Assessment of completeness of reporting in intervention studies using livestock: an example from pain mitigation interventions in neonatal piglets

O'Connor, A; Anthony, R; Bergamasco, L; Coetzee, JF; Dzikamunhenga, RS; Johnson, AK; Karriker, LA; Marchant-Forde, JN; Martineau, GP; Millman, ST; Pajor, EA; Rutherford, KMD; Sprague, M; Sutherland, MA; von Borell, E; Webb, SR

Published in:
Animal

DOI:
10.1017/S1751731115002323

First published: 01/01/2015

Document Version
Publisher's PDF, also known as Version of record

An assessment of completeness of reporting in studies evaluating pain management in the neonatal piglet during routine management procedures (review).

A. O'Connor[1], R. Anthony[2], L. Bergamasco[3], J.F. Coetzee[1], R.S. Dzikamunhenga[1], A.K. Johnson[4], L.A. Karriker[10], J.N. Marchant-Forde[5], G.P. Martineau[6], S.T. Millman[1,7], E.A. Pajor[13], K. Rutherford[8], M. Sprague[10], M.A. Sutherland[9], E. von Borell[11] and S.R. Webb[12]

[1] *Department of Veterinary Diagnostic and Production Animal Medicine, College of Veterinary Medicine, Iowa State University, Ames, Iowa, USA*

[2] *Department of Philosophy, University of Alaska Anchorage, Anchorage, Alaska, USA*

[3] *Department of Animal and Poultry Sciences, College of Agriculture and Life Science, Virginia Tech, Blacksburg, VA 24060*

[4] *Department of Animal Science, College of Agriculture and Life Sciences, Iowa State University, Ames, IA, USA*

[5] *USDA-ARS, Livestock Behavior Research Unit, West Lafayette, IN, USA*

[6] *Swine clinic, Department of Animal Production, National Veterinary School, Toulouse, France*

[7] *Department of Biomedical Science, College of Veterinary Medicine, Iowa State University, Ames, IA, USA*

[8] *Scotland'd Royal College (SRUC), Edinburgh, United Kingdom*

[9] *AgResearch Ltd, Ruakura Research Centre, Hamilton, New Zealand*

[10] *American Association of Swine Veterinarians (AASV), Perry, Iowa, USA,*

[11] *Department of Animal Husbandry and Ecology, Institute of Agricultural and Nutritional Sciences, Martin-Luther-University Halle-Wittenberg, Germany*

[12] *National Pork Board, Des Moines, Iowa, USA.*

[13] *Department of Production Animal Health, Faculty of Veterinary Medicine, University of Calgary, Canada*

Corresponding author: Annette O'Connor, Email: oconnor@iastate.edu, Telephone: +1 515 294 5012, Fax: +1 515 294 1072.

Short title: *Complete reporting of welfare studies*

Abstract

Accurate and complete reporting of study methods, results, and interpretation are essential components for any scientific process, allowing end-users to evaluate the internal and external validity of a study. When animals are used in research, excellence in reporting is expected as a matter of continued ethical acceptability of animal use in the sciences. Our primary objective was to assess completeness of reporting for a series of studies relevant to mitigation of pain in neonatal piglets undergoing routine management procedures. Our second objective was to illustrate how authors can report the items in the REFLECT statement using examples from the animal welfare science literature. Fifty-two studies from 40 articles were evaluated using a modified REFLECT statement. No single study reported all REFLECT checklist items. Seven studies reported specific objectives with testable hypotheses. Six studies identified primary or secondary outcomes. Randomization and blinding were considered to be partially reported in 21 and 18 studies respectively. No studies reported the rationale for sample sizes. Several studies failed to report key design features such as units for measurement, means, standard deviations, standard errors for continuous outcomes or comparative characteristics for categorical outcomes expressed as either rates or proportions. In the discipline of animal welfare science, authors, reviewers, and editors are encouraged to use available reporting guidelines to ensure that scientific methods and results are adequately described and free of misrepresentations and inaccuracies. Complete and accurate reporting increases the ability to apply the results of studies to the decision-making process and prevent wastage of financial and animal resources.

55 Keywords: animal welfare, data collection, piglets, pain, reviews

56 **Implications**

57 Authors have an ethical responsibility to report the study design and results in a manner
58 that enables reproduction of results and assessment of bias. In this paper we discuss
59 approaches for comprehensive reporting in animal welfare studies. Checklists such as
60 the REFLECT statement provide guidance for reporting studies. Such standards
61 represent the current minimum for reported standards.

62

Introduction

Complete reporting of study conduct and results has always been an important part of the scientific process, however in recent years there has been a renewed focus on the importance of complete and accurate reporting. Driving forces behind this focus include (1) an increased scrutiny of scientific findings, (2) the manner in which scientific information is applied to the decision-making process, and (3) concerns over wastage of animals and resources used in research endeavors (O'Connor *et al.*, 2010, Sargeant and O'Connor, 2013, Ioannidis *et al.*, 2014, Macleod *et al.*, 2014). The increased use of formal research synthesis techniques, such as risk assessment, systematic reviews and meta-analysis, in the decision-making process of public policy makers and for regulatory purposes also places greater importance on the incorporation of primary research into these methods. These explicit uses of research data have led to efforts that ensure accurate estimates of the magnitude of the effect and that potential for biases are incorporated into research synthesis techniques. If studies are incompletely reported, then the results may not be useable for secondary purposes, and the financial resources are wasted and the ethical value of the animals is unappreciated.. In order to avoid waste of recourses and to appropriately recognize the ethical value of animal research subjects, authors have an ethical obligation to provide as complete and as accurate a report as possible and editors and peer-reviewers have an obligation to ensure that the authors do so.

A common research question used for policy development is the assessment of interventions designed to mitigate an adverse outcome. Numerous recently developed

guidelines exist for identifying what a complete account of an intervention assessment study represents (Kilkenny *et al.*, 2010, Moher *et al.*, 2010, O'Connor *et al.*, 2010, Sargeant *et al.*, 2010, Schulz *et al.*, 2010, Campbell *et al.*, 2012). In areas where the reporting of intervention assessments have been evaluated, reporting has frequently been identified as incomplete (Anttila *et al.*, 2006, Sargeant *et al.*, 2009, Schulz *et al.*, 2010).

We are unaware of other studies that have assessed the completeness of reporting in studies focused on interventions for animal welfare outcomes. The primary objective was to assess completeness of reporting interventions designed to mitigate pain in neonatal piglets undergoing routine management procedures. Our second objective was to illustrate how authors can report the items recommended by a single/uniform reporting guideline framework using examples from existing animal welfare science literature. We sought to identify aspects of study design, analysis, and results that were inadequately reported and provide examples so that education of animal welfare science researchers could be targeted to improve reporting in the future.

Methods and Materials

Study population:

This project used literature identified for a systematic review to identify research gaps and develop recommendations related to pain mitigation in the neonatal piglet undergoing castration, tail docking or ear notching (2009, Dzikamunhenga *et al.*, 2014, O'Connor *et al.*, 2014). Details about the protocol, search, screening process to identify

relevant studies, and resulting review are available elsewhere (Dzikamunhenga *et al.*, 2014, O'Connor *et al.*, 2014). For the assessment of comprehensive reporting, we used the studies relevant to the original review. The unit of concern for reporting was a study/trial. Two or more studies/trial were occasionally reported in a single article. An intervention study/trial must have at least 2 arms (treatment groups).

Reporting consistent with REFLECT (Reporting guidElines For randomized controLled trials for livEstoCk and food safety) guidelines

The REFLECT statement is a reporting guideline for randomized controlled trials that assess interventions for food-producing animals such a swine and is therefore suitable for this topic area (http://www.REFLECT-statement.org) (O'Connor *et al.*, 2010, Sargeant *et al.*, 2010). The REFLECT statement comprises 22 checklist items (***Table 1***), of which we assessed the reporting of 17. The rationale for including these items in a publication is provided by Sargeant *et al.* (2010). The reporting of five REFLECT checklist items was not assessed (***Table 1***). We did not assess study flow (REFLECT checklist item 13) because we expected that studies relevant to the interventions were of such short duration that it was unlikely any loss to follow-up would occur i.e., few piglets would leave the study because the outcome could be assessed. We did not assess REFLECT checklist items 2 (Introduction and Background), 20 (Discussion and Interpretation), 21 (Generalizability) and 22 (Overall Evidence) because they are more prone to subjective assessment.

For REFLECT checklist item 3 (Methods and Participants), we extracted the country in which a study was conducted if it was explicitly reported in the article. Otherwise, the reviewer scored location as “not reported” and the item was “partially reported.” For REFLECT checklist item 5 (Objectives) to be considered “completely reported,” the objectives had to be associated with a hypothesis that related to the outcomes. For REFLECT checklist item 6 (Outcomes), we considered for studies that assessed only one outcome that this was the primary outcome. Otherwise, we expected the authors to designate a primary outcome or this checklist item was considered “incompletely reported.” We also added one item to assess if the studies reported random allocation to group. This was necessary because the REFLECT statement makes the *a priori* assumption that studies are randomized. Based on the assumption that the study is randomized, the REFLECT asks for information about the steps in the randomization approach for assessment of its validity. i.e., sequence generation, allocation concealment and implementation. If a study doesn’t randomize to group, then the steps of randomization will not be reported and listed as missing from the report.

We assessed the reporting of statistical analyses (REFLECT checklist item 12) using the guidelines by Lang and Altman (2014). We considered statistical analyses fully reported if *all* of the following were provided:

1. A full description of the main methods for analyzing the primary and/or secondary objectives of the study;
2. Clear methodology used for each analysis, rather than just listing in one place all the statistical methods used;

3. Confirmation that data conformed to assumptions of the test used to analyze them. In particular, if the analyses specified that 1) skewed data were analyzed with nonparametric tests, 2) paired data were analyzed with paired tests, and 3) the underlying relationship analyzed with linear regression models was linear;
4. Whether and how any allowance or adjustments were made for multiple comparisons (performing multiple hypotheses tests on the same data) when the reported results suggested such adjustment was necessary. For example, when studies reported comparison of multiple time points or trials with 3+ trial arms in the results we expected a report of the approach to adjusting for such pairwise comparisons, i.e., Tukey's, Bonferroni's, etc. If authors did not report the approach, but did report that adjustment was conducted, this was considered "complete reporting";
5. For t-tests only, whether tests were one- or two-tailed and justification for the use of one-tailed tests;
6. Description of the alpha level (e.g., 0.05) that defined statistical significance;
7. The name of the statistical package or program used in the analyses. In this situation we considered reporting complete even if only the program, rather than the package, was reported, i.e., both SAS® and SAS® PROC MIXED were considered "complete reporting".

If at least one but not all of the above were reported, then we considered statistical analyses "partially reported".

The presence or absence of each REFLECT checklist item was independently evaluated by two reviewers. Disagreements were initially resolved by one of the reviewers. Where there was disagreement between reviewers about the presence of a checklist item, one reviewer would re-evaluate the article. If this approach did not resolve the conflict, then the item was discussed with a third reviewer. As with any assessment of comprehensive reporting, quality assessments were not made. For example, we did not assess if the method used to allocate piglets to treatment groups reduced bias, rather we assessed if the approach to allocation was reported.

Reporting of procedures, trial characteristics, study design features and summary measures

REFLECT checklist items are very general, and as some sources of heterogeneity are domain-specific, we also determined if specific aspects of some checklist items were reported. We specifically assessed if the following were reported: type of production system (i.e., all in/all out or continuous flow or not reported), and facility types where the research was conducted (i.e., university-owned farm or laboratory/research facility or privately owned/commercial operation or not reported). We extracted specifics about the reporting of the interventions. We also evaluated reporting of descriptors of the study design: number of animals enrolled in the trial, and number of animals enrolled in trial arms. The inclusion in the report of statistical descriptions of the outcomes, including effect sizes and measures of precision were also evaluated.

Results

A total of 622 articles were identified by original search and of those, 52 studies from 40 articles met the eligibly criteria for the review and were eligible for assessment of the approach to reporting (Dzikamunhenga *et al.*, 2014, O'Connor *et al.*, 2014). All the studies were experimental and therefore should have been randomized trials; no relevant cohort studies were identified. The characteristics of the studies assessed are provided in **Supplementary Table 1**. A summary of the completeness of reporting of items from the REFLECT checklist is shown in ***Table 1***. No single study reported all of the REFLECT checklist items evaluated in this analysis. None of the studies assessed, reported the selection criteria for farms or animals, the approach to allocation to group, the sample size rationale, complete description of statistical methods, baseline data by group for animals enrolled, complete description of the results, information about ancillary analyses or the occurrence of adverse evens by group. Other checklist items were only reported by some of the studies (***Table 1***).

The reporting of the information that would enable end-users to understand the relevance of the study population to a target population was poor. Often, eligibility criteria for the farms and animals used were missing. The frequency of reporting country of conduct and study setting is shown in **Supplementary Table S1.**

Specific intervention information (REFLECT checklist item 4) was reasonably well reported; all studies provided at least some information about the interventions assessed. **Supplementary Table S2** provides reporting examples for the studies that assessed non-steroidal anti-inflammatory drug interventions. In the interest of space the

other interventions are not included. In ***Table 2*** we provide a simple summary of basic outcome measures: means (or proportions) and measures of precision and trial arm sample size; frequently this information was not reported. In **Table 3, Table 4** and **Supplementary Table S2** we provide examples where the REFLECT items were well reported from the studies included in the review. In a few situations, no examples could be found in the 52 studies and examples were drawn from other animal studies. **Table 3** focuses on the description of the methods and materials, while **Table 4** focuses on presentation of the results. The material in the **Supplementary Table S3** relates to the introduction and discussion in a manuscript. The three tables should be used together when preparing a manuscript.

Reporting of REFLECT items that relate to objectives and hypotheses

In the remaining part of the manuscript, we discuss the rationale for a select few REFLECT checklist items so authors are aware of how the information is used by readers; however, a full explanation of the rationale for each REFLECT item is available in Sargeant *et al.* (2010).

Although the objective of the study and sometimes a secondary objective were often provided, very few studies translated the objective into a testable hypothesis that included the metric to be measured (REFLECT checklist item 5). Translating the objective to a hypothesis with a specific metric is important because some metrics may be more valid for specific objectives than others. Therefore knowing the exact metric that will be tested is important. For example, an objective of a study may be to assess the impact of the intervention on pain mitigation, and this would be assessed using a

comparison of the mean Hertz of vocalizations in piglets receiving the anesthetic intervention compared to the mean Hertz of vocalizations in piglets without the anesthetic, i.e., H_0 = mean1 - mean2 = 0. Clarification of the hypothesis ensures the end user knows which metric is being used to assess the objective, and should facilitate identification of the primary outcome.

Reporting of REFLECT items that relate to outcomes and sample size

A clear description of which outcomes were primary or secondary was never explicitly reported by authors who assessed multiple outcomes (REFLECT checklist item 6). The only studies that received a “yes” for this item reported only one outcome. Another item poorly reported was the primary outcome. Knowledge of the primary outcome is necessary to assess the power of the study. Unless explicitly declaring that a study is a pilot or making use of animals used for another purpose, assessments of interventions should be hypothesis-driven. The hypothesis should be specific enough to enable determination that the number of animals enrolled should be sufficient to enable detection of a clinically meaningful difference in the outcome. Researchers therefore should prospectively design and justify the sample size, which requires knowledge of the primary outcome. Further, if authors do not have an *a priori* hypothesis about a primary outcome, the potential to “data mine” for a statistically significant outcome and selective reporting bias is high.

No studies reported the rationale for the sample size (Checklist item 7). This was surprising, as all studies seemed to purposefully assess the effect of an intervention on an outcome and, therefore, the number of animals needed to detect the magnitude of

effect of interest is a prerequisite step in study design. Although reduction of animals included in studies is an important principle of animal research, this concept does not negate the need for sufficient power to detect clinical meaningful changes in the outcome. There are numerous papers devoted to the need for adequately powered animals studies (Cohen, 1997, Hofmeister *et al.*, 2007, Chapman and Seidel, 2008).

Reporting of REFLECT items that relate to confounding - allocation to group/ randomization

REFLECT checklist items 8 through 10 (Sequence Generation, Allocation Concealment, and Implementation, respectively) are based on the assumption that the study is randomized. A description of the method of developing the randomization for the sequence generation, allocation concealment, and implementation, was not provided in any study. Thirty-three of 52 studies used the term “randomly” or “randomized” or “random” in their description of piglet allocation to treatment group. Occasionally it was unclear if the approach used was truly random, despite a description as such. For example, one study described randomly assigning 245 clinically healthy piglets to one of 12 experimental groups. However, the sample sizes in each of the seven relevant arms were very different, suggesting a method other than random allocation. Several studies reported using restrictions of randomization. Blocking by continuous covariates or stratification by categorical covariates was reported in 39 studies. Covariates used were weight, litter, weight and litter, sow or weight, or litter and adoption. No study that controlled for weight using blocking explicitly reported the size of the block. Details about the approach to allocation are part of reporting that enables assessment of

internal validity as they relate to the exchangeability of groups. If it cannot be determined that groups are exchangeable then it is unclear if the observed differences can be attributed to the intervention. Furthermore, without details of the randomization approach, approaches that are haphazard [lacking any obvious principle of organization] or convenient may be incorrectly reported as random. The importance of random allocation is highlighted by authors of the CONSORT statement which we quote here *“Random assignment is the preferred method; it has been successfully used regularly in trials for more than 50 years. (reference in original text) Randomisation has three major advantages (reference in original text). First, when properly implemented, it eliminates selection bias, balancing both known and unknown prognostic factors, in the assignment of treatments. Without randomisation, treatment comparisons may be prejudiced, whether consciously or not, by selection of participants of a particular kind to receive a particular treatment. Second, random assignment permits the use of probability theory to express the likelihood that any difference in outcome between intervention groups merely reflects chance. (reference in original text) Third, random allocation, in some situations, facilitates blinding the identity of treatments to the investigators, participants, and evaluators, possibly by use of a placebo, which reduces bias after assignment of treatments. (reference in original text) Of these three advantages, reducing selection bias at trial entry is usually the most important. (reference in original text) “* (Moher *et al.*, 2010) As many welfare studies are small, it is reasonable that authors would employ restricted randomization tools such as stratification and blocking to increase the power of studies. Regardless of the approach

to randomization, it should be described fully so that end users can assess the potential for bias.

Reporting of REFLECT items that relate to performance and measurement/information bias- blinding

Of the 52 studies, 18 reported blinding as part of their protocol; however, none provided a full description of the approach used to blind the study (REFLECT checklist item 11). Blinding, whether for allocation of treatments or interventions or assessment, was infrequently reported by authors. As blinding is designed to reduce measurement/information bias, it is important to know if outcome assessment is biased. There is some evidence in veterinary science and animal welfare that absence of blinding is associated with more positive outcomes (Burns and O'Connor, 2008, Tuyttens *et al.*, 2014).

Reporting of REFLECT items related to Statistical Methods

Statistical methods (REFLECT checklist item 12) were not reported in 8 studies. In the remaining 44 studies, statistical methods were considered partially reported because they failed to meet all the criteria described above. Assessment of comprehensive reporting of statistical methods is very difficult; the measure of comprehensiveness is that a reasonably informed individual would be able to assess the validity, although what is “reasonable” might appear itself subjective. We would encourage authors to consult with documents published previously that describe what should be included in a description of statistical methods. (Lang and Altman, 2014).

Reporting of REFLECT items related to setting, study population characteristics

Dates relevant to the study recruitment and performance were described in only 6 studies (REFLECT checklist item 14). Although it is difficult to envision how year or season could affect the response of piglets to pain mitigation, such information is very relevant for other topics especially those that seek to understand the influence of season or year on an outcome. The same principle can be inferred for study location (i.e., country or region or production system).

Baseline demographics and clinical characteristics of each group were generally poorly reported (REFLECT checklist item 15). When weight and age information were presented as summary measures for all enrolled pigs together, we considered this to be partially reported. It is recommended that authors provide demographic information about the groups separately, so that end users can assess if the groups are comparable especially given the absence of reporting of allocation methods. Demographic information was frequently reported in the methods section and not explicitly in the results section. REFLECT and other statements make the distinction that the methods and materials could, and potentially should, be written before the study is started, therefore the demographic information of the study groups such as the mean age and mean weight (including standard deviations) are a result and should be presented in the results section.

Reporting of REFLECT items related to results of analyses

The actual number of piglets that contributed to data analyses (REFLECT item 16) was frequently not reported. Presumably authors felt that reporting the number of enrolled animals would suffice because the potential for loss-to-follow up in the subject matter studied was low. For this topic area, this assumption may be valid and failure to report that no loss-to follow up occurred may not be a source of bias. Sometimes the unit of analysis was not the same as the number of animals in the study. This was particularly important for the behavior data, which could be reported as number of pigs that demonstrated an activity or the number of time periods when an event was observed. These clearly have different denominators. Similarly, some outcomes appeared to be measured only on a subset of enrolled animals, perhaps because testing all animals was time consuming or expensive. Supplementary Figure S1 is an excerpt of a table (Sutherland *et al.*, 2011) that provides the number of animals included in the analysis.

Effect measures regarding outcomes were often poorly reported. Supplementary Table S3 provides examples of the information missing from some studies. Such information would be needed to assess the magnitude of effect so that the balance of benefits and harms could be evaluated (which cannot be evaluated by p-values). If only the p-value is reported, it is not possible to know the magnitude or direction of the effect (i.e., whether the intervention increased or decreased the outcome). Furthermore, measures of variation were often not reported or not reported clearly, especially in figures where it was not always possible to discern if the error bar represented a SEM, a SD, or a confidence interval. In studies that used random effects variables to control for clustering, the variance components were never reported, despite their importance for future study design and interpretation.

Ancillary analyses (REFLECT checklist Item 17, Outcomes and Estimation) were not reported in any study, as no *a priori* primary and secondary outcomes were reported and no sample size justifications were provided. The rationale for reporting ancillary analyses is to give end-users knowledge of potentially interesting results that arise from data exploration and are therefore hypothesis generating rather than hypothesis testing.

Proactive reporting of adverse events was expected in order for end-users to balance the benefits and harms of using pain-mitigating interventions in neonatal piglets. Harms are often rare and therefore often only detected using secondary analyses. Such information would have allowed us to understand whether the reported mortality rate was excessive compared to baseline trends in production. Sometimes adverse events were reported in a way that we could not determine the group to which the animals that experienced the intervention were allocated. Knowledge of the group to which the animal was assigned is vital to interpreting harms. For example, reporting that 10 animals died in the study is not informative, compared to reporting five animals died in each group or one animal died in the control group and nine in the treated group.

Discussion

In the area of animal welfare research we found that, as in other disciplines related to veterinary and animal sciences, reporting of intervention studies was frequently incomplete (Burns and O'Connor, 2008, Sargeant *et al.*, 2011). Overall many studies failed to report information that would be needed to assess internal and external validity. There are both ethical and, in some countries, legal reasons for ensuring that scientists using animals must not only adhere to adequately justified methodology but that they

should also be able to articulate it according to high reporting standards to their peers and the public. The privilege given to scientists to use research animals entails adherence to rigorous reporting standards that help to ensure compliance with national and international policies that protect the welfare of all research animals.

Some of the journals we have assessed might not be considered as truly “scientific" as these (mostly national/local) journals are periodical magazines intended to inform practitioners on new developments. As an example, the journal "Der Praktische Tierarzt" has a different audience than Journal of Animal Science. However, if such journals do choose to publish primary research then it seems that the standards of reporting would still apply. Another reason for omissions may be lack of awareness of the need for comprehensive reporting due to the multidisciplinary nature of many projects.

Publication of the results of a scientific study is not the end of the scientific process (Sargeant and O'Connor, 2013). Presumably researchers publish with the intent that the results of a study will enable generation of new hypotheses, validate current hypotheses, or influence decision-making. These secondary uses of primary data rely on the validity of the original study design, analyses and such assessment of validity can only be determined if the report is transparent, accurate and comprehensive (O'Connor *et al.*, 2010, Sargeant and O'Connor, 2013). Further, if incomplete reporting casts doubt over the results of studies then the monetary and ethical value of the animal and financial resources used to generate the data were not fully realized. Also animals may have suffered unnecessarily. If the study is considered important enough that the

information is needed for decision-making, it may even be necessary to repeat the study (Ioannidis *et al.*, 2014, Macleod *et al.*, 2014). In situations where the reporting is so incomplete that useful data cannot be extracted from the original experiment and the study must be repeated, this would be incongruous with the 3Rs (replacement, reduction and refinement) and not be a good use of already scarce research funding. Professionally, the credibility of the authors of the original study could be called into question.

It is unclear why reporting is incomplete. Some might suggest that this is because of lack of awareness of reporting guidelines. However, the concepts of reproducible research and reporting are manner that reflects the experiment is not new or novel, so lack of awareness cannot explain all of the incomplete reporting. (Grindlay *et al.*, 2014)

. It is imperative that research reporting be complete to enable reproducibility, assessment of the internal and external validity of the study and knowledge translation. Given that animal welfare science is an discipline that often involves interventions that may be perceived as unpleasant to the animal, attention to the quality of reporting is especially critical to advance the field. Comprehensive reporting is an ethical responsibility for researchers undertaking this type of research. For intervention studies, the reader should be able to understand the magnitude and precision of the estimated effect of the intervention, and the probability that the effect is consistent with the null hypothesis. The reader should also be able to assess the potential for bias.

We would encourage authors to consider using reporting guidelines to improve reporting. Consistent with the 3Rs, in particular the reduction principle, using reporting

guidelines can maximise information from the animals used in the study and minimise the risk of unnecessary studies, therefore reducing further animal use. We are aware that the omission of this information as well as important design characteristics, analyses, or results is often unintentional. Also, we are well aware what constitutes a complete report is not a static list. As knowledge and technology change, the standards for how science is conducted and reported should be expected to change. Given these changing standards, however, the most recent checklists represent minimum current standards. This would not preclude authors from including or editors and peer-reviewers from requesting additional information. Checklists provide guidance for reporting but researchers should adhere to the underlying reporting principles to provide a report that facilitates reuse of the data and enables assessment of bias. With the growing frequency of multiple collaborators involved in manuscript preparation, the final editor may not be aware of all the aspects required for reporting. One reason for an incomplete report might be a lack of knowledge of what and how items should be reported. However, resources are becoming increasingly available to mitigate this problem. Documents specific to animal studies like the REFLECT statement for livestock trials, the ARRIVE guidelines specific to biomedical uses of animals and a European Food Safety Authority document specific to euthanasia are included in this list of available resources (Kilkenny *et al.*, 2010, O'Connor *et al.*, 2010, Sargeant *et al.*, 2010, EFSA, 2013). The ARRIVE guidelines are designed for animals used in experimental settings with a focus on animal populations where the independence assumption is valid. The REFLECT statement is more specific for livestock and provides more focus on non-independent populations such as occur in housed animals. As

reporting guidelines are relatively new, the impact on reporting has not bee assessed yet. For example, REFLECT was unavailable when many of the papers in this review were published. The standards of reporting observed here are therefore not reflective of the impact of reporting guidelines. The examples provided in ***Table 3***, **Table 4** and **Supplementary Table S3** can be used as a guideline for how some of the studies we reviewed effectively reported the information requested by the checklist. All three tables should be used together, and are broken into sections here for presentation purposes. Use of a reporting checklist might help reduce the number of items not reported.

Conclusion

The overall conclusion after assessing these studies using REFLECT, is that there is (1) an opportunity to improve the reporting and (2) a need to raise awareness of the importance in providing a complete report of how animal welfare studies are conducted. The continued ethical and legal acceptability of using animals is contingent upon accurate and complete reporting. Accurate and complete reporting, in most cases, relates to both high quality research and responsible conduct in animal research.

Acknowledgement

This project was funded by the National Pork Board (NPB) Grant no. 12-186. The authors have no conflicts of interest to declare.

References

2009. In Recognition and Alleviation of Pain in Laboratory Animals, Washington (DC).

Al-Shahi Salman R, Beller E, Kagan J, Hemminki E, Phillips RS, Savulescu J, Macleod M, Wisely J and Chalmers I 2014. Increasing value and reducing waste in biomedical research regulation and management. Lancet 383, 176-185.

Anttila H, Malmivaara A, Kunz R, Autti-Ramo I and Makela M 2006. Quality of reporting of randomized, controlled trials in cerebral palsy. Pediatrics 117, 2222-2230.

Beirendonck Sv, Driessen B, Verbeke G and Geers R 2011. Behavior of piglets after castration with or without carbon dioxide anesthesia. Journal of Animal Science 89, 3310-3317.

Burns MJ and O'Connor AM 2008. Assessment of methodological quality and sources of variation in the magnitude of vaccine efficacy: A systematic review of studies from 1960 to 2005 reporting immunization with Moraxella bovis vaccines in young cattle. Vaccine 26, 144-152.

Campbell MK, Piaggio G, Elbourne DR, Altman DG and Group C 2012. Consort 2010 statement: extension to cluster randomised trials. BMJ 345, e5661.

Chan AW, Song F, Vickers A, Jefferson T, Dickersin K, Gotzsche PC, Krumholz HM, Ghersi D and van der Worp HB 2014. Increasing value and reducing waste: addressing inaccessible research. Lancet 383, 257-266.

Chapman PL and Seidel GE, Jr. 2008. Experimental design, power and sample size for animal reproduction experiments. Reproduction, Fertility, and Development 20, 33-44.

Coetzee JF, Edwards LN, Mosher RA, Bello NM, O'Connor AM, Wang B, Kukanich B and Blasi DA 2012. Effect of oral meloxicam on health and performance of beef steers relative to bulls castrated on arrival at the feedlot. Journal of Animal Science 90, 1026-1039.

Cohen J 1997. Statistical power analysis for the behavioral science. Academic Press, New York, NY, USA.

Dewell RD, Millman ST, Gould SA, Tofflemire KL, Whitley RD, Parsons RL, Rowe EW, Liu F, Wang C and O'Connor AM 2014. Evaluating approaches to measuring ocular pain in bovine

calves with corneal scarification and infectious bovine keratoconjunctivitis-associated corneal ulcerations. Journal of Animal Science 92, 1161-1172.

Dzikamunhenga RS, Anthony R, Coetzee J, Gould S, Johnson A, Karriker L, McKean J, Millman ST, Niekamp SR and O'Connor AM 2014. Pain management in the neonatal piglet during routine management procedures. Part 1: A systematic review of randomized and non-randomized intervention studies. Animal Health Research Reviews 15, 14-38.

EFSA 2013. Guidance on the assessment criteria for studies evaluating the effectiveness of stunning interventions regarding animal protection at the time of killing. EFSA Journal 11, 41.

Funk L, O'Connor AM, Maroney M, Engelken T, Cooper VL, Kinyon J and Plummer P 2009. A randomized and blinded field trial to assess the efficacy of an autogenous vaccine to prevent naturally occurring infectious bovine keratoconjunctivis (IBK) in beef calves. Vaccine 27, 4585-4590.

Glynn HD, Coetzee JF, Edwards-Callaway LN, Dockweiler JC, Allen KA, Lubbers B, Jones M, Fraccaro E, Bergamasco LL and KuKanich B 2013. The pharmacokinetics and effects of meloxicam, gabapentin, and flunixin in postweaning dairy calves following dehorning with local anesthesia. Journal of Veterinary Pharmacology and Therapeutics 36, 550-561.

Gould S, Dewell R, Tofflemire K, Whitley RD, Millman ST, Opriessnig T, Rosenbusch R, Trujillo J and O'Connor AM 2013. Randomized blinded challenge study to assess association between Moraxella bovoculi and Infectious Bovine Keratoconjunctivitis in dairy calves. Veterinary Microbiology 164, 108-115.

Grindlay DJ, Dean RS, Christopher MM and Brennan ML 2014. A survey of the awareness, knowledge, policies and views of veterinary journal Editors-in-Chief on reporting guidelines for publication of research. BMC Veterinary Research 10, 10.

Hansson M, Lundeheim N, Nyman G and Johansson G 2011. Effect of local anaesthesia and/or analgesia on pain responses induced by piglet castration. Acta Veterinaria Scandinavica 53, 34.

527 Hofmeister EH, King J, Read MR and Budsberg SC 2007. Sample size and statistical power in
528 the small-animal analgesia literature. Journal of Small Animal Practice 48, 76-79.
529 Ioannidis JP, Greenland S, Hlatky MA, Khoury MJ, Macleod MR, Moher D, Schulz KF and
530 Tibshirani R 2014. Increasing value and reducing waste in research design, conduct, and
531 analysis. Lancet 383, 166-175.
532 Kilkenny C, Browne WJ, Cuthill IC, Emerson M and Altman DG 2010. Improving bioscience
533 research reporting: the ARRIVE guidelines for reporting animal research. PLoS Biology 8,
534 e1000412.
535 Kreuder AJ, Coetzee JF, Wulf LW, Schleining JA, KuKanich B, Layman LL and Plummer PJ
536 2012. Bioavailability and pharmacokinetics of oral meloxicam in llamas. BMC Veterinary
537 Research 8, 85.
538 Lahrmann KH, Kmiec M and Stecher R 2006. Piglet castration with ketamine/azaperone-
539 anaesthesia: concurring with animal welfare, practical, but economic? Die Saugferkelkastration
540 mit der Ketamin/Azaperon-Allgemein-anästhesie: tierschutzkonform, praktikabel,
541 aberwirtschaftlich? Praktische Tierarzt 87, 802-809.
542 Lang TA and Altman DG 2014. Statistical Analyses and Methods in the Published Literature:
543 The SAMPL Guidelines*. In Guidelines for Reporting Health Research: A User's Manual, pp.
544 264-274, John Wiley & Sons, Ltd.
545 Macleod MR, Michie S, Roberts I, Dirnagl U, Chalmers I, Ioannidis JP, Al-Shahi Salman R,
546 Chan AW and Glasziou P 2014. Biomedical research: increasing value, reducing waste. Lancet
547 383, 101-104.
548 Marx G, Horn T, Thielebein J, Knubel B and von Borell E 2003. Analysis of pain-related
549 vocalization in young pigs. Journal of Sound and Vibration 266, 687-698.
550 Moher D, Hopewell S, Schulz KF, Montori V, Gotzsche PC, Devereaux PJ, Elbourne D, Egger
551 M, Altman DG and Consolidated Standards of Reporting Trials G 2010. CONSORT 2010

552 Explanation and Elaboration: Updated guidelines for reporting parallel group randomised trials.
553 Journal of Clinical Epidemiology 63, 1-37.
554 O'Connor A, Anthony R, Bergamasco L, Coetzee J, Gould S, Johnson AK, Karriker LA,
555 Marchant-Forde JN, Martineau GS, McKean J, Millman ST, Niekamp S, Pajor EA, Rutherford K,
556 Sprague M, Sutherland M, von Borell E and Dzikamunhenga RS 2014. Pain management in the
557 neonatal piglet during routine management procedures. Part 2:Grading the quality of evidence
558 and the strength of recommendations. Animal Health Research Reviews 15, 39-62.
559 O'Connor AM, Sargeant JM, Gardner IA, Dickson JS, Torrence ME, Consensus Meeting P,
560 Dewey CE, Dohoo IR, Evans RB, Gray JT, Greiner M, Keefe G, Lefebvre SL, Morley PS,
561 Ramirez A, Sischo W, Smith DR, Snedeker K, Sofos J, Ward MP and Wills R 2010. The
562 REFLECT statement: methods and processes of creating reporting guidelines for randomized
563 controlled trials for livestock and food safety by modifying the CONSORT statement. Zoonoses
564 Public Health 57, 95-104.
565 Reiner G, Schollasch F, Hillen S, Willems H, Piechotta M and Failing K 2012. Effects of
566 meloxicam and flunixin on pain, stress and discomfort in male piglets during and after surgical
567 castration. Berliner und Munchener Tierarztliche Wochenschrift 125, 305-314.
568 Sargeant JM and O'Connor AM 2013. Issues of reporting in observational studies in veterinary
569 medicine. Preventive Veterinary Medicine 113, 323-330.
570 Sargeant JM, O'Connor AM, Gardner IA, Dickson JS, Torrence ME and Consensus Meeting P
571 2010. The REFLECT statement: reporting guidelines for Randomized Controlled Trials in
572 livestock and food safety: explanation and elaboration. Zoonoses Public Health 57, 105-136.
573 Sargeant JM, Saint-Onge J, Valcour J, Thompson A, Elgie R, Snedeker K and Marcynuk P
574 2009. Quality of reporting in clinical trials of preharvest food safety interventions and
575 associations with treatment effect. Foodborne Pathogens and Disease 6, 989-999.

576 Sargeant JM, O'Connor AM, Renter DG, Kelton DF, Snedeker K, Wisener LV, Leonard EK,
577 Guthrie AD and Faires M 2011. Reporting of methodological features in observational studies of
578 pre-harvest food safety. Preventive Veterinary Medicine 98, 88-98.
579 Schulz KF, Altman DG, Moher D and Group C 2010. CONSORT 2010 Statement: updated
580 guidelines for reporting parallel group randomized trials. Open Medicine 4, e60-68.
581 Sutherland MA, Davis BL and McGlone JJ 2011. The effect of local or general anesthesia on
582 the physiology and behavior of tail docked pigs. Animal 5, 1237-1246.
583 Sutherland MA, Davis BL, Brooks TA and Coetzee JF 2012. The physiological and behavioral
584 response of pigs castrated with and without anesthesia or analgesia. Journal of Animal Science
585 90, 2211-2221.
586 Tenbergen R 2012. Investigation of the use of analgesics at the time of castration and tail-
587 docking and following parturition for improving performance and reducing pain in pigs.
588 University of Guelph.
589 Tuyttens FAM, de Graaf S, Heerkens JLT, Jacobs L, Nalon E, Ott S, Stadig L, Van Laer E and
590 Ampe B 2014. Observer bias in animal behaviour research: can we believe what we score, if we
591 score what we believe? Animal Behaviour 90, 273-280.
592
593

***Table 1**:* Checklist for REFLECT statement and the frequency of reporting of REFLECT checklist items

Location in paper	Item number	REFLECT checklist descriptor	Reported	Partially reported	Not reported
Title and abstract	1	How study units were allocated to interventions (i.e. "random allocation," "randomized," or "randomly assigned" or “weight matched”)	5	0	47
Methods Participants	3	Eligibility criteria for owner/managers and study units at each level of the organizational structure, and the settings and locations where the data were collected.	0	10	42
Interventions	4a	Precise details of the interventions intended for each group, the level at which the intervention was allocated, and how and when interventions were actually administered.	21	31	0
Objectives	5	Specific objectives and hypotheses.	7	8	37
Outcomes	6	Clearly defined primary and secondary outcome measures and the levels at which they were measured and, when applicable, any methods used to enhance the quality of measurements (e.g., multiple observations, training of assessors).	6	5	41
Sample Size	7	How sample size was determined and, when applicable, explanation of any interim analyses and stopping rules.	0	0	52
Randomization—Sequence generation	8	Method used to generate the random allocation sequence at the relevant level of the organizational structure, including details of any restrictions (e.g.,	0	0	52

		blocking, stratification)			
Randomization—Allocation concealment	9	Method used to implement the random allocation sequence at the relevant level of the organizational structure (e.g., numbered containers or central telephone), clarifying whether the sequence was concealed until interventions were assigned.	0	0	52
Randomization—Implementation	10	Who generated the allocation sequence, who enrolled study units, and who assigned study units to their groups at the relevant level of the organizational structure?	0	0	52
Blinding (masking)	11	Whether or not participants administering the interventions, caregivers, and those assessing the outcomes were blinded to group assignment. If done, how the success of blinding was evaluated. Provide justification for not using blinding if it was not used.	0	18	34
Statistical methods	12	Statistical methods used to compare groups for all outcome(s); clearly state the level of statistical analysis and methods used to account for the organizational structure, where applicable; methods for additional analyses, such as subgroup analyses and adjusted analyses.	0	44	8
Results Recruitment	14	Dates defining the periods of recruitment and follow-up.	6	0	46
Baseline data	15	Baseline demographic and clinical characteristics of each group	0	5	47
Numbers analysed	16	Number of study units (denominator) in each group	1	7	44

		included in each analysis and whether the analysis was by "intention-to-treat." State the results in absolute numbers when feasible (e.g., 10/20, not 50%).			
Outcomes and estimation	17	For each primary and secondary outcome, a summary of results for each group, accounting for each relevant level of the organizational structure, and the estimated effect size and its precision (e.g., 95% confidence interval).	0	11	41
Ancillary analyses	18	Address multiplicity by reporting any other analyses performed, including subgroup analyses and adjusted analyses, indicating those pre-specified and those exploratory.	0	0	52
Adverse events	19	All important adverse events or side effects in each intervention group.	0	22	30

The following were not assessed, Introduction (REFLECT Item 2), Study flow (REFLECT Item 13) Discussion (REFLECT Item 20), Generalizability (REFLECT Item 21) and Overall evidence (REFLECT Item 22).

597 ***Table 2:*** Reporting means and measures of precision, and arm sample size in studies evaluated for complete reporting.

Outcome assessed[1]	Number of relevant study arms	Arms for which data was extracted from figures	Arms with missing summary features	Description of missing summary measures
Intervention: General Anesthesia (CO_2/O_2)				
Cortisol 0-60 minutes	8	4	3	2 means, 3 SDs
Cortisol 1-24 hours	6	2	3	2 means, 3 SDs
β-endorphins 0-60 minutes	9	2	2	2 means, 2 SDs
β-endorphins 1-24 hours	3	1	2	2 means, 2 SDs, 2 arm sample size
Norepinephrine 0-60 minutes	2	1	1	Arm sample size
Pain-like behaviors 0-60 minutes	8	4	2	1 mean and 2 SDs
Intervention: Local Anesthesia (Lidocaine)				
Cortisol 0-60 minutes	8	7	7	6 SDs and 1 arm sample size
Cortisol 1-24 hours	6	6	6	6 SDs and 3 arm sample size
Norepinephrine 0-60 minutes	1	0	1	1 mean, 1 SD and 1 arm sample size
Frequency 0-60 minutes[2]	4	0	3	3 SDs and 1 arm sample size
Energy 0-60 minutes[2]	4	2	2	1 SD and 2 arm sample size
Rate 0-60 minutes[2]	8	0	7	7 SDs and 3 arm sample size
Pain-like behaviors 0-60 minutes	3	0	2	1 mean, 2 SDs and 2 arm sample size
Pain-like behaviors 1-24 hours	1	0	1	mean, SD, arm sample size
Intervention: NSAID (Carprofen, flunixin, ketoprofen, meloxicam)				
Cortisol 0-60 minutes	15	10	2	1 mean, 2 SDs and 1 arm sample size

Cortisol 1-24 hours	10	4	3	2 means, 3 SDs
Energy 0-60 minutes	5	1	3	1 mean, 1 SD and 3 totals
Pain-like behaviors 0-60 minutes	2	0	2	1 SD and 1 arm sample size
Pain-like behaviors 1-24 hours	5	0	2	1 SD and 1 arm sample size

598 SD=standard deviation

599 [1] For more details of exact outcomes measured refer to Dzikamunhenga *et al.* (2014)

600

601 Examples for the following not included: Introduction (REFLECT Item 2), Study flow (REFLECT Item 13) Discussion

602 (REFLECT Item 20), Generalizability (REFLECT Item 21) and Overall evidence (REFLECT Item 22).

603 ***Table 3*:** Examples of reported "Methods" items from the trials reported consistent with REFLECT guidelines

Paper section and topic	Item	Example from review studies
Participants	3	"Sows were housed in commercial farrowing crates on a commercial farm in Saxony-Anhalt, Germany" (Marx *et al.*, 2003).
Interventions	4	"Two groups were treated with Flunixin (5 mg); the group termed Flu-30 received an i.m. injection of Flunixin 30 min before castration and of 0.1 ml NaCl (0.9%) immediately before castration, the group termed Flu-0 received 0.1 ml NaCl (0.9%) 30 min before castration and Flunixin immediately before castration" (Reiner *et al.*, 2012).
Objectives	5	"The objective of this study was to evaluate the effect of providing CO_2 anesthesia before castration on the behavior of piglets for up to 8 d after castration in comparison with piglets castrated without anesthesia.... The hypothesis of the above study is that piglets will experience less pain and discomfort after castration when anesthetized with CO_2 before castration, thus improving their overall welfare"(Beirendonck *et al.*, 2011).
Outcomes	6	Defining outcomes- "The primary outcome was Infectious Bovine Keratoconjunctivitis (IBK) cumulative incidence over the study period. The secondary outcome was weaning weight" (Funk *et al.*, 2009)*. Methods of measurement and level of assessment"... and the behavior of each individual pig was recorded using 1 min scan samples (direct observations) for 120 min" (Sutherland *et al.*, 2011).
Sample size	7	"Prior to conducting the study, it was determined that twelve animals per group were required to obtain 80% power to detect a 60% difference in IBK risk between groups based on an expected 10% IBK risk in controls and at least 70% IBK risk in inoculated animals. The test was based on a one sided difference in proportions test for independent binomial data with significance level 0.05. Thus, our aim was to enrol 36 animals. No stopping rules or interim analyses were planned or conducted" (Gould *et al.*, 2013).*
Randomization -- Sequence generation	8	"Forty steers were randomly assigned to one of five treatment groups as described in Table 1. Calves were ranked in ascending order of bodyweight, blocked into cohorts of five calves, and within each cohort, calves were assigned a random number (Excel, Microsoft Works 2010; Microsoft, Redmond, WA, USA). Random numbers were then assigned (Excel, Microsoft Works 2010, Microsoft) to treatment groups to ensure that bodyweight was equally distributed between treatment groups"(Glynn *et al.*, 2013).*
Randomization -- Allocation concealment	9	"The individual who generated the allocation sequence was not involved in assessment of eligibility or the outcome" (Gould *et al.*, 2013).*

Paper section and topic	Item	Example from review studies
Randomization -- Implementation	10	"The allocation status, based on eye and calf, on day 0 was confirmed by two people prior to the allocation. The allocation status of the eye was concealed from the individual responsible for scarification and inoculation process" (Gould *et al.*, 2013).*
Blinding (masking)	11	"Two technicians, who were not blind to the treatments due to practical reasons, performed all measurements. The measurements were split between the two technicians with each technician performing the same measurements in all herds" (Hansson *et al.*, 2011).
Statistical methods	12	"Least square mean estimates for each treatment group and the corresponding estimated SE are reported. Pairwise comparisons were conducted using Bonferroni's method to adjust for multiple comparisons and avoid inflation of Type I error rate. Statistical significance for these multiple comparisons was designated a priori as a P-value ≤ 0.05" (Coetzee *et al.*, 2012)*. "For physiological measures, the main fixed effects were treatment and time. Litter was a random effect. The interactions between treatment by time and treatment by litter were included in the model" (Sutherland *et al.*, 2012).

*Example not selected from the study set

606

607 ***Table 4:*** Examples of reported "Results" items from the trials reported consistent with REFLECT guidelines

Paper section and topic	Item	Example from review studies
Study flow	13	"Nine calves (9/19) in Trial 2 had missing data on d +10 because practical constraints prevented collection of PA-MNT data around scheduled ophthalmic exams and euthanasia. One calf in Trial 1 developed severe respiratory disease and was euthanized on d +7." (Dewell *et al.*, 2014)*
Recruitment	14	"The studies were conducted in two piglet breeding operations (Unit A 550 breeding sows, two-week production cycle; Unit B 560 sows, four-week production cycle) from February 2003 to May 2003" (Lahrmann *et al.*, 2006).
Baseline data	15	
Numbers analyzed	16	See Supplementary Figure 1
Outcomes and estimation	17	See Supplementary Figure 1
Ancillary analyses	18	"There was not a significant difference between treatment groups with respect to mortality rate. Piglets receiving meloxicam had a mortality rate of 3.18% and piglets receiving the placebo had a mortality rate of 3.84% (P=0.33). Piglets receiving ketoprofen had a mortality rate of 2.91% whereas piglets receiving the placebo had a mortality rate of 3.94% (P=0.27)" (Tenbergen, 2012).
Adverse events	19	"No adverse effects were noted after IV or oral meloxicam administration" (Kreuder *et al.*, 2012)*.

608 *Example not selected from the study set

609