Veterinary Record Open

OPEN ACCESS

Incomplete reporting of experimental studies and items associated with risk of bias in veterinary research

Maxime Rufiange,[1] Frédérik Rousseau-Blass,[1] Daniel S J Pang[1,2]

To cite: Rufiange M, Rousseau-Blass F, Pang DSJ. Incomplete reporting of experimental studies and items associated with risk of bias in veterinary research. *Veterinary Record Open* 2019;**6**:e000322. doi:10.1136/vetreco-2018-000322

Received 8 October 2018
Revised 25 January 2019
Accepted 6 February 2019

[1]Clinical Sciences, Université de Montréal, Saint-Hyacinthe, Quebec, Canada
[2]Veterinary Clinical & Diagnostic Sciences, University of Calgary, Calgary, Alberta, Canada

Correspondence to
Dr Daniel S J Pang;
danielpang17@hotmail.com

ABSTRACT

In in vivo research, the reporting of core items of study design is persistently poor, limiting assessment of study quality and study reproducibility. This observational cohort study evaluated reporting levels in the veterinary literature across a range of species, journals and research fields. Four items (randomisation, sample size estimation, blinding and data exclusion) were assessed as well as availability of study data in publicly accessible repositories. From five general and five subject-specific journals, 120 consecutively published papers (12 per journal) describing in vivo experimental studies were selected. Item reporting was scored using a published scale (items ranked as fully, partially or not reported) according to completeness of reporting. Papers in subject-specific journals had higher median reporting levels (50.0 per cent vs 33.3 per cent, P=0.007). In subject-specific journals, randomisation (75.0 per cent vs 41.7 per cent, P=0.0002) and sample size estimation (35.0 per cent vs 16.7 per cent, P=0.025) reporting was approximately double that of general journals. Blinding (general 48.3 per cent, subject-specific 50.0 per cent, P=0.86) and data exclusion (general 53.3 per cent, subject-specific 63.3 per cent, P=0.27) were similarly reported. A single paper made study data readily accessible. Incomplete reporting remains prevalent in the veterinary literature irrespective of journal type, research subject or species. This impedes evaluation of study quality and reproducibility, raising concerns regarding wasted financial and animal resources.

INTRODUCTION

A key component of high-quality studies is complete and transparent reporting.[1] Limited reporting impedes interpretation of studies and experimental reproducibility.[2–4] Disturbingly, the limited reporting of items associated with a risk of bias in study design has been associated with inflated effect sizes, reflecting an association between reporting and study quality. This has contributed to failures of translational research, unnecessary animal use and financial waste.[2–8]

Bias in research is broadly defined as a systematic error in results or inferences.[9] Careful study design attempts to minimise the introduction of factors leading to bias and transparent reporting allows evaluation of such factors.[8] Common sources of bias are failure to randomise, lack of blinding and undisclosed or unexplained exclusion of data from analysis.

Proper randomisation provides internal validity and limits selection bias, while blinding prevents detection and performance biases with investigators/care givers potentially influencing observations.[10 11] Data handling (decisions surrounding data inclusions and exclusions) shapes the analysis, results and conclusions of a study, which supports external validity; the generalisability of findings to other populations.[11] Therefore, the rationale for including or excluding subjects or data should be explicitly described. Sample size estimation is a critical component of study design. Smaller studies are less precise due to greater sampling variation and a sample size that is too small to identify an important treatment effect is more likely to lead to a false negative result.[9 11 12]

Despite the acceptance of these items as indicators of study quality, their reporting is poor in both laboratory animal and veterinary studies.[7 13–17] The introduction and widespread endorsement of reporting guidelines has yielded limited improvements in reporting quality and the risk of bias remains high.[13 14 16–20] It has been proposed that focusing on a universal set of core reporting standards could increase adoption by users and facilitate study evaluation.[1] Such an approach, in conjunction with an editorial policy of enforced adherence to reporting standards, may improve reporting standards.[21]

Furthermore, focusing on the core items of randomisation, blinding, data exclusions and sample size estimation facilitates comparisons between studies employing different species across research domains.

The primary objectives of this study were to: (1) examine a current cross-section of the veterinary literature regardless of species or field of research and with a focus on key items reflective of the potential for bias (randomisation, blinding, data exclusions) and

completeness of reporting (sample size estimation) and (2) to compare these reporting levels between general and subject-specific journals. Secondary objectives were to evaluate the accessibility of study data and explore the relationship between journal impact factor and item reporting. An observational cohort study was designed to test the hypotheses that items associated with completeness of reporting and risk of bias would be poorly reported overall (<50.0 per cent)[22] and that there would be no significant difference between papers published in general and subject-specific journals.

MATERIALS AND METHODS

Literature search methods

An a priori sample size estimate was calculated using commercial software (Sergeant, ESG, 2018, Epitools epidemiological calculators, Ausvet, available at: http://epitools.ausvet.com.au). The calculation was performed for the comparison between journal types, based on detecting a difference between two proportions (difference of 25 percentage points between journal types). Alpha level was set at 0.05 with 80 per cent power using a two-tailed test. Sample size was set at 60 papers per journal type to allow for a whole number of papers to be selected from the predefined list of journals.

One hundred and twenty papers were selected from 10 veterinary journals: 5 general and 5 subject-specific journals. Journals of interest were selected from the Veterinary Sciences category of the Journal Citation Reports (2017 Journal Citation Reports (Clarivate Analytics, 2017), accessed October 3, 2017), with journals selected semi-objectively, taking into account impact factor and citation counts (preference for journals with higher values of each) and publication of clinical trials. The five general journals were: *Equine Veterinary Journal*, *The Veterinary Journal*, *Preventive Veterinary Medicine*, *BMC Veterinary Research* and *Veterinary Record*. The five subject-specific journals were: *Journal of Veterinary Internal Medicine*, *Veterinary Surgery*, *Veterinary Anaesthesia and Analgesia*, *Veterinary Dermatology* and *Journal of Veterinary Emergency and Critical Care*.

From each journal, 12 papers were selected, beginning the search with the most recently available (ie, chronological rather than randomised), including those published online as early access (accessed October 5, 2017). MR screened the titles and the abstracts according to predetermined inclusion/exclusion criteria. Full texts were retrieved if there was uncertainty about fulfilment of these criteria. The search continued until 12 qualifying papers were identified for each journal.

Inclusion and exclusion criteria

Papers were included if they were in English and described an in vitro experimental design (parallel or cross-over) with a comparison group (from either client-owned or research animals with natural or induced diseases). Reviews, descriptive/observational studies, case reports or series and in vivo experiments were excluded from analysis. No restrictions were applied to the field of research. As this study was based on published literature, ethical approval was not sought.

Paper evaluation

Papers were evaluated using a published operationalised checklist, applied to assess the items of interest (randomisation, blinding, sample size estimation and data exclusions).[23] The published checklist was applied, with minor adaptations to reflect the application to clinical or experimental animals (the original checklist was designed for biomedical (laboratory animal) in vivo and in vitro research, table 1). Randomisation, blinding and sample size estimation items were categorised as either fully, partially or not reported. The data exclusion item was evaluated as three subitems (table 1). Papers were not assessed for methodological quality, that is, assessment was limited to evaluating completeness of reporting. For example, blinding would be classified as fully reported if there was a statement that blinding was not possible. In addition to the four core items, the availability of study data was evaluated ('data deposition' item, table 1).

A cohort of 15 of the selected papers was initially evaluated independently by two raters (MR and FRB) using the operationalised checklist. Their evaluations were compared in a group meeting with a third investigator (DP) who was blinded to paper authors and journal, and differences resolved by consensus. All remaining papers (full text, including any supplemental material) were then assessed by both raters independently. Raters were not blinded to paper authors or journal. Following review, any differences were resolved by consensus discussion with the third investigator (DP).

Statistical analysis

Data were tested for normality (D'Agostino-Pearson test) and appropriate parametric or non-parametric analyses applied. To create a picture of overall reporting, reporting of the four items was considered together (for each paper, a proportion was calculated from the number of items reported out of all possible items) and described (median percentage) for all papers combined. In cases where Overall reporting levels between journal types (general vs subject-specific) were compared with a Mann-Whitney U test (data not normally distributed). Individual items and subitems were compared between journal types with a z-test. Due to low reporting prevalence, a chi-squared test was used to assess 'data deposition' and the subitem 'pre-establishing exclusion criteria'.

Reporting of individual journals was limited to descriptive statistics as the planned sample size (12 papers/journal) was insufficient to make statistical comparisons. The relationship between journal impact factor and item reporting was evaluated with a Pearson's correlation coefficient. Statistical software was used for analyses (GraphPad Prism V.6.00 for Windows, GraphPad Software, San Diego, California, USA and SAS V.9.3,

Table 1 Checklist used to evaluate completeness of reporting for randomisation, sample size estimation, blinding and data exclusions and data availability

Item title	Classification	Descriptor
1. Randomisation	Fully reported	► If a method of randomisation for allocating samples or animals to experimental groups is described. ► If there is a statement describing that randomisation was not possible.
	Partially reported	► If randomisation is not described for each experiment performed.* ► If randomisation is mentioned but the method used is not described.
	Not reported	► If there is no statement of randomisation.
2. Sample size estimation	Fully reported	► If sample size is justified based on having adequate power to detect a predetermined difference for the identified primary outcome(s) of interest, including at least three out of four elements (alpha, beta, variability, difference of interest) required to calculate sample size.
	Partially reported	► If sample size estimate is mentioned without an explicit statement of power or difference to be detected. ► If sample size is not estimated for each primary outcome previously identified, or if primary outcomes have not been specifically identified. ► If fewer than three elements required to calculate sample size are provided.
	Not reported	► No mention of sample size estimation in the paper or deviations from the fully or partially reported descriptors.
3. Blinding	Fully reported	► If blinding is reported for group allocation and/or when assessing the outcome(s) for each experiment. ► If there is a statement that blinding was not possible.
	Partially reported	► If a general, non-specific statement of blinding is made, eg. 'this was a blinded study', without specifying blinding was to group allocation or outcome assessment, or both. ► Reported blinding is incomplete: does not include all allocations/outcomes.
	Not reported	► No statement on blinding in the manuscript.
4a. Exclusion of samples or animals from the analysis	Fully reported	► If there is a statement that a sample or animal was excluded from analysis. ► If there is a statement that no data were excluded from analysis.
	Partially reported	► If the number of animals (or samples) reported in the results matches the description in the methods, but there is no explicit statement regarding data exclusion.
	Not reported	► If the number of animals from which data were collected is described in the results and this differs from the number enrolled/included. ► If the number of animals from which data were collected is described in the results without reporting numbers enrolled in the methods.
4b. Defining exclusion criteria	Fully reported	► If there is a description of why, or in which situation(s), data would be excluded.
	Partially reported	► Not applicable to this item.
	Not reported	► If an exclusion is described without explanation. ► If the total number of animals from which data could be collected changes from the methods to the results without explanation. ► If there is no explanation of why, or in which situation(s), data were excluded.
4c. Pre-establishing exclusion criteria	Not applicable	► If the response to 4b is not reported.
	Fully reported	► If an explicit statement of exclusion criteria being pre-established is made.
	Partially reported	► Not applicable to this item.
	Not reported	► If there is no explicit statement that exclusion criteria were pre-established.

Continued

Table 1 Continued

Item title	Classification	Descriptor
5. Data deposition	Fully reported	► If data are freely available (no need to contact author(s) for access), for example, data repository.
	Partially reported	► If data are described as being available by contacting the author(s).
	Not reported	► If there is no mention of data availability.

Checklist adapted from Cramond and others used to score 120 papers from a cross-section of experimental studies in the veterinary literature.[23] Papers were selected from general (n=5 journals, 12 papers sampled per journal) and subject-specific (n=5 journals, 12 papers sampled per journal) veterinary journals and completeness of reporting for these items assessed.
*Applies if there are multiple experiments/trials included in the same manuscript.

SAS Institute, Cary, North Carolina, USA). Values of P<0.05 were considered to be statistically significant and 95 per cent CIs are presented for differences between journal types. The data supporting the results are available in the Harvard Dataverse: https://doi.org/10.7910/DVN/O1XFGR

RESULTS

Overall reporting levels were low (all papers combined, n=120 papers), with a median of 50.0 per cent of items fully reported and 18.4 per cent of items partially reported (figure 1A and B). This reflected the low levels of fully-reported individual items and subitems, which ranged from 4.9 per cent to 68.3 per cent. Levels of partial reporting were also low, less than 40.0 per cent for all items and subitems (figure 1).

Fourteen papers did not fully report any of the four items, but all papers partially reported at least one item. A single paper fully reported all items. The highest and lowest median values of fully-reported items were in the *Journal of Veterinary Internal Medicine* (75.0 per cent) and *BMC Veterinary Research* (26.7 per cent), respectively (figure 2).

For all items combined, full reporting occurred more often in papers published in subject-specific journals (50.0 per cent) than those published in general journals (33.3 per cent, P=0.007).

Consequently, partial reporting levels were greater for general journals (20.0 per cent; subject-specific: 16.7 per cent, P=0.048).

Items

In comparing the reporting of individual items between subject-specific and general journals, randomisation and sample size estimation were fully reported approximately twice as often in subject-specific journals. In contrast, blinding was fully reported to a similar degree for both journal types (table 2). Partial reporting levels were similar between journal types for sample size estimation and blinding. Randomisation was reported approximately twice as often in general journals (table 2).

Reporting standards were broadly similar in both journal types for data exclusion subitems, with full reporting in approximately half to three-quarters of papers for the subitems 'exclusion of samples or animals from the analysis' and 'defining exclusion criteria' (table 2). The subitem 'pre-establishing exclusion criteria' was reported in fewer than five papers (table 2).

Data deposition was low, with 0.83 per cent (95 per cent CI 0.02 per cent to 4.6 per cent) of all papers meeting the criteria for full reporting and 9.2 per cent (95 per cent CI 4.7 per cent to 15.8 per cent) meeting the criteria for partial reporting. Examining journal types revealed

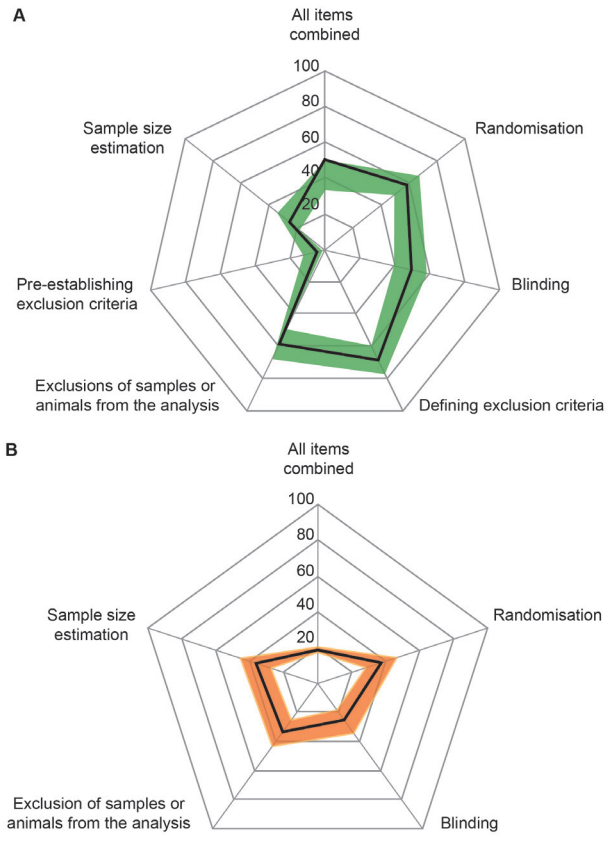

Figure 1 Radar plots illustrating proportions of fully (A) and partially reported (B) items reflective of a risk of bias and reporting completeness, and data deposition. Data are from 120 veterinary clinical trials (2016–2017) published in general (n=5) and subject-specific (n=5) journals (12 papers per journal). Data are median percentages for all bias items combined and proportions (%) for individual items, with 95% CIs indicated by the shaded regions.

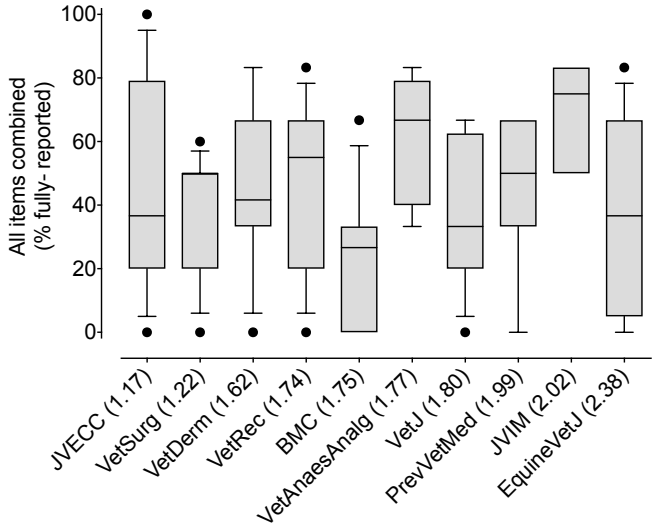

Figure 2 Proportion of fully reported items associated with a risk of bias and completeness of reporting (randomisation, blinding, sample size estimation, data exclusions) in papers published in selected veterinary journals. One-hundred and twenty veterinary clinical trials (2016–2017) published in general (n=5) and subject-specific (n=5) journals (12 papers per journal) were assessed. Box and whisker plots: box limits are interquartile ranges, horizontal lines within boxes the median and whiskers the 10th–90th percentile. Solid circles are outliers. JVECC, *Journal of Veterinary Emergency and Critical Care*; VetSurg, *Veterinary Surgery* (the median value is 50.0%); VetDerm, *Veterinary Dermatology*; VetRec, *Veterinary Record*; BMC, *BMC Veterinary Research*; VetAnaesthAnalg, *Veterinary Anaesthesia and Analgesia*; VetJ, *Veterinary Journal*; PrevVetMed, *Preventive Veterinary Medicine*; JVIM, *Journal of Veterinary Internal Medicine*; EquineVetJ, *Equine Veterinary Journal.* Journal impact factors from 2016 are in parenthesis.

that no papers in subject-specific journal (0/60) and one paper in a general journal (1/60) fully-reported data deposition (P=0.99, 95 per cent CI −1.6 per cent to 4.9 per cent). Partial reporting occurred more frequently in general journals (11/60 papers, subject-specific: 0/60 papers, P=0.0006, 95 per cent CI 8.5 per cent to 28.1 per cent).

No significant correlation was identified between journal impact factor and the percentage of fully-reported items (r=0.057, r^2=0.003, P=0.54, figure 2).

DISCUSSION

This study showed that the frequency of full reporting of items that reflect a risk of bias and reporting completeness borders between poor and moderate, using a proposed threshold of 50.0 per cent.[22] Unexpectedly, statistically significant differences in reporting were identified between general and subject-specific journals. Considering the importance of complete reporting, these observed differences should not be overinterpreted in the face of suboptimal reporting levels.

Complete reporting of randomisation, including a description of allocation method, was similar to the level reported in a recent study focusing on reporting of randomisation in veterinary clinical trials, in which approximately half of trials identifying themselves as randomised did not report the method of randomisation.[10] These findings suggest an improvement from earlier studies of veterinary clinical trials (published between 1989–1990[14] and 2006–2008[16 17]) in dogs, cats and livestock, in which only 12 per cent–20 per cent of trials in which randomisation was applied reported the allocation method used. Of concern, where purported

Table 2 Proportions of full and partial reporting of items reflective of a risk of bias and reporting completeness

Item	Subject-specific full reporting (n/60)	General full reporting (n/60)	P value (95% CI of the difference)	Subject-specific partial reporting (n/60)	General partial reporting (n/60	P value (95% CI of the difference)
1. Randomisation	75.0% (45)	41.7% (25)	<0.01 (15.0% to 51.0%)	20.0% (12)	53.3% (32)	<0.01 (16.0% to 50.0%)
2. Sample size estimation	35.0% (21)	16.7% (10)	0.03 (2.3% to 33.7%)	38.3% (23)	33.3% (20)	0.57 (–12.0% to 22.0%)
3. Blinding	50.0% (30)	48.3% (29)	0.86 (–16.2% to 19.6%)	26.7% (16)	23.3% (14)	0.67 (–12.2% to 18.8%)
4a. Exclusion of samples or animals from the analysis	63.3% (38)	53.3% (32)	0.27 (–7.7% to 28%)	28.3 (17)	38.3 (23)	0.24 (–6.8% to 27.0%)
4b. Defining exclusion criteria	75.0% (45)	61.7% (37)	0.13 (–3.6% to 30.0%)	N/A	N/A	N/A
4c. Pre-establishing exclusion criteria	6.7% (3/45)*	2.7% (1/37)*	0.62 (–5.4% to 13.0%)	N/A	N/A	N/A

Full and partial reporting for individual items associated with risk of bias and completeness of reporting. Data are from 120 papers of experimental studies in the veterinary literature. Papers were selected from general (n=5 journals, 12 papers sampled per journal) and subject-specific (n=5 journals, 12 papers sampled per journal) veterinary journals and assessed according to a checklist adapted from Cramond and others.[23] Data are proportions (%) with 95% CI of the difference between journal types. Reported P values are for comparisons between journal types.
*Proportion calculated from number of papers describing exclusion criteria (item 4b).
N/A, not applicable (these items could only be classified as fully or not reported).

randomisation methods are described, 13 per cent of trials (8/62) used methods that are non-random.[10] This is similar to the proportion of non-random randomisation methods reported 10 years earlier, suggesting that many veterinary researchers remain unfamiliar with core concepts of study design, emphasising the importance of explicitly stating randomisation methods.[24] The rates observed here were approximately six times higher than those observed in in vivo biomedical studies, in which the same assessment scale was applied.[21]

Full reporting of sample size estimation has improved compared with the low rates (0 per cent-5 per cent) observed in reports from the veterinary literature published during the previous three decades[14 16 17]; however, the results presented here, similar to those of Giuffrida, highlight that much remains to be *done*.[25] Where study results are negative, the absence of any discussion of sample size estimation prevents interpretation of the findings, greatly limiting the value of such studies alongside the potential waste of resources (financial and animal).[12 25]

The reporting of blinding in this study approximated the upper end of the range reported previously for the veterinary small animal and livestock clinical trial literature (25.0 per cent–60.0 per cent).[14 16 17] Again, this reflects limited improvement despite the repeated demonstration of poor reporting quality.

For data exclusions, the rates observed here were in line with those previously reported, although reporting rates vary, perhaps reflecting differences in study methodology and population sampled.[14 16 17] Reporting of blinding and data exclusions from analysis were approximately double those reported by Macleod.[21] In comparison with the study by Macleod, the higher rates of reporting observed here for the four items suggests a systematic difference in reporting behaviour between veterinary clinical trials and in vivo biomedical research.[21] The reasons for this are unclear but could reflect a pressure to focus on novel findings in biomedical research, with an emphasis on the substantial volume of data often presented to the detriment of space devoted to reporting methods, or a reluctance to provide detailed supplementary materials.

The obvious consequences of incomplete reporting are to limit reproducibility in research and impede critical evaluation of published work. Additionally, and of particular concern, is evidence that incomplete reporting of items with a risk of bias is associated with inflated effect sizes.[3 5 8 26–28] That is, the failure to report an item associated with a risk of bias can be an indication of a deficit in study design and conduct. Evidence for inflated effect sizes is limited in the veterinary literature, although an association between non-reporting of items associated with a risk of bias and an increase in positive results has been reported.[16 17 29] This raises important questions regarding the ethical use of animals in research and fiscal responsibility.

Numerous reporting guidelines have been developed to address reporting deficits and they have received widespread support from biomedical and veterinary journals.[30]

For example, the ARRIVE (Animals in Research: Reporting In Vivo Experiments) and CONSORT (Consolidated Standards of Reporting Trials) guidelines apply to many veterinary studies and the REFLECT (Reporting Guidelines for Randomized Controlled Trials for Livestock and Food Safety) and STROBE-Vet (Strengthening the Reporting of Observational Studies in Epidemiology-Veterinary) guidelines are specific to veterinary medicine.[30–33] Unfortunately, despite the number of guidelines available adherence to reporting guidelines is low, indicating that journal support or endorsement, without some mechanism to enforce adherence, is insufficient.[13 15 34–37] To the authors' knowledge, introduction of a mandatory reporting checklist is the only approach that has been shown to improve reporting quality.[20 38]

Data accessibility reflects transparency in research, supports verification of results and analysis, facilitates systematic reviews and meta-analyses and is increasingly requested by biomedical journals as a condition of publication. There may be instances when data access should be limited (risk of revealing personal or security-related information or data with commercial value), but these limitations seldom apply to publicly funded research.[39] Based on author guidelines (accessed April 23, 2018), four of the journals studied (*BMC Veterinary Research*, *Veterinary Journal*, *Preventive Veterinary Medicine* and *Veterinary Record*) encouraged data access through the use of repositories, although it was not mandatory. Furthermore, while data repositories can be easily found online, veterinary journals could do more to suggest repositories that meet their data policy requirements, including being recognised and trusted by the scientific community. Approximately 18.0 per cent of papers (all published in *BMC Veterinary Research*) included a statement that data were available on contacting the author; however, author compliance to requests for data access may be low.[40]

Journal impact factor is calculated from the ratio between citations received (to articles published in the preceding two years) to the total number of articles published over the same period. It is said to reflect the mean number of citations received by a paper published in that journal, but this misrepresents the skewed citation distribution observed in journals, leading to the common misconception that journal impact factor reflects the quality of individual papers, a case of judging a book's contents by its cover.[41–48] The small sample of papers from each journal limits interpretation, although there was no discernible correlation between reporting quality and journal impact factor and this is consistent with the findings of larger studies.[41 44] Interestingly, the only paper that fully reported all items was published in the journal with the lowest journal impact factor.

Limitations

The four items evaluated in this study represent a minimal requirement, suggested as universally applicable

to experimental studies and allowing for a rapid assessment of the risk of bias and error. Focusing on these items should not be viewed as detracting from the use of more complete guidelines. The raters were not blinded to author name(s) and institution(s) when evaluating papers but it is unlikely that knowledge of the identity of the authors and institutions had an impact on evaluations as papers were reviewed independently using the checklist and both raters are trainees in the early stages of their careers, with limited knowledge of the authors and institutions represented.

In studies where treatment effects are markedly different, efforts to blind observers could be limited. Maintaining blinding during data analysis could offset resulting bias, particularly if the person performing the analysis was not involved in data collection. This was not assessed in this study as the authors adhered to the published checklist used.

Two papers of the 120 selected (1.7 per cent) were made available online as uncorrected proofs following acceptance. It is possible that further changes, influencing the assessment of reporting, were made to these papers before final publication, although it is highly unlikely that they would affect interpretation of the findings. The small sample of papers from each journal and the narrow impact factor range across veterinary medicine may have limited identifying a link between journal impact factor and reporting quality; however, it does not appear that journal impact factor has an important influence on reporting quality.[39]

CONCLUSION

The quality of reporting across veterinary medicine remains low, with limited improvements in reporting standards over the last three decades. Within the context of this well-established problem and considering the ready availability of reporting guidelines, the observed differences in reporting between general and subject-specific journals are inconsequential. These findings are concerning as they reveal a considerable lack of transparency in study reporting, the consequences of which are to limit evaluation of published work and attempts to reproduce results. These results should not be interpreted as a comment on the quality of the studies evaluated; however, a potential link between poor reporting and inflated effect sizes deserves further study considering the implications (financial and ethical) of animal research.

Acknowledgements The authors would like to thank Dr Guy Beauchamp for statistical support and Vivian Leung for graphical design (Faculty of Veterinary Medicine, Université de Montréal).

Funding Discovery Grant from the Natural Sciences and Engineering Research Council of Canada (ID: 424022-2013, awarded to DSJP). MR receives a stipend from the Fondation J-Louis Lévesque.

Disclaimer The funders had no role in study design, data collection and analysis or decision to publish.

Competing interests None declared.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement Data are freely available via link at end of methods section.

REFERENCES

1. Landis SC, Amara SG, Asadullah K, *et al*. A call for transparent reporting to optimize the predictive value of preclinical research. *Nature* 2012;490:187–91.
2. Crossley NA, Sena E, Goehler J, *et al*. Empirical evidence of bias in the design of experimental stroke studies: a metaepidemiologic approach. *Stroke* 2008;39:929–34.
3. Macleod MR, van der Worp HB, Sena ES, *et al*. Evidence for the efficacy of NXY-059 in experimental focal cerebral ischaemia is confounded by study quality. *Stroke* 2008;39:2824–9.
4. Page MJ, Higgins JP, Clayton G, *et al*. Empirical Evidence of Study Design Biases in Randomized Trials: Systematic Review of Meta-Epidemiological Studies. *PLoS One* 2016;11:e0159267.
5. Chalmers TC, Celano P, Sacks HS, *et al*. Bias in treatment assignment in controlled clinical trials. *N Engl J Med* 1983;309:1358–61.
6. Freedman LP, Cockburn IM, Simcoe TS. The Economics of Reproducibility in Preclinical Research. *PLoS Biol* 2015;13:e1002165.
7. Kilkenny C, Parsons N, Kadyszewski E, *et al*. Survey of the quality of experimental design, statistical analysis and reporting of research using animals. *PLoS One* 2009;4:e7824.
8. Schulz KF, Chalmers I, Hayes RJ, *et al*. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995;273:408–12.
9. Higgins JPT, Altman DG, Sterne JAC. Assessing risk of bias in included studies. In: Higgins JPT, Green S, eds. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011], 2011. The Cochrane Collaboration. www.handbook.cochrane.org. (accessed 17 May 2018).
10. Di Girolamo N, Giuffrida MA, Winter AL, *et al*. In veterinary trials reporting and communication regarding randomisation procedures is suboptimal. *Vet Rec* 2017;181:195.
11. Lund EM, James KM, Neaton JD. Clinical trial design: veterinary perspectives. *J Vet Intern Med* 1994;8:317–22.
12. Hofmeister EH, King J, Read MR, *et al*. Sample size and statistical power in the small-animal analgesia literature. *J Small Anim Pract* 2007;48:76–9.
13. Baker D, Lidster K, Sottomayor A, *et al*. Two years later: journals are not yet enforcing the ARRIVE guidelines on reporting standards for pre-clinical animal studies. *PLoS Biol* 2014;12:e1001756.
14. Lund EM, James KM, Neaton JD. Veterinary randomized clinical trial reporting: a review of the small animal literature. *J Vet Intern Med* 1998;12:57–60.
15. Leung V, Rousseau-Blass F, Beauchamp G, *et al*. ARRIVE has not ARRIVEd: Support for the ARRIVE (Animal Research: Reporting of in vivo Experiments) guidelines does not improve the reporting quality of papers in animal welfare, analgesia or anesthesia. *PLoS One* 2018;13:e0197882.
16. Sargeant JM, Elgie R, Valcour J, *et al*. Methodological quality and completeness of reporting in clinical trials conducted in livestock species. *Prev Vet Med* 2009;91:107–15.
17. Sargeant JM, Thompson A, Valcour J, *et al*. Quality of reporting of clinical trials of dogs and cats and associations with treatment effects. *J Vet Intern Med* 2010;24:44–50.
18. Muir WW, Ueyama Y, Noel-Morgan J, *et al*. A Systematic Review of the Quality of IV Fluid Therapy in Veterinary Medicine. *Front Vet Sci* 2017;4:127.
19. Totton SC, Cullen JN, Sargeant JM, *et al*. The reporting characteristics of bovine respiratory disease clinical intervention trials published prior to and following publication of the REFLECT statement. *Prev Vet Med* 2018;150:117–25.
20. Grindlay DJC, Dean RS, Christopher MM, *et al*. A survey of awareness, knowledge, policies and views of veterinary journal Editors-in-Chief on reporting guidelines for publication of research. *BMC Vet Res* 2014:10:10.
21. Macleod MR. The NPQIP Collaborative Group. Findings of a retrospective controlled cohort study of the impact of a change in Nature journals' editorial policy for life sciences research on the

completeness of reporting study design and execution. *bioRxiv* 2017.
22. Delgado-Ruiz RA, Calvo-Guirado JL, Romanos GE. Critical size defects for bone regeneration experiments in rabbit calvariae: systematic review and quality evaluation using ARRIVE guidelines. *Clin Oral Implants Res* 2015;26:915–30.
23. Cramond F, Irvine C, Liao J, *et al*. Protocol for a retrospective, controlled cohort study of the impact of a change in Nature journals' editorial policy for life sciences research on the completeness of reporting study design and execution. *Scientometrics* 2016;108:315–28.
24. Brown DC. Control of selection bias in parallel-group controlled clinical trials in dogs and cats: 97 trials (2000-2005). *J Am Vet Med Assoc* 2006;229:990–3.
25. Giuffrida MA. Type II error and statistical power in reports of small animal clinical trials. *J Am Vet Med Assoc* 2014;244:1075–80.
26. Savović J, Jones HE, Altman DG, *et al*. Influence of reported study design characteristics on intervention effect estimates from randomized, controlled trials. *Ann Intern Med* 2012;157:429–38.
27. Vesterinen HM, Sena ES, ffrench-Constant C, *et al*. Improving the translational hit of experimental treatments in multiple sclerosis. *Mult Scler* 2010;16:1044–55.
28. Burns MJ, O'Connor AM. 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* 2008;26:144–52.
29. EQUATOR. Network: Enhancing the QUAlity and Transparency Of health Research. 2008 http://www.equator-network.org (accessed 17 May 2018).
30. Kilkenny C, Browne WJ, Cuthill IC, *et al*. Improving bioscience research reporting: the ARRIVE guidelines for reporting animal research. *PLoS Biol* 2010;8:e1000412.
31. Schulz KF, Altman DG, Moher D, *et al*. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *PLoS Med* 2010;7:e1000251.
32. O'Connor AM, Sargeant JM, Gardner IA, *et al*. The REFLECT statement: methods and processes of creating reporting guidelines for randomized controlled trials for livestock and food safety. *J Vet Intern Med* 2010;24:57–64.
33. Sargeant JM, O'Connor AM, Dohoo IR, *et al*. Methods and Processes of Developing the Strengthening the Reporting of Observational Studies in Epidemiology-Veterinary (STROBE-Vet) Statement. *J Food Prot* 2016;79:2211–9.
34. Avey MT, Moher D, Sullivan KJ, *et al*. The Devil Is in the Details: Incomplete Reporting in Preclinical Animal Research. *PLoS One* 2016;11:e0166733.
35. Collins FS, Tabak LA. Policy: NIH plans to enhance reproducibility. *Nature* 2014;505:612–3.
36. Liu Y, Zhao X, Mai Y, *et al*. Adherence to ARRIVE Guidelines in Chinese Journal Reports on Neoplasms in Animals. *PLoS One* 2016;11:e0154657.
37. Turner L, Shamseer L, Altman DG, *et al*. Does use of the CONSORT Statement impact the completeness of reporting of randomised controlled trials published in medical journals? A Cochrane review. *Syst Rev* 2012;1:60.
38. Han S, Olonisakin TF, Pribis JP, *et al*. A checklist is associated with increased quality of reporting preclinical biomedical research: A systematic review. *PLoS One* 2017;12:e0183591.
39. Pilat D, Fukasaku Y. OECD Principles and Guidelines for Access to Research Data from Public Funding. *Data Sci J* 2007;6:OD4–OD11.
40. Wicherts JM, Bakker M, Molenaar D. Willingness to share research data is related to the strength of the evidence and the quality of reporting of statistical results. *PLoS One* 2011;6:e26828.
41. Garfield E. The history and meaning of the journal impact factor. *JAMA* 2006;295:90–3.
42. Macleod MR, Lawson McLean A, Kyriakopoulou A, *et al*. Risk of Bias in Reports of In Vivo Research: A Focus for Improvement. *PLoS Biol* 2015;13:e1002273.
43. Colquhoun D. Challenging the tyranny of impact factors. *Nature* 2003;423:479. discussion 480.
44. Munafò MR, Stothart G, Flint J. Bias in genetic association studies and impact factor. *Mol Psychiatry* 2009;14:119–20.
45. Seglen PO. Why the impact factor of journals should not be used for evaluating research. *BMJ* 1997;314:497–502.
46. Pang DSJ. Misconceptions surrounding the relationship between journal impact factor and citation distribution in veterinary medicine. *Vet Anaesth Analg* 2018.
47. Christopher MM. Weighing the impact (factor) of publishing in veterinary journals. *J Vet Cardiol* 2015;17:77–82.
48. Anon. San Francisco Declaration on Research Assessment. 2012 https://sfdora.org (Accessed 22 January 2019).