XVIII Congreso de Especialidades Veterinarias
26-27 de Abril de 2019 - Palacio de Congresos - ZARAGOZA

¿DEBERÍA DE CREER TODO LO QUE LEO EN LOS ARTÍCULOS CIENTÍFICOS?

(Evaluating a scientific paper - should I believe what I read?)

Daniel SJ Pang
Université de Montréal
1500 Avenue des Vétérinaires, Saint-Hyacinthe, Québec, Canada J2S 7C6

The aim of this session is to provide attendees with some general criteria that can be used to evaluate the research literature through critical reading of papers. These criteria are based on assessments of reporting completeness and measures of validity that are found in high quality studies. Additionally, common sources of confusion regarding the interpretation of statistics and evidence will be presented. To fully understand and appreciate the results of a study, the study methods must be clearly presented and containing key information. The importance of different aspects of the methods will vary to some extent between different subjects of research, but there are core items that should be reported. Existing deficits in study reporting in both clinical and experimental animal research are significant (Kilkenny et al. 2009, Lund et al. 1998, Di Girolamo et al. 2017). Such deficits have been associated with failures in study reproducibility, translational research and artificial inflation of treatment effects (Chalmers et al. 1983, Freedman et al. 2015, Percie du Sert and Rice 2014, Macleod et al. 2008).

Kilkenny et al. (2009), in a study of 271 experimental animal research papers, found that fewer than half the papers reported age and weight and one quarter did not report sex. Furthermore a little over 10% of papers reported using randmisation and none of the papers explained the sample size used. Following these findings, the same group developed a set of reporting guidelines, the ARRIVE (Animals in Research: Reporting In Vivo Experiments) Guidelines. Unfortunately, the impact of the ARRIVE guidelines appears to have been limited (Baker et al. 2014, Leung et al. 2018). Recent work by Leung et al. (2018) comparing the quality of reporting before and five years after publication of the ARRIVE guidelines found that reporting standards remained low (Fig. 1), particularly for those items associated with bias in study design (randomisation, blinding, data handling). The lowest reporting level was associated with sample size estimation, which was reported in 14% of papers. These findings indicate that recommending and drawing attention to various reporting guidelines is an ineffective means of improving reporting and some degree of enforcement may be required (Macleod et al. 2017). Furthermore, a criticism of reporting guidelines, such as ARRIVE, is their complexity and specificity to different study types. Nevertheless, complete reporting is necessary to fully evaluate what was done and how the findings may apply in different environments (generalisability). Simple questions that are useful to consider are: were scientific instruments calibrated and sufficiently accurate and precise? Were appropriate control groups included? Does the study population breed, sex and age translate to my practice and are they relevant to the disease process being studied?

An alternative approach, which may be useful to make a rapid assessment of study quality, is to focus on a few key items. These are randomisation, blinding, data handling (unexplained or unreported data inclusions/exclusions) and sample size estimation (Landis et al. 2012). Randomisation, blinding and data handling are items associated with biased study design. Randomisation and blinding reflect internal validity, where bias can affect how treatments are assigned or observations influenced. Data handling has a direct impact on the results and eventual interpretation of a study's findings, a reflection of external validity (generalisability to other populations). Appropriate sample size is critical to the power of a study to detect a statistically significant difference. An insufficient sample size may result in a false negative finding. Large sample sizes are more likely to result in a statistically significant result, but it must then be decided if this represents a relevant or important difference. The pressure to focus on achieving significance leads to over-interpretation of study data. Historically, reporting of these core items in veterinary clinical trials has been poor, with low reporting rates for randomisation and sample size estimation (Lund et al. 1998). Furthermore, core concepts, such as randomisation, may not be fully understood by authors (Di Girolamo et al. 2017). Even when studies identify as being randomised, this term may be used inappropriately, reflecting ongoing misconceptions around the term and how it should be applied. Di Girolamo et al. (2017), in a study of 114 randomised veterinary trials, found that 7% of studies used a non-random method to allocate

treatment and close to half the trials did not explain the mechanism of randomisation.

A recent study of 120 papers from 5 general and 5 subject-specific veterinary journals reporting animal research found that reporting rates were better in subject-specific than general journals, but that reporting rates were highly variable according to individual items (Rufiange et al. 2019). Randomisation was fully reported by 42-75% of papers, with blinding and data exclusion criteria fully reported in approximately 50% of papers and sample size estimation fully reported in just 17-35% of papers. Only one paper fully reported all four items and there was no apparent relationship between reporting level and impact factor.

There is now convincing evidence in the biomedical literature that failure to report these key items (randomisation, blinding, data handling and sample size estimation) is associated with inflated treatment effects (Macleod et al. 2008, Savović et al. 2012, Vesterinen et al. 2010). On the basis of these inflated effect sizes, putative therapeutic drugs have progressed to human clinical trials before failing, with a substantial associated animal and financial cost. While such evidence in veterinary clinical trials is limited, there is no reason to suspect that the situation will be different (Burns et al. 2008). This has the potential to directly and negatively impact clinical practice and future research. As a result the findings of studies not reporting these items should be interpreted with caution.

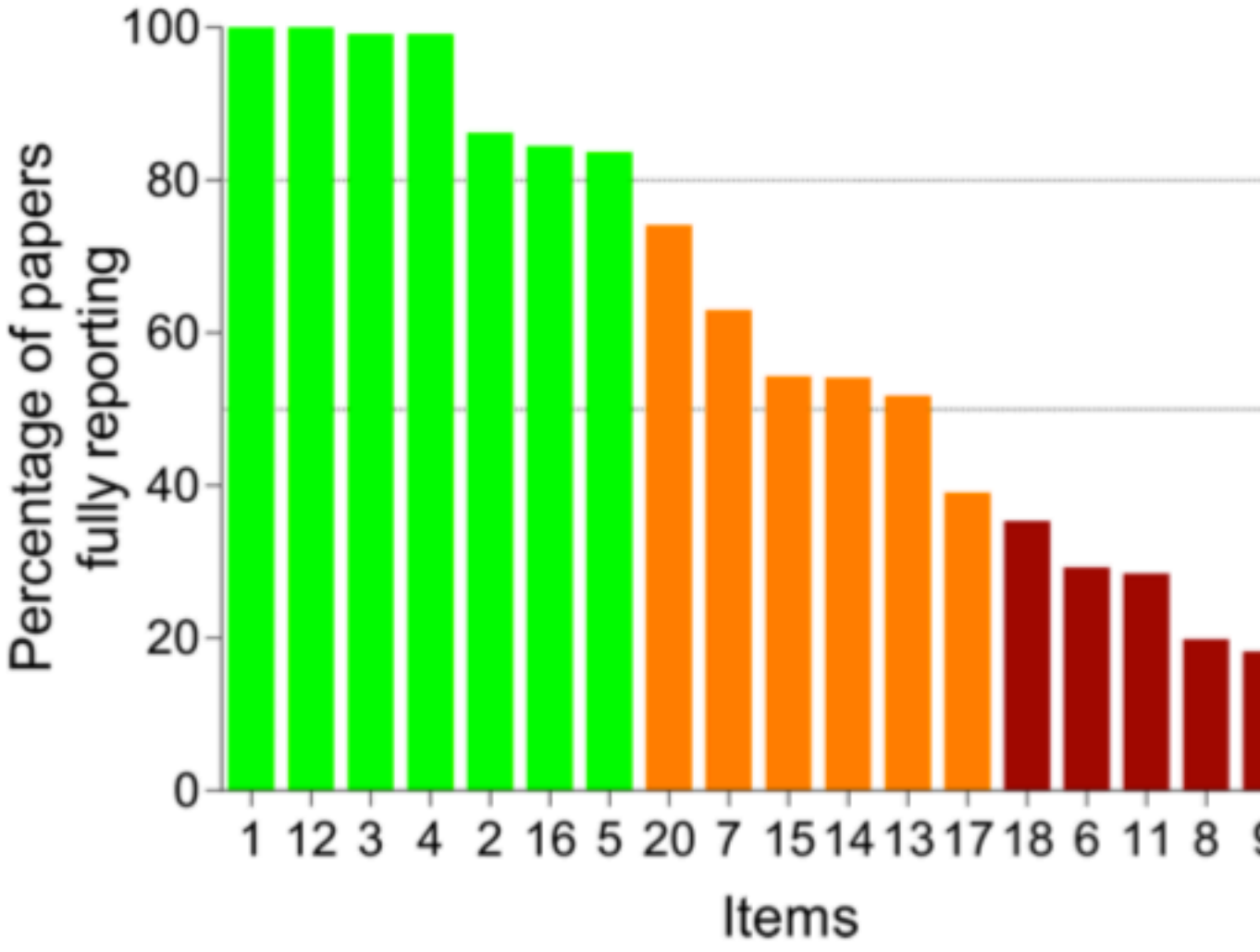

Figure 1: The percentage of papers (n = 116) published in 2015 (5 years after publication of the ARRIVE guidelines) reporting ARRIVE guideline items (Leung et al. 2018).

References:

Baker 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

Burns and O'Connor. 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.

Chalmers et al. Bias in treatment assignment in controlled clinical trials. *N Engl J Med* 1983;309:1358–61

Freedman et al. The Economics of Reproducibility in Preclinical Research. *PLoS Biol* 2015;13:e1002165

Percie du Sert and Rice. Improving the translation of analgesic drugs to the clinic: animal models of neuropathic pain. *B J Pharm* 2014; 171:2951-2963.

Kilkenny et al. Survey of the quality of experimental design, statistical analysis and reporting of research using animals. *PLoS One* 2009;4:e7824

Kilkenny et al. Improving bioscience research reporting: the ARRIVE guidelines for reporting animal research. PLoS Biol. 2010; 8: e1000412

Di Girolamo et al. In veterinary trials reporting and communication regarding randomisation procedures is suboptimal. *Vet Rec* 2017;181:195

Landis et al. A call for transparent reporting to optimize the predictive value of preclinical research. *Nature* 2012;490:187–91

Leung 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

Lund et al. Veterinary randomized clinical trial reporting: a review of the small animal literature. *J Vet Intern Med* 1998;12:57–60

Macleod et al. Evidence for the efficacy of NXY-059 in experimental focal cerebral ischaemia is confounded by study quality. *Stroke* 2008;39:2824–9

Macleod et al. 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*. 2017187245

Rufiange et al. Incomplete reporting of experimental studies and items associated with risk of bias in veterinary research. *BMJ Vet Rec Open* 2019 In press.

Savović et al. Influence of reported study design characteristics on intervention effect estimates from randomized, controlled trials. *Ann Intern Med* 2012;157:429–38

Vesterinen et al. Improving the translational hit of experimental treatments in multiple sclerosis. *Mult Scler* 2010;16:1044–55.