

Scotland's Rural College

## 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](https://doi.org/10.1017/S1751731115002323)

Print publication: 01/04/2016

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

[Link to publication](#)

### Citation for published version (APA):

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. (2016). Assessment of completeness of reporting in intervention studies using livestock: an example from pain mitigation interventions in neonatal piglets. *Animal*, 10(4), 660 - 670.  
<https://doi.org/10.1017/S1751731115002323>

### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

### Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

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

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

8 <sup>1</sup> *Department of Veterinary Diagnostic and Production Animal Medicine, College of*  
9 *Veterinary Medicine, Iowa State University, Ames, Iowa, USA*

10 <sup>2</sup> *Department of Philosophy, University of Alaska Anchorage, Anchorage, Alaska, USA*

11 <sup>3</sup> *Department of Animal and Poultry Sciences, College of Agriculture and Life Science,*  
12 *Virginia Tech, Blacksburg, VA 24060*

13 <sup>4</sup> *Department of Animal Science, College of Agriculture and Life Sciences, Iowa State*  
14 *University, Ames, IA, USA*

15 <sup>5</sup> *USDA-ARS, Livestock Behavior Research Unit, West Lafayette, IN, USA*

16 <sup>6</sup> *Swine clinic, Department of Animal Production, National Veterinary School, Toulouse,*  
17 *France*

18 <sup>7</sup> *Department of Biomedical Science, College of Veterinary Medicine, Iowa State*  
19 *University, Ames, IA, USA*

20 <sup>8</sup> *Scotland'd Royal College (SRUC), Edinburgh, United Kingdom*

21 <sup>9</sup> *AgResearch Ltd, Ruakura Research Centre, Hamilton, New Zealand*

22 <sup>10</sup> *American Association of Swine Veterinarians (AASV), Perry, Iowa, USA,*

23 <sup>11</sup> *Department of Animal Husbandry and Ecology, Institute of Agricultural and Nutritional*

24 *Sciences, Martin-Luther-University Halle-Wittenberg, Germany*

25 <sup>12</sup> *National Pork Board, Des Moines, Iowa, USA.*

26 <sup>13</sup> *Department of Production Animal Health, Faculty of Veterinary Medicine, University of*

27 *Calgary, Canada*

28

29 Corresponding author: Annette O'Connor, Email: [aconnor@iastate.edu](mailto:aconnor@iastate.edu), Telephone: +1

30 515 294 5012, Fax: +1 515 294 1072.

31 Short title: *Complete reporting of welfare studies*

32

33 **Abstract**

34 Accurate and complete reporting of study methods, results, and interpretation are  
35 essential components for any scientific process, allowing end-users to evaluate the  
36 internal and external validity of a study. When animals are used in research, excellence  
37 in reporting is expected as a matter of continued ethical acceptability of animal use in  
38 the sciences. Our primary objective was to assess completeness of reporting for a  
39 series of studies relevant to mitigation of pain in neonatal piglets undergoing routine  
40 management procedures. Our second objective was to illustrate how authors can report  
41 the items in the REFLECT statement using examples from the animal welfare science  
42 literature. Fifty-two studies from 40 articles were evaluated using a modified REFLECT  
43 statement. No single study reported all REFLECT checklist items. Seven studies  
44 reported specific objectives with testable hypotheses. Six studies identified primary or  
45 secondary outcomes. Randomization and blinding were considered to be partially  
46 reported in 21 and 18 studies respectively. No studies reported the rationale for sample  
47 sizes. Several studies failed to report key design features such as units for  
48 measurement, means, standard deviations, standard errors for continuous outcomes or  
49 comparative characteristics for categorical outcomes expressed as either rates or  
50 proportions. In the discipline of animal welfare science, authors, reviewers, and editors  
51 are encouraged to use available reporting guidelines to ensure that scientific methods  
52 and results are adequately described and free of misrepresentations and inaccuracies.  
53 Complete and accurate reporting increases the ability to apply the results of studies to  
54 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

## 63 **Introduction**

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

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

85 guidelines exist for identifying what a complete account of an intervention assessment  
86 study represents (Kilkenny *et al.*, 2010, Moher *et al.*, 2010, O'Connor *et al.*, 2010,  
87 Sargeant *et al.*, 2010, Schulz *et al.*, 2010, Campbell *et al.*, 2012). In areas where the  
88 reporting of intervention assessments have been evaluated, reporting has frequently  
89 been identified as incomplete (Anttila *et al.*, 2006, Sargeant *et al.*, 2009, Schulz *et al.*,  
90 2010).

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

## 100 **Methods and Materials**

### 101 *Study population:*

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

106 relevant studies, and resulting review are available elsewhere (Dzikamunhenga *et al.*,  
107 2014, O'Connor *et al.*, 2014). For the assessment of comprehensive reporting, we used  
108 the studies relevant to the original review. The unit of concern for reporting was a  
109 study/trial. Two or more studies/trial were occasionally reported in a single article. An  
110 intervention study/trial must have at least 2 arms (treatment groups).

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

113 The REFLECT statement is a reporting guideline for randomized controlled trials that  
114 assess interventions for food-producing animals such a swine and is therefore suitable  
115 for this topic area (<http://www.REFLECT-statement.org>) (O'Connor *et al.*, 2010,  
116 Sargeant *et al.*, 2010). The REFLECT statement comprises 22 checklist items (**Table**  
117 **1**), of which we assessed the reporting of 17. The rationale for including these items in a  
118 publication is provided by Sargeant *et al.* (2010). The reporting of five REFLECT  
119 checklist items was not assessed (**Table 1**). We did not assess study flow (REFLECT  
120 checklist item 13) because we expected that studies relevant to the interventions were  
121 of such short duration that it was unlikely any loss to follow-up would occur i.e., few  
122 piglets would leave the study because the outcome could be assessed. We did not  
123 assess REFLECT checklist items 2 (Introduction and Background), 20 (Discussion and  
124 Interpretation), 21 (Generalizability) and 22 (Overall Evidence) because they are more  
125 prone to subjective assessment.

126

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

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

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

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

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

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

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

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

191 **Results**

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

205

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

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

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

#### 224 *Reporting of REFLECT items that relate to objectives and hypotheses*

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

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

236 comparison of the mean Hertz of vocalizations in piglets receiving the anesthetic  
237 intervention compared to the mean Hertz of vocalizations in piglets without the  
238 anesthetic, i.e.,  $H_0 = \text{mean}_1 - \text{mean}_2 = 0$ . Clarification of the hypothesis ensures the  
239 end user knows which metric is being used to assess the objective, and should facilitate  
240 identification of the primary outcome.

241 *Reporting of REFLECT items that relate to outcomes and sample size*

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

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

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

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

265 REFLECT checklist items 8 through 10 (Sequence Generation, Allocation Concealment,  
266 and Implementation, respectively) are based on the assumption that the study is  
267 randomized. A description of the method of developing the randomization for the  
268 sequence generation, allocation concealment, and implementation, was not provided in  
269 any study. Thirty-three of 52 studies used the term “randomly” or “randomized” or  
270 “random” in their description of piglet allocation to treatment group. Occasionally it was  
271 unclear if the approach used was truly random, despite a description as such. For  
272 example, one study described randomly assigning 245 clinically healthy piglets to one of  
273 12 experimental groups. However, the sample sizes in each of the seven relevant arms  
274 were very different, suggesting a method other than random allocation. Several studies  
275 reported using restrictions of randomization. Blocking by continuous covariates or  
276 stratification by categorical covariates was reported in 39 studies. Covariates used were  
277 weight, litter, weight and litter, sow or weight, or litter and adoption. No study that  
278 controlled for weight using blocking explicitly reported the size of the block. Details  
279 about the approach to allocation are part of reporting that enables assessment of

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

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

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

306 Of the 52 studies, 18 reported blinding as part of their protocol; however, none provided  
307 a full description of the approach used to blind the study (REFLECT checklist item 11).

308 Blinding, whether for allocation of treatments or interventions or assessment, was  
309 infrequently reported by authors. As blinding is designed to reduce  
310 measurement/information bias, it is important to know if outcome assessment is biased.

311 There is some evidence in veterinary science and animal welfare that absence of  
312 blinding is associated with more positive outcomes (Burns and O'Connor, 2008,  
313 Tuytens *et al.*, 2014).

314 *Reporting of REFLECT items related to Statistical Methods*

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

323 *Reporting of REFLECT items related to setting, study population characteristics*

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

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

342 *Reporting of REFLECT items related to results of analyses*

343

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

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

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

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

## 382 **Discussion**

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

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

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

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

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

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

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

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

433 guidelines can maximise information from the animals used in the study and minimise  
434 the risk of unnecessary studies, therefore reducing further animal use. We are aware  
435 that the omission of this information as well as important design characteristics,  
436 analyses, or results is often unintentional. Also, we are well aware what constitutes a  
437 complete report is not a static list. As knowledge and technology change, the standards  
438 for how science is conducted and reported should be expected to change. Given these  
439 changing standards, however, the most recent checklists represent minimum current  
440 standards. This would not preclude authors from including or editors and peer-reviewers  
441 from requesting additional information. Checklists provide guidance for reporting but  
442 researchers should adhere to the underlying reporting principles to provide a report that  
443 facilitates reuse of the data and enables assessment of bias. With the growing  
444 frequency of multiple collaborators involved in manuscript preparation, the final editor  
445 may not be aware of all the aspects required for reporting. One reason for an  
446 incomplete report might be a lack of knowledge of what and how items should be  
447 reported. However, resources are becoming increasingly available to mitigate this  
448 problem. Documents specific to animal studies like the REFLECT statement for  
449 livestock trials, the ARRIVE guidelines specific to biomedical uses of animals and a  
450 European Food Safety Authority document specific to euthanasia are included in this list  
451 of available resources (Kilkenny *et al.*, 2010, O'Connor *et al.*, 2010, Sargeant *et al.*,  
452 2010, EFSA, 2013). The ARRIVE guidelines are designed for animals used in  
453 experimental settings with a focus on animal populations where the independence  
454 assumption is valid. The REFLECT statement is more specific for livestock and provides  
455 more focus on non-independent populations such as occur in housed animals. As

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

#### 464 Conclusion

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

471

#### 472 Acknowledgement

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

475  
476

#### References

477 2009. In Recognition and Alleviation of Pain in Laboratory Animals, Washington (DC).

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

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

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

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

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

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

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

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

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

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

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

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

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

510 Funk L, O'Connor AM, Maroney M, Engelken T, Cooper VL, Kinyon J and Plummer P 2009. A  
511 randomized and blinded field trial to assess the efficacy of an autogenous vaccine to prevent  
512 naturally occurring infectious bovine keratoconjunctivitis (IBK) in beef calves. *Vaccine* 27, 4585-  
513 4590.

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

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

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

525 Hansson M, Lundeheim N, Nyman G and Johansson G 2011. Effect of local anaesthesia and/or  
526 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 Lahrman 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, Sisco 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 Tuytens 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

594 **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

595 The following were not assessed, Introduction (REFLECT Item 2), Study flow (REFLECT Item 13) Discussion (REFLECT  
 596 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 <sup>1</sup>	Number of relevant study arms	Arms for which data was extracted from figures	Arms with missing summary features	Description of missing summary measures
<b>Intervention: General Anesthesia (CO<sub>2</sub>/O<sub>2</sub>)</b>				
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
<b>Intervention: Local Anesthesia (Lidocaine)</b>				
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 <sup>2</sup>	4	0	3	3 SDs and 1 arm sample size
Energy 0-60 minutes <sup>2</sup>	4	2	2	1 SD and 2 arm sample size
Rate 0-60 minutes <sup>2</sup>	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
<b>Intervention: NSAID (Carprofen, flunixin, ketoprofen, meloxicam)</b>				
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 <sup>1</sup> 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 <i>et al.</i> , 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 <i>et al.</i> , 2012).
Objectives	5	“The objective of this study was to evaluate the effect of providing CO <sub>2</sub> 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 <sub>2</sub> before castration, thus improving their overall welfare”(Beirendonck <i>et al.</i> , 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 <i>et al.</i> , 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 <i>et al.</i> , 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 <i>et al.</i> , 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 <i>et al.</i> , 2013).*
Randomization -- Allocation concealment	9	"The individual who generated the allocation sequence was not involved in assessment of eligibility or the outcome” (Gould <i>et al.</i> , 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 <i>et al.</i> , 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 <i>et al.</i> , 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 $\leq 0.05$ ” (Coetzee <i>et al.</i> , 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 <i>et al.</i> , 2012).

604 \*Example not selected from the study set

605

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 <i>et al.</i> , 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” (Lahrman <i>et al.</i> , 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 <i>et al.</i> , 2012)*.

608 \*Example not selected from the study set

609