![COSMOS - Navigate the Compliance Universe]

Report on Research Compliance Volume 20, Number 10. September 21, 2023
OLAW-Supported ARRIVE 2.0 Guidelines Can Ensure Quality of Research Using Animals

By Jane Anderson

Following the ARRIVE 2.0 guidelines on animal research can help weed out poorly designed studies, potentially saving research dollars and curbing the collateral costs involved in misused research: multitudes of animals wasted and potentially human beings harmed.

However, Penny Reynolds, assistant professor of anesthesiology at the University of Florida and a co-author of the ARRIVE 2.0 revised guidelines, said in a recent Office of Laboratory Animal Welfare (OLAW) webinar describing the elements of ARRIVE that not enough researchers are using them.[1]

ARRIVE stands for Animal Research: Reporting of In Vivo Experiments. The guidelines include "what has been agreed on by international consensus as best practice for reporting animal-based research," Reynolds said. "The whole goal is to improve it so it's more useful and has a longer shelf life. The entire theme is this increased emphasis on rigorous, well-described methodology and shifting the emphasis away from sexy, splashy results which may have no substance to them."

NIH has been emphasizing the importance of the concepts in ARRIVE. In February, NIH published a notice encouraging the use of the ARRIVE Essential 10 Checklist in all publications featuring animal-based research involving vertebrates and cephalopods.[2]

In addition, an Aug. 14 blog post from Devon Crawford, program director for the Office of Research Quality at the National Institute of Neurological Disorders and Stroke, noted that "transparent publications follow established guidelines to ensure that important research practices are reported."[3]

Established guidelines include "the CONSORT statement for clinical trials, ARRIVE guidelines for animal studies, and PRISMA [Preferred Reporting Items for Reviews and Meta-Analyses] statement for systematic reviews," Crawford explained. "It is difficult to assess the rigor and robustness of studies that do not fully follow these guidelines. Yet, many papers do not report important practices."

New Version Released in 2020

According to Reynolds, "well over 1,000 journals have now officially endorsed these guidelines [but] they still don't seem to have had sufficient traction, and reporting standards are still universally quite poor." The ARRIVE 2.0 guidelines—developed by an international working group convened in 2017—attempted to address this by updating and streamlining the guidelines to make them more user-friendly, she said.

The original ARRIVE guidelines were published in 2010; however, more recent research shows that the majority of animal studies still do not report basic metrics, indicating poor experimental design, along with "broken checks and balances" in editorial processes and peer review, Reynolds said.

The ARRIVE 2.0 update, published in 2020, includes the guidelines and the ARRIVE Essential 10 Checklist.[4]

The checklist is two-tiered, with a "Recommended 11" following the "Essential 10." The "Essential 10" standards include the "minimum information required for assessing rigor and reproducibility," while the "Recommended 11" standards include "information required for assessing study-specific context," Reynolds said.

Reynolds stated the guidelines should be followed throughout the entire research process, and the checklist is intended to make using the guidelines easier. ARRIVE can be used to design experiments, identify and record information that otherwise might have been missed, and report information in the manuscript, she said.

Reynolds noted that use of the ARRIVE Essential 10 Checklist is encouraged but not mandated. "However, when NIH suggests something, it's probably a good idea to pay attention, especially since the funding climate doesn't show any signs of improving any time soon," she added.

The Essential 10 Checklist is presented not in rank order but in workflow order to reflect "the natural flow of an experimental process," Reynolds said.

This article will review the first five items on the checklist: study design, sample size, inclusion and exclusion criteria, randomization and blinding. The November issue of *RRC* will discuss the final five: outcome measures, statistical methods, experimental animals, experimental procedures and results.

1. *Checklist item: Study design*

   1. For each experiment, provide brief details of study design, including:

   a. *The groups being compared, including control groups. If no control group has been used, the rationale should be stated.*

   b. *The experimental unit (e.g., a single animal, litter, or cage of animals).*

   "So, what it means [is], what are you comparing?" Reynolds explained. "And this is a formal statistical structuring of your predictor variables. What is being compared? The experimental unit is your unit of analysis." Study design is the backbone of good research, she said. "It details how your data are collected [and] what data are collected. It determines the statistical analyses for sure and also how the results are to be interpreted. As a result, it increases power, reduces noise, and increases the information you can get from the study. It also reduces animal numbers." A design can't be imposed after data is collected, Reynolds said. "So, it matters because this is the single biggest obstacle to improving the quality of research overall—that people don't understand what the study design is. Often, I see it conflated with a method of analysis such as a t-test or an analysis of variance. The clue is in the name. Those are methods of analysis of the data, which are predicated on the assumption you have an underlying design to begin with." A study that hasn't been designed is "grossly inefficient and highly wasteful," Reynolds said.

2. *Checklist item: Sample size*

   a. *Specify the exact number of experimental units allocated to each group, and the total number in each experiment. Also indicate the total numbers of animals used.*

   b. *Explain how the sample size was decided. Provide details of any a priori sample size calculation, if done.*

   These numbers need to add up, but "it's also part of numbers justification," Reynolds explained. "Are the numbers of animals used in the study or the experimental units used in the study adequate to answer the research question in the first place? So, are the numbers feasible? Are they verifiable, and are they ethical?" This matters because sample size is "the number one reproducibility item" and is also "the number one

defining principle for any use of animals," she said. "Unfortunately, [in] the majority of published studies and more than 95% (it's probably closer to 98%), they neither justify the numbers that they used or even report the numbers in such a way you can actually figure out how many animals were used in the first place."

3. *Checklist item: Inclusion and exclusion criteria*

   a. *Describe any criteria used for including and excluding animals (or experimental units) during the experiment, and data points during the analysis. Specify if these criteria were established a priori. If no criteria were set, state this explicitly.*

   b. *For each experimental group, report any animals, experimental units, or data points not included in the analysis and explain why. If there were no exclusions, state so.*

   c. *For each analysis, report the exact value of n in each experimental group.*

"You need consistent a priori criteria for including or disqualifying both the animals and their data," Reynolds said. "It matters because not only do you need to define the subject pool for obtaining the best positive data so that your sample is truly representative of the defined study population, [but] it also minimizes the bias that results from arbitrary decisions as to whether or not to include or exclude data," she said. "On more than one occasion, I've gone into a lab where an experiment was being conducted and overheard the project leader saying, 'Oh, well, this animal doesn't seem to be doing so well. We're not going to include it,' or 'We'll make it a control.' That's cherry-picking data and results. It's dishonest. It is borderline unethical, and it is beginning to skirt research misconduct, which is kind of harsh, but at the very best, all you're doing is producing a highly biased and nonrepresentative set of results," Reynolds said.

4. *Checklist item: Randomisation*

   a. *State whether randomisation was used to allocate experimental units to control and treatment groups. If done, provide the method used to generate the randomisation sequence.*

   b. *Describe the strategy used to minimise potential confounders such as the order of treatments and measurements, or animal/cage location. If confounders were not controlled, state this explicitly.*

Randomization is a formal technical process based on a method of probability and assessment of assigning interventions to the experimental units, Reynolds said. "The best way of doing it is by computer algorithm because it's unbiased, and also you can use it to provide an audit trail for your methods, so you need to explain what the method is and the particular algorithm that you used," she said. "Randomization is the number one item for validity. If sample size is the number one element for reproducibility, this is the one for validity," Reynolds said. "Randomization minimizes systematic bias, and that's what you'll see most often cited in the literature. What people don't understand is that most of the basic statistical hypothesis tests are predicated on the fundamental assumption that randomization was performed. If it's not performed, your statistical hypothesis tests are actually invalid. You really don't know what it is that your results are being compared to, and there are no good reasons not to randomize."

5. *Checklist item: Blinding*

   Describe who was aware of the group allocation at the different stages of the experiment:

   - *during the allocation,*

- *the conduct of the experiment,*

- *the outcome assessment, and*

- *the data analysis*

"Now 'blinding' is kind of an old-fashioned term. It's a bit biased and ableist and sort of discriminatory, so I actually prefer the more descriptive term 'allocation concealment,'" Reynolds said. "What it is, is that you're hiding from some or all of the personnel involved in the experiment which treatment was received by which subject or experimental unit. This is logistics. You can't do allocation concealment after the data are collected, although it can be imposed at any or all stages, preferably all four."

This matters because cognitive biases are always present, Reynolds said. "You may not even know that you have cognitive biases, but it's especially critical for outcomes where any sort of subjective evaluation is required, like histology or assessing behavior or clinical progress of an animal," she said. "The tendency is to be biased in favor of whatever intervention that you prefer, say a test over control. You really want your test to work, so you'd be more inclined to judge results favorably if you knew which treatment it had already received."