Theriogenology 85 (2016) 254–266

Contents lists available at ScienceDirect

Theriogenology

journal homepage: www.theriojournal.com

A multilaboratory study on the variability of bovine semen analysis

Leonardo F.C. Brito

ABS Global Inc., DeForest, Wisconsin, USA

ARTICLE INFO

Article history:
Received 17 February 2015
Received in revised form 18 April 2015
Accepted 16 May 2015

Keywords:
Semen evaluation
Andrology laboratory
Quality assurance

ABSTRACT

To evaluate the variability of semen analysis, five replicates of 10 different bovine frozen semen batches were coded with different identification numbers and submitted to various laboratories for evaluation. Three studies were conducted: study I included eight laboratories in semen processing centers in the United States; study II included one laboratory in one semen processing center and five veterinary university laboratories in the United States; and study III included five veterinary university laboratories in Brazil. Evaluation methodology, sample classification criteria, and reporting format varied considerably among laboratories. There were laboratory effects ($P < 0.05$) on sperm concentration, motility, and morphology results in all studies. When Bland–Altman plots were evaluated, differences in sperm concentration were approximately between −5 and +5 × 10^6 sperm/mL in study I, when the same method of evaluation was used by all laboratories but ranged between −30 and +30 × 10^6 sperm/mL in studies II and III. Differences in the proportions of motile sperm were approximately −30% to +30%, and differences in the proportion of normal sperm were −15% to +15% in studies I and II; these differences were −15% to +15% and −10% to +10%, respectively, in study III. Mean absolute (one tail) proportional differences in estimates across all laboratories ranged from 9% to 31%, 16% to 37%, and 9% to 14% for sperm concentration, motile sperm, and normal sperm across studies; much larger (48%–86%) differences were observed for sperm abnormality categories. Intralaboratory and interlaboratory precision varied considerably across laboratories and seemed to be at least in part related to methods used for evaluation; precision was better when the NucleoCounter was used for evaluation of sperm concentration, whereas the use of computer-assisted sperm analysis for evaluation of sperm motility resulted in greater precision in some but not all laboratories. None of the laboratories that classified samples as satisfactory or unsatisfactory achieved complete consistency for all replicates within all batches. In addition, consistent classification among laboratories was observed for just three batches in studies II and III. These observations put the reliability of semen analysis in check and make it very difficult, if not impossible, to meaningfully interpret evaluation results.

1. Introduction

Semen analysis constitutes the most important clinical laboratory test currently available to evaluate male fertility. Analysis results are used to attest breeding soundness, to determine the adequacy of insemination doses, to diagnose specific causes of infertility, and to guide clinical and management decisions. Despite the profound economical and emotional implications related to semen analysis, a pervasive attitude exists among veterinarians, livestock producers, and sire owners that semen analysis is a trivial task and very little attention have been paid to the quality of the analysis itself.

Corresponding author. Tel.: +1 608 509 3808.
E-mail address: lfcbrito@lycos.com.

0093-691X/$ – see front matter © 2016 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.theriogenology.2015.05.027

Although in theory, semen analysis is a simple task, in practice, obtaining accurate and precise results for sperm concentration, motility, and morphology requires knowledge of the principles and peculiarities of different tests and equipment, technical expertise, and implementation of comprehensive laboratory quality assurance programs. Several multicenter studies involving human andrology laboratories have reported large variations in semen evaluation results within and across laboratories [1–4]. The reasons for these large variations are likely the lack of standard methods, uniform training, and quality control procedures and have led some specialists to refer to semen analysis as the "neglected test" [5].

Because similar studies involving veterinary andrology laboratories have not been reported, the present study was conducted with the objective of evaluating the variability in semen analysis using frozen–thawed bovine samples as a model.

2. Materials and methods

2.1. Semen batches and laboratories

Bovine semen samples were extended in egg yolk–glycerol extender, packaged into 0.5-mL straws, frozen, and shipped to participating laboratories. Ten different batches were used, and five replicates of the same batch were coded with different identification numbers. Therefore, each laboratory received 50 samples for evaluation.

In study I, samples were evaluated by eight laboratories (laboratories A–H) in semen processing centers in the United States as part of an informed project. The laboratories were asked to use their usual methods of evaluation and report sperm concentration, motility (% progressively motile), and morphology (% normal sperm). The laboratories were also asked to provide a brief description of the methods used for evaluation. In study II, samples were submitted to five veterinary university laboratories that offered semen evaluation services in the United States; samples were also evaluated at a semen processing center (laboratories I–N). In study III, samples were submitted to five veterinary university laboratories (laboratories O–S) that offered semen evaluation services in Brazil. University laboratories were unaware of the study and were asked to provide their standard semen analysis report.

2.2. Statistical analysis

Statistical analysis was performed using Statistix 8 (Analytical Software, Tallahassee, FL, USA). General linear models were used to detect and locate laboratory effects on semen analysis results with replicates nested within batch. Means across laboratories calculated for each batch and differences from individual replicates were described in Bland–Altman plots. Absolute (one tail) proportional differences for individual results were also calculated considering the average across all laboratories for each batch as 100%. The effects of batch characteristics on observed absolute proportional differences (dependent variable) were determined by linear regression using mean batch concentration, motility, and morphology as independent variables. In addition, intralaboratory and interlaboratory coefficients of variation (CVs) were calculated for each of the 10 batches, and means within and across laboratories were calculated.

3. Results

3.1. Methods used for semen evaluation

Laboratories were assumed to have used their standard methods of semen thawing, handling, and analysis, and not all methodology details were available. Among the laboratories included in study I, laboratory E reported using the SQA-Vb system for determining sperm concentration, motility, and morphology in a single step. All other laboratories reported using the NucleoCounter SP-100 for determining sperm concentration and wet-mount preparations with either phase-contrast or differential interference contrast optics for evaluating sperm morphology. Laboratories B, D, and H reported using × 1000 magnification for morphology evaluation, whereas × 400 or × 600 was used by laboratories F and A, respectively; two laboratories did not report the magnification at which morphology was evaluated. A total of 100 sperm per sample were examined in five laboratories, whereas two laboratories did not report the number of sperm evaluated for morphology. Sperm motility was determined subjectively in five laboratories, whereas IVOS or Ceros computer-assisted sperm analysis (CASA) systems were used in laboratories A and H, respectively.

In studies II and III, only one laboratory (laboratory J) provided a comprehensive description of the methods used for semen evaluation as part of the report, including description of equipment, materials, dilutions, number of evaluated sperm, and so forth. Other laboratories simply mentioned the method used for some of the evaluations included in the report (e.g. "hemocytometer" or "wet-mount with phase contrast"), whereas others did not include any description of the methods used to obtain the results.

Among the laboratories included in study II, laboratory K reported using the SQA-Vb system for determining sperm concentration, laboratories I and L reported using the NucleoCounter SP-100, and the other laboratories reported using hemocytometers. Sperm motility was determined subjectively in laboratories K and M, whereas CASA systems were used in the other laboratories (two laboratories reported using either IVOS or Ceros, and two laboratories did not specify the system). Laboratories I, K, and N reported using wet-mount preparations with phase contrast for sperm morphology evaluation, whereas laboratory J reported using eosin/nigrosin–stained slides. Only laboratories I and N reported the magnification (× 1000) used for morphology evaluation. A total of 100 sperm per sample were examined in laboratories I, J, and K, whereas 200 sperm were examined in laboratory M; two laboratories did not report the number of sperm evaluated for morphology.

Among the laboratories included in study III, laboratory S reported using the Neubauer hemocytometer for determining sperm concentration, but other laboratories did not