Invited review: Recommendations for reporting intervention studies on reproductive performance in dairy cattle: Improving design, analysis, and interpretation of research on reproduction


*Scibus, Camden, New South Wales, Australia 2570
†Faculty of Veterinary Science, University of Sydney, Camden, New South Wales, Australia 2570
‡Department of Animal Sciences, University of Missouri, Columbia 65211
§Department of Animal Sciences, Washington State University, Pullman 99164-6310
¶Department of Animal Sciences and Industry, Kansas State University, Manhattan 66506
‖‖Department of Animal and Range Sciences, Montana State University, PO Box 172900, Bozeman 59717-2900
**Jemora Pty Ltd., PO Box 2277, Geelong, Victoria, Australia 3220
††Dapto, Horsley, NSW 2530, Australia
‖‖Department of Animal Sciences, University of Florida, Gainesville 32611-0910
§§Population Medicine, Ontario Veterinary College, University of Guelph, Guelph, Ontario, Canada N1G 2W1

ABSTRACT

Abundant evidence from the medical, veterinary, and animal science literature demonstrates that there is substantial room for improvement of the clarity, completeness, and accuracy of reporting of intervention studies. More rigorous reporting guidelines are needed to improve the quality of data available for use in comparisons of outcomes (or meta-analyses) of multiple studies. Because of the diversity of factors that affect reproduction and the complexity of interactions between these, a systematic approach is required to design, conduct, and analyze basic and applied studies of dairy cattle reproduction. Greater consistency, clarity, completeness, and correctness of design and reporting will improve the value of each report and allow for greater depth of evaluation in meta-analyses. Each of these benefits will improve understanding and application of current knowledge and better identify questions that require additional modeling or primary research. The proposed guidelines and checklist will aid in the design, conduct, analysis, and reporting of intervention studies. We propose an adaptation of the REFLECT (Reporting Guidelines for Randomized Controlled Trials for Livestock and Food Safety) statement to provide guidelines and a checklist specific to reporting intervention studies in dairy cattle reproduction. Furthermore, we provide recommendations that will assist investigators to produce studies with greater internal and external validity that can more often be included in systematic reviews and global meta-analyses. Such studies will also assist the development of models to describe the physiology of reproduction.

Key words: reporting guidelines, reproduction, study design, meta-analysis, metabolic disease definition

INTRODUCTION

Our goal is to improve the completeness, clarity, and correctness of design, analysis, reporting, and interpretation of studies on reproduction in dairy cattle. This process will improve interpretation of individual studies. More consistent and better study design and reporting will maximize the ability of researchers to use studies of reproduction in dairy cattle in systematic reviews and meta-analyses and improve understanding of the causes of variability in results from apparently similar studies.

These recommendations are directed to animal and veterinary scientists interested in improving reproductive efficiency of dairy cattle, who would use the checklists and guidelines when designing studies and preparing reports for publication. The audience includes new and established investigators, as well as reviewers and
Suboptimal reproductive performance of dairy cattle has been well documented in economic and production system efficiency terms (Lucy, 2001; Norman et al., 2009). Reproductive performance is influenced by many factors, including genetics, young stock rearing, nutritional management, cow housing and comfort, management of the transition period, season, weather, and disease. For good reason, many of these factors have been studied in isolation, with little accounting for interacting or confounding variables, but that is changing. Research is increasingly assessing relationships for interaction and accounting for confounding of these varying effects on reproduction. Meta-analyses of nutritional effects on reproduction have been conducted (Lean et al., 2012; Rodney et al., 2015), identifying a negative effect of soluble protein and positive effects of fatty acids on reproductive performance in lactating dairy cows, respectively. Several systems models of reproduction and the integration of genetics, nutrition, and reproduction have been recently published (Boer et al., 2011; Martin et al., 2012; McNamara and Shields, 2013).

The low estimates of heritability of reproductive performance reflect the fact that fertility traits are complex and multifactorial; however, the estimated genetic merit for reproductive traits is increasing (Berry et al., 2014). Important explanatory environmental variables; that is, nongenetic effects that influence reproduction, are unmeasured in most genetic analyses and most physiology and nutrition experiments. Numerous environmental variables likely contribute considerably more compared with genetics to reproductive phenotype directly (Bello et al., 2012). The environmental effects that need to be accounted for might be poorly and inconsistently measured, and definitions of fertility measures can vary. Such problems lead to the possibility of substantial unmeasured genotype by environment interactions. The compilation of studies for further detailed analyses by systematic review or meta-analysis may help to identify the presence and sources of heterogeneity of results among investigations of a similar research question (Bisinotto et al., 2015), but only if the primary studies are sufficiently valid, completely reported, and consistent in definitions and conduct to be included; unfortunately, this is rarely the case (Haimerl et al., 2013; Rodney et al., 2015).

Relatively few recent publications have addressed study design in cattle research using randomized controlled trials, but examples include St-Pierre (2007) and Tempelman (2009). Those authors stressed the need for adequacy of sample size and the need to avoid pseudo-replication, and they discussed appropriate approaches to study design. In conducting meta-analyses on reproductive studies examining hormonal and dietary interventions, the following failings were noted in many of the studies evaluated: randomization or allocation methods were not described; the blinding of researchers to treatment allocation was not stated; the use of placebo treatments was not stated; some studies lacked clarity in regard to whether nutritional treatments were allocated to individual cattle or groups of cattle; and relatively few papers provided or correctly analyzed time-to-event outcomes; for example, time to first service or pregnancy (Amann, 2005; Haimerl et al., 2013; de Boer et al., 2014; Rodney et al., 2015). Finally, notwithstanding earlier attempts to standardize reporting of reproductive outcomes (Fetrow et al., 1990), there has been inconsistent use of outcome measures and terminology in studies of reproductive performance. These failings are not unique to reproductive studies; others have documented deficiencies in study design and reporting in human and veterinary medicine (DerSimonian et al., 1982; Elbers and Schukken, 1995; Moher et al., 2001; O’Connor, 2010). Further, the problems of reproductive research study design are not isolated to cattle (Kastelic and Gandolfi, 2005). Substantial flaws in reproductive study design for studies involving horses, cattle, swine, companion, laboratory, and wild animals have been noted (Amann, 2005; Simoneit et al., 2011). The need to use research resources efficiently and to maximize the information gleaned from experiments on reproductive performance of dairy cattle provided the major stimulus to develop these guidelines. Reporting guidelines are one useful tool to improve the quality of published research (O’Connor, 2010). The “CONSORT statement,” checklist and explanation document (Consolidated Standards of Reporting Trials; Moher et al., 2001) was developed primarily for human medicine to provide guidelines for design, conduct, and reporting of randomized controlled trials and contain standard methods and guidelines for reporting measures. Veterinary medical and food safety specialists adapted CONSORT to produce the “REFLECT statement” on Reporting Guidelines for Randomized Controlled Trials (O’Connor et al., 2010; Sargeant et al., 2010) for livestock trials, and guidelines for animal research have also been produced (Kilkenny et al., 2010). Papers in medical journals that endorse the CONSORT guidelines have more complete reporting of randomized controlled trials, although after many years, there is still room for improvement (Turner et al., 2012).
A CHECKLIST AND GUIDELINES FOR STUDIES IN DAIRY CATTLE REPRODUCTION

Awareness of the limitations in the capacity of any individual research group to address the challenges of improving reproductive performance of cattle grew among researchers in the fields of reproduction, nutrition, genetics, and disease. Additionally, evaluations of study designs and reporting in reproductive research encountered in the process of research syntheses (e.g., Boer et al., 2011) stimulated discussion among researchers on an informal basis. The need to improve reproductive study design and reporting became clear. Meetings were organized in 2011, 2012, and 2013 in conjunction with the American Dairy Science Association®–American Society of Animal Science annual meetings, and an ad hoc subgroup of participants undertook to develop recommendations on the conduct and analyses of reproductive intervention studies, using a process of discussion and iterative evaluation of information, based on the principles and documents developed in the REFLECT (O’Connor et al., 2010; Sargeant et al., 2010) and CONSORT statements (Altman et al., 2001, 2012). The rationale for the development of specific guidelines was that there were sufficient matters of importance and uniqueness to warrant separate guidelines and discussion for reproductive intervention studies using dairy cattle.

We adapted the REFLECT checklist to apply to dairy cattle reproduction intervention studies. Table 1 provides a checklist and brief rationale for the best practices of study design. Table 2 describes in more detail our suggested definitions and correct statistical analyses of reproductive performance outcome measures. Table 2 is supported by Supplemental Table S1 (http://dx.doi.org/10.3168/jds.2015-9445), which provides detail of less commonly used reproductive measures. It is important to describe environmental conditions. Consequently, suggestions to provide consistent measures and methods to describe environment and nutrition are provided in some detail. Supplemental Table S2 (http://dx.doi.org/10.3168/jds.2015-9445) summarizes standardized animal, environmental, and nutritional variables for study design and reporting. It is clearly not our intent that studies should report all of these items; instead, this compilation is intended to encourage inclusion of as much as is feasible of this contextual information by standardized inclusion of these covariates. Any given study would certainly not manipulate or even be able to control for all these factors, but as complete a description and definition of the circumstances of the study as possible will improve interpretation and utility of the study. We acknowledge that funding or logistical constraints will limit the size and scope of studies and the covariates that can be measured, and that smaller-scale preliminary studies are often needed to begin to generate data to address a research question. However, all studies will make their best contribution when the points in Table 1 are considered in study design and comprehensively reported. Table 2 provides guidelines for definitions of disease. Consistent use of clear, validated terminology, case definitions, and diagnostic criteria will help researchers understand the effects and interactions of disease on reproduction.

The remainder of the paper describes elements of the reporting guidelines in more detail.

ELABORATION OF THE CHECKLIST

Hypotheses and Objectives (Table 1, Item 5)

The statement of concise hypotheses must be included with an explicit statement of the objectives of the study; primary objectives must be clearly distinguished from secondary objectives. The hypothesis is a critical attribute of any study, and failure to state a hypothesis is a critical failure. It is common in reproductive studies to examine multiple outcomes. Primary outcomes are the outcomes of greatest interest; these variables should determine the study sample size. Other outcomes are, by definition, less important and should not determine the sample size. Investigators should be aware of the potential to increase risk of type I statistical error by multiple hypothesis testing without adequate adjustment and should aim to limit the hypotheses tested to those that are consistent with the stated objectives of the study.

Example:

“Therefore, the aim of this study was to characterize the phenotypic performance of cows with similar genetic merit for milk production traits and similar proportions of Holstein genes, but divergent genetic merit for fertility traits (genotypes Fert+ and Fert−). Specifically, we tested the hypothesis that cows with negative estimated breeding value (EBV) for calving interval would have superior fertility performance, would be less reliant on body reserve mobilization, and would have detectable differences in circulating concentrations of metabolic hormones and metabolites compared with cows that had positive EBV for calving interval (Cummins et al., 2012).”

It is important to consider the appropriate type of hypothesis and the effects on interpretation of the study and sample size. Four possible types of statistical
### Table 1. Checklist and brief rationale for design, conduct and reporting of research experiments in dairy cattle reproduction, adapted from the REFLECT guidelines (Sargeant et al., 2010)

<table>
<thead>
<tr>
<th>Item no.</th>
<th>REFLECT paper section</th>
<th>Reproduction reporting guideline and importance to study quality</th>
<th>Checklist: Reported on page ___ of the paper</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Title and abstract</td>
<td>The design of the study is explicitly stated (e.g., randomized controlled trial) in the title and abstract.</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Introduction—Background</td>
<td>The introduction clearly and concisely explains the problem or question that the study will address, and cites key literature that, if relevant, supports or contradicts the hypothesis of the study.</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Methods—Participants</td>
<td>Inclusion and exclusion criteria are stated explicitly. The experimental unit is described clearly [e.g., the cow (with or without repeated measures), estrus cycle, insemination, single or multiple rounds of synchronization, or herd]. Genetic identity is ideally characterized. <strong>Critical to study quality.</strong></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Interventions (a)</td>
<td>Precise details of the interventions intended and how, when, and at what level (herd, cow, estrus cycle or insemination) treatments were actually administered for each group are reported. <strong>Critical to study quality.</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Interventions (b)</td>
<td>Details of interventions in each group are described, including the nature of the control (placebo or no treatment, or basal treatment or positive control). <strong>Important to study quality.</strong></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Objectives</td>
<td>A concise hypothesis statement is included. The objectives are clearly stated and distinguished as primary (i.e., that drive the design) or secondary. <strong>Critical to study quality.</strong></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Outcomes</td>
<td>The primary (i.e., that drive the sample size) and secondary outcomes are explicitly stated. The follow-up or observation period is stated (e.g., for how long after intervention cows were monitored or eligible to be recorded). The handling of data from cattle that did not experience the outcome of interest (e.g., were not inseminated or did not become pregnant during the observation period) is explicitly stated. Causes of censoring are reported. The use or eligibility for, methods and intensity of estrus detection are described. <strong>Important to study quality.</strong> The method and timing (days after insemination or end of mating) of pregnancy diagnosis are described. The criteria used for pregnancy (e.g., membrane slip by palpation; visualization of fluid, embryo or heartbeat by ultrasound; the cut-point used for chemical tests) are described. If a chemical test for pregnancy is used, a reference to validation data for the test is provided.</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Sample size</td>
<td>The basis of the sample size (for each outcome, if relevant) is clearly stated, including the nature of the outcome variable (binary; categorical; continuous; comparison of proportions, means, counts, or time-to-event), the expected outcome or difference, the expected variance, and the chosen confidence (α) and power (1−β). Nonindependence (clustering; e.g., correlation of inseminations within a cow, or cows within a herd) is accounted for. If there are multiple levels (e.g., insemination, cow, or herd) of outcomes or covariates, that is accounted for. <strong>Critical to study quality.</strong></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Randomization—Sequence generation</td>
<td>The level (farm, breeding period, cow, insemination) at which treatments are applied is stated and is congruent with the stated objectives and with the unit of statistical analysis. <strong>Critical to study quality.</strong> Descriptive data, including a measure of variance, is provided to demonstrate the success of matching and/or of random assignment to treatment across important covariates (e.g., if cows were matched on previous lactation milk yield, the mean and SD yields; similarly (as appropriate) covariates (e.g., parity, DIM at treatment, BCS, inseminator, fertility index of bulls used) are shown to be similar (balanced) between treatment groups.</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Randomization—Allocation concealment</td>
<td>The method used to generate and apply random treatment assignments is described. The method of allocation of experimental units to treatment is described, including who assigned and applied treatments and whether the same personnel were involved in making observations. Use of random treatment assignment is explicitly stated in the abstract and in the methods. <strong>Critical to study quality.</strong></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Randomization—Implementation</td>
<td>The method of allocation of experimental units to treatment is described, including who assigned and applied treatments and whether the same personnel were involved in making observations. Use of random treatment assignment is explicitly stated in the abstract and in the methods. <strong>Critical to study quality.</strong></td>
<td></td>
</tr>
</tbody>
</table>

*Continued*
Table 1 (Continued). Checklist and brief rationale for design, conduct and reporting of research experiments in dairy cattle reproduction, adapted from the REFLECT guidelines (Sargeant et al., 2010)

<table>
<thead>
<tr>
<th>Item no.</th>
<th>REFLECT paper section</th>
<th>Reproduction reporting guideline and importance to study quality</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>Blinding (masking)</td>
<td>Blinding (hiding of treatment assignments to personnel and assessors) is described at the levels of allocation to treatment, care and management of animals, observation of outcomes and covariates, and statistical analysis. Blinding should be as complete as possible; for example, caregivers or assessors should not be aware that cattle or pens are treatment A or B, even if they do not know what each represents. If blinding is not done or not complete, justification is provided. Inherent to this will be use of “sham” (placebo) treatments for control cattle. <strong>Important to study quality.</strong></td>
</tr>
<tr>
<td>12</td>
<td>Statistical methods</td>
<td>The statistical methods are explicitly linked to the objectives and the experimental unit. The statistical methods are correct for the nature of the data (see Table 2). <strong>Critical to study quality.</strong> Adjusted analyses and the covariates included in mathematical models and adjusted results (e.g., least squares means) are described. Testing of model assumptions (e.g., normality; linearity of continuous predictors) and of assessment of fit of models are described. When outcomes are not independent due to possible clustering (e.g., cows with a herd in a multi-herd study or repeated measures), these effects are explicitly addressed in the analysis. The use of fixed, random or mixed effects in the study for herds, pens, animals, and time (for repeated measures) should be explicitly stated and justified. P-values are quantified to &lt;0.001 (i.e., $P = 0.71$ or $P = 0.003$, not “$P &lt; 0.05$” or “NS”). It is preferable to report 95% confidence intervals or equivalents around measures of effect. <strong>Critical to study quality.</strong></td>
</tr>
<tr>
<td>13</td>
<td>Results–Study flow</td>
<td>The numbers of animals examined, enrolled/treated, excluded, withdrawn, lost to follow-up or having missing or unusable data, and analyzed are provided for each group, normally in a diagram or table.</td>
</tr>
<tr>
<td>14</td>
<td>Recruitment</td>
<td>The geographic location of the study site(s), dates of the study, housing/management system, and general environmental conditions are described.</td>
</tr>
<tr>
<td>15</td>
<td>Baseline data</td>
<td>Where relevant or feasible, the diets fed are described (see Supplemental Table S2; 10.3168/jds.2015-9445). If intake is not directly measured, estimation of a surrogate descriptor of intake (e.g., pen-level feed delivery and refusals) is desirable. Descriptive data (n, mean, median, or proportion, and a measure of variance) are provided for each treatment group on breed, age or parity, stage of lactation, body condition, milk production.</td>
</tr>
<tr>
<td>16</td>
<td>Numbers analyzed</td>
<td>For each result presented, the number of observations in the analysis, the effect estimate, and a measure of its variance (SD, SE, or confidence interval) is presented. <strong>Critical to study quality.</strong> The numbers included in the analysis are described as being based on intention to treat or on final inclusion criteria. Specifically, the number and proportion of animals initially recruited or enrolled but not included in each final analysis is stated explicitly. The covariates in multivariable models are listed, or their effects are presented as for main effects. When relative measures of effect (e.g., odds ratio, relative risk, or hazard ratio) are presented, a description of absolute performance and its variance is also included (e.g., proportion of animals inseminated in a specified period; probability of pregnancy per insemination; median days to pregnancy). <strong>Critical to study quality.</strong></td>
</tr>
<tr>
<td>17</td>
<td>Outcomes and estimation</td>
<td>If multiple comparisons of treatment are made on the same data, adjustments of methods or interpretation are described. If analysis of a subgroup is performed, the justification for it is clearly stated (e.g., based on an a priori hypothesis, which should be reflected in the study design, or detection of an interaction).</td>
</tr>
<tr>
<td>18</td>
<td>Ancillary analyses</td>
<td>The number of animals withdrawn for protocol violations, welfare reasons, disease, or death is reported for each group. It is critical, in intervention studies, to note and report any adverse reaction as these can be related to treatment. Pharmacovigilance is an important role of research. Important to study quality. Adverse or side effects reasonably attributable to treatment are reported or the absence of these is stated.</td>
</tr>
<tr>
<td>19</td>
<td>Adverse events</td>
<td><strong>Continued</strong></td>
</tr>
</tbody>
</table>
null hypotheses can be tested: no difference, inferiority, superiority, and nonequivalence (i.e., either superior or inferior). No difference is the most commonly tested null hypothesis, but some studies test the inferiority, nonequivalence (Donner 1984), or, rarely, superiority, of one treatment relative to another. A noninferiority alternative hypothesis is often used for evaluating the efficacy of an experimental treatment against a positive control when it is hypothesized that the experimental treatment is clinically not inferior in effectiveness (D’Agostino et al., 2003; Piaggio et al., 2012) and may be quicker, cheaper, or simpler.

Participants and Study Unit (Table 1, Item 3)

Authors should accurately and completely describe cattle with the taxonomic identifier, breed(s), sex(es), age(s), and productive status(es). Enhanced definitions of identity may be useful for literature searching using computer algorithms because, for example, a computer search may not recognize that a cow and a heifer both refer to *Bos taurus*. With increasing availability of information on single nucleotide polymorphisms, it may be prudent to retain samples of hair (including follicles) or other biological material suitable for subsequent genotyping. Authors should provide gene accession numbers for all genes referenced in the paper, according to the National Center for Biotechnology Information’s Gene database (http://www.ncbi.nlm.nih.gov/gene).

Authors must clearly state the experimental unit and the unit of statistical analysis. The unit of analysis may be the insemination, the 21-d breeding period (when analyzing 21-d estrus detection and 21-d pregnancy risks, for example), the lactation, or the cow (see Supplemental Table S1; http://dx.doi.org/10.3168/jds.2015-9445). Units can also be aggregated to the group within herd or to the herd level. The experimental unit can often differ from the unit of analysis. For example, the experimental unit may be a pen but the unit of analysis may be a cow. In intervention trials, the experimental unit can be defined as the smallest system or unit of experimental material to which a single treatment (or treatment combination) is assigned by the investigator (Hurlbert, 1984). Treatments should be applied and replicated such that the treatment applied to one experimental unit cannot affect another experimental unit. For reproductive studies in which group interactions are important (e.g., sexually active groups may influence the sensitivity of detection of estrus of other cattle), investigators should think carefully about study designs that account for these effects. Pseudo-replication can occur where there is a failure to distinguish between the experimental unit and the units of analysis within an experimental unit (Hurlbert, 1984; St-Pierre, 2007).

Treatments (Table 1, Item 4)

Provide specific details of interventions (treatments) used. Terms such as “applied per labeled instructions,” “as per manufacturer’s instructions,” “standard industry practices,” or “routine treatment” do not provide sufficient detail to ensure they are able to be repeated in subsequent studies (Sargeant et al., 2010) and must be avoided. In particular, for studies that involve synchronization of estrus or ovulation, the allocation and
Table 2. Suggested measures to be used in reproductive studies—units of analysis, inclusion criteria, definitions, and statistical methods

<table>
<thead>
<tr>
<th>Measure</th>
<th>Unit of analysis</th>
<th>Inclusion criteria</th>
<th>Event to be defined</th>
<th>Analyses should account for¹</th>
<th>Statistical method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interval from calving to an event:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>To first ovulation</td>
<td>Cow²</td>
<td>All cows</td>
<td>First ovulation date</td>
<td>Clustering of data</td>
<td>Survival analysis with right censoring of cows not yet ovulated when observations ceased</td>
</tr>
<tr>
<td>To first observed estrus</td>
<td>Cow²</td>
<td>All cows</td>
<td>First observed estrus date</td>
<td>Clustering of data</td>
<td>Survival analysis with right censoring of cows not yet observed in estrus when observations ceased</td>
</tr>
<tr>
<td>Interval from either calving or start of breeding period date to an event:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>To first insemination</td>
<td>Cow²</td>
<td>All cows still in group on their start of breeding period date and eligible for insemination; exclude cows inseminated during their voluntary waiting period (VWP)</td>
<td>First insemination date</td>
<td>Clustering of data; VWP duration</td>
<td>Survival analysis with right censoring of cows not yet inseminated when inseminations ceased</td>
</tr>
<tr>
<td>To pregnancy</td>
<td>Cow²</td>
<td>All cows still in group on their start of breeding period date and eligible for insemination; exclude cows inseminated during their VWP</td>
<td>Pregnancy date</td>
<td>Clustering of data; VWP duration</td>
<td>Survival analysis with right censoring of cows that did not become pregnant. For the latter, the interval ends when the cow left the study (died, sold, completed the experiment, became ineligible for breeding)</td>
</tr>
<tr>
<td>Binary outcome variables³</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pregnancy to first insemination</td>
<td>Cow²</td>
<td>Only inseminated cows that were pregnancy tested long enough after first insemination to detect any resulting pregnancy</td>
<td>Pregnancy date</td>
<td>Clustering of data; calving to first insemination interval</td>
<td>Logistic or Poisson regression</td>
</tr>
<tr>
<td>Pregnant at the end of mating period</td>
<td>Cow²</td>
<td>All cows in group and eligible for insemination from their start of breeding period date until the first Pregnancy date or end of mating date plus time to pregnancy test; exclude cows inseminated during their VWP</td>
<td>Pregnancy date</td>
<td>Clustering of data; VWP duration; duration of breeding period</td>
<td>Logistic or Poisson regression</td>
</tr>
<tr>
<td>Pregnant at the end of the study</td>
<td>Cow²</td>
<td>All cows in group and eligible for insemination from their start of breeding period date until the first Pregnancy date or end of study date plus time to pregnancy test; exclude cows inseminated during their VWP</td>
<td>Pregnancy date (lengths of time from start of breeding period dates to last pregnancy diagnosis should be described.)</td>
<td>Clustering of data; VWP duration; duration of breeding period</td>
<td>Logistic or Poisson regression</td>
</tr>
<tr>
<td>Pregnancy loss</td>
<td>Cow²</td>
<td>Only cows that were diagnosed as pregnant and subsequently monitored for pregnancy loss. The method by which pregnancy was first diagnosed and by what day post insemination; and the days after confirmation of pregnancy was pregnancy loss diagnosed should be provided.</td>
<td>Pregnancy loss</td>
<td>Clustering of data. If binary variable (i.e., loss = yes or no), times in which pregnant cows were observed for pregnancy loss</td>
<td>If binary variable (loss = yes or no), logistic or Poisson regression. If time from confirmed pregnancy to assumed date of pregnancy loss, use survival analysis with right censoring of pregnant cows that did not lose their pregnancy at last time that any loss would have been detected.</td>
</tr>
</tbody>
</table>

¹Adjustment required if the stated variable varies between cows; other potential confounders should also be assessed.
²Or lactation where multiple lactations from the same cow are included.
³For binary outcomes, Poisson regression to estimate risk ratio rather than the odds ratio from logistic regression is acceptable or even preferable (see Ospina et al., 2012).
timing of treatments should detail both the intervention and the animals receiving it with regard to time from calving and parity and any other factors determining when treatments are administered. In many cases, a clear diagram that details the precise timing of interventions is useful. Details on the doses and source of different preparations used should also be provided. For pharmaceutical interventions, the minimum description should include the compound name, the concentration, the dose, the delivery matrix, and the route and the frequency of administration (Sargeant et al., 2010).

Particular attention is needed to describe the trial design and at which level (e.g., farm, pen, animal, or a lower level) that treatment was applied. Details should be provided to ensure that differences in management among treatment groups are clearly identified, especially details of environment, nutrition, grouping, and other differences among subgroups of cattle. Allocation of semen may not be balanced across cattle of different production or genetic value and could be a source of bias. Consequently, the method of allocating semen should be described. Further, differences in the performance of inseminators are possible and the allocation of inseminators across treatment groups should be accounted for in study design and analysis. The method of controlling for potential confounding effects of inseminators should be described. If randomization is conducted within subgroups, this should be clearly stated; for example, “heifers in each herd were randomized to treatment 1 or 2 within farm”; or “pens were allocated within each farm to treatment A or B.” It must be clear whether the allocation of treatment was made at the level of the individual or at the group. This is particularly important for nutritional interventions; ambiguity on this point is common. For example, “each cow was individually fed with XX g of product Y per day or with XX g of placebo as a top-dress for the total mixed ration to which only that cow had access; individual daily feed intakes were recorded.” Full details of dietary interventions, ingredients and chemical composition, and placebo should be provided; approaches to describing rations are provided below.

Bacterial or viral challenge studies are rarely used in reproductive research but examples do exist (McGowan et al., 1993); for a detailed discussion of the use of these, readers are directed to Sargeant et al. (2010). Although not strictly challenge studies in the typical sense, studies of metabolic clearance (e.g., progesterone catabolism in circulation, or hyperinsulinemic-euglycemic clamps) should preferably be conducted in animals that are relevant to the objective of the study; that is, nonlactating or ovariectomized heifers or cows may not be representative of dairy cows in early lactation.

**Blinding (Table 1, Item 11)**

There are obvious advantages in blinding assessors of outcomes to treatment allocation. Unlike in human medicine, there is no need to obscure the nature of the treatment from the patients themselves to avoid a psychological “placebo effect,” particularly on self-reported outcomes; rather the focus is on those who conduct the study. However, “sham” treatments with placebos (as distinct from no treatment) may be required to maintain blinding to avoid differential measurement error (i.e., measurement error that varies by treatment). For example, if treated cows are handled or restrained, their feeding and social routines disrupted, or a possibly unpleasant experience imposed (e.g., receiving an injection), these actions should equally apply to the control group even if they are not strictly necessary to deliver “untreated” status, unless that comparison is the goal of the study. In a meta-analysis of human experiments comparing blinded and nonblinded assessments of binary outcomes in patients, despite a generally high degree of agreement between assessors, Hróbjartsson et al. (2012) identified an optimistic, average exaggeration of the nonblinded odds ratio by 36% relative to the blinded odds ratio. It is unlikely that objective outcomes such as pregnancy diagnosis would be influenced by conscious or unconscious biases if treatment allocation were known, or that animal identifications and treatment assignments would be remembered when analyzing plasma samples in a large study. However, outcomes such as detection of estrus based on secondary signs or measurement of the size of ovarian structures by ultrasound, or assignment of BCS are susceptible to differential bias if observers are not blinded to treatment status. It may be sufficient to cause bias if observers are aware that animals with odd-numbered ear tags or in a certain group received a different treatment than other animals, even if they do not know which treatment a particular group received.

Whereas attention to randomization is designed to provide the best opportunity for allocation of treatments to otherwise similar groups, blinding should ensure that groups are treated equally. In many cases, placebos that look, smell, or feel like the test articles are required to ensure that there is blinding. In some cases, especially with nutritional interventions, it can be difficult to develop a sham treatment. If blinding is not used, a rationale for this decision and discussion of possible biases should be provided in the text. Sargeant et al. (2010) provided a detailed outline and rationale for the use of blinding and suggested that a detailed description of how treatments were blinded from those involved in assessing outcomes would im-
prove understanding of the risk of bias due to awareness of treatment. Recommendations were similar to those in CONSORT (Altman et al., 2001). Haahr and Hróbjartsson (2006) reviewed 200 human experiments published after CONSORT (in 2001) and found that, of 156 articles described as “double blind,” only 2% of the articles explicitly described the “blinding” status of patients, health care providers, and data collectors. Haimerl et al. (2013) examined the design of studies used to examine prostaglandin treatment of endometritis in cattle. Differences in estimates of treatment effects based on study design were found and, despite a limited study base, there was evidence of failure to randomize or blind studies.

**Outcomes (Table 1, Items 5 and 6)**

It is critical to define the case or reproductive event and outcomes (see Table 2 and Supplemental Table S1; http://dx.doi.org/10.3168/jds.2015-9445) and to specifically define the populations that are eligible for inclusion in calculations of each measure. Definitions of events are vital to controlling sources of variation in studies. Authors should clearly state the voluntary waiting period (VWP) applied to each experimental unit when insemination or pregnancy is an outcome. The breeding period start date should be defined as calving date plus voluntary waiting period + 1, where VWP is the postpartum period during which no inseminations are allowed, even when cows are detected in estrus. It is acceptable to assume that this is the same for all cows in a group as long as it reflects the actual decisions about when first insemination occurs.

In all cases, the period of putative treatment effect should be defined. For example, an intervention may have a putative effect on the outcome at a single mating or throughout the breeding period. The time at which cows may be eligible for treatment is important. Statements such as “only cows less than 150 days in milk at the time of the first injection were enrolled” and “the follow-up period for determination of pregnancy was 60 days after the last mating, and pregnancy was reconfirmed at 7 months of pregnancy by rectal palpation” provide clarity.

For seasonal and split-calving herds, all or most cows are allocated the same calendar date for their start of breeding period date (also known as the herd’s mating start date or planned start of mating). Thus, the VWP for these cows ends 1 d before the herd’s mating start date. The breeding period duration for each cow is defined as the last date at which insemination (artificial or by bulls) would be allowed should estrus be detected — start of breeding period date + 1.

Standard operating procedures should be in place and described for all procedures used and previously validated systems of measurement should be referenced; for example, validated body condition scoring, lameness measures, or cut-points for chemical determinations of pregnancy should be used.

**Sample Size and Power Calculations (Table 1, Item 7)**

A power \((1 − \beta)\) of 80% and type 1 error risk of 5% \((\alpha)\) are often used in calculations of sample size. For studies examining continuous outcomes, an estimate of the magnitude of the treatment effect is the effect size of a study, which is the difference in the experimental group mean minus the control group mean, divided by the pooled standard deviation of the two groups. Most reproductive studies evaluate binary outcomes (Table 2; Supplemental Table S1; http://dx.doi.org/10.3168/jds.2015-9445), and effect size estimates for these studies can be based on the differences between proportions in outcomes for the experimental and control groups. The effect size difference that can be detected will be inversely related to sample size: the smaller the difference, the larger the required group sizes. The sample size is also determined by the type of alternative hypothesis investigated; that is, no difference, inferiority, superiority, and nonequivalence, and for the latter 3, the size of the margin used to determine the difference between groups. Numerous texts, papers, and software (Donner, 1984; Nakagawa and Cuthill, 2007; Tempelman, 2009; Piaggio et al., 2012) discuss and provide formulas concerning methods and considerations for power or sample size calculation, so these details will not be discussed further here.

For illustration, to detect an intervention effect that increases the probability of pregnancy at an insemination from 35 to 40% with a power \((1 − \beta)\) = 0.8 and \(\alpha = 0.05\) requires 1,471 cows per group. Additional cows may be required to account for clustering (i.e., lack of statistical independence) of cows within groups and herds. Sample size calculations should be made considering exclusions after enrollment and losses to follow-up that might influence testing of the primary hypothesis; for example, deaths or culls during the period from before calving to start of the breeding period would reduce the number of cattle available to test differences among treatment groups. Although, ideally, losses are controlled by careful study management, large studies and longitudinal studies (sometimes over several years) will have loss of study units. It is important to describe losses of study units and the reasons for these losses because these may influence the validity of the study. It is common for 5 to 10% of cows to be culled...
between calving and the start of the breeding period, and a larger proportion may experience some form of disease that might be an exclusion criterion. For example, if the study excludes cattle with any clinical disease events before or during the study or explores a subgroup analysis of only cyclic or successfully synchronized cattle, these results will not be applicable to herds as a whole.

**Randomization and Blocking (Table 1, Items 8 to 11)**

A statement such as “cows were randomly assigned to treatment” does not contain sufficient detail to ensure that a truly randomized assignment of study units occurred. The method used to assign study units should be described in sufficient detail to understand how the random assignment occurred and how any stratification or blocking of subpopulations was managed. The defining principle of this process is to ensure that within each study stratum, cattle have an equal opportunity to be assigned to each group and that the treatment to which a specific study unit is allocated is determined by a chance process and cannot be predicted (Sargeant et al., 2010). Ad hoc allocation is not a random process; the study should only be called a randomized trial if a formal random allocation process was used. In reproductive studies, there is a considerable risk if awareness of group allocation exists for farm workers or managers to manipulate the allocation of treatment, if there is a perceived benefit for high-producing or valuable cows.

The application of blocking should be done with caution but it can reduce the risk of having imbalanced groups by chance in terms of obvious confounding factors such as breed, calving date, parity or age, location, or production level. Blocking can be most useful when there is large variation between study units in strong determinants of reproductive performance and when sample sizes are small. Consideration should be given to previous treatments in herds in which research is commonly conducted because prior treatments might influence responses in the study in question.

A simple, explicit statement of who was responsible for generating the random numbers sequence, specific details of who managed the method of allocation of treatments to the study units, and how the sequence was concealed from participants, including staff in the study, will provide evidence of study rigor. Terms such as “blind” or “double blind” alone should be avoided, as these are not explicit.

**Periods of Recruitment and Follow-Up (Table 1, Item 13)**

The period over which study units were enrolled into a study, as well as a clear description of the duration of observation of covariates and outcomes, provides an understanding of seasonal influences and the similarity of exposure to environmental conditions for the cattle throughout a study. Marked differences may exist in environmental and nutritional exposures for cattle that calve several weeks apart. These differences should be reported, and it may be appropriate to control for this effect either by blocking or by statistical means. Further, extended periods of follow-up can influence reproductive studies because cows that are slow to conceive may have more opportunities for breeding and provide data that are skewed to the right in terms of time-to-event data or increase the numerator of an outcome such as services per conception. Adverse or extreme weather conditions may influence responses to treatment and should be reported. Defining the period of eligibility for observation of estrus or for mating is critical in reproductive studies examining time to estrus or pregnancy, respectively.

**Laboratory Analytical Methods**

If assays are validated to standard analytes of interest, reference to that validation is sufficient to satisfy this aspect of the reporting requirements. The Food and Drug Administration (2001) provides a sound reference that should be consulted for additional detail on the use of analytical methods. Analytic tests (RIA or ELISA) should be validated for the species and sample type; for example, serum, plasma, or other. If plasma is used, the assay should be validated for the particular anticoagulant in the plasma. A description of assay validation includes the primary and secondary antibodies used (species and source) and the dilution or concentration of the primary and secondary antibodies.

One important feature of the primary antibody is its specificity with respect to other closely related antigens. Cross-reactivity is classically defined as the ratio of the concentration of the unlabeled antigen to the concentration of the cross-reactant at 50% displacement of the standard curve and is typically less than 1%. The species and source of the standard should be included in the assay validation. For RIA, a description of the tracer, including the method of radioactive labeling and specific activity (mCi/mg) of the tracer, should be included. The total number of disintegrations per minute (dpm) added to each assay tube should be reported. The percentage of radioactivity bound at the specified antibody dilution and the percent of non-specific binding should be reported. The composition of assay buffers, incubation times and temperatures, the method for separation of bound and free, and the method for measuring radioactive decay (i.e., counting) within the bound or free fractions should be explained.
The stability of samples should be tested by evaluation of 3 samples at high, middle, and low concentrations through 3 freeze-thaw cycles (Food and Drug Administration, 2001). A standard curve for the assay should be described in terms of the mass of hormone at each point of the standard curve. It is important to include high-, middle-, and low-concentration quality control samples in each assay run. These should be prepared in triplicate based on the number of anticipated assay runs. The test should be designed so that the bulk of the assay determinations reside within the middle of the mass range, where the accuracy and precision of the test are the greatest.

The analytical sensitivity or lower limit of quantification of the test should be reported as the minimum hormone concentration or mass that can be distinguished from zero. Some authors use a value of average percent binding of the zero tube (no antigen) minus 2 SD as a reasonable approximation of the percent binding at the minimum sensitivity of the assay. One validation criterion is that assaying a serial dilution of serum or plasma should yield a displacement curve that is parallel to the standard curve (often referred to as “parallelism” or “linearity”). An additional expectation of a valid assay is that adding a known amount of mass to a sample will yield a recovery of the added mass of 100% (often referred to as a test for “recovery” or “added mass”). If 1 ng of hormone is added to a serum sample containing 1 ng, for example, then the assay result should be 2 ng (100% recovery of added mass). For assays that require extraction of the antigen from serum or plasma, an “extraction efficiency” should be included (percent of antigen mass removed from the sample via the extraction method). The extraction efficiency is typically estimated by spiking a known amount of radiolabeled antigen, performing the extraction, and then measuring the amount of radioactivity in the extracted tube. Samples are typically analyzed in duplicate at a given volume and a correction factor is used to convert the mass measured within a given assay tube to a concentration (mass per unit volume of serum or plasma). The average intra (within)-assay coefficient of variation (variation among duplicate samples measured in the same assay) as well as inter (between)-assay coefficient of variation (variation among duplicate samples measured in different assays) should be reported. If possible, assay results should be compared with other existing tests. A final test involves demonstrating that the assay measurements are consistent with known physiological events. For example, an assay for bovine LH should detect relatively low and high concentrations of LH, respectively, before and after an injection of GnRH in cattle.

The procedures for assaying hormones in serum or plasma have evolved over time. An important change over the last 2 decades has been the commercialization of assay kits for the measurement of hormones and metabolites. The new kits (typically in an ELISA format) are convenient and may require less time for assay completion compared with traditional RIA. Unfortunately, a commercially available kit may have not been validated for the target species or the biological fluid (serum, plasma, or otherwise) that is used in the assay. When ELISA is used to measure a hormone, then there should be a minimum expectation for assay validation. This minimum expectation should include the specificity of the primary antibody, a description of the standard curve in terms of the mass of hormone at each point, the sensitivity of the assay, parallelism or linearity of diluted samples, percent recovery of the added mass, the intra- and interassay CV, and comparison with existing RIA or ELISA or a demonstration of hormone changes during known physiological events.

Statistical Methods and Data Analysis (Table 1, Item 10)

The methods recommended for analysis of some common reproductive measures are outlined in Table 2, which provides details of outcomes, unit of analysis, inclusion criteria, and suggested methods of analysis; information for further outcome measures are in Supplemental Table S1 (http://dx.doi.org/10.3168/jds.2015-9445). The principles behind these recommendations are fundamental to the nature of the data being analyzed. New statistical techniques may eventually improve on those recommended here, but these represent the most correct currently available methods of analysis of reproductive data. Errors in analysis include evaluation of dichotomous outcomes as if they were continuous and normally distributed (i.e., using linear regression). Failure to use a correct technique for the distribution of the outcome (e.g., pregnant to an insemination; yes or no) will, at least, typically underestimate variance, leading to incorrect application of statistical approaches and increased possibility of type I error. For dichotomous outcomes, logistic regression is a correct method and produces an odds ratio. When the outcome is common (experienced by >10 to 20% of study units (e.g., probability of pregnancy per insemination), the odds ratio will be more extreme than the risk ratio. Application of Poisson regression to estimate the risk ratio may be preferable and produces effect estimates that are more intuitively expressed and understood (Barros and Hirakata, 2003; Ospina et al., 2012; O’Connor, 2013).
When the outcome is the interval to an event (insemination, pregnancy, pregnancy loss, or culling), the correct method is time-to-event (survival) analysis (Lee et al., 1989). A common error is to use linear regression to assess the mean time to an event, including only those cows that experience that event. This introduces a potentially substantial bias by excluding all cows that failed to achieve the endpoint of interest (i.e., ignoring all cows that did not become pregnant despite being eligible and under study for weeks or months), and by having incorrect variance for the data that are analyzed.

When herd-level analysis is undertaken, the distribution of binary cow- or lactation-level variables aggregated to the herd level may be treated as continuous data. Although one cow is either pregnant or not, the annual 21-d pregnancy rates for many herds might range from 30 to 50%, with essentially all values between the represented range (or at least possible). In cases of such herd-level data, linear regression may be adequate; this can enhance communication of results but some caution is required. The underlying binomial distribution is heteroscedastic, although this is less pronounced if all observed percentages are between, for example, 30 and 70%, and fitted values can be outside 0 to 100%. Testing that residuals are normally distributed and that all predicted values lie between 0 and 100%, and undertaking transformations for analysis if necessary, are important.

Statistical methods must account for any clustering of the outcome variable, because assumptions of statistical independence of observations often do not hold (e.g., cows within a pen or herd). Clustering can occur when the analysis includes multiple inseminations (or other observations) from the same lactation, multiple lactations from the same cow, or multiple cows from the same herd. The effect of farm or pen needs to be considered in multi-site studies or when multiple pens within a site are used. The clustering effects should be accounted for using either fixed effects or, more often, random effect models (McDermott and Schukken, 1994; Schukken et al., 2003). References on study design that examine the unit of analysis and discuss the importance of clustering in general (Dohoo et al., 2003) and in cattle studies (St-Pierre, 2007; Tempelman, 2009) are available.

It is strongly recommended that investigators seek statistical advice in both the design and analysis phase of the study to ensure that appropriate designs, sample sizes, and statistical analyses are used. The statistical methods applied to each analysis should be described and, critically, the unit of analysis should be stated. When a reference group is used, the reference group must be clearly stated for each variable analyzed.

Censoring

For time-to-event variables (e.g., time to pregnancy), reasons and times of right censoring must be stated, and numbers of study units for each reason reported. Right censoring results when cattle die or are otherwise lost from a study before the end of the study period (e.g., culled before pregnancy status is determined). When modest numbers of lactations are right censored, the sensitivity of effect estimates to the 2 extreme deviations from independent censoring assumptions should be assessed and reported. These extreme deviations are (1) that all censored records experienced the event at the time they were censored, and (2) that no censored records would have experienced the event in a long period had they not been censored.

Example: Survival Analysis.

“For the analysis of time to pregnancy, cows that were not diagnosed pregnant were considered right censored. A right censored observation meant that pregnancy had not occurred by the end of study or by the time observation of the subject had ceased. Overall, 42 of the total of 212 cows were censored for time to pregnancy. Similarly, if the first ovulation was not observed by 80 DIM, the observation was right censored; 8 of the 76 cows [from Bernal-Santos et al. (2003) and Castaneda-Gutierrez et al. (2005)] were censored (de Veth et al., 2009).”

For each result presented, the number of observations used in the analysis should be reported and discrepancies among outcomes and between planned and actual numbers explained; for example, due to omission or inability of collection of a sample or based on exclusion of a study unit (including the reason for exclusion). The numbers included in the analysis should be described as being based on intention-to-treat or on final inclusion criteria. Normally, the intention-to-treat basis should be included. For example, this would mean providing results for all animals enrolled in the study, and not exclusively for those that, for example, were detected in estrus or were retrospectively determined to be cyclic or to have been successfully synchronized by the treatment. The latter subgroup analyses are legitimate for exploring hypotheses and probing the nature of responses to treatment but should not be the only analyses reported. Specifically, the numbers and proportions of animals initially eligible but not enrolled or enrolled but not included in each final analysis should be reported explicitly by treatment group, along with reasons for these exclusions.
For the primary outcomes and each secondary outcome, at least crude descriptive statistics of the absolute performance and measures of variance (e.g., mean and SD or median and interquartile range) must be reported. The process, methods, and cut-points for screening of covariates to be assessed in multivariable models and for checking collinearity among predictors must be described. This should include details of the procedure (e.g., “the association of each binary covariate with the outcome was assessed in a $2 \times 2$ contingency table using the Chi-square statistic, and those with $P < 0.2$ were offered to the multivariable model,” and “correlation among continuous covariates was assessed with Pearson correlation coefficients and if $r > |0.8|$, the more biologically plausible variable was retained and the other not analyzed further.”) Statements such as “variables were screened with SAS” are insufficient to provide an understanding of the approach used in the analyses. It is not necessarily advantageous or prudent to test all possible interactions, but in a randomized trial, at least interactions between treatment and significant covariates in the final model, or those putative interactions with treatment expected and accounted for in the study design should be tested in the statistical model. It may be helpful to describe the statistical “materials” (e.g., Proc Mixed in SAS) but it is necessary to report the method (e.g., “mixed linear regression, including a random effect of herd and accounting for repeated measures of plasma progesterone”). The structure of accounting for correlation of outcomes (repeated measures or clustering) and the means by which the structure was chosen should be described [e.g., “the correlation structure that produced the lowest Akaike information criterion for the final model was selected such that repeated measures of the outcome were accounted for with an autoregressive (type 1) correlation structure, while the clustering of cows within herds was accounted for with a compound symmetry structure”]. The method of model building (e.g., manual or automated; forward selection or step-wise; backward elimination or step-wise) must be specifically described (as these terms are used with varying meanings), and steps taken to identify confounding effects should be specified. The effect measures must include a measure of variance (standard error of differences, confidence interval, or standard errors of means). If treatment effects or results in each group are reported as least squares means, the covariates accounted for in the least squares means must be clearly specified. There are mixed views about fitting baseline covariates (i.e., measures from each subject taken before treatments are first administered) in statistical models from randomized controlled trials. Models with no such covariates fitted should be reported along with any other models.

**Results: Flow of Study Units (Table 1, Items 17 and 18)**

It is almost inevitable that cattle will be lost to death or removal or lost to follow-up during studies, and a critical part of understanding the rigor with which a study is conducted is to understand how many subjects were lost, the causes of loss, and the timing of that loss. A table or figure outlining these losses by group is ideal for addressing this requirement (Sargeant et al., 2010).

Deviations to the study protocol and the number of animals to which this applied should be listed and reasons for these should be detailed. Similarly, if there are no deviations to protocol, this should also be noted (Sargeant et al., 2010).

**Baseline Demographics (Table 1, Item 15)**

A brief statement or table outlining the demographics of each treatment group provides essential information contributing to the understanding of a study. The information should be reported to ensure that numbers of study units are easily identified, that information for each relevant level of the organizational structure is present, and that the form of presentation provides an opportunity for further analysis (Sargeant et al., 2010). Reproductive studies will be influenced by many factors, including the age, parity, weight, milk production, and nutritional status of the cattle involved. At the herd level, weather, nutrition, stocking density, type of housing, or pasture system could influence responses. Consequently, baseline information on these should be provided.

When studies are used as data in subsequent studies (e.g., meta-analysis), herd- or study-level descriptions of the environment and nutrition of the group may help to explain the heterogeneity of variance observed among studies. Substantial heterogeneity has been observed in studies examining the effects of hormonal interventions designed to increase conception at a synchronized estrus. Consequently, we provide suggestions for providing information on the cows, environment, and nutrition (Supplemental Table S2; http://dx.doi.org/10.3168/jds.2015-9445).

**Baseline Data: Environment and Nutrition (Table 1, Item 15)**

Guidelines to environmental descriptions are provided in Supplemental Table S2 (http://dx.doi.org/10.3168/jds.2015-9445). These guidelines are not in themselves part of the reporting guideline checklist but elaborate on the general requirement to describe the environmental, management, and nutritional context of the study.
These recommendations are modified from those developed by Lean and Lean (2010) and are based on the information required for the CPMDairy model (CPM-Dairy, version 3.08; http://cahpwww.vet.upenn.edu/node/77). The extent to which conditions are described will depend on the nature of the hypotheses. Studies that have extended periods during which reproductive outcomes may be influenced or that involve multiple commercial herds and do not have a primary focus on effects or interactions of nutrition or the physical environment (weather or housing) may provide less detail on nutritional management or weather conditions. It is important to report in detail the assays used to make determinations of feed value as the assay methods can vary among laboratories used for testing. We recommend testing of individual feed components rather than total mixed rations, although testing pooled samples of these obtained directly from mixer feed-out can be useful.

The challenges outlined above for the execution of nutritional intervention studies (in particular, the difficulty of inevitable confounding for larger interventions resulting from the additions of lipids, proteins, or carbohydrates) require that descriptions of diets be detailed. In particular, the intervention under study should be fully described. If a fat is the intervention studied, detail on the fatty acid profile and level of ruminal protection, if present, should be provided. Similarly, protein-based interventions should have amino acid profiles provided. Descriptions such as “used according to manufacturer’s recommendations” or stating the composition according to the guaranteed analysis is not enough—guaranteed analyses are usually stated as minimum or maximum guarantees and may not adequately reflect the nutrients that a supplement provides. Confounding arises if the addition of one ingredient to a ration results in an incremental increase in DMI or, if DMI is decreased or static to controls, the loss of intake of another nutrient or nutrients. Attention must be paid to changes in diet over time.

Pasture feeding poses particular challenges to ensuring that the description of environmental conditions is adequate, especially for nutritional inputs. Estimations of pasture mass can be made using physical cuts of pasture, electronic density measures, plate meters, or geospatial methods. Alkanes and other markers can be used to estimate individual intakes of cattle. However, the precision of these methods is very variable; therefore, the precision with which dry matter estimation is made is low compared with daily measures in confined systems, and not frequently made at the level of the individual. Similarly, pasture varies in composition seasonally and during a day and careful thought needs to be given to the type of study attempted and level of precision of pasture nutritive values required to support the hypotheses studied. Care should be taken to ensure that pasture studies are not pseudo-replicated. Specifically, care should be taken to ensure that there are replicated groups of cattle grazing on pasture and not simply two herds, even if these have randomly assigned pastures. A good example of a replicated study design for cattle on pasture is provided by Auldist et al. (2013). Cattle on pastures are exposed to internal and external parasite loads or challenge; a description of parasite status, and possibly treatment to remove these, before or during the trial may be needed.

**Baseline Data: Disease (Table 1, Item 15)**

The periparturient period has the highest incidence of disease for adult cattle. We provide clear and reproducible descriptions of disease (Supplemental Table S3; http://dx.doi.org/10.3168/jds.2015-9445), whether recorded as an outcome or a covariate. In cases of metabolic disease, the comments of Lean et al. (2009) concerning definitions of these disorders are useful in providing a series of postulates by which the suitability of metabolic disease definitions can be assessed. In some of the cases of disease outlined in Supplemental Table S3 (http://dx.doi.org/10.3168/jds.2015-9445), the postulates provided by Lean et al. (2009) have not been met; however, the definitions represent the best available evidence.

**Adverse Effects (Table 1, Item 19)**

Adverse or side effects including injury, disease, or death that could be reasonably associated with treatment must be reported and the absence of these should be stated. These may include apparently direct (e.g., hypersensitivity reaction to an injected treatment) or indirect (abscess at an injection site or physical injury in the course of treatment administration) effects of treatment.

**Discussion and Interpretation (Table 1, Item 20)**

The study’s novelty, strengths, benefits, limitations, weaknesses, and harms should be explicitly described and discussed both relative to the hypothesis and to other published work. There should be a critical assessment of whether the objectives of the study were met, and what the implications of the results are, both for application in the field and for future investigation. Rather than a methodical review of each experiment or result, the discussion should include a synthesis of the methods and results to address the points above. Thoughtful critique of what could have been done differ-
Validity and Generalizability (Table 1, Item 21)

The concepts of external and internal study validity are central to the discussion and interpretation of data. The greatest internal validity of an observation results when experimental conditions are tightly controlled (e.g., under laboratory conditions) and there is little bias or measurement error. However, responses observed in tightly controlled studies may not reflect the response or variability of responses that can be expected in the field with more uncontrolled or unmeasured covariates. In other words, the laboratory study may have a low external validity. The study designs with the greatest external validity are randomized controlled clinical trials, especially multicenter studies, and meta-analytical studies derived from these. Sargeant et al. (2010) noted that a randomized trial conducted at a single site may not be representative of the variety of clinical situations possible, and authors should acknowledge this. The discussion should explicitly include comments on the implications of the study design, subjects, and environmental context on the generalizability of the results (i.e., the external validity or the extent to which the results and inferences can be extrapolated and applied beyond the study animals and site).

Many of the guideline sections above reflect the need to achieve excellent internal validity of a study. For example, the section on assays outlines the methods required to ensure that hormonal measures undertaken are both accurate and precise. If these are well conducted and documented, the internal validity of such measures should be good, leading to confidence in the results. The external validity of a study, however, reflects the extent to which a study can be applied beyond the population in which it was conducted. The target population for a study must be considered and inferential errors can result from studies that fail to sufficiently consider the context in which a study was conducted, or extend the relevance beyond the population in which the study was conducted. An example might be to conclude that the results achieved by use of an intervention in maiden heifers would be applicable to lactating cow populations.

The atomistic and ecological fallacies should be considered in study design and in interpretation and discussion of results. The atomistic fallacy occurs when a relationship observed at a lower level is incorrectly assumed to occur also at a higher level (Diez-Roux, 1998; Diez Roux, 2002). The ecological fallacy occurs when a relationship observed at a higher level is incorrectly assumed to occur also at a lower level (Diez-Roux, 1998; Diez Roux, 2002). These fallacies can occur only when a relationship at one level differs from the corresponding relationship at a higher (or lower) level. For example, higher-producing herds may have overall better reproductive performance, but within herds, the higher-producing cows may or may not have worse reproductive performance (Bello et al., 2012). A treatment that increases the probability of pregnancy to first insemination may be associated with a difference in the interval to AI. Consequently, pregnancy at the group level may not be altered. If, for example, the treatment reduces probability of detecting estrus, or reduces the probability of pregnancy at subsequent inseminations, or delays the next ovulation in nonpregnant cows, then pregnancy at the group level may be reduced.

General Interpretation of Results (Table 1, Item 22)

The general interpretation of results should allow the reader to understand the significance of the work conducted in the context of the effects on industry and society in general. The work should be discussed with regard to existing or similar studies; if similar studies do not exist, authors should indicate the limitation that this represents to broader interpretation (Sargeant et al., 2010).

Conflict of Interest

A simple statement outlining funding sources that provided assistance or that sponsored the study must be provided. If the investigators have any perceived conflicts of interest in producing the document, these should be outlined.

CONCLUSIONS

With developments in laboratory analyses, statistical methods, and the increased scale of animal agriculture internationally, research on animal reproduction needs
to evolve. It is our hope that the checklist detailed here can contribute to improving the quality of reporting, reproducibility, and applicability of research findings. Furthermore, more consistent reporting of experimental conditions can reduce barriers to the conduct of comprehensive meta-analyses, which will help the research community explore the most complex interactions influencing reproduction.

ACKNOWLEDGMENTS

This study received support from Dairy Australia (Melbourne, Australia), Church and Dwight Inc. (Princeton, NJ) and Scibus (Camden, NSW, Australia). Alex Crosbie contributed to the development of Supplemental Table S3. We thank Mike Overton (Elanco Animal Health, Athens, GA) and Robert Van Saun (Pennsylvania State University, University Park) for contributions. The authors have no conflict of interest in producing this document.

REFERENCES


INVITED REVIEW: GUIDELINES FOR REPRODUCTIVE RESEARCH


