

# Evaluation of the diagnostic performance of semen morphologic examination in the bull breeding soundness evaluation using Bayesian latent class analysis

T.G. Gunderson,<sup>1</sup> DVM, MS, ACVPM-Epidemiology; K.K. Shuck<sup>2</sup>; B.L. Vander Ley,<sup>2</sup> DVM, PhD, ACVPM; D.R. Smith,<sup>3</sup> DVM, PhD, ACVPM-Epidemiology

<sup>1</sup>Department of Clinical Sciences, College of Veterinary Medicine, Kansas State University, Manhattan, KS 66506

<sup>2</sup>Great Plains Veterinary Educational Center, School of Veterinary Medicine and Biomedical Sciences, University of Nebraska-Lincoln, Clay Center, NE 68933

<sup>3</sup>Department of Pathobiology and Population Medicine, College of Veterinary Medicine, Mississippi State University, Mississippi State, MS 39762

\*Corresponding author: Dr. David Smith, [DSmith@cvm.msstate.edu](mailto:DSmith@cvm.msstate.edu)

## Abstract

The objective of this study was to estimate the sensitivity and specificity of the semen morphology examination in the bull breeding soundness evaluation (BSE) for classifying *Bos taurus* bulls as subfertile. Semen, collected during BSE from beef bulls at the U.S. Meat Animal Research Center in Clay Center, Neb., was evaluated using eosin-nigrosin stained slides at 1000X magnification by an experienced technician (Evaluator A) and an experienced veterinarian (Evaluator B). Bulls with slides that had < 70% morphologically normal sperm were classified as subfertile. Inter-evaluator agreement was assessed using Cohen's kappa coefficient. Sensitivity, specificity and true prevalence were estimated using Bayesian latent class analysis. Sensitivity was defined as the probability of an evaluator correctly identifying a bull as subfertile, and specificity was defined as the probability of correctly identifying a bull as satisfactory. Of 478 bulls evaluated for semen morphology, 68 (14%) were classified as subfertile by Evaluator A and 96 (20%) were classified as subfertile by Evaluator B. The kappa coefficient for inter-evaluator agreement was 0.5. The estimated true prevalence of subfertile bulls was 16% (95% C.I. 6-32%). The estimated sensitivity and specificity for Evaluator A were 52% (95% C.I. 22-84%) and 93% (95% C.I. 86-99%) respectively, and for Evaluator B were 62% (95% C.I. 31-89%) and 88% (95% C.I. 80-97%). These results show similar test sensitivity and specificity between the two evaluators. Diagnostic value of the morphologic examination was slightly greater than knowledge of bull fertility status prior to testing. Given a practitioner's assessment of pre-test probability that a bull is subfertile, these diagnostic parameters can be used to predict the probability that the morphology portion of the BSE accurately predicts subfertility.

**Key words:** bull breeding soundness diagnostic performance

## Introduction

The breeding soundness evaluation (BSE), as defined by the Society for Theriogenology (SFT), is a diagnostic test that classifies bulls as either satisfactory potential breeders, or some form of non-satisfactory potential breeder (i.e., unsatisfactory potential breeder, or classification deferred).<sup>1</sup> Due to factors that influence the ability of a bull to sire calves that cannot

practically be evaluated on a BSE, such as libido, serving capacity, and social dominance rank, it has long been recognized that the main value of the BSE lies not in its ability to identify fertile bulls, but to identify subfertile and infertile bulls.<sup>2-4</sup> The rationale behind having multiple non-satisfactory categories is that some manifestations of subfertility are not permanent, and that a bull may return to a state of being a satisfactory potential breeder with time and/or treatment.<sup>2-4</sup> Inherent to this rationale is the recognition that the predictive value of a non-satisfactory diagnosis is imperfect.<sup>3,4</sup>

Multiple studies have demonstrated the existence of bulls that produce relatively few calves, despite multiple breeding opportunities, demonstrating that fertility in bulls follows a gradient and is not a binary matter of being able to impregnate females or not.<sup>5-8</sup> Furthermore, many bulls that fail a BSE are still capable of impregnating healthy cycling females, especially if given enough time and mating opportunities.<sup>7,9</sup>

Perhaps a more useful way to define fertility as a binary outcome would be to specify that a bull be able to impregnate a certain number of females within a defined time frame. The authors of a 2008 review stated that the standards of the BSE, as defined by the SFT, were intended to assess the likelihood of a bull establishing pregnancy in  $\geq 25$  healthy, cycling females in a 65-70-day breeding season.<sup>3</sup> To the best of our knowledge, the BSE being based on this standard is not the official position of the SFT<sup>a</sup>. Notwithstanding, this is a potentially useful definition of fertility, as a cow:bull ratio of approximately 25:1 is commonly used in pasture breeding systems.<sup>10,11</sup>

Based on this standard, a bull that successfully sires 25 calves within a breeding season would be considered fertile, and one that is incapable of doing so would be considered subfertile. Given this definition of bull fertility, the predictive value of the BSE can be assessed by first determining the test's sensitivity, the probability of the BSE correctly identifying truly subfertile bulls, and test specificity, the probability of the BSE correctly identifying truly fertile bulls; in essence determining the sensitivity and specificity of the BSE for detecting subfertile bulls. Once the sensitivity and specificity of the BSE have been determined, the predictive values of a positive or negative test result can be estimated based on pre-test probability of a bull being subfertile.

Sensitivities and specificities have traditionally been calculated by creating  $2 \times 2$  contingency tables that compare a test result to a known health condition. The known health condition is often determined by a theoretical “gold standard” test result. In the case of bull fertility, the gold standard is difficult to assess. While it is possible to determine the calf output from individual bulls, it is often impractical in multi-sire pastures, and is only a useful for assessing the diagnostic performance of the BSE if bulls are turned out with cows regardless of the test result. Furthermore, because there are multiple factors that affect calf output that are independent of a bull’s inherent fertility, failure of a bull to sire 25 calves within a defined breeding season does not necessarily mean he was incapable of doing so, and therefore it cannot be used as a gold standard.

One approach to evaluating the performance of diagnostic tests in the absence of a true gold standard test is Bayesian latent class analysis (BLCA).<sup>12</sup> Bayesian latent class analysis uses Bayes’ theorem to determine certain probabilities (e.g., sensitivity and specificity) that are latent within the comparison of results between at least 2 different diagnostic tests attempting to categorize the same outcome. These tests may be independent of each other, or dependent. However, if results for both tests are known and informative prior distributions for the latent parameters exist, Bayes’ theorem can be used as a framework for determining the sensitivity and specificity of both tests, along with the prevalence of the condition. While the resulting values would only be an estimate for the latent parameters, taking thousands of random samples from the posterior distribution of the model (i.e., Markov Chain Monte Carlo simulation) will create a distribution of estimates that will center around the most likely estimates and also enable the calculation of credible intervals around these mode values.<sup>12-15</sup>

It is necessary to have results from at least 2 tests to complete a BLCA. In this case, BLCA could be conducted from the results of 2 different evaluators performing the BSE, or critical aspects of it, independently of each other. Because semen morphology is the factor that results in the vast majority of bulls being classified as non-satisfactory on their BSE, and because it is highly correlated with semen motility and scrotal circumference, this would seem like a reasonable candidate for independent evaluation.<sup>10,16</sup> Therefore, the objective of this study was to use BLCA to estimate the test performance of the semen morphology evaluation portion of the BSE from 2 independent evaluators using separate stained slides.

## Materials and methods

### Data collection

The target population for this study was *Bos taurus* beef bulls. Our sampling frame was the population of beef bulls at the U.S. Meat Animal Research Center (USMARC) in Clay Center, Neb. The vast majority of bulls in this population are of *Bos taurus* lineage, though a few (5-10%) have some *Bos indicus* influence. The research described in this study was covered by USMARC Institutional Animal Care outline number 154.1. We chose this population of bulls because it represented our target population and had a sufficient number of bulls to provide a reasonable likelihood of model convergence.<sup>14,15</sup> We collected semen samples from these bulls in April, 2023 as part of their annual BSE testing program. Semen from bulls that met physical exam criteria were smeared onto microscope slides and stained with eosin-nigrosin stain; progressive sperm

motility and scrotal circumference were also recorded.<sup>1,2</sup> A registered veterinary technician with 40 years of experience evaluating semen morphology (Evaluator A) evaluated one set of smears on-site and recorded the number of normal spermatozoa out of 100 counted. A second set of semen smears from each bull was cover-slipped using a xylene-based mountant and evaluated in like fashion off-site by a veterinarian with 10 years of experience evaluating semen morphology (Evaluator B). The slides evaluated by Evaluator A were not cover-slipped at the time of evaluation.

### Data analysis

The data from both evaluators were entered and categorized in spreadsheet software<sup>b</sup>. The agreement in percent morphologically normal sperm recorded by both evaluators was assessed using Spearman’s rank correlation. Semen morphology results from the 2 evaluators were categorized as satisfactory and non-satisfactory based on whether the samples had a minimum of 70% morphologically normal spermatozoa (i.e., semen samples with a percentage of morphologically normal spermatozoa below 70% were categorized as non-satisfactory). The categorical results from both evaluators were cross-tabulated, and agreement between evaluators was calculated using Cohen’s kappa coefficient.<sup>17</sup> Only the morphology assessment was used for categorizing the bulls (i.e., bull semen morphology was categorized as satisfactory or non-satisfactory regardless of sperm motility and scrotal circumference). Bayesian latent class analysis was used to estimate the diagnostic performance of the semen morphologic portion of the BSE (i.e., sensitivity to detect subfertile bulls, specificity to detect fertile bulls, and prevalence of subfertility). The cross tabulated results were uploaded into an open-source statistical software program<sup>c</sup>. For purposes of the latent class analysis, the on-site evaluations were considered “Test A” and the off-site evaluations were considered “Test B”. Because both sets of semen slides were prepared with semen from the same ejaculate, they were paired samples and the results of the morphology assessment were considered to be dependent. Beta distributions were selected as priors for the sensitivity and specificity of the test and were generated from published data from a population of *Bos indicus/Bos taurus* cross bulls in Texas.<sup>7</sup> In that study, bulls’ pre-breeding semen morphology and their calf-output as determined by DNA parentage verification were reported.<sup>7</sup> To estimate the sensitivity and specificity of the test from this data, we categorized bulls as fertile if they sired a minimum of 25 calves during their breeding season, and as subfertile if they failed to sire 25 calves. Given that a bull was subfertile, sensitivity was defined as the probability that a bull had a pre-breeding percentage of normal semen < 70%. Given that a bull was fertile, specificity was defined as the probability a bull had a pre-breeding percentage of normal semen  $\geq 70\%$ . To estimate the prior distribution for prevalence of subfertile bulls, we used published data from a study performed in western Canada on *Bos taurus* bulls.<sup>16</sup> In that study, bulls were categorized as subfertile if they failed a BSE. The shape hyperparameters for the prior distribution were obtained using a betabuster function within the statistical software package<sup>c</sup>. A Gibbs sampler program (JAGS) was used to construct the Bayesian graphical model and estimate the latent parameters by performing 3 separate Markov chain Monte Carlo simulations with 101,000 iterations each<sup>c</sup>. The JAGS model used a burn-in of 1,000 iterations. The modes of the distributions of estimates produced by the Markov chain Monte Carlo simulations were used as point estimates for the

latent parameters, and the credible intervals were the 2.5<sup>th</sup> and 97.5<sup>th</sup> percentiles of these distributions.

A sensitivity analysis was performed to assess how changing the hyperparameters of the prior distribution affected the outcome. The modes for the prior distribution of sensitivity and specificity were shifted 10 percentage points in either direction, and the 2.5<sup>th</sup>/97.5<sup>th</sup> percentiles were shifted away from the mode by up to 20 percentage points (depending on the skewness of the distribution and proximity to the extremes).

The positive and negative predictive values for each evaluator were estimated over the full range of possible pre-test probability of subfertility.

## Results

Out of 500 bulls evaluated, we obtained semen morphology observations from both evaluators on 478 bulls (96%). Bulls that had disqualifying physical defects, had below minimum scrotal circumference, or had very dilute semen were generally not evaluated for morphology. However, there was 1 yearling bull with a scrotal circumference of 30 cm that received a satisfactory morphologic evaluation by both evaluators. Out of the 488 bulls evaluated for semen motility by Evaluator A, 14 had less than 30% progressive sperm motility. Of these, 13 received semen morphology assessments by both evaluators. Out of these 13 samples, 2 were classified as having satisfactory morphology by both evaluators, 1 was classified as having satisfactory morphology by Evaluator A but not Evaluator B, and 1 was classified as having satisfactory morphology by Evaluator B but not Evaluator A. The remaining 9 were classified as non-satisfactory on morphology by both evaluators.

Of 478 bulls evaluated for semen morphology by both evaluators, 68 (14%) were classified as subfertile by Evaluator A and 96 (20%) were classified as subfertile by Evaluator B. Figure 1 shows the contingency table of semen classifications from both evaluators based solely on morphology. The Spearman rank correlation coefficient for both evaluators was 0.56, and the Cohen's kappa coefficient for inter-evaluator agreement was 0.5. The estimated true prevalence of subfertile bulls in the population studied was 16% (95% C.I. 6-32%). The estimated sensitivity and specificity for Evaluator A were 52% (95% C.I. 22-84%) and 93% (95% C.I. 86-99%), and for Evaluator B were 62% (95% C.I. 31-89%) and 88% (95% C.I. 80-97%), respectively. Figures 2, 3 and 4 show the informative prior distributions and conditional posterior distributions for the model's prevalence, sensitivity and specificity parameters, respectively. Sensitivity analysis of the results showed that the modes for the latent parameters (i.e., prevalence, sensitivity, and specificity) did not change meaningfully when the modes of the prior distributions were changed by 10% in either direction. Table 1 shows the values obtained for each latent parameter during the sensitivity analysis. The plots for the positive and negative predictive values for each evaluator are presented in Figure 5. At the prevalence of subfertility estimated by BLCA (~16%), the positive predictive value of a subfertile classification based on semen morphology was ~45-55% depending on the evaluator, and the negative predictive value was ~90-95%.

## Discussion

An important tenet of practicing evidence-based medicine is understanding how well a diagnostic test predicts an outcome of interest. In this study, the diagnostic test was the

morphology portion of the BSE; the condition being tested was evidence for subfertility due to abnormal sperm morphology. Many studies have evaluated the statistical association of semen morphology with calf output. However, few, if any, attempt to estimate the discrete probability of the BSE predicting subfertility.<sup>7-9,18-20</sup>

Furthermore, some of these studies demonstrate that the semen morphology portion of the BSE may misclassify the fertility status of some bulls.<sup>7-9,18</sup> This is relevant because BSE misclassification can have far reaching financial implications for cow-calf producers. The largest potential cost of BSE misclassification is failure of a bull to impregnate cows, even though he was classified as a satisfactory potential breeder. However, the opportunity cost of replacing bulls that are culled due to BSE misclassification is another significant financial burden for cow-calf producers that far exceeds the explicit cost of paying an evaluator to perform the test.<sup>21</sup> Additionally, the culling of potentially valuable sale bulls that were misclassified as subfertile constitutes a particularly frustrating loss for bull breeders. Conversely, buyers who purchase sale bulls that were misclassified as satisfactory may also incur large losses. For these reasons, it is important that evaluators have some understanding of the probability for the morphological portion of the BSE to misclassify bulls as either subfertile or fertile.

To help achieve this goal, this study estimated the diagnostic parameters of sensitivity and specificity of the semen morphology portion of the BSE for detecting subfertile bulls, as well as the prevalence of subfertile bulls within the study population. The point estimates for the sensitivity and specificity of the morphology portion of the BSE for detecting subfertility are unique to the individual evaluators that participated in this study, and may not be perfectly generalizable to other evaluators. However, the results of a 2011 study suggest that when practitioners perform the morphology portion of the BSE according to SFT standards, their classifications largely agreed with those of a reference lab.<sup>22</sup> It should be noted that this study assessed crude agreement, and did not account for agreement that may have occurred by chance. A more recent study compared the classification of bulls between 3 different expert evaluators based on the morphology portion of the BSE. Using Cohen's Kappa coefficient, this study revealed substantial variability in inter-evaluator agreement.<sup>23</sup> Notwithstanding, the current analysis suggests that, in general, the morphology portion of the BSE is more specific than it is sensitive for detecting subfertile bulls and that these parameters did not differ substantially between evaluators.

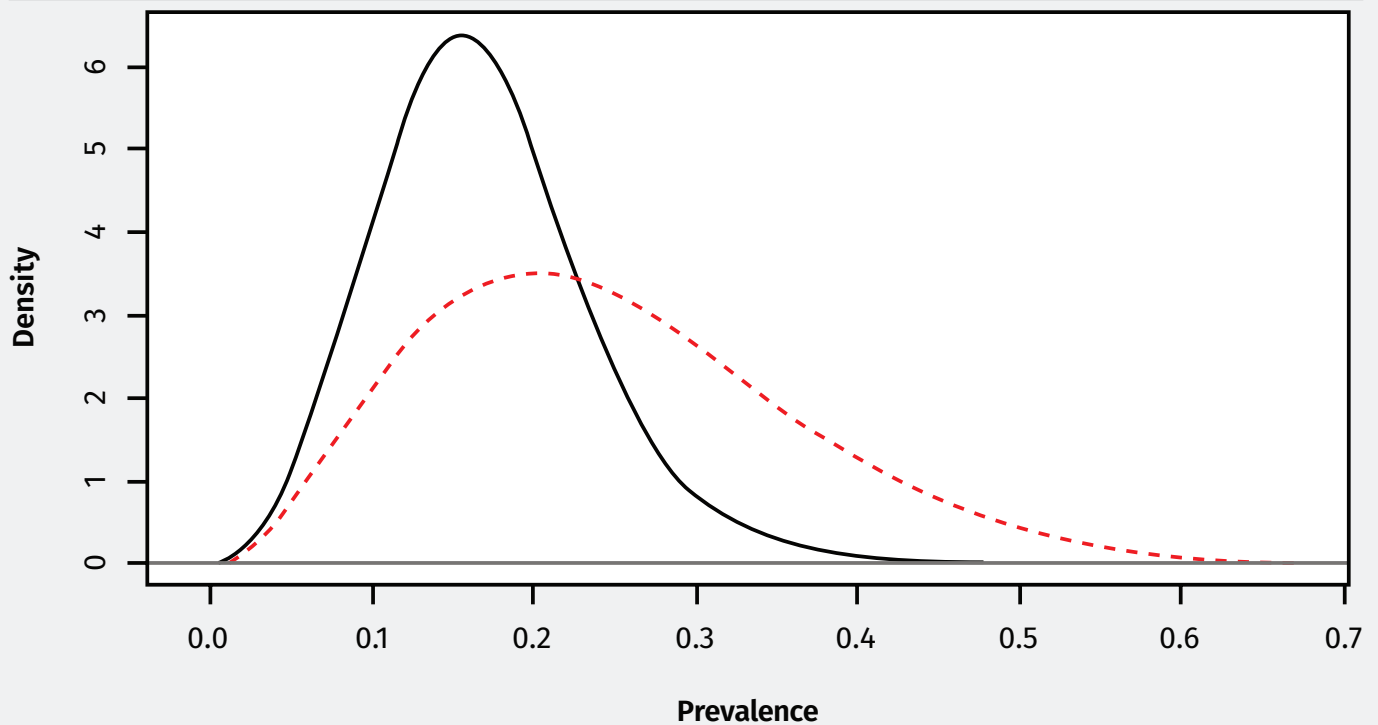
There are studies that attempt to estimate the prevalence of subfertility in bull populations, but in these studies, prevalence of subfertility was defined as either the proportion of bulls that fail a BSE, or the proportion of bulls that fail to sire a certain number of calves.<sup>5,7,9,16,22</sup> The point estimate for prevalence of subfertile bulls in the present analysis suggests that the proportion of bulls that fail the BSE may overestimate the true prevalence of subfertile bulls in a population, while the proportion of bulls that pass the morphological exam may underestimate the prevalence of potentially fertile bulls in a population.

It has been long held that the BSE may not be able to predict a bull's ability to sire calves as well as it can predict a bull's lack of ability. One way to assess the ability of a bull to sire calves, or lack of ability, is known pregnancy status of cows exposed

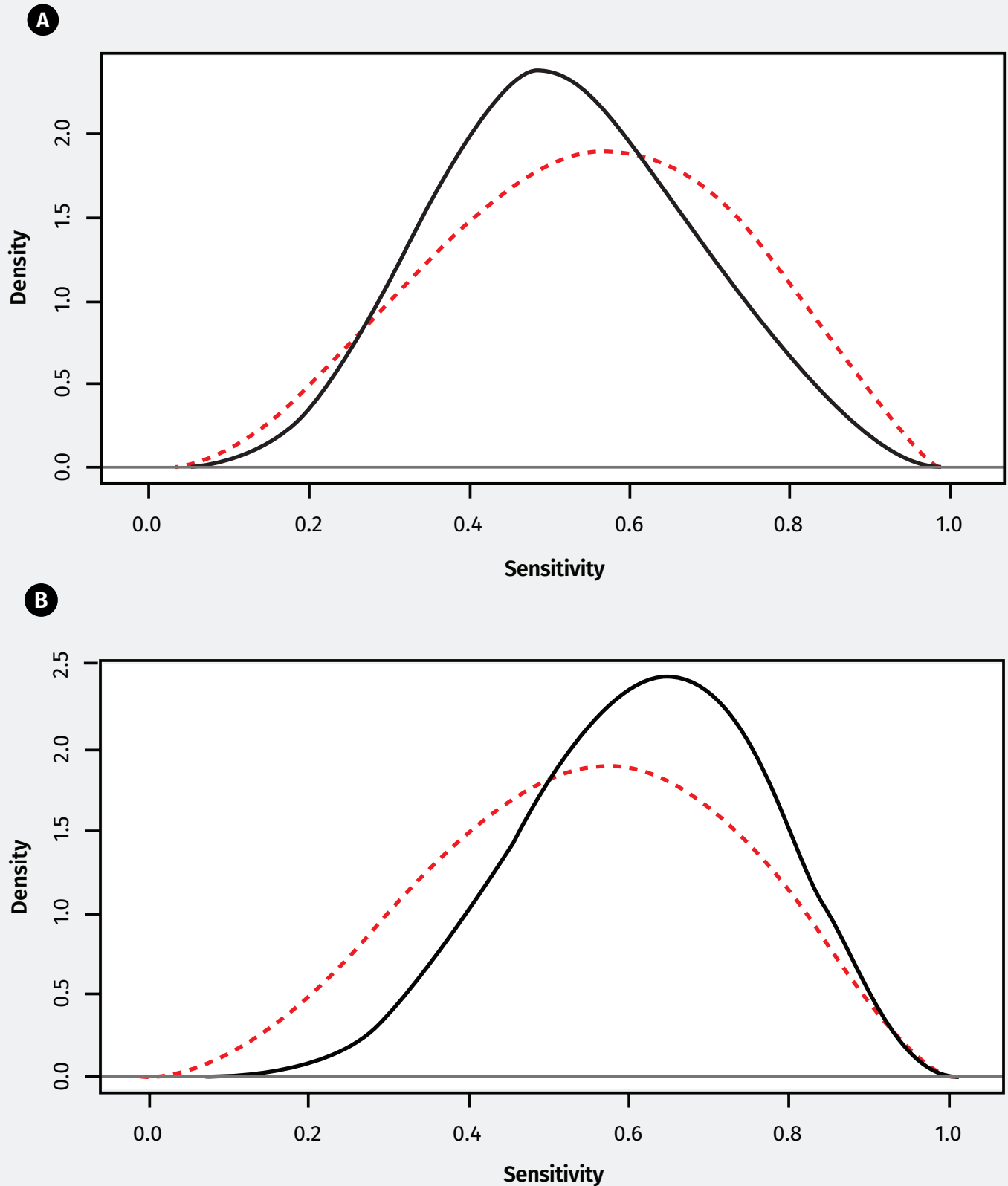
**Figure 1:** Contingency table comparing results from 2 separate evaluators who classified 478 bulls from the U.S. Meat Animal Research Center in Nebraska as satisfactory or non-satisfactory based on percentage of morphologically normal sperm.

	Non-satisfactory Evaluator A	Satisfactory Evaluator A	
Non-satisfactory Evaluator B	<b>48</b>	<b>48</b>	<b>96</b>
Satisfactory Evaluator B	<b>20</b>	<b>362</b>	<b>382</b>
	<b>68</b>	<b>410</b>	<b>478</b>

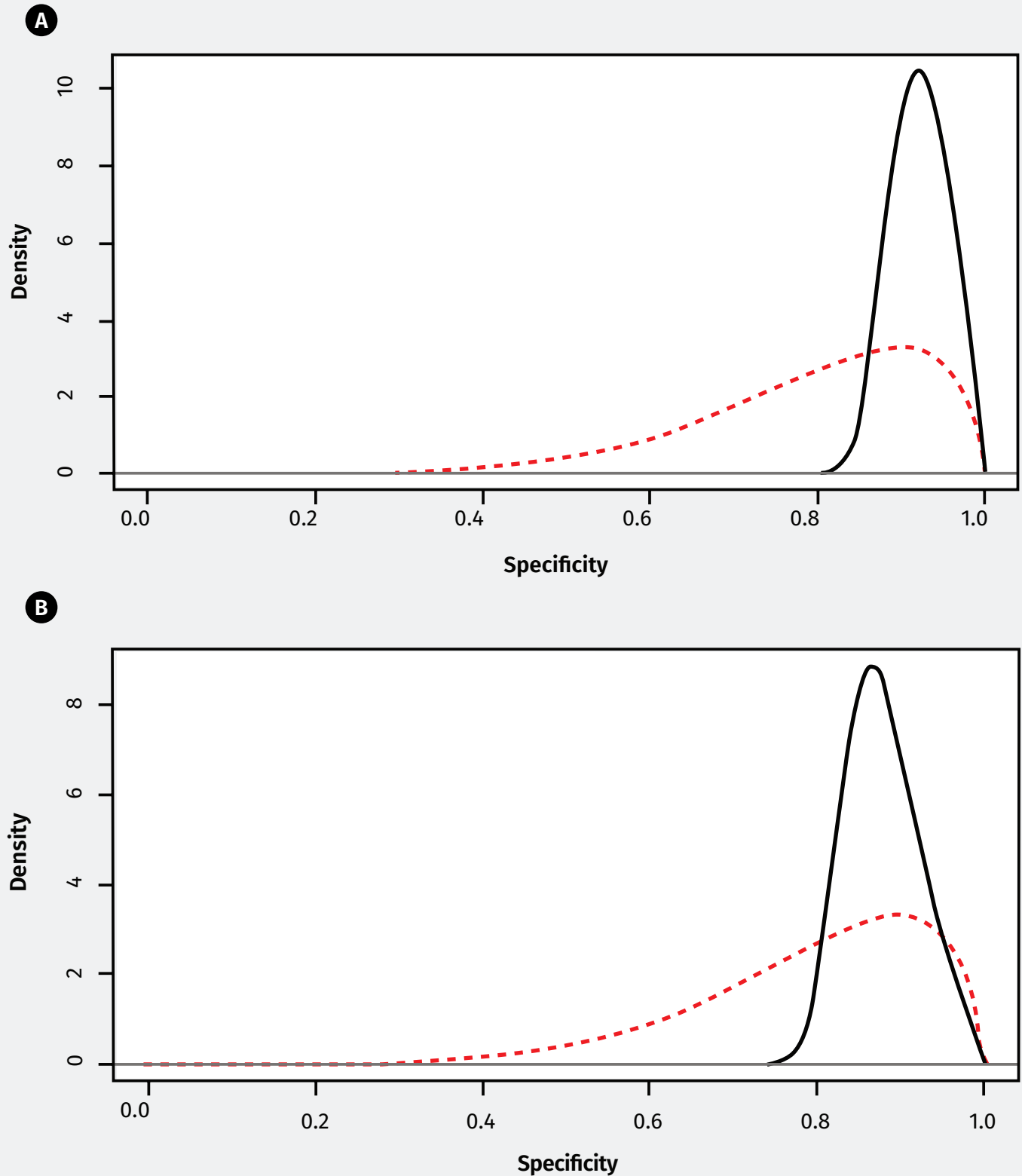
**Figure 2:** Prevalence of subfertile breeders in 478 predominately *Bos taurus* bulls from the U.S. Meat Animal Research Center in Nebraska based on Bayesian latent class analysis. The solid, black line represents the posterior distribution, with the peak representing the estimated prevalence. The red, dashed line represents the prior distribution used for Bayesian latent class analysis.



**Figure 3:** Sensitivity of the BSE to detect subfertile breeders for Evaluator A and Evaluator B based on Bayesian latent class analysis of semen morphology results from a population of 478 predominately *Bos taurus* bulls from the U.S. Meat Animal Research Center in Nebraska. The solid, black lines represent the posterior distributions, with the peaks representing the estimated sensitivity. The red, dashed lines represent the prior distribution used for Bayesian latent class analysis.



**Figure 4:** Specificity of the BSE to detect subfertile breeders for Evaluator A and Evaluator B based on Bayesian latent class analysis of semen morphology results from a population of 478 predominately *Bos taurus* bulls from the U.S. Meat Animal Research Center in Nebraska. The solid, black lines represent the posterior distributions, with the peaks representing the estimated sensitivity. The red, dashed lines represent the prior distributions used for Bayesian latent class analysis.



**Table 1:** Sensitivity analysis of the results of Bayesian latent class analysis to determine the diagnostic performance of the morphology portion of bull breeding soundness evaluations of 478 predominately *Bos taurus* bulls from the U.S. Meat Animal Research Center in Nebraska. The baseline diagnostic parameters represent the estimated prevalence of subfertile bulls in the population, as well as sensitivity and specificity of the semen morphology assessment of two separate evaluators for detecting subfertile bulls. Each column represents a separate sensitivity analysis wherein the mode of a prior distribution was changed by 10 percentage points and the 2.5th/97.5th percentiles were changed by up to 20 percentage points from the mode. Point estimates within columns represent the modes of the posterior distributions for each analysis, with the corresponding 95% credible intervals in parentheses.

Latent parameter	Baseline	Increase prior sn by 10%	Decrease prior sn by 10%	Increase prior sp by 10%	Decrease prior sp by 10%	Increase prevalence by 10%	Decrease prevalence by 10%
Prevalence	16% (6-32%)	16% (5-33%)	16% (5-35%)	19% (7-34%)	16% (5-31%)	19% (6-38%)	14% (2-30%)
Sensitivity of Evaluator A	52% (22-84%)	52% (13-93%)	48% (5-95%)	54% (24-85%)	51% (20-85%)	51% (21-84%)	54% (21-87%)
Sensitivity of Evaluator B	62% (31-89%)	65% (23-96%)	63% (15-98%)	63% (33-89%)	60% (28-88%)	60% (28-88%)	61% (27-89%)
Specificity of Evaluator A	93% (86-99%)	92% (85-99%)	91% (83-99%)	94% (86-100%)	92% (85-98%)	93% (86-99%)	92% (85-99%)
Specificity of Evaluator B	88% (80-97%)	88% (79-97%)	88% (78-97%)	90% (81-99%)	87% (79-96%)	88% (80-98%)	86% (79-96%)

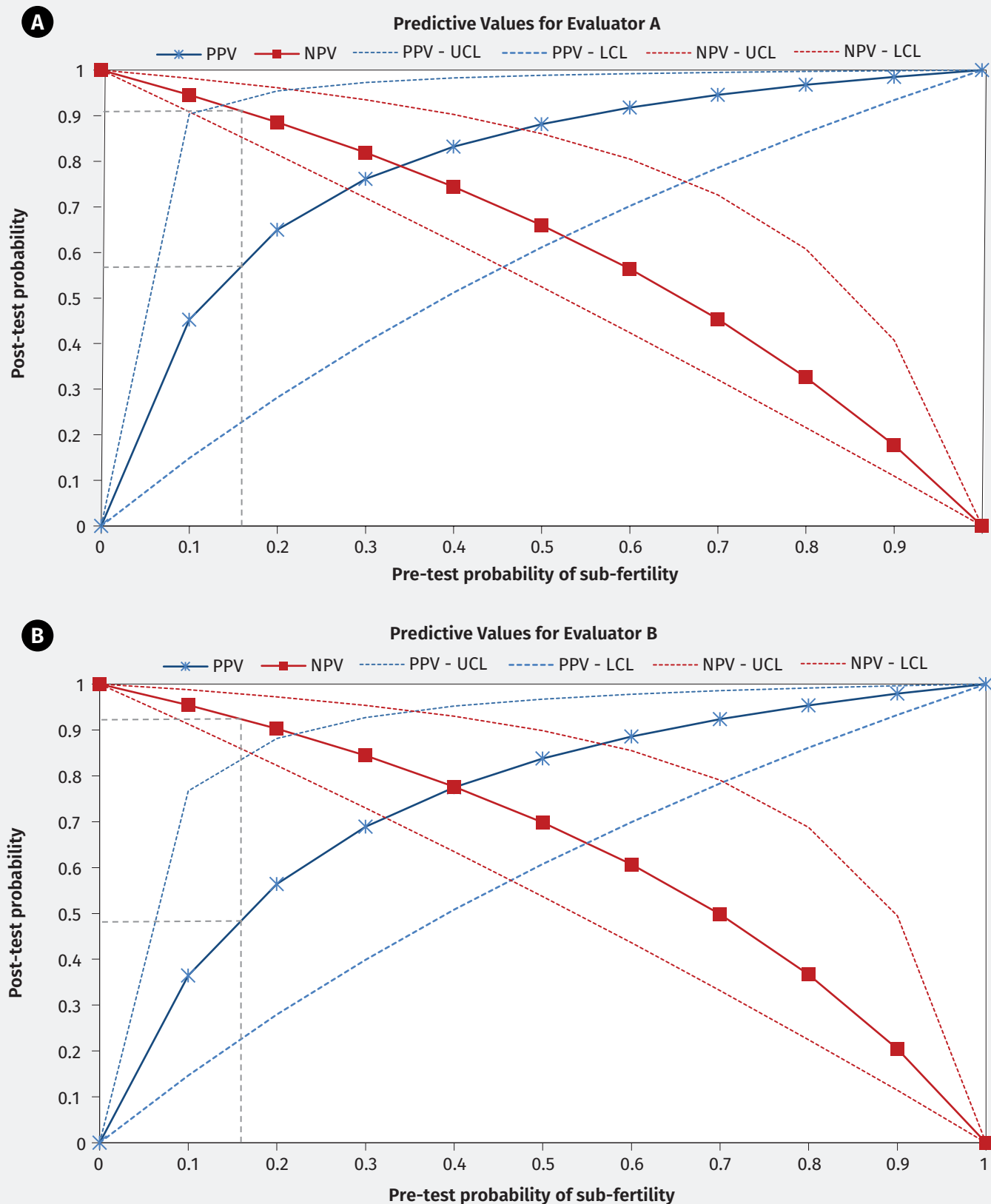
to the bull. There are several ways to detect pregnancy, and in a single-sire pasture, this might be an effective way to evaluate the fertility of a bull, assuming that the cows were fertile prior to breeding. However, in multi-sire pastures it is not feasible to determine the parentage of each fetus at pregnancy diagnosis. Therefore, in this setting, which is common to many cow-calf production models, it is difficult to assess the fertility of individual bulls. Waiting until the calf is born and determining parentage either from a DNA sample or a blood type, is one way to address this problem. This method for assessing fertility has been used by other studies, and has also been used to show associations between pre-breeding assessment of bulls and future calf-output.<sup>7-9</sup> However, none of these studies evaluated the BSE in terms of its sensitivity and specificity for identifying subfertile, or fertile, bulls. Furthermore, one of the issues with making a comparison between pre-breeding BSE classification and calf output is that there are many other factors that influence calf output aside from bull fertility. This is why bulls that meet the requirements of the BSE are classified as satisfactory *potential* breeders. In other words, whereas calf output is a useful way to assess fertility, it is not in the strictest sense a gold standard, and making a contingency table that compares BSE classification to being able to sire at least 25 calves is not the most accurate way to estimate the ability of the BSE to identify subfertile bulls.

Furthermore, in order for such a comparison to be made, the bulls would need to be categorized prior to breeding, then turned out to breed cows regardless of BSE classification. Obtaining the necessary data from a sufficient sample of bulls that are bred to cows regardless of their BSE classification would be challenging to say the least. Operations that are willing to incur the expense and difficulty of hiring a veterinarian to perform BSEs on their bulls are less likely to use bulls that are categorized as non-satisfactory, otherwise why perform the test? This was an issue in a large study in western Canada that sought to evaluate the association between various aspects of the BSE and pregnancy outcomes in cows bred

to bulls that were evaluated by participating veterinarians.<sup>10</sup> While the study did not find an association between BSE classification and pregnancy outcomes, only ~5% of the bulls that were exposed to cows in the study were categorized as non-satisfactory, resulting in relatively low statistical power to detect differences in outcomes between bulls that passed the BSE and those that did not. In summary, because there are other factors that determine calf-output that are not assessed as part of the standard BSE, and there may be extraneous factors that result in an otherwise fertile bull failing to impregnate an allotted number of cows, neither pregnancy diagnosis nor calf parentage can be considered true gold standards for the inherent subfertility of a bull.

Bayesian latent class analysis provides a means for overcoming these challenges in determining the true diagnostic performance of the BSE. To identify the latent parameters, this method compared the results of 2 dependent tests that evaluated separate diagnostic samples collected from the same ejaculate. This may partially explain why the Spearman rank-correlation and Cohen's kappa coefficient showed only moderate agreement between evaluators. Even if both evaluators had identical criteria for classification of individual spermatozoa, because their respective slides came from different samples of the ejaculate, we would expect there to be some variation in each evaluator's classification of normal vs. abnormal cells. Notwithstanding, because classification of spermatozoa is subjective, we expect some of this disagreement to come from differences in the methods of evaluation between evaluators. For example, a previously mentioned study assessed the inter-evaluator agreement of 3 board certified theriogenologists evaluating the same stained slides of smeared bull semen and classifying bulls as satisfactory or non-satisfactory using the same cut-off for abnormal sperm morphology as the present study. The Cohen's kappa coefficients for agreement between these 3 evaluators ranged from 0.32 to 0.64.<sup>23</sup> It should be noted that, while the evaluators in the aforementioned study examined the same slides, they did not necessarily evaluate the

**Figure 5:** The positive and negative predictive value of the morphology portion of the BSE over the full range of possible prevalence of subfertility in a theoretical population of predominately *Bos taurus* bulls. Dashed lines represent the 95% confidence intervals for the estimates. Vertical gray lines represent the estimated true prevalence of subfertile breeders in the population studied. Horizontal gray lines represent the estimated post-test probability for a positive (i.e., bull diagnosed as subfertile) and negative (i.e., bull diagnosed as not subfertile) test result in the population studied.



same regions of the slide and so some of this variation still may have come from sampling. It is also worth noting that this same study evaluated whether increasing the number of spermatozoa evaluated per slide from 100 to 400 had any effect on the classification of bulls, and did not find an effect.

A potential weakness of the current study was a lack of identifiability in the model. In BLCA, identifiability occurs when the analysis has at least as many independent combinations of observable outcomes as latent parameters being estimated. Our analysis compared data from 2 dependent tests. A 2 × 2 contingency table comparing the results of these 2 diagnostic tests produced 4 combinations of observable outcomes. However, because any one of these combinations of outcomes is constrained by the total number of observations, only 3 of the combinations may be considered independent. Our model estimated 7 latent parameters including prevalence of subfertile bulls in our population, sensitivity of evaluators A and B for classifying bulls as subfertile, specificity of evaluators A and B for classifying non-subfertile bulls, and positive and negative covariation between tests that had to be accounted for due to the assumption of dependence between the 2 tests. Because the model lacked identifiability, to develop a converging model we had to use informative priors for 5 of these latent parameters.

Lack of identifiability is a common issue in BLCA, and creates some level of uncertainty in the accuracy of the model.<sup>12,13,24</sup> However, there are various ways to deal with a lack of identifiability; this study used 2 different methods. The first was the use of non-uniform, informative prior distributions that were based on published estimates for calf output as a function of pre-breeding results of the semen morphology portion of the BSE.<sup>7</sup> These distributions provided estimates for sensitivity and specificity that were reasonable and consistent with a similar study.<sup>9</sup> The second was to perform a sensitivity analysis to determine how influential changes in these informative priors were on the posterior distributions. The sensitivity analysis performed on this model showed that a shift of 10 percentage points in the mode of the prior distributions produced shifts in the mode of the posterior distributions of at most 4 percentage points, and usually only 2-3 percentage points. In many cases, no shift in the mode of the posterior distribution was observed (Table 1). There are few, if any, objective guidelines for how much of a shift in posterior distributions is acceptable relative to shifts in prior distributions, but in general it is recommended that the shifts be relatively small.<sup>12,25</sup> The changes seen in the posterior distributions of this model during sensitivity analysis were no more than half as large as the shifts in the prior distributions, and in most cases were much smaller than half. In no case were they large enough to substantially affect the interpretation of the results.

## Conclusions

The estimated sensitivity and specificity of the semen morphology portion of the BSE did not differ meaningfully between evaluators. In general, the morphologic examination portion of the BSE appears to have high specificity and low sensitivity for identifying subfertile bulls. Across the range of pre-test probability for subfertility, the morphologic examination of semen slightly improved what was known before the test was conducted. Therefore, the diagnostic parameters of the morphologic portion of the BSE can be used to improve the accuracy of a practitioner's assessment of a bull's inherent ability, or lack of ability, to sire his quota of calves.

## Endnotes

<sup>a</sup> Personal communication, Kastelic JP, Koziol JH, & Barth AD, September 2023

<sup>b</sup> Microsoft Excel, Microsoft Corporation, Redmond, WA

<sup>c</sup> R version 4.1.2, R Foundation for Statistical Computing, Vienna, Austria; R packages used in the model were *epiR*, *R2jags*, *coda*, *mcmcplots*, *ggplot2*

## Acknowledgements

This study was a contribution of the Beef Cattle Population Health and Reproduction Program at Mississippi State University. We would like to thank Dr. Will Davis for reviewing the manuscript and study design, and Dr. Alexis Thompson for her assistance with running Bayesian latent class models in R. We appreciate and acknowledge the contribution of samples from U.S. Meat Animal Research Center livestock.

## Funding

This work was supported by Agricultural and Food Research Initiative 2018-69003-28706 from the USDA National Institute of Food and Agriculture. USDA is an equal opportunity provider, employer and lender.

## Conflicts of Interest

The authors of this study have no conflicts of interest to report. Artificial intelligence was not used for data collection or analysis, nor was it used as a primary tool for manuscript preparation. Any opinions, findings, conclusions or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the view of the U.S. Department of Agriculture.

## Author contributions

All authors contributed to the design of the study, the acquisition of data, and/or analysis and interpretation of data. Furthermore, all authors contributed to drafting and/or revising the manuscript, and approved the final version for publication.

## References

1. Society for Theriogenology. Online Store: Paper Forms. Accessed February 3, 2024. <https://www.therio.org/store/ListProducts>
2. Koziol JH, Armstrong CL. *Society for Theriogenology Manual for Breeding Soundness Examination of Bulls 2nd Edition*. Society for Theriogenology; 2018.
3. Kastelic J, Thundathil J. Breeding soundness evaluation and semen analysis for predicting bull fertility. *Reprod Domest Anim*. 2008;43:368-373. <https://doi.org/10.1111/j.1439-0531.2008.01186.x>
4. Kastelic JP, Thundathil JC, Brito LFC. Bull BSE and semen analysis for predicting bull fertility. *Clin Theriogenology*. 2012;4(3):277-287.
5. Abell KM, Theurer ME, Larson RL, et al. Calving distributions of individual bulls in multiple-sire pastures. *Theriogenology*. 2017;93:7-11. <https://doi.org/10.1016/j.theriogenology.2017.01.010>

6. Holroyd RG, Doogan VJ, Faveri JD, et al. Bull selection and use in northern Australia 4. Calf output and predictors of fertility of bulls in multiple-sire herds. *Anim Reprod Sci.* 2002;15(7):67-79. [https://doi.org/10.1016/s0378-4320\(02\)00026-x](https://doi.org/10.1016/s0378-4320(02)00026-x)
7. Whitworth WA, Forrest DW, Sprott LR, Holloway JW, Warrington BG. Case study: influence of seminal, physical, and mating behavior traits of bulls on number of calves sired per bull in a multisire herd. *Prof Anim Sci.* 2008;24(2008):184-188. [https://doi.org/10.15232/S1080-7446\(15\)30834-2](https://doi.org/10.15232/S1080-7446(15)30834-2)
8. Coulter GH, Kozub GC. Efficacy of methods used to test fertility of beef bulls used for multiple-sire breeding under range conditions. *J Anim Sci.* 1989;67(7):1757-1766. <https://doi.org/10.2527/jas1989.6771757x>
9. Fitzpatrick LA, Fordyce G, McGowan MR, et al. Bull selection and use in northern Australia Part 2. Semen traits. *Anim Reprod Sci.* 2002;71(1-2):39-49. [https://doi.org/10.1016/s0378-4320\(02\)00024-6](https://doi.org/10.1016/s0378-4320(02)00024-6)
10. Waldner CL, Kennedy RI, Palmer CW. A description of the findings from bull breeding soundness evaluations and their association with pregnancy outcomes in a study of western Canadian beef herds. *Theriogenology.* 2010;74(5):871-883. <https://doi.org/10.1016/j.theriogenology.2010.04.012>
11. Menegassi SRO, Barcellos JOJ, Lampert V do N, Borges JBS, Peripolli V. Bioeconomic impact of bull breeding soundness examination in cow-calf systems. *Rev Bras Zootec.* 2011;40(2):441-447. <https://doi.org/10.1590/S1516-35982011000200028>
12. Johnson WO, Jones G, Gardner IA. Gold standards are out and Bayes is in: Implementing the cure for imperfect reference tests in diagnostic accuracy studies. *Prev Vet Med.* 2019;167:113-127. <https://doi.org/10.1016/j.prevetmed.2019.01.010>
13. Dufour S, Arango Sabogal JC. Bayesian latent class models for prevalence estimation and diagnostic test evaluation. Bayesian latent class models for prevalence estimation and diagnostic test evaluation. August 3, 2022. Accessed February 3, 2024. <https://lcm-dx.netlify.app/>
14. Finch WH, Bronk KC. Conducting Confirmatory Latent Class Analysis Using MPlus. *Struct Equ Model Multidiscip J.* 2011;18(1):132-151. <https://doi.org/10.1080/10705511.2011.532732>
15. Wurpts IC, Geiser C. Is adding more indicators to a latent class analysis beneficial or detrimental? Results of a Monte-Carlo study. *Front Psychol.* 2014;5. <https://doi.org/10.3389/fpsyg.2014.00920>
16. Barth AD, Waldner CL. Factors affecting breeding soundness classification of beef bulls examined at the Western College of Veterinary Medicine. *Can Vet J.* 2002;43(4):274-284.
17. Tang W, Hu J, Zhang H, Wu P, He H. Kappa coefficient: a popular measure of rater agreement. *Shanghai Arch Psychiatry.* 2015;27(1):62-67. <https://doi.org/10.11919/j.issn.1002-0829.215010>
18. Makarechian M, Farid A. The relationship between breeding soundness evaluation and fertility of beef bulls under group mating at pasture. *Theriogenology.* 1985;23(6):887-898. [https://doi.org/10.1016/0093-691X\(85\)90007-X](https://doi.org/10.1016/0093-691X(85)90007-X)
19. Wiltbank JN, Parish NR. Pregnancy rate in cows and heifers bred to bulls selected for semen quality. *Theriogenology.* 1986;25(6):779-783. [https://doi.org/10.1016/0093-691X\(86\)90093-2](https://doi.org/10.1016/0093-691X(86)90093-2)
20. Farin PW, Chenceth PJ, Tomky DF, Ball L, Pexton JE. Breeding soundness, libido and performance of beef bulls mated to estrus synchronized females. *Theriogenology.* 1989;32(5):717-725. [https://doi.org/10.1016/0093-691x\(89\)90460-3](https://doi.org/10.1016/0093-691x(89)90460-3)
21. Gunderson TG, Kim KN, Coatney KT, Smith DR. To BSE or not to BSE: a capital budgeting analysis of the use of the bull breeding soundness evaluation (BSE) to improve reproductive efficiency and profitability in cow-calf herds. *Transl Anim Sci.* 2025;9:txaf024. <https://doi.org/10.1093/tas/txaf024>
22. Menon AG, Thundathil JC, Wilde R, Kastelic JP, Barkema HW. Validating the assessment of bull sperm morphology by veterinary practitioners. *Can Vet J.* 2011;52:407-408.
23. Reeves A, Klabnik J, Strickland L, et al. Bull sperm morphology assessment varied by evaluator. *Clin Theriogenology.* 2022;14(1):11-17. <https://doi.org/10.58292/ct.v14.9293>
24. Dufour S, Durocher J, Dubuc J, Dendukuri N, Hassan S, Buczinski S. 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.* 2017;140:122-133. <https://doi.org/10.1016/j.prevetmed.2017.03.008>
25. Roos M, Martins TG, Held L, Rue H. Sensitivity analysis for Bayesian hierarchical models. *Bayesian Anal.* 2015;10(2). <https://doi.org/10.1214/14-BA909>

