Translational roadblock: are publication guidelines the way around it?

Author:
Stefanie Schindler,
Animalfree Research,
Bern, Switzerland

Address of correspondence:
Dr. Dr. Stefanie Schindler
Animalfree Research
Postgasse 15
Postfach 817
e-mail: schindler@animalfree-research.org phone: 0041 (0) 44 422 7070

Abstract:

With regards to new insights, the scientific community has to rely on transparent and accurate reporting of design, conduct and outcomes of studies in order to allow readers, peer reviewers and editors an assessment of methodology and relevance. In animal experimentation, it was found that, to a disturbing extent, vital information such as the method of randomization and blinding is not or inadequately disclosed in publications. To make matters worse, a sizable number of animal studies is not reported at all, leading to an overstatement of intervention effects, unnecessary repetitions and premature clinical studies. Both factors contribute to translational problems where clinical trials do not reflect the results of preclinical findings and seriously impede therapy development. In addition, unnecessary in vivo studies present a considerable animal welfare issue. In 2010, two sets of guidelines were published with the aim of improving on project design and reporting, making meta- analyses and systematic reviews more feasible and reducing animal use: the Gold Standard Publication Checklist (GSPC) and the Animal Research: Reporting In Vivo Experiments (ARRIVE) guidelines. This mini-review focuses on the ARRIVE-Guidelines.

Keywords: Guidelines, bias, reproducibility, translation, CONSORT, ARRIVE

1. Introduction

The difficulties of translation from animal studies to human clinical trials ("bench to bedside") have been increasingly recognized over the past decades with the most frequent cause of failure being lack of efficacy. In drug development, the success rates have fallen from 28 to 18% (Arrowsmith, 2011); with regards to translation into clinical studies, it was found that drug candidates from preclinical studies in cancer research fail to 95% in Phase I (Kola & Landis, 2004), and attrition rates in stroke drug development are at 99% (O'Collins et al., 2006).

While the necessity of proper conduct and reporting as well as the avoidance of publication bias has long been recognized for human clinical trials, spurring several initiatives such as Consolidated Standards of Reporting Trials (CONSORT) (Begg et al., 1996), Stroke Therapy Academic Industry Roundtable (STAIR) (Saver, Albers, Dunn, Johnston, & Fisher, 2009), Standards for the Reporting of Diagnostic accuracy studies (STARD) (Christopher, 2007), to name but a few, the importance of adequate design, conduct, and reporting of preclinical trials has taken longer to catch on. In recent years, several authors have pointed out the fact that poor quality of conduct and reporting have a direct influence on the success rate of clinical trials. A systematic review conducted in 2007 compared the treatment effects of 6 interventions in humans (benefit or harm) to the corresponding animal experiments and found that all animal studies were of poor quality regarding for example lack of reporting of allocation concealment, considerable heterogeneity of outcomes, evidence of publication bias or poor experimental design. Indeed, clinical trials (started after encouraging results from animal experimentation) resulted in an increased health risk in 2 interventions (Perel et al., 2007). The poor quality of animal studies with resulting unsatisfactory clinical outcomes has been documented by a multitude of authors (for review please refer to (Hartung, 2013), such as (Horn, de Haan, Vermeulen, Luiten, & Limburg, 2001), (Pound, Ebrahim, Sandercock, Bracken, & Roberts, 2004), (Hackam & Redelmeier, 2006), or (E. S. Sena, Briscoe, et al., 2010).

One of the major causes has been identified as biased reporting. Randomization, investigator blinding and sample size calculation are the key approaches to assess and reduce bias in scientific work using animals (Krauth, Woodruff, & Bero, 2013). Non-blinded studies and non-randomized studies report exaggerated efficacies (Macleod et al., 2008), (van der Worp, Sena, Donnan, Howells, & Macleod, 2007), but both are mentioned in few publications (Kilkenny et al., 2009), (van der Worp, de Haan, Morrema, & Kalkman, 2005). The reporting of sample size calculation is very close to nil (E. Sena, van der Worp, Howells, & Macleod, 2007), (Kilkenny et al., 2009), (Macleod et al., 2015).

The need to place more emphasis on design, conduct and reporting was voiced by many, e. g. (Ioannidis et al., 2014), (Kilkenny, Browne, Cuthill, Emerson, & Altman, 2010), and (Landis et al., 2012). Landis et al. have particularly specified the need to disclose details on the estimation and handling of data, (e.g. when and why data collection was stopped, criteria for inclusion/exclusion of data), on the question on how outliers are defined and handled and emphasized that primary endpoints are to be selected prospectively and not after the conclusion of the study. They find that, as a first step, journals and/or funders should provide

reviewers with clear guidance on the cores of study design and provide a minimum set of standards.

There is no possible doubt that the publication bias is disastrous for meta-analyses and systematic reviews. Indeed, as Briel et al. phrased it, it is one of the major threats to their validity (Briel et al., 2013). This is, to say the least, unfortunate, since these are indispensable tools: summarizing quantitatively the results of individual studies (meta- analyses) or identifying and evaluating existing evidence (systematic reviews).

In the light of these findings, it is worrying (to say the least) that five years after the publication of the Gold Standard Publication Checklist (GSPC) by (Hooijmans, Leenaars, & Ritskes-Hoitinga, 2010) and the Animal Research: Reporting In Vivo Experimentation (ARRIVE) Guidelines by (Kilkenny et al., 2010) - the latter having been officially adopted into the author´s guidelines of more than 400 scientific journals - there is still low reporting of measures to reduce this bias, as was recently found in the field of rheumatology (Ting, Hill, & Whittle, 2015) and in vivo research in general (Macleod et al., 2015).

2. The CONSORT Statement

The Consolidated Standards of Reporting Trials (CONSORT) Statement was published in 1996, and was amongst the first responding to the poor reporting of randomized clinical trials (RCTs). Cochrane´s observation that "when humans make observations there is always the possibility of bias" was the lynchpin for the CONSORT initiative (Begg et al., 1996) in order to provide guidance on what to report in a research article. It was designed to assist authors in writing, editors and reviewers in reviewing, and readers in critically appraising articles while helping to identify potentially biased results.

5 years after the publication of the CONSORT guidelines, it was evident that the reporting of RCTs still needed improvement. For example, a review of 122 published RCTs that evaluated selective serotonin uptake inhibitors (SSRI) for depression found that only a single one (0.8%) described the method of randomization adequately (Hotopf, Lewis, & Normand, 1997). As a consequence, the CONSORT Guidelines were revised in 2001 (Altman et al., 2001) (Moher 2001), resulting in a version containing only the items absolutely necessary to reporting, leaving a 22-item checklist and flow diagram.

Nevertheless, and in spite of that, the effects on publication quality appear difficult to assess: (Hill, LaValley, & Felson, 2002) looked at altogether 240 RCTs on adult rheumatic diseases published in two time periods before and after the introduction of CONSORT (1987-88 and 1997-98), their results hinting that CONSORT has led to improvements. In a similar vein, a comparison of publications dating from 2000 and 2006, respectively, reported improvements but stated that quality remains comfortably beneath acceptable levels (Hopewell, Dutton, Yu, Chan, & Altman, 2010). In a systematic review by (Plint et al., 2006) the authors addressed the question of whether the CONSORT checklist has improved the quality of randomized trials. The results appear encouraging with regards to sequence generation and allocation concealment, but can provide no definitive answer due to the very low number of 8 out of the 248 studies retrieved e.g. in MEDLINE, EMBASE and others that met the inclusion criteria.

Another study on altogether 253 RCTs showed that in 15 journals endorsing CONSORT reporting has not improved consistently, with several essentials still being suboptimal (Mills, Wu, Gagnier, Heels-Ansdell, & Montori, 2005). Altman, who performed a review of 167 high-impact medical journals (Altman, 2005) found that only 22% mention CONSORT in their Guidelines to Authors, and of those, a quarter referred to the obsolete 1996 version.

By 2010, more than 400 journals supported CONSORT and a further update of the statements was performed (Schulz, Altman, & Moher, 2010). Still, the findings by Mills et al. were confirmed by Turner, although the group stated that there are relative improvements when CONSORT is endorsed by journals (Turner, Shamseer, Altman, Schulz, & Moher, 2012).

3. Reporting of animal experimentation

The problems with reporting on the results of animal experimentation are twofold: inadequate disclosure of vital measures to reduce bias such as randomization and blinding; and the "disappearing" of animal studies through no publication at all, both of which are detrimental to the field of science and to the development of effective therapies. Evidence on how exactly publication bias affects animal research is presently scarce: Sena et al. estimated a non- publication rate of 13.6% in the area of non-ischemic stroke (E. S. Sena, van der Worp, Bath, Howells, & Macleod, 2010), but non-publication of negative results appears to be prevalent. In 2012, ter Riet et al. performed a survey on 454 participants working in animal research in 2012 – the first survey on animal experimenters focusing on publication bias. Despite a low return rate, and the survey being conducted in only one country, the data were complete enough to evaluate. From those 421 that were working "non-profit", the answer was that 35-70% of conducted experiments on animals were published. Of those remaining, (n=21) responded that 5-50% (average 10) are published. Notable was a discrepancy in the species of animals: Own work with large animals was published in 79-100% (average 90%). Overall, the authors of the paper estimate that on average 50% of animal experiments are published, but it may be far less (ter Riet et al., 2012).

The authors come to the conclusion that, if these results are representative, this could mean that the collective literature is biased.

More recently, Tsilidis et al. have looked at 160 meta-analyses of animal experiments from the Collaborative Approach to Meta-Analysis and Review of Animal Data in Experimental Studies (CAMARADES) in neurological disorders and found that the number of published studies with positive (statistically significant) results is too large to be true (Tsilidis et al., 2013). They found evidence of excess significance in studies across all 6 areas of neurological diseases which the authors attribute to reporting biases. In addition, the observed "positive" results were three times the number that could be reasonably expected.

Excess significance had previously been found by other authors as well (E. S. Sena, van der Worp, et al., 2010), and publication bias has been associated with exaggerated effect size (Macleod, O'Collins, Howells, & Donnan, 2004) as well as deficiencies in randomization and blinding (Crossley et al., 2008), (Bebarta, Luyten, & Heard, 2003).

Further authors who have expressed concern about distorted results and/or to

insufficient reporting of animal studies have been (Horn et al., 2001) (Hackam & Redelmeier, 2006), (Dirnagl, 2006), (Knight, 2003), (Mignini & Khan, 2006), (Bracken, 2009a), (Bracken, 2009b), and (Korevaar, Hooft, & ter Riet, 2011). In conclusion, the importance of these issues in laboratory animal research cannot be overstressed.

Before this background, it seems almost incredible that adherence to the respective guidelines appears to be so difficult to achieve. As (Macleod et al., 2009) reviewed nearly 10 years after establishing the STAIR Guidelines, only a handful of researchers in the respective scientific community came even close to complying with them. There was still no effective neuroprotective drug in sight and the original hypothesis - that these guidelines would lead to improvements in translation - still remained to be proven. Incidentally, adherence to them was associated with a marked decrease in reported effect size, as O´Collins noted in a study on 1,026 publications on acute stroke (O'Collins et al., 2006). The non-reporting of animal numbers or sample size calculations is a major failure, since underpowered studies may lead to overstatement of effects (Vesterinen et al., 2010), (Macleod et al., 2015) and, more dramatically and importantly, erode the scientific database and can result in the downfall of an entire research area - as published by Button et al. for the field of Neurosciences (Button et al., 2013).

By 2014, there were still few studies reporting blinding, randomization and allocation concealment as well as statistical design (E. S. Sena, Currie, McCann, Macleod, & Howells, 2014). This is in concordance with Baker et al., who also could not notice significant improvements in reporting after the adoption of ARRIVE (Baker, Lidster, Sottomayor, & Amor, 2014). The non-compliance to guidelines was already noted by Altman for the CONSORT Guidelines (Altman, 2005), and poor effect of the guidelines on the quality of reporting was confirmed by (Turner et al., 2012). The same was seen by Macleod for STAIR (Macleod et al., 2009), and by (Coppus, van der Veen, Bossuyt, & Mol, 2006) for STARD, indicating that there appears to be an acceptance and implementation problem that is not specific to animal experimentation. Reporting Guidelines are also underused by peer reviewers (Hirst & Altman, 2012) and lack of adherence is overlooked by reviewers and editors (Coppus et al., 2006)

In a survey by Grindlay (Grindlay, Dean, Christopher, & Brennan, 2014), whose aim was to assess the knowledge and views of veterinary editors-in-chief on reporting guidelines in general, 2 specific guidelines (REFLECT and ARRIVE) were considered especially relevant for the veterinary field. Of the altogether 68 responses the authors received, 47.1% (32/68) had no clue what a reporting guideline is; 24 had the information that specific guidelines exist, and of those, 23 specified which guidelines in particular they were familiar with.

ARRIVE came in second (after CONSORT), with 16 respondents. The survey included an open question part. Amongst the reasons for not adopting the ARRIVE guidelines were fear of losing submission, an increased workload for all involved, and a sense of tradition.

4. The ARRIVE Guidelines

Several authors had published on the reporting (or lack of reporting) of statistical methods in animal experimentation, although not in the form of a systematic survey, e.g. (Alfaro, 2005), (Smith, Birke, & Sadler,

1997), when Macleod et al. summarized in 2009 that it would be reasonable to suggest that preclinical testing in animal models should adopt similar standards in order to improve decision-making and reduce wasting resources (Macleod et al., 2009).

In the same year, a large and comprehensive survey on publications of animal-based studies was commissioned by the National Centre for the Replacement, Refinement and Reduction of Animal in Research (NC3Rs), (Kilkenny et al., 2009) which resulted in the recognition that what was needed was exactly what Macleod et al. had proposed. It reviewed 271 randomly chosen publications dating almost exclusively from 2003 - 2005 on research on live rats, mice and non-human primates (NHPs), covering a wide variety of scientific areas, but being restricted to publicly funded research in the UK and the USA.

In 4% of all 271 publications, the number of animals used was not disclosed. 87% did not report randomization and 86% did not report blinding. Of the 12% reporting random allocation, 9% provided details. Of 48 randomly chosen studies out of the 271 which did report on animal numbers, none reported sample size calculation.

As a consequence of this report, an expert working group involving animal researchers, journal editors, statisticians and funders, developed the ARRIVE Guidelines in June 2009, using CONSORT as a model. The agreed-on guidelines were then used for a wider consultation, including e.g. the Medical Research Council, the Wellcome Trust, the Royal Society and others. The ARRIVE guidelines were published in 2010 (Kilkenny et al., 2010). All items promote high-quality, comprehensive reporting.

As first suggestions on guidelines for animal experimentation, Macleod et al. had proposed in 2009 that they include, for example, data on animals (species, strain, substrain, source) and measures to avoid bias (sample size calculation, inclusion/exclusion criteria, randomization, allocation concealment, reporting on animals excluded from the analysis, blinded assessment of outcome and reporting on potential conflict of interest and study funding). Accordingly, the ARRIVE checklist has 20 items, including details on the experimental animals (species, strain, source, age, weight, sex) as well as on their environment (housing, e.g. bedding material; number of cage companions; husbandry conditions such as breeding programme, light/dark cycle, ambient temperature, type of food, access to food and water, environmental enrichment) and welfare-related assessments and interventions that were carried out prior to, during, or after the experiment. The reasons for assigning such vital importance to this kind of information may not instantly be plausible, but evidence is emerging that these data may indeed be crucial for reproducibility, as is outlined in the next chapter.

4.1. The scientific rationale behind the ARRIVE guidelines

In recent years, there has been an increasing amount of data indicating that environmental conditions influence the outcomes of animal experiments to an unsuspected degree. A complete description of these data would constitute an entire review in itself, which is why, here, only a few examples can be provided. Cao et al. reported a striking influence on tumor growth in mice, reducing tumor weight (Melanoma) by 77.2% in the group that lived for 6 weeks in an enriched environment prior to cancer cell injection as compared to the group in

standard laboratory housing (Cao et al., 2010). With regards to tumor development, further data on lung cancer exist about the influence of ambient temperature by (Kokolus et al., 2013).

Ad libitum feeding has been questioned by Martin et al., raising the question whether such laboratory rodents are at all useful, them being typically overweight, insulin-resistant and hypertensive (Martin, Ji, Maudsley, & Mattson, 2010). It was noted before that the influence of comorbidities in laboratory animals is not being included into the evaluation of study outcomes (Perel et al., 2007). Indeed, reducing food intake leads to decreased cancer rates (Ross & Bras, 1971), (Albanes, 1987) and type-2 diabetes (Masoro, 2009). Standard ad libitum feeding can decrease the life span of rats by an average of 1.6 years (2.4 vs. 4 years). The maximum lifespan of ad libitum fed rats was 2.9 years, while the restricted rats lived for 4.6 years. (Weindruch & Sohal, 1997).

Further examples comprise the group of Nevalainen, who found that pair housing of rabbits over a period of several months reduces both variance in growth and serum alkaline phosphatase (Nevalainen, Nevalainen, Guhad, & Lang, 2007), and Salvarrey-Strati et al. who discovered (by accident) that changes in the environmental enrichment led to an unexpected decrease in osteoarthritis in male mice with surgically induced knee defects (Salvarrey-Strati, Watson, Blanchet, Lu, & Glasson, 2008). The importance of circadian rhythms on observed toxicity was investigated by (Ben-Cherif et al., 2013). A recent review of the influence of housing and husbandry conditions on rats has been reviewed by (Prager, Bergstrom, Grunberg, & Johnson, 2011). That these details must be disclosed (in the electronic supplement) is only the logical consequence of these insights in order to ensure that animal experiments are at all reproducible. In this light, the Kilkenny results of a notable lack of disclosure of details on the animals: 43% reported age, 46% weight, both data were reported by 13%, and 24% reported neither) appear all the more critical (Kilkenny et al., 2009).

Shortly after Kilkenny et. al. published the ARRIVE guidelines, Vesterinen et al. confirmed the main findings of their 2009 survey with an analysis of 156 original publications in one volume of the Journal of Cerebral Blood Flow and Metabolism, one of the top journals in the respective field, for study design, statistics and analysis and reporting (Vesterinen et al., 2011). The review was, unlike the work from Kilkenny et al., not restricted to animal studies, but included in vitro work and studies on humans. Still, when singling out the animal studies, the main deficiencies found by Kilkenny et al. could be fully confirmed with regards to randomization, blinding, and reporting/information about the rationale of the sample size. In a more recent 2015 study on rheumatology, key items that were poorly reported were randomization and blinding (17.1% and 29.3%, respectively) and also underreported were animal species and strain. None gave information about sample size calculation or details of allocation, and attrition rate was not reported in 80.5% of the cases. Only 9.8% gave details of all adverse effects (Ting, Hill, & Whittle, 2015).

In the field of cerebrovascular research, major translational problems are particularly evident, with only few treatments of proven efficiency available despite promising clinical trials, with several authors hinting that this may be one of the major contribution to the roadblock that exists in the field (Dirnagl, 2006), (E. Sena et al., 2007), (Fisher et al.,

2009), (Phillips, MacLehose, & Kaufman, 2008), (Crossley et al., 2008), (Jerndal et al., 2010). Vesterinen et al. conclude that overcoming the translational roadblock through the ARRIVE Guidelines is a first step.

5. Discussion and Outlook

The selective reporting of animal experiments has not been widely assessed, partly because of the difficulties in retrieving unpublished results. Scientific literature represents an incomplete and biased subset of research findings. As Ioannidis put it in 2005: Most published research findings are false (Ioannidis, 2005).

In 2006, the Journal of Blood Flow and Metabolism (JCBM) had been amongst the first to alert for a quest for quality preclinical research (Dirnagl, 2006), and adopted the ARRIVE Guidelines in 2011 (Dirnagl & Lauritzen, 2011).

In the meantime, the ARRIVE Guidelines have been translated into Mandarin, Italian, Portuguese and Spanish (https://www.nc3rs.org.uk/arrive-guidelines, (accessed on 8th of September 2015), and officially endorsed by over 400 journals. Nevertheless, there is the question how effectively they are enforced. As the data of Baker et al. indicate, there is a definite lack of enforcement after endorsement (Baker et al., 2014). More recently, using the ARRIVE Checklist, and evaluating the publications of two leading journals in the field of rheumatology in 2012, Ting et al found that, of altogether 41 papers, 1 published a negative result. None reported sample size calculation. In their 2015 paper, Macleod et al. focused again on measures to reduce the risk of bias through reporting, and, once again, out of a 146 sample size of randomly selected publications, none reported sample size calculation and 4 (3%) reported blinded assessment of outcomes (Macleod et al., 2015). Although things have improved over the course of the years (1941-2012), the authors found that sample size calculation remains the "foster child", and that a journal´s impact factor has no influence on the frequency of reporting on that piece of information.

Although some editors state, that the guidelines are nothing to adhere rigidly to, and although it has been stated that a full-blown implementation of the ARRIVE guidelines clearly requires major changes in the reporting norms of biology (Baker, Gerritsen, Rundle, & Amor, 2011), (Schwarz, Iglhaut, & Becker, 2012); others, such as (Drummond, Paterson, & McGrath, 2010) would encourage even more detail, since e.g. the method of killing as well as the means of providing care after procedures, such as analgesia, may be highly relevant.

There is a high bias in research through unreported studies that make no contribution to scientific knowledge, or, even worse, give results a false accountability and - in passing - wreck systematic reviews/meta-analyses. Chalmers has commented very critically on non-publication, demanding obligatory reporting on the outcomes of animal based studies (Chalmers, 1990). In the same vein, Varga et al. propose the establishment of an online database where all approved animal numbers are registered (similar to the documentation of human clinical trials), in order to avoid that animal studies "disappear" and to reduce publication bias (Varga, Hansen, Sandoe, & Olsson, 2010).

It is important to note that guidelines are supposed to provide support for authors, peer reviewers and editors. They are not a

suffocation of ingenuity, but to the contrary, even the most ingenious finding must be supported by reproducibility and comprehensible communication. In the publication by Ting et al., the authors acknowledge that it will probably take more than the adoption of the Arrive Guidelines to make a real difference, and that the ARRIVE Guidelines are just a first step. This is in complete concordance with Vesterinen et al. (Ting, Hill, & Whittle, 2015).

What is to be done? While welcoming the ARRIVE Guidelines to their journal "Environmental Health Perspectives", Tilson and Schroeder state that education on experimental design is currently not sufficient and courses on general principles of study design, biostatistics etc. are just as important as those on ethics. The urgent need for education on the significance of these reporting guidelines is also suggested (Christopher, 2007), and (Erb, 2010). The editor of "Disease Models and Mechanisms" (Siegel, 2011) expresses her surprise that so many papers lead their readers on a "wild goose chase" by failing to report sufficient detail. An important point is that authors leave out what other authors leave out in their paper, again emphasizing the requirement for education and good leadership.

In response to this, SYRCLE (SYstematic Review Centre for Laboratory animal Experimentation) has developed SYRCLE's RoB tool in order to counteract this problem (Hooijmans et al., 2014).

In conclusion, researchers have to be made aware of the vital importance of proper conduct and reporting in their education: on the one hand, for the benefit of their own work, and for being able to assess the work of others. Journals have to send a clear message via their reviewers and editors that they are not only adopting, but enforcing the ARRIVE guidelines. And, last but not least, funders (who are often overlooked as players!) must make it mandatory to publish a study, irrespective of the outcome, and to publish it in a way that enables those who read it to assess its value and relevance.

There has to be a consensus in the scientific community that the culture of publication has to change – and fast.

6. Conflict of interest: The author declares no conflict of interest.

7. References

Albanes, D. (1987). Total calories, body weight, and tumor incidence in mice. *Cancer Res, 47*(8), 1987-1992.

Alfaro, V. (2005). Specification of laboratory animal use in scientific articles: current low detail in the journals' instructions for authors and some proposals. [Review]. *Methods Find Exp Clin Pharmacol, 27*(7), 495-502. doi: 10.1358/mf.2005.27.7.921309

Altman, D. G. (2005). Endorsement of the CONSORT statement by high impact medical journals: survey of instructions for authors. [Research Support, Non-U.S. Gov't]. *BMJ, 330*(7499), 1056-1057. doi: 10.1136/bmj.330.7499.1056

Altman, D. G., Schulz, K. F., Moher, D., Egger, M., Davidoff, F., Elbourne, D., . . . Lang, T. (2001). The revised CONSORT statement for reporting randomized trials: explanation and elaboration. [Consensus Development Conference Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, P.H.S. Review]. *Ann Intern Med, 134*(8), 663- 694.

Arrowsmith, J. (2011). Trial watch: phase III and submission failures: 2007-2010. [News]. *Nat Rev Drug Discov, 10*(2), 87. doi: 10.1038/nrd3375

Baker, D., Gerritsen, W., Rundle, J., & Amor, S. (2011). Critical appraisal of animal models of multiple sclerosis. [Research Support, Non-U.S. Gov't Review]. *Mult Scler, 17*(6), 647-657. doi: 10.1177/1352458511398885

Baker, D., Lidster, K., Sottomayor, A., & Amor, S. (2014). Two years later: journals are not yet enforcing the ARRIVE guidelines on reporting standards for pre-clinical animal studies. *PLoS Biol, 12*(1), e1001756. doi: 10.1371/journal.pbio.1001756

Bebarta, V., Luyten, D., & Heard, K. (2003). Emergency medicine animal research: does use of randomization and blinding affect the results? [Comparative Study Review]. *Acad Emerg Med, 10*(6), 684-687.

Begg, C., Cho, M., Eastwood, S., Horton, R., Moher, D., Olkin, I., . . . Stroup, D. F. (1996). Improving the quality of reporting of randomized controlled trials. The CONSORT statement. [Research Support, Non-U.S. Gov't]. *JAMA, 276*(8), 637-639.

Ben-Cherif, W., Dridi, I., Aouam, K., Ben-Attia, M., Reinberg, A., & Boughattas, N. A. (2013).Circadian variation of Valproic acid pharmacokinetics in mice. [Research Support, Non-U.S. Gov't]. Eur J Pharm Sci, 49(4), 468-473. doi: 10.1016/j.ejps.2013.05.009

Bracken, M. B. (2009a). Why animal studies are often poor predictors of human reactions to exposure. *J R Soc Med, 102*(3), 120-122. doi: 10.1258/jrsm.2008.08k033

Bracken, M. B. (2009b). Why are so many epidemiology associations inflated or wrong? Does poorly conducted animal research suggest implausible hypotheses? *Ann Epidemiol, 19*(3), 220-224. doi: 10.1016/j.annepidem.2008.11.006

Briel, M., Muller, K. F., Meerpohl, J. J., von Elm, E., Lang, B., Motschall, E., . . . Bassler, D. (2013). Publication bias in animal research: a systematic review protocol. [Research Support, Non-U.S. Gov't]. *Syst Rev, 2*, 23. doi: 10.1186/2046-4053-2-23

Button, K. S., Ioannidis, J. P., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S., & Munafo, M. R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. [Research Support, Non-U.S. Gov't]. *Nat Rev Neurosci, 14*(5), 365- 376. doi: 10.1038/nrn3475

Cao, L., Liu, X., Lin, E. J., Wang, C., Choi, E. Y., Riban, V., . . . During, M. J. (2010). Environmental and genetic activation of a brain-adipocyte BDNF/leptin axis causes cancer remission and inhibition. [Research Support, N.I.H., Extramural]. *Cell, 142*(1), 52-64. doi: 10.1016/j.cell.2010.05.029

Chalmers, I. (1990). Underreporting research is scientific misconduct. [Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, P.H.S.]. *JAMA, 263*(10), 1405-1408.

Christopher, M. M. (2007). Improving the quality of reporting of studies of diagnostic accuracy: let's STARD now. [Editorial]. *Vet Clin Pathol, 36*(1), 6.

Coppus, S. F., van der Veen, F., Bossuyt, P. M., & Mol, B. W. (2006). Quality of reporting of test accuracy studies in reproductive medicine: impact of the Standards for Reporting of Diagnostic Accuracy (STARD) initiative. [Evaluation Studies Research Support, Non-U.S. Gov't]. *Fertil Steril, 86*(5), 1321-1329. doi: 10.1016/j.fertnstert.2006.03.050

Crossley, N. A., Sena, E., Goehler, J., Horn, J., van der Worp, B., Bath, P. M., . . . Dirnagl, U. (2008). Empirical evidence of bias in the design of experimental stroke studies: a metaepidemiologic approach. [Evaluation Studies Research Support, Non-U.S. Gov't]. *Stroke, 39*(3), 929-934. doi: 10.1161/STROKEAHA.107.498725

Dirnagl, U. (2006). Bench to bedside: the quest for quality in experimental stroke research. [Research Support, Non-U.S. Gov't Review]. *J Cereb Blood Flow Metab, 26*(12), 1465-1478. doi: 10.1038/sj.jcbfm.9600298

Dirnagl, U., & Lauritzen, M. (2011). Improving the quality of biomedical research: guidelines for reporting experiments involving animals. [Editorial Introductory]. *J Cereb Blood Flow Metab, 31*(4), 989-990. doi: 10.1038/jcbfm.2010.219

Drummond, G. B., Paterson, D. J., & McGrath, J. C. (2010). ARRIVE: new guidelines for reporting animal research. [Editorial]. *Exp Physiol, 95*(8), 841. doi: 10.1113/expphysiol.2010.053785

Erb, H. N. (2010). Changing expectations: Do journals drive methodological changes? Should they? *Prev Vet Med, 97*(3-4), 165-174. doi: 10.1016/j.prevetmed.2010.09.011

Fisher, M., Feuerstein, G., Howells, D. W., Hurn, P. D., Kent, T. A., Savitz, S. I.,

& Lo, E. H. (2009). Update of the stroke therapy academic industry roundtable preclinical recommendations. [Review]. *Stroke, 40*(6), 2244-2250. doi: 10.1161/STROKEAHA.108.541128

Grindlay, D. J., Dean, R. S., Christopher, M. M., & Brennan, M. L. (2014). A survey of the awareness, knowledge, policies and views of veterinary journal Editors-in-Chief on reporting guidelines for publication of research. *BMC Vet Res, 10*, 10. doi: 10.1186/1746-6148-10-10

Hackam, D. G., & Redelmeier, D. A. (2006). Translation of research evidence from animals to humans. [Letter Research Support, Non-U.S. Gov't Review]. *JAMA, 296*(14), 1731-1732. doi: 10.1001/jama.296.14.1731

Hill, C. L., LaValley, M. P., & Felson, D. T. (2002). Secular changes in the quality of published randomized clinical trials in rheumatology. [Evaluation Studies Research Support, Non-U.S. Gov't]. *Arthritis Rheum, 46*(3), 779-784. doi: 10.1002/art.512

Hirst, A., & Altman, D. G. (2012). Are peer reviewers encouraged to use reporting guidelines? A survey of 116 health research journals. [Research Support, Non-U.S. Gov't]. *PLoS One, 7*(4), e35621. doi: 10.1371/journal.pone.0035621

Hooijmans, C. R., Leenaars, M., & Ritskes-Hoitinga, M. (2010). A gold standard publication checklist to improve the quality of animal studies, to fully integrate the Three Rs, and to make systematic reviews more feasible. [Research Support, Non-U.S. Gov't]. *Altern Lab Anim, 38*(2), 167-182.

Hooijmans, C. R., Rovers, M. M., de Vries, R. B., Leenaars, M., Ritskes-Hoitinga, M., & Langendam, M. W. (2014). SYRCLE's risk of bias tool for animal studies. [Research Support, Non-U.S. Gov't]. *BMC Med Res Methodol, 14*, 43. doi: 10.1186/1471-2288- 14-43

Hopewell, S., Dutton, S., Yu, L. M., Chan, A. W., & Altman, D. G. (2010). The quality of reports of randomised trials in 2000 and 2006: comparative study of articles indexed in PubMed. [Research Support, Non-U.S. Gov't]. *BMJ, 340*, c723. doi: 10.1136/bmj.c723

Horn, J., de Haan, R. J., Vermeulen, M., Luiten, P. G., & Limburg, M. (2001). Nimodipine in animal model experiments of focal cerebral ischemia: a systematic review. [Review]. *Stroke, 32*(10), 2433-2438.

Hotopf, M., Lewis, G., & Normand, C. (1997). Putting trials on trial--the costs and consequences of small trials in depression: a systematic review of methodology. [Review]. *J Epidemiol Community Health, 51*(4), 354-358.

Ioannidis, J. P. (2005). Why most published research findings are false. *PLoS Med, 2*(8), e124. doi: 10.1371/journal.pmed.0020124

Ioannidis, J. P., Greenland, S., Hlatky, M. A., Khoury, M. J., Macleod, M. R., Moher, D., . . . Tibshirani, R. (2014). Increasing value and reducing waste in research design, conduct, and

analysis. *Lancet, 383*(9912), 166-175. doi: 10.1016/S0140-6736(13)62227-8

Jerndal, M., Forsberg, K., Sena, E. S., Macleod, M. R., O'Collins, V. E., Linden, T., . . .

Howells, D. W. (2010). A systematic review and meta-analysis of erythropoietin in experimental stroke. [Meta-Analysis Research Support, Non-U.S. Gov't Review]. *J Cereb Blood Flow Metab, 30*(5), 961-968. doi: 10.1038/jcbfm.2009.267

Kilkenny, C., Browne, W. J., Cuthill, I. C., Emerson, M., & Altman, D. G. (2010). Improving bioscience research reporting: the ARRIVE guidelines for reporting animal research. [Guideline Research Support, Non-U.S. Gov't Review]. *PLoS Biol, 8*(6), e1000412. doi: 10.1371/journal.pbio.1000412

Kilkenny, C., Parsons, N., Kadyszewski, E., Festing, M. F., Cuthill, I. C., Fry, D., . . . Altman, D. G. (2009). Survey of the quality of experimental design, statistical analysis and reporting of research using animals. [Research Support, Non-U.S. Gov't]. *PLoS One, 4*(11), e7824. doi: 10.1371/journal.pone.0007824

Knight, J. (2003). Negative results: Null and void. [News]. *Nature, 422*(6932), 554-555. doi: 10.1038/422554a

Kokolus, K. M., Capitano, M. L., Lee, C. T., Eng, J. W., Waight, J. D., Hylander, B. L., . . .

Repasky, E. A. (2013). Baseline tumor growth and immune control in laboratory mice are significantly influenced by subthermoneutral housing temperature. [Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't]. *Proc Natl Acad Sci U S A, 110*(50), 20176-20181. doi: 10.1073/pnas.1304291110

Kola, I., & Landis, J. (2004). Can the pharmaceutical industry reduce attrition rates? *Nat Rev Drug Discov, 3*(8), 711-715. doi: 10.1038/nrd1470

Korevaar, D. A., Hooft, L., & ter Riet, G. (2011). Systematic reviews and meta-analyses of preclinical studies: publication bias in laboratory animal experiments. *Lab Anim, 45*(4), 225-230. doi: 10.1258/la.2011.010121

Krauth, D., Woodruff, T. J., & Bero, L. (2013). Instruments for assessing risk of bias and other methodological criteria of published animal studies: a systematic review. [Research Support, N.I.H., Extramural Review]. *Environ Health Perspect, 121*(9), 985-992. doi: 10.1289/ehp.1206389

Landis, S. C., Amara, S. G., Asadullah, K., Austin, C. P., Blumenstein, R., Bradley, E. W., . . . Silberberg, S. D. (2012). A call for transparent reporting to optimize the predictive value of preclinical research. [Research Support, N.I.H., Extramural]. *Nature, 490*(7419), 187-191. doi: 10.1038/nature11556

Macleod, M. R., Fisher, M., O'Collins, V., Sena, E. S., Dirnagl, U., Bath, P. M., . . . Howells, D. W. (2009). Good laboratory practice: preventing introduction of bias at the bench. *Stroke, 40*(3), e50-52. doi:

10.1161/STROKEAHA.108.525386

Macleod, M. R., Lawson McLean, A., Kyriakopoulou, A., Serghiou, S., de Wilde, A., Sherratt, N., . . . Sena, E. S. (2015). Risk of Bias in Reports of In Vivo Research: A Focus for Improvement. [Research Support, Non-U.S. Gov't]. *PLoS Biol, 13*(10), e1002273. doi: 10.1371/journal.pbio.1002273

Macleod, M. R., O'Collins, T., Howells, D. W., & Donnan, G. A. (2004). Pooling of animal experimental data reveals influence of study design and publication bias. [Meta- Analysis Research Support, Non-U.S. Gov't]. *Stroke, 35*(5), 1203-1208. doi: 10.1161/01.STR.0000125719.25853.20

Macleod, M. R., van der Worp, H. B., Sena, E. S., Howells, D. W., Dirnagl, U., & Donnan, G. A. (2008). Evidence for the efficacy of NXY-059 in experimental focal cerebral ischaemia is confounded by study quality. [Meta-Analysis Research Support, Non- U.S. Gov't]. Stroke, 39(10), 2824-2829. doi: 10.1161/STROKEAHA.108.515957

Martin, B., Ji, S., Maudsley, S., & Mattson, M. P. (2010). "Control" laboratory rodents are metabolically morbid: why it matters. [Research Support, N.I.H., Intramural]. *Proc Natl Acad Sci U S A, 107*(14), 6127-6133. doi: 10.1073/pnas.0912955107

Masoro, E. J. (2009). Caloric restriction-induced life extension of rats and mice: a critique of proposed mechanisms. [Review]. *Biochim Biophys Acta, 1790*(10), 1040-1048. doi: 10.1016/j.bbagen.2009.02.011

Mignini, L. E., & Khan, K. S. (2006). Methodological quality of systematic reviews of animal studies: a survey of reviews of basic research. [Comparative Study Meta-Analysis Research Support, Non-U.S. Gov't]. *BMC Med Res Methodol, 6*, 10. doi: 10.1186/1471-2288-6-10

Mills, E., Wu, P., Gagnier, J., Heels-Ansdell, D., & Montori, V. M. (2005). An analysis of general medical and specialist journals that endorse CONSORT found that reporting was not enforced consistently. *J Clin Epidemiol, 58*(7), 662-667. doi: 10.1016/j.jclinepi.2005.01.004

Nevalainen, T. O., Nevalainen, J. I., Guhad, F. A., & Lang, C. M. (2007). Pair housing of rabbits reduces variances in growth rates and serum alkaline phosphatase levels. [Research Support, Non-U.S. Gov't]. *Lab Anim, 41*(4), 432-440. doi: 10.1258/002367707782314247

O'Collins, V. E., Macleod, M. R., Donnan, G. A., Horky, L. L., van der Worp, B. H., & Howells, D. W. (2006). 1,026 experimental treatments in acute stroke. [Comparative Study Evaluation Studies Research Support, Non-U.S. Gov't]. *Ann Neurol, 59*(3), 467-477. doi: 10.1002/ana.20741

Perel, P., Roberts, I., Sena, E., Wheble, P., Briscoe, C., Sandercock, P., . . . Khan, K. S. (2007). Comparison of treatment effects between animal experiments and clinical trials: systematic review. [Comparative Study Research Support, Non-U.S.

Gov't Review]. *BMJ, 334*(7586), 197. doi: 10.1136/bmj.39048.407928.BE

Phillips, C. V., MacLehose, R. F., & Kaufman, J. S. (2008). (Errors in statistical tests) 3. *Emerg Themes Epidemiol, 5*, 9. doi: 10.1186/1742-7622-5-9

Plint, A. C., Moher, D., Morrison, A., Schulz, K., Altman, D. G., Hill, C., & Gaboury, I. (2006). Does the CONSORT checklist improve the quality of reports of randomised controlled trials? A systematic review. [Review]. *Med J Aust, 185*(5), 263-267.

Pound, P., Ebrahim, S., Sandercock, P., Bracken, M. B., & Roberts, I. (2004). Where is the evidence that animal research benefits humans? [Review]. *BMJ, 328*(7438), 514-517. doi: 10.1136/bmj.328.7438.514

Prager, E. M., Bergstrom, H. C., Grunberg, N. E., & Johnson, L. R. (2011). The importance of reporting housing and husbandry in rat research. *Front Behav Neurosci, 5*, 38. doi: 10.3389/fnbeh.2011.00038

Ross, M. H., & Bras, G. (1971). Lasting influence of early caloric restriction on prevalence of neoplasms in the rat. *J Natl Cancer Inst, 47*(5), 1095-1113.

Salvarrey-Strati, A., Watson, L., Blanchet, T., Lu, N., & Glasson, S. S. (2008). The influence of enrichment devices on development of osteoarthritis in a surgically induced murine model. *ILAR J, 49*(4), 23-30.

Saver, J. L., Albers, G. W., Dunn, B., Johnston, K. C., & Fisher, M. (2009). Stroke Therapy Academic Industry Roundtable (STAIR) recommendations for extended window acute stroke therapy trials. [Congresses Research Support, N.I.H., Extramural]. *Stroke, 40*(7), 2594-2600. doi: 10.1161/STROKEAHA.109.552554

Schulz, K. F., Altman, D. G., & Moher, D. (2010). CONSORT 2010 statement: updated guidelines for reporting parallel group randomized trials. [Guideline Research Support, Non-U.S. Gov't]. *Obstet Gynecol, 115*(5), 1063-1070. doi: 10.1097/AOG.0b013e3181d9d421

Schwarz, F., Iglhaut, G., & Becker, J. (2012). Quality assessment of reporting of animal studies on pathogenesis and treatment of peri-implant mucositis and peri-implantitis. A systematic review using the ARRIVE guidelines. [Research Support, Non-U.S. Gov't Review]. *J Clin Periodontol, 39 Suppl 12*, 63-72. doi: 10.1111/j.1600-051X.2011.01838.x

Sena, E., van der Worp, H. B., Howells, D., & Macleod, M. (2007). How can we improve the pre-clinical development of drugs for stroke? [Research Support, Non-U.S. Gov't Review]. *Trends Neurosci, 30*(9), 433-439. doi: 10.1016/j.tins.2007.06.009

Sena, E. S., Briscoe, C. L., Howells, D. W., Donnan, G. A., Sandercock, P. A., & Macleod, M. R. (2010). Factors affecting the apparent efficacy and safety of tissue plasminogen activator in thrombotic occlusion

models of stroke: systematic review and meta- analysis. [Meta-Analysis Research Support, Non-U.S. Gov't Review]. *J Cereb Blood Flow Metab, 30*(12), 1905-1913. doi: 10.1038/jcbfm.2010.116

Sena, E. S., Currie, G. L., McCann, S. K., Macleod, M. R., & Howells, D. W. (2014). Systematic reviews and meta-analysis of preclinical studies: why perform them and how to appraise them critically. [Review]. *J Cereb Blood Flow Metab, 34*(5), 737-742. doi: 10.1038/jcbfm.2014.28

Sena, E. S., van der Worp, H. B., Bath, P. M., Howells, D. W., & Macleod, M. R. (2010). Publication bias in reports of animal stroke studies leads to major overstatement of efficacy. [Research Support, Non-U.S. Gov't]. *PLoS Biol, 8*(3), e1000344. doi: 10.1371/journal.pbio.1000344

Siegel, V. (2011). Reproducibility in research. [Editorial]. *Dis Model Mech, 4*(3), 279-280. doi: 10.1242/dmm.008037

Smith, J. A., Birke, L., & Sadler, D. (1997). Reporting animal use in scientific papers. *Lab Anim, 31*(4), 312-317.

ter Riet, G., Korevaar, D. A., Leenaars, M., Sterk, P. J., Van Noorden, C. J., Bouter, L. M., . .. Hooft, L. (2012). Publication bias in laboratory animal research: a survey on magnitude, drivers, consequences and potential solutions. *PLoS One, 7*(9), e43404. doi: 10.1371/journal.pone.0043404

Ting, K. H., Hill, C. L., & Whittle, S. L. (2015). Quality of reporting of interventional animal studies in rheumatology: a systematic review using the ARRIVE guidelines. *Int J Rheum Dis, 18*(5), 488-494. doi: 10.1111/1756-185X.12699

Tsilidis, K. K., Panagiotou, O. A., Sena, E. S., Aretouli, E., Evangelou, E., Howells, D. W., . . . Ioannidis, J. P. (2013). Evaluation of excess significance bias in animal studies of neurological diseases. *PLoS Biol, 11*(7), e1001609. doi: 10.1371/journal.pbio.1001609

Turner, L., Shamseer, L., Altman, D. G., Schulz, K. F., & Moher, D. (2012). Does use of the CONSORT Statement impact the completeness of reporting of randomised controlled trials published in medical journals? A Cochrane review. [Review]. *Syst Rev, 1*, 60. doi: 10.1186/2046-4053-1-60

van der Worp, H. B., de Haan, P., Morrema, E., & Kalkman, C. J. (2005). Methodological quality of animal studies on neuroprotection in focal cerebral ischaemia. [Research Support, Non-U.S. Gov't Review]. *J Neurol, 252*(9), 1108-1114. doi: 10.1007/s00415- 005-0802-3

van der Worp, H. B., Sena, E. S., Donnan, G. A., Howells, D. W., & Macleod, M. R. (2007). Hypothermia in animal models of acute ischaemic stroke: a systematic review and meta-analysis. [Meta-Analysis Research Support, Non-U.S. Gov't Review]. *Brain, 130*(Pt 12), 3063-3074. doi: 10.1093/brain/awm083

Varga, O., Hansen, A. K., Sandoe, P., & Olsson, I. A. (2010). Improving transparency and ethical

accountability in animal studies: three ways to link ethical approvals to publications. *EMBO Rep*, *11*(7), 500-503. doi: 10.1038/embor.2010.91

Vesterinen, H. M., Egan, K., Deister, A., Schlattmann, P., Macleod, M. R., & Dirnagl, U. (2011). Systematic survey of the design, statistical analysis, and reporting of studies published in the 2008 volume of the Journal of Cerebral Blood Flow and Metabolism. [Research Support, Non-U.S. Gov't]. *J Cereb Blood Flow Metab*, *31*(4), 1064-1072. doi: 10.1038/jcbfm.2010.217

Vesterinen, H. M., Sena, E. S., ffrench-Constant, C., Williams, A., Chandran, S., & Macleod, M. R. (2010). Improving the translational hit of experimental treatments in multiple sclerosis. [Meta-Analysis Research Support, Non-U.S. Gov't Review]. *Mult Scler*, *16*(9), 1044-1055. doi: 10.1177/1352458510379612

Weindruch, R., & Sohal, R. S. (1997). Seminars in medicine of the Beth Israel Deaconess Medical Center. Caloric intake and aging. [Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, P.H.S. Review]. *N Engl J Med*, *337*(14), 986-994. doi: 10.1056/NEJM199710023371407

