Research Design. Two of NIH's recent areas of emphasis are (1) rigorous experimental design that will produce robust and unbiased results and (2) consideration of relevant biological variables in such design. The genesis of these emphases stems from NIH's realization that not all of the results produced with its funding have been reproducible. A comprehensive overview of what NIH recommends to correct this problem can be found at https://grants.nih.gov/reproducibility/index.htm.

Extra requirements with respect to experimental design and the consideration of biological variables may be imposed by individual Institutes and Centers, either on their website or in Funding Opportunity Announcements that they issue. Those specific requirements, if they exist, take precedence over the general instructions that follow.

Aids to understanding how you can enhance reproducibility of your own work are in video format. You can access them at the following links:

- Reproducibility of Data Collection and Analysis – Modern Technologies in Cell Biology: Potentials and Pitfalls (11-24-2014)
  https://videocast.nih.gov/summary.asp?Live=15277&bhcp=1
- Reproducibility of Data Collection and Analysis – Modern Technologies in Structural Biology: Potentials and Pitfalls (03-13-2015)
  https://videocast.nih.gov/summary.asp?Live=15910&bhcp=1
- Reproducibility of Data Collection and Analysis – Modern Technologies in Genome Technology: Potentials and Pitfalls (06-04-2015)
  https://videocast.nih.gov/summary.asp?Live=16381&bhcp=1
- NIH Workshop on Reproducibility in Cell Culture Studies
  09-28-2015 Day 1: https://videocast.nih.gov/summary.asp?Live=16876&bhcp=1
  09-29-2015 Day 2: https://videocast.nih.gov/Summary.asp?file=19196&bhcp=1
- Improving Openness and Reproducibility of Scientific Research (10-26-2015)
  https://videocast.nih.gov/summary.asp?live=17454&bhcp=1
- Clearinghouse for Training Modules to Enhance Data Reproducibility
  https://www.nigms.nih.gov/training/pages/clearinghouse-for-training-modules-to-enhance-data-reproducibility.aspx

<u>Rigorous Experimental Design for Robust and Unbiased Results</u>

One of the problems that NIH has identified is that some – many? – investigators have not received sufficient training in "strict application of the scientific method to ensure robust and unbiased experimental design, methodology, analysis, interpretation and reporting of results." As a consequence, as noted above, results of their research may not be replicable when the "same" experiments are repeated using appropriate experimental design. If you are one of those persons, it is relatively simple to catch up. For example, since changes in application requirements were announced in March of 2016, NIH has published aids that are designed to enhance rigor and reproducibility (e.g., https://nexus.od.nih.gov/all/2016/07/31/your-one-page-guide-to-rigor-and-reproducibility and https://grants.nih.gov/reproducibility/documents/grant-guideline.pdf). You can also find many texts and journal articles that describe rigorous experimental design for qualitative, quantitative and mixed-methods research online. They range from ones that are general/philosophical (e.g., http://www.sfn.org/Advocacy/Policy-Positions/Research-Practices-for-Scientific-Rigor and http://mbio.asm.org/content/7/6/e01902-16.full) to others (e.g., http://www.stat.cmu.edu/~hseltman/309/Book/Book.pdf) that are point-by-point guides. Many of

the publications would appear to be discipline specific. However, the principles and fundamentals of good experimental design and analysis are generally applicable across disciplinary boundaries. As an alternative approach to finding the sought-after resources, consult your reference librarian.

NIH will also address the problem of insufficient grounding in experimental design and transparency prospectively. Beginning "as early as 2017," NIH (and AHRQ) will require that their trainees receive formal instruction in rigorous experimental design and transparency to enhance reproducibility (see https://grants.nih.gov/grants/guide/notice-files/NOT-OD-16-034.html).

Unfortunately, rigorous design alone won't get you all the way to where you need to be. How you implement the design is also important, particularly if you are doing bench research. If you aren't doing so already, adhering to principles of Good Laboratory Practice as closely as is practicable is something that we recommend you consider. Although the original principles were applied to improve studies of drug safety, in our opinion they extend to any laboratory in which investigators want to produce results that are replicable. As defined by the Medicines and Healthcare Products Regulatory Agency (UK), Good Laboratory Practice procedures provide "a framework within which laboratory studies are planned, performed, monitored, recorded, reported, and archived. … GLP helps assure 'regulatory authorities' [read as 'NIH'] that the data submitted are a true reflection of the results obtained …" You can find manuals describing Good Laboratory Practice online. For example, the World Health Organization offers a very helpful publication (http://www.who.int/tdr/publications/documents/glp-handbook.pdf) that is titled *Handbook – Good Laboratory Practice (GLP)*. Stating in a grant application that you adhere to applicable principles of Good Laboratory Practice would be a strong indicator that you are serious about the issue of reproducibility. Many of the practices are relatively simple to implement and can make a <u>big</u> difference with respect to others being able to replicate your work. For example, having a standardized format and worksheets for record keeping is essential, as is a full set of standard operating procedures for your research group. The latter is even more important if you are proposing a multi-laboratory effort, as would be the case for a Research Program Project (P01) application. Little things, like checking the accuracy of your scale on a routine basis, or storing hygroscopic chemicals in either a desiccator or under a vacuum, or putting dates on reagent containers so that you can monitor shelf life may seem tedious at first, but will prove to be well worth the trouble in the end.

Another approach to standardizing how you conduct research is to adopt applicable methods that PCORI's (Patient-Centered Outcomes Research Institute) Methods Committee has created or endorsed (http://www.pcori.org/research-results/research-methodology/pcori-methodology-standards).

<u>Consideration of Relevant Biological Variables</u>

Experimental design should also take into account biological characteristics that could cause the results of investigations to vary. Sex is a particularly important and often overlooked source of variation. Full consideration of sex in the design and implementation of research requires much more than just the inclusion of both sexes. NIH suggests that applicants and reviewers should additionally consider (i) the extent to which the influence of sex, if any, has been included in the review of literature, (ii) the formulation of research questions and the design of experiments; (iii) stratified randomization of males and females has been included in the experimental design; (iv) the experimental design allows for disaggregation of data, so that results obtained from males and females can be analyzed separately and compared; (v) treatment or toxicity effects can be assessed for each sex separately; (vi) the potential influence

of sex has been included in the interpretation of results; and (vii) the extent to which projected generalizations from the study are appropriate, given the scope of the actual results. (Adapted from FAQs answer IV.2 at https://grants.nih.gov/reproducibility/faqs.htm).

With respect to learning how sex should be factored into research design, we recommend that you take advantage of online training modules that have been posted by NIH's Office of Research on Women's Health (https://orwh.od.nih.gov/research/sex-gender/methods-and-techniques) and the Canadian Institutes of Health Research's (CIHR's) Institute of Gender and Health (http://www.cihr-irsc.gc.ca/e/49347.html). One thing that you will learn is that it is not always necessary to include an equal number of males and females in every experiment. This point, as well as a number of others, have been made in two must-read publications. The first of these (http://www.fasebj.org/content/early/2015/10/28/fj.15-279554.full.pdf+html) was written by Dr. Janine Austin Clayton. She is Director of the Office of Research on Women's Health at NIH. In it, she makes the very important point that "considering sex as a biological variable is not the same as looking for sex differences." Clayton goes on to make the point that the results of rigorously designed preliminary studies can be used to justify why there is no need to include both sexes in subsequent definitive investigations – or, by contrast, that such data could reveal the existence of a potential difference. Part of the definition of "rigorously designed" in the preceding sentence is that the data from each sex be kept separate, so that they can be compared. If such a comparison would suggest a difference, then and only then would there be the need to consider in depth whether (or not) a sex difference exists. One course of action thereafter would be to study each sex in numbers sufficient to reach statistically valid conclusions. Alternatively, if the investigator would elect not to perform additional rigorous, in-depth investigations (not recommended), at the least, in the name of reporting transparency, s/he should describe in related publications that a sex difference could exist. Doing so would appropriately alert subsequent investigators.

The second publication (https://bsd.biomedcentral.com/articles/10.1186/s13293-016-0066-x) includes Dr. Clayton as a coauthor. It extends discussion to metrics that can be used by peer reviewers to assess the appropriateness and extent to which sex has been considered as a biological variable in preclinical studies that propose the use of either human subjects or non-human vertebrate animals. A table of twelve questions is presented that peer reviewers are encouraged to ask. The questions represent a consensus of the NIH and the CIHR, which are the principal funders of biomedical research in North America. CIHR's reviewer checklist, which we find very instructive and helpful, can be downloaded as a PDF at http://www.cihr-irsc.gc.ca/e/documents/igh-checklist-integrating-fund-initiatives-bio-en.pdf.

If you are proposing human or animal experiments in which it would be impossible to include both sexes, you need to state why, even if it is intuitively obvious. For example, it would not be possible to include both sexes in a study of ovarian dysfunction.

If it would be possible but unnecessary to include both sexes, in your opinion, you must include strong justification for that approach. You are required to provide similar justification for not considering other potential causes of biological variation, e.g., age, from your design.

Although NIH does not require it at this time, it encourages investigators to report whether their cell lines have male or female karyotypes and to consider the possible influence of sex on the analysis of results obtained from such cell lines (see FAQs answers IV.3 at and IV.4 at https://grants.nih.gov/reproducibility/faqs.htm).

There are many other biological variables than sex that could potentially cause variation in the results obtained. Three more mentioned by NIH include weight, age, and health status. Others might include race, ethnicity, species/strain of laboratory animal, and diet, for example.