI)**
Serum ELISA	Sensitivity*	0.926 (0.664, 0.997)	0.994 (0.985, 0.999)
	Specificity*	0.976 (0.762, 1.0)	0.974 (0.947, 0.992)
Milk ELISA	Sensitivity*	0.926 (0.664, 0.997)	0.992 (0.982, 0.998)
	Specificity*	0.976 (0.762, 1.0)	0.934 (0.897, 0.961)
Serum x milk dependence	Sensitivity	Uniform†	9.1×10^{-4} (5.7×10^{-6}, 4.0×10^{-3})
	Specificity	Uniform‡	-3.9×10^{-5} (-2.6×10^{-3}, 4.4×10^{-3})
Veterinarian A TRUS	Sensitivity	0.926 (0.664, 0.997)	0.778 (0.720, 0.830)
	Specificity	0.976 (0.762, 1.0)	0.947 (0.762, 1.0)
Veterinarian B TRUS	Sensitivity	0.926 (0.664, 0.997)	0.905 (0.849, 0.946)
	Specificity	0.976 (0.762, 1.0)	0.976 (0.888, 1.0)
Veterinarian C TRUS	Sensitivity	0.926 (0.664, 0.997)	0.864 (0.802, 0.913)
	Specificity	0.976 (0.762, 1.0)	0.998 (0.976, 1.0)
Veterinarian D TRUS	Sensitivity	0.926 (0.664, 0.997)	0.780 (0.719, 0.833)
	Specificity	0.976 (0.762, 1.0)	0.994 (0.930, 1.0)

Veterinarian E TRUS	Sensitivity	0.926 (0.664, 0.997)	0.846 (0.755, 0.914)
	Specificity	0.976 (0.762, 1.0)	0.947 (0.887, 0.983)
Farm 1 pregnant cows	Proportion	0.736 (0.522, 0.892)	0.933 (0.896, 0.961)
Farm 2 pregnant cows	Proportion	0.735 (0.490, 0.907)	0.820 (0.758, 0.872)
Farm3 pregnant cows	Proportion	0.422 (0.228, 0.635)	0.578 (0.517, 0.637)
Farm 4 pregnant cows	Proportion	0.628 (0.395, 0.826)	0.866 (0.819, 0.906)
Farm 5 pregnant cows	Proportion	0.326 (0.129, 0.579)	0.432 (0.359, 0.507)

PI = probability interval.

*Reported as the weighted estimates rather than stratum-specific values that could be calculated through the addition and subtraction of conditional dependence terms included in the final Bayesian model.

†Distribution varied based on the acceptable range of values calculated as: maximum (-(1-Se[serum])*(1-Se[milk]) and -Se[serum]*Se[milk]) to minimum (Se[serum]*(1-Se[milk]) and Se[serum]*(1-Se[milk])); where Se = sensitivity at each iteration.

‡ Distribution varied based on the acceptable range of values calculated as: maximum (-(1-Sp[serum])*(1-Sp[milk]) and -Sp[serum]*Sp[milk]) to minimum (Sp[serum]*(1-Sp[milk]) and Sp[serum]*(1-Sp[milk])); where Sp = specificity at each iteration.