Published in final edited form as:

Anesthesiology. 2016 April ; 124(4): 763–765. doi:10.1097/ALN.0000000000001044.

Reporting of preclinical research in ANESTHESIOLOGY: Transparency and Enforcement

James C. Eisenach, MD,

Department of Anesthesiology, Wake Forest University School of Medicine, Winston-Salem, NC

David S. Warner, MD, and

Duke University School of Medicine, Durham, NC

Timothy T. Houle, PhD

Department of Anaesthesiology and Critical Care, Massachusetts General Hospital, Harvard University Medical Center, Boston, MA.

This editorial announces requirements for reporting experiments in animals, cells, molecules, or other biological foci that we will term “preclinical” in this editorial. In order for reviewers, editors, and readers to better gauge the quality of research, journals often endorse reporting guidelines developed by consensus methods and promulgated by organizations focused on improving quality of research conduct and reporting, such as **Enhancing the QUALity and Transparency Of health Research** (EQUATOR: www.equator-network.org). At their site, you will find consensus recommendations for reporting a wide variety of research designs including randomized clinical trials, observational studies, and systematic reviews. Included in that list, you will find recommendations for reporting preclinical studies as described in **Animal Research: Reporting of In Vivo Experiments** (ARRIVE).¹ Based on the ARRIVE guideline, at Anesthesiology we will require all investigators to:

1. Describe the experiments adequately to allow other researchers to replicate them,
2. Report whether measures to reduce bias were used, including random allocation and blinding, and how they were performed
3. Report how the sample size was determined, and
4. Report the data analysis plan.

Following are descriptions of why we require these elements and details for each.

Correspondence: James C. Eisenach, MD, Department of Anesthesiology, Wake Forest School of Medicine, Medical Center Boulevard, Winston-Salem, NC 27157. Phone: 336–716–4182, Fax: 336–716–0288, Editor-in-Chief@Anesthesiology.org.

Competing interest: Dr. Eisenach is the Editor-in-Chief of Anesthesiology and his institution receives salary support from the ASA for this position. Dr. Houle is the statistical Editor of Anesthesiology and his institution receives salary support from the ASA for this position.

The Problem

Imagine reading a clinical study where investigators gave patients a drug thought to speed recovery from sedation after anesthesia or a placebo. The description and results of the study in the article includes these statements:

Patients received either study drug (n=22) or placebo (n=23) and sedation was assessed using standard questionnaires and a battery of motor tasks known to be affected by sedation at 30 min after admission to the recovery room. The primary outcome was speed to perform a finger tracking task. Groups were compared using Student's t-test with $p < 0.05$ considered significant. Results showed that patients receiving the study drug recovered significantly faster after anesthesia by the primary outcome ($p = 0.048$).

What the investigators actually did was the following:

The investigators had not performed these tests before, so decided to give the first 20 patients the placebo. The results were consistent with other studies of recovery from sedation, so they then gave the next 20 patients the active drug. They examined the results and noted that only one of the outcomes, speed of finger tracking, showed a large, but variable drug effect in the anticipated direction, but only at 30 min after surgery (measurements were actually made at 15, 30, 45, and 60 min after surgery). They used several statistical tests to compare groups for this outcome, and the one that was closest to statistical significance showed $p = 0.09$ after they excluded one patient receiving active study drug who had a longer time than the others. Based on these promising results, the investigators enrolled 2 more patients per group. This resulted in $p = 0.06$, so they enrolled 1 more patient per group and observed a statistically significant effect ($p = 0.048$); they rejected the null hypothesis and stopped the study.

Had the investigators completely reported their actual methods, it's unlikely that a journal would accept such an article, or that a reader would put much stock in its results. To ensure adequate reporting for clinical trials, Anesthesiology requires submitted research to conform to the **CON**solidated **S**tandards of **R**eporting **T**rials (CONSORT) guidelines. Among many reporting elements, CONSORT requires including **1)** an adequate description of the experiments to allow other researchers to replicate them, **2)** report of the measures used to reduce bias, including whether and how random allocation and blinding methods were used, **3)** how the sample size was determined, and **4)** the data analysis plan. These are the same reporting elements that we will now require in all preclinical studies.

It took decades for clinical investigators to embrace these elements as critical to interpretable, reproducible, and actionable science. Given the extent that modern preclinical research lacks rigor regarding these elements, the reporting quality in such studies is "reminiscent of the situation in clinical research about 50 years ago".² Reporting these elements is only present a minority of the time (or not at all for sample size calculations), even in journals which strongly endorse the ARRIVE statement.^{3,4} The lack of translation of reporting quality to preclinical research may reflect many causes, but the consequences of poor reporting can be readily observed. The lack of reporting rigor may underlie the inability

of independent industry laboratories to replicate a majority of landmark studies from academic laboratories performing cancer, cardiovascular, and stroke research.^{2,5,6} Failure of clinical translation and of replication of preclinical research were cited by leaders of the National Institute of Neurologic Diseases and Stroke⁷ and the National Institutes of Health⁸ when they called on journals, investigators, and funders to improve education in good scientific design and in transparent reporting of essential research design elements.

Required reported elements of ARRIVE for Anesthesiology

Authors are encouraged to review the full ARRIVE guidelines¹ (in addition to the citation, they can be directly accessed at www.nc3rs.org.uk/arrive-guidelines) prior to submission of preclinical studies to Anesthesiology. However, the following items will be particularly scrutinized in research submissions.

1. Describe the experiments adequately to allow other researchers to replicate them

This is unchanged from our current requirement, and investigators are encouraged to report the key aspects of the experiments that would allow an experienced investigator outside of their laboratory to attempt replication of the study. All studies that were performed should be reported, not just those which support the hypothesis, including the number of animals in these studies and the statistical analysis. Pilot studies used to define conditions should be described only to the extent that they would aid in replication.

2. Report whether measures to reduce bias were used, including random allocation and blinding, and how they were performed

Some investigators argue that random allocation is not necessary because they are studying inbred or highly homogeneous animal populations, and that blinding is not necessary because the animal is effectively blinded to treatment. However, the need for these procedures is underscored by changes in animal behavior due to seasonal changes in the source of the protein in commercial animal chow⁹ and large inter-individual animal variability in behaviors prior to and after surgery.¹⁰ Additionally, we now know that environmental influences can alter subsequent biology and physiology *via* epigenetic and other mechanisms, despite presumed identical genomes. Similarly, experimenter blinding is essential whenever possible given that unintentional experimenter bias can influence measurements as evidenced by the fact that effect sizes of interventions are lower in studies when blinding is performed.¹¹

3. Report how the sample size was determined

Although many preclinical articles include multiple experiments, it should be reported for each experiment whether there was an *a priori* defined primary outcome measure and sample size based on estimates of variance and minimum biologically meaningful effect sizes. We recognize the need

for exploratory science, and it is quite likely that unblinded, non-randomized experiments might be included in an article as preliminary observations. Very small sample sizes in preclinical research may result in a high likelihood of false results and in mis-estimation of the true effect size, and the ethics of such unreliable research has been questioned.¹² Concerns over the unreliability of small sample size has led at least one journal to only accept studies with a minimum sample size of 5.¹³ Thus, in addition to a power calculation, at very small sample sizes, the reliability of the observation should be considered.

4. Report the data analysis plan

Prospective definition of primary outcome(s) and an analysis plan are needed to design a high quality study that has a good chance of being replicated in future studies. In clinical research, prospective documentation of these design aspects is required through trial registration. Although trial registration is not required for preclinical research, the authors should state whether primary outcomes and an analysis plan were established before the study started, and to declare what elements of the analysis were derived after examination of the data (i.e., post hoc). Clinical research investigators report the number of subjects recruited into the trial, randomized into conditions, and the number excluded from the analysis, as well as the reasons for exclusion. This same practice should be reported for each experiment involving animals. Although there may be cases where a majority of animals are excluded from data analysis due to technical failures, providing this information is extremely valuable to other investigators who wish to replicate the experiment or method. Whether any data were excluded as outliers should also be reported, including how outliers were defined and whether this was done prospectively and prior to unblinding. Often, it is advisable to report the analysis with and without outliers to allow a reader to evaluate the data in both contexts.

Enforcement

As noted, despite journal endorsement of these and other elements of the ARRIVE guidelines for reporting preclinical research, articles in these journals report the elements only a small minority of the time. Furthermore, there has been little improvement in reporting practices over the past 3 years and little difference between journals with high or low impact factors.^{3,4} For the past several years Anesthesiology has scanned all clinical trials with custom designed software to identify elements of CONSORT which are not included, and we will do the same for preclinical research for these elements of ARRIVE. The goal of these efforts is not to reduce the amount of preclinical research we publish or to discourage authors from considering Anesthesiology for publication of their preclinical research. Rather, the goal of these efforts is to enhance trust by our readers in the quality of

the science we publish, and to enhance trust by investigators that this published work is more likely to replicated and perhaps translated into improved care of patients.

Acknowledgments

Supported in part by grant R37-GM48085 from the National Institutes of Health, Bethesda, MD

References

1. Kilkenny C, Browne WJ, Cuthill IC, Emerson M, Altman DG. Improving bioscience research reporting: the ARRIVE guidelines for reporting animal research. *PLoS Biol.* 2010; 8:e1000412. [PubMed: 20613859]
2. Begley CG, Ellis LM. Drug development: Raise standards for preclinical cancer research. *Nature.* 2012; 483:531–3. [PubMed: 22460880]
3. Macleod MR, Lawson McLean A, Kyriakopoulou A, Serghiou S, de Wilde A, Sherratt N, Hirst T, Hemblade R, Bajor Z, Nunes-Fonseca C, Potluru A, Thomson A, Baginskaite J, Egan K, Vesterinen H, Currie GL, Churilov L, Howells DW, Sena ES. Correction: Risk of Bias in Reports of In Vivo Research: A Focus for Improvement. *PLoS Biol.* 2015; 13:e1002301. [PubMed: 26556632]
4. Baker D, Lidster K, Sottomayor A, Amor S. Two years later: journals are not yet enforcing the ARRIVE guidelines on reporting standards for pre-clinical animal studies. *PLoS Biol.* 2014; 12:e1001756. [PubMed: 24409096]
5. Prinz F, Schlange T, Asadullah K. Believe it or not: how much can we rely on published data on potential drug targets? *Nat Rev Drug Discov.* 2011; 10:712. [PubMed: 21892149]
6. Warner DS, James ML, Laskowitz DT, Wijdicks EF. Translational research in acute central nervous system injury: lessons learned and the future. *JAMA Neurol.* 2014; 71:1311–8. [PubMed: 25111291]
7. Landis SC, Amara SG, Asadullah K, Austin CP, Blumenstein R, Bradley EW, Crystal RG, Darnell RB, Ferrante RJ, Fillit H, Finkelstein R, Fisher M, Gendelman HE, Golub RM, Goudreau JL, Gross RA, Gubitza AK, Hesterlee SE, Howells DW, Huguenard J, Kelner K, Koroshetz W, Krainc D, Lazic SE, Levine MS, Macleod MR, McCall JM, Moxley RT III, Narasimhan K, Noble LJ, Perrin S, Porter JD, Steward O, Unger E, Utz U, Silberberg SD. A call for transparent reporting to optimize the predictive value of preclinical research. *Nature.* 2012; 490:187–191. [PubMed: 23060188]
8. Collins FS, Tabak LA. Policy: NIH plans to enhance reproducibility. *Nature.* 2014; 505:612–613. [PubMed: 24482835]
9. Shir Y, Ratner A, Seltzer Z. Diet can modify autotomy behavior in rats following peripheral neurectomy. *Neurosci Lett.* 1997; 236:71–4. [PubMed: 9404814]
10. Peters CM, Hayashida KI, Suto T, Houle TT, Aschenbrenner CA, Martin TJ, Eisenach JC. Individual Differences in Acute Pain-induced Endogenous Analgesia Predict Time to Resolution of Postoperative Pain in the Rat. *Anesthesiology.* 2015; 122:895–907. [PubMed: 25581910]
11. Sena E, van der Worp HB, Howells D, Macleod M. How can we improve the pre-clinical development of drugs for stroke? *Trends Neurosci.* 2007; 30:433–9. [PubMed: 17765332]
12. Button KS, Ioannidis JP, Mokrysz C, Nosek BA, Flint J, Robinson ES, Munafò MR. Power failure: why small sample size undermines the reliability of neuroscience. *Nat Rev Neurosci.* 2013; 14:365–376. [PubMed: 23571845]
13. Curtis MJ, Bond RA, Spina D, Ahluwalia A, Alexander SPA, Giembycz MA, Gilchrist A, Hoyer D, Insel PA, Izzo AA, Lawrence AJ, MacEwan DJ, Moon LDF, Wonnacott S, Weston AH, McGrath JC. Experimental design and analysis and their reporting: new guidance for publication in *BJP*. *British Journal of Pharmacology.* 2015; 172:3461–3471. [PubMed: 26114403]

Figure.