Accelerating Biomedical Discoveries through Rigor and Transparency

Judith A. Hewitt, Liliana L. Brown, Stephanie J. Murphy, Franziska Grieder, and Shai D. Silberberg

Judith A. Hewitt, PhD, is the Chief of the Research Resources Section in the Office of Biodefense, Research Resources and Translational Research, in the Division of Microbiology and Infectious Diseases at the National Institute of Allergy and Infectious Diseases in Bethesda, MD. Liliana L. Brown, PhD, is a Program Officer in the Office of Genomics and Advanced Technologies, in the Division of Microbiology and Infectious Diseases at the National Institute of Allergy and Infectious Diseases in Bethesda, MD. Stephanie J. Murphy, VMD, PhD, is the Director of the Division of Comparative Medicine within the Office of Infrastructure Programs in the Division of Program Coordination, Planning, and Strategic Initiatives in the Office of the Director at the National Institutes of Health in Bethesda, MD. Franziska Grieder, DVM, PhD, is the Director of the Office of Infrastructure Programs in the Division of Program Coordination, Planning, and Strategic Initiatives in the Office of the Director at the National Institutes of Health in Bethesda, MD. Shai D. Silberberg, PhD, is the Director of Research Quality at the National Institute of Neurological Disorders and Stroke in Bethesda MD.

Address correspondence and reprint requests to Dr. Judith A. Hewitt, Office of Biodefense, Research Resources and Translational Research, National Institute of Allergy and Infectious Diseases, 5601 Fishers Lane, Room 8G51, Bethesda MD 20892-9825 or email jhewitt@nih.gov.

Abstract

Difficulties in reproducing published research findings have garnered a lot of press in recent years. As a funder of biomedical research, the National Institutes of Health (NIH) has taken measures to address underlying causes of low reproducibility. Extensive deliberations resulted in a policy, released in 2015, to enhance reproducibility through rigor and transparency. We briefly explain what led to the policy, describe its elements, provide examples and resources for the biomedical research community, and discuss the potential impact of the policy on translatability with a focus on research using animal models. Importantly, while increased attention to rigor and transparency may lead to an increase in the number of laboratory animals used in the near term, it will lead to more efficient and productive use of such resources in the long run. The translational value of animal studies will be improved through more rigorous assessment of experimental variables and data, leading to better assessments of the translational potential of animal models, for the benefit of the research community and society.

Key words: animal models; bias; qualification; quality; reproducibility; rigor; statistics; transparency

Introduction

Advances and discoveries in science are built upon prior findings. Results that can be reproduced by others and that stand the test of time serve as a foundation for future discoveries, while those results that are not upheld are abandoned during the course of the "self-correcting" scientific process (Collins and Tabak 2014). Strong results in biomedical research form the basis of translational, and ultimately,

clinical development of new therapeutic interventions and diagnostic tests.

As a major funder of biomedical research, the NIH expects that highly rigorous research will be conducted. However, there is growing public concern that the seemingly low reproducibility in biomedical research is an indication of low-quality research. For example, recent attempts by industry to reproduce findings published by the biomedical academic community revealed that 64–89% of the major findings were not reproducible (Begley and Ellis 2012; Prinz et al. 2011). A poll of scientists showed that more than 70% of researchers have been unable to reproduce another scientist's experiments (Baker 2016). Similar issues with reproducibility were raised in the field of psychology (Glasziou et al. 2008; Hartshorne and Schachner 2012; Open Science Collaborative 2015; Vasilevsky et al. 2013). Recently, NIH has taken actions to clarify and formalize the expectations for rigor and transparency in the research that it funds. In 2012, the NIH leadership established a working group to discuss potential contributing factors to the lack of reproducibility and to propose possible actions (Collins and Tabak 2014). This led a number of NIH Institutes and Centers to conduct pilots that tested some new approaches for more rigorous and transparent applications and reviews. The results of these pilots were then evaluated by a number of NIH-wide committees, which provided recommendations to the NIH leadership for ways to increase rigor in NIH-funded research for broad implementation. All of this activity resulted in a new policy released in 2015 (NOT-OD-16-011; see Table 1) that updated the grant application and review requirements for applications received starting in 2016.

Contributing Factors to Low Reproducibility

Several factors can lead to results that cannot readily be reproduced. For one, complex innovative experimental techniques developed in one laboratory may require extensive training before successfully being employed by others. Until this expertise is gained, it may appear that the original findings cannot be reproduced. Second, at times, variables that affect the results are unknown or not recognized and therefore not reported or described in sufficient detail. These confounding variables can prevent other investigators from reproducing the results if the unknown variables happen to be different between research groups. Both of these factors are part and parcel of scientific discovery, but their impact on scientific progress can be better controlled if the design, conduct, and analysis of studies are described in sufficient detail to allow others to build on the findings. Two other important factors are insufficient attention to measures that minimize unconscious and unintentional bias, as well as issues with the identity and/or integrity of cell lines, antibodies, and other experimental resources, discussed in further detail below.

Numerous publications have examined the reporting quality of peer-reviewed, published literature, many of which focused on preclinical research (Howells and Macleod 2013; Macleod et al. 2015; Pouwels et al. 2016; Sena et al. 2014). Hackam and Redelmeier (2006), for example, found that of 76 highly cited animal studies testing interventions published in the top 7 impact-factor journals, only 49% were rated as "good methodological quality," defined as a score of 50% or greater on a list of 10 standards defined by the Stroke Therapy Academic Industry Roundtable (STAIR 1999). Notably, in this study only 20% of the publications reported whether the experiments were conducted blind and only 12% reported whether the animals were randomized to comparison groups. Another survey of 271 publications reported on individual quality characteristics, ranging from 74% reporting the sex of animals used, to 43% and 46% reporting age or weight, respectively, to 14% of studies reporting blinding, 12% reporting randomization, and 2% reporting sample size calculation (Kilkenny et al. 2009). Importantly, systematic reviews found that studies that reported fewer quality measures tended to display a greater intervention effect, suggesting that lack of reporting is associated with lack of practice, leading to biased results (Macleod et al. 2005, 2008).

Rigor and Transparency to Increase the Predictive Value of Animal Studies

Animal models are crucial to the development of preventative, therapeutic, or diagnostic interventions, since they can establish a reasonable expectation of safety and efficacy in humans, that is, translation. However, translational failures are commonplace, where an intervention appears to hold great promise during animal testing, but fails in clinical trials. At least one-half of all suspended Phase II and Phase III trials are due to lack of efficacy (Hay et al. 2014). Possible reasons for these translational failures include animals inadequately representing the complexity of human diseases, greater heterogeneity among human subjects in comparison to research animals, significant differences in outcome measures between preclinical and clinical trials (beyond the scope of this article), and poor methodology in the animal experiments, among other potential factors (Hay et al. 2014; Ioannidis 2006; Perel et al. 2007). The accumulating evidence for insufficient attention to measures designed to minimize chance observations as well as unconscious and unintentional bias led the National Institute of Neurological Disorders and Stroke (NINDS) to hold a workshop in 2012 with academic researchers, journal editors, reviewers, funding agencies, disease advocacy groups, and the pharmaceutical industry to discuss issues surrounding the lack of transparency in reporting preclinical (i.e., nonclinical) research. The attendees of the workshop recognized that all stakeholders share responsibility for improving the rigor and transparency of research, resulting in recommendations of a minimal set of reporting parameters that were endorsed by the majority of participants (Landis et al. 2012). These guidelines were adopted by several funding and publishing organizations.

Expanding on this effort, NIH, *Nature*, and *Science* organized a second workshop in 2014 with journal editors representing over 30 basic/preclinical science journals, resulting in guidelines endorsed by over 135 journals ([No authors listed] 2014; McNutt 2014). These recommendations list the critical details that should be included in publications of preclinical research data (Table 2). The NIH anticipates that by emphasizing rigor and transparency, the new policy will accelerate biomedical discoveries and translatability where inadequate study design was responsible for the lack of predictive outcomes. While this policy is still too new to measure any impact, we will describe the elements of the policy, provide examples and resources for the biomedical research community, and discuss the potential impact of the policy on the translatability of animal models.

As discussed above, the issue of low reproducibility has multiple potential contributing factors. When NIH tackled the issue, it became clear that these multiple inputs should be addressed by a single policy. Additionally, the NIH grant policy and the preclinical reporting guidelines overlap considerably in content, yet emphasize different stages of research, namely planning and publishing, respectively. These complementary

Table 1 NIH guide notices relating to rigor and transparency (as of October 1, 2016)

Notice number	Release date	Title	Purpose	Link
NOT-OD-15-102	June 9, 2015	Consideration of Sex as a Biological Variable in NIH-funded Research	This notice focuses on NIH's expectation that scientists will account for the possible role of sex as a biological variable in vertebrate animal and human studies.	http://grants.nih.gov/grants/guide/notice-files/NOT-OD-15-102.html
NOT-OD-15-103	June 9, 2015	Enhancing Reproducibility through Rigor and Transparency	In this notice, the NIH Office of Extramural Research plans to clarify and revise application instructions and review criteria to enhance reproducibility of research findings through increased scientific rigor and transparency.	http://grants.nih.gov/grants/guide/notice-files/NOT-OD-15-103.html
NOT-OD-16-004	October 13, 2015	NIH and AHRQ Announce Upcoming Changes to Policies, Instructions and Forms for 2016 Grant Applications	This notice informs the biomedical and health services research communities of planned changes to policies, forms, and instructions for grant applications submitted in 2016. Rigor and transparency are one of several areas that the planned changes will focus on.	http://grants.nih.gov/grants/guide/notice-files/NOT-OD-16-004.html
NOT-OD-16-005	October 13, 2015	NIH and AHRQ Announce Upcoming Changes to Post-Award Forms and Instructions	This notice informs the biomedical and health services research communities of planned changes to policies, forms and instructions for interim and final progress reports, and other post-award documents associated the monitoring, oversight, and closeout of an award. Changes include addition of clarifying rigor language to the PHS Research Performance Progress Report.	http://grants.nih.gov/grants/guide/notice-files/NOT-OD-16-005.html
NOT-OD-16-011	October 9, 2015	Implementing Rigor and Transparency in NIH and AHRQ Research Grant Applications	This notice informs the biomedical research community of specific updates to application instructions and review language for research grant applications intended to enhance the reproducibility of research findings through increased scientific rigor and transparency.	http://grants.nih.gov/grants/guide/notice-files/NOT-OD-16-011.html
NOT-OD-16-012	October 13, 2015	Implementing Rigor and Transparency in NIH and AHRQ Career Development Award Applications	This notice informs the biomedical research community of specific updates to application instructions and review language for career development awards intended to enhance the reproducibility of research findings through increased scientific rigor and transparency.	http://grants.nih.gov/grants/guide/notice-files/NOT-OD-16-012.html
NOT-OD-16-031	December 15, 2015 (Effective Date: January 25, 2016)	Updates to NIH and AHRQ Research Performance Progress Reports (RPPR) to Address Rigor and Transparency	This notice informs the biomedical and health services research community of planned changes to address rigor and transparency to the PHS RPPR instructions for all annual non-competing (Type 5) NIH and AHRQ awards that support research activities.	http://grants.nih.gov/grants/guide/notice-files/NOT-OD-16-031.html
NOT-OD-16-034	December 17, 2015	Advanced Notice of Coming Requirements for Formal Instruction in Rigorous Experimental Design and Transparency to Enhance Reproducibility: NIH and AHRQ Institutional Training Grants, Institutional Career Development Awards, and Individual Fellowships	This notice informs the biomedical and health services research communities of NIH and AHRQ plans to require formal instruction in scientific rigor and transparency to enhance reproducibility for all individuals supported by institutional training grants, institutional career development awards, or individual fellowships.	http://grants.nih.gov/grants/guide/notice-files/NOT-OD-16-034.html
NOT-OD-16-058	January 22, 2016	Reminder: NIH and AHRQ Grant Application Changes for Due Dates On or After January 25, 2016	This notice reminds the biomedical and health services research communities of announced changes to grant application	http://grants.nih.gov/grants/guide/notice-

Continued

Table 1 Continued

Notice number	Release date	Title	Purpose	Link
			policies and instructions for due dates on or after January 25, 2016. Rigor and transparency in research grant applications (including small business and complex research grant applications) are one of the areas that the planned changes will focus on.	files/NOT-OD-16-058.html
NOT-OD-16-081	March 23, 2016	Reminder: NIH and AHRQ Grant Application Changes for Due Dates On or After May 25, 2016	This notice reminds the biomedical and health services research communities of announced changes to grant application policies and instructions for due dates on or after May 25, 2016. Rigor and transparency in research grant applications (including small business and complex research grant applications) are one of the areas that the planned changes will focus on.	http://grants.nih.gov/grants/guide/notice-files/NOT-OD-16-081.html

efforts will be more effective in producing the needed changes than either policy or guidelines alone. In addition, the NIH encourages the scientific community to establish and publish consensus standards and best practices that may vary by scientific field and can be cited by applicants as well as support peer review. In the absence of established standards or best practices, researchers should transparently report on what they have done so that consensus can emerge. Importantly, changes in how researchers are trained in principles of rigor and transparency were viewed by NIH as separate from the practice of research and will be the subject of a subsequent policy.

Summary of the NIH Policy on Enhancing Reproducibility through Rigor and Transparency

There are four elements emphasized in the new NIH grant policy, and each is described below. The policy is written broadly to apply to the many areas of biomedical research funded by NIH, though translational failures from preclinical research provided the impetus for the NIH policy. This discussion focuses mainly on preclinical animal studies, though many of the principles considered here are broadly applicable. Table 1 compiles the NIH Guide Notices relevant to this policy, and this paper reflects the authors' interpretation of the policy.

Scientific Rigor

Scientific rigor, or rigorous experimental design, has an objective-based definition in order to fit the myriad of scientific fields supported by NIH. Scientific rigor is the strict application of the scientific method to ensure robust and unbiased experimental design, methodology, analysis, interpretation, and reporting of results. This includes full transparency in reporting experimental details, data, and results so that others may reproduce and extend the findings. NIH expects applicants to describe how they will achieve meaningful results when describing the experimental design and proposed methods. Meaningful results are obtained using methods designed to avoid bias and can be reproduced under well-controlled and reported experimental conditions (for further details, see section on experimental bias below). Describing how one will achieve scientific rigor itself is not an issue, though doing so within grant application page limits is perceived as a challenge. Many journals now devote more space to methods and supplemental data as part of their policy on reproducibility, and the journal guidelines (NIH 2016b) are a good starting point for deciding what details are important to consider and include in a grant application.

Scientific Premise

Scientific premise is the rigor, or strength, of the key data supporting the proposed research. Scientific premise is *not* simply the hypothesis or the perceived importance of the research. For grant applications, it is expected that investigators present a careful consideration of the strengths and weaknesses of the most pertinent prior research put forward in the application in support of the working hypothesis, whether published or not. This consideration may include examination of the other three elements of the policy in a retrospective manner. For example, were prior studies performed with sufficient rigor to minimize biases? Do prior results warrant moving to a new series of experiments, or should they be repeated and/or refined? Were *relevant* biological variables considered? Were *key* resources

Table 2 Examples of online training and other resources for addressing rigor and transparency in biomedical research

Topic	Name of resource	Description	Website*
General—Rigor and Reproducibility	NIH Rigor and Reproducibility Web-portal and Training Modules	This NIH **web portal** provides information about the efforts underway by NIH to enhance rigor and reproducibility in scientific research. This site includes **online training modules** developed by NIH to focus on integral aspects of rigor and reproducibility in the research endeavor, such as bias, blinding, and exclusion criteria. The modules are not meant to be comprehensive, but rather are intended as a foundation to build on and a way to stimulate conversations, which may be facilitated by the use of the accompanying discussion materials.	http://www.nih.gov/science/reproducibility https://www.nih.gov/research-training/rigor-reproducibility/training
	NIH Office of Extramural Research Web Portal and General Policy Overview	This **web portal** from the NIH Office of Extramural Research has up-to-date information and tools for the extramural community on the policy. OER created a general policy overview on reproducibility that is publicly available. This 30-minute **narrated slide presentation** presents the rationale behind the policy and is a good starting point for anyone trying to gain familiarity with the issues.	http://grants.nih.gov/reproducibility/index.htm http://grants.nih.gov/reproducibility/module_1/presentation.html
	NINDS Workshop: Optimizing the Predictive Value of Preclinical Research	This website has a **summary, agenda, presentations, and recommendations** from the June 2012 workshop organized by NINDS on Optimizing the Predictive Value of Preclinical Research, attended by academics, journal editors, funding agencies, industry, and disease advocacy groups.	http://www.ninds.nih.gov/funding/areas/channels_synapses_and_circuits/rigor_and_transparency/
	NIGMS Clearinghouse for Training Modules to Enhance Data Reproducibility	This clearinghouse will provide all of the **training modules** to enhance data reproducibility, developed or funded by NIH. As additional modules are completed, they will be added to this site.	https://www.nigms.nih.gov/training/pages/clearinghouse-for-training-modules-to-enhance-data-reproducibility.aspx
	American Physiological Society: Reproducibility Journal Club	**Journal club activity** to gain insight into the challenges of improving scientific rigor	http://www.the-aps.org/mm/SciencePolicy/Agency-Policy/Reproducibility/Reproducibility-Toolkit/Journal-Club-Activity.html
	Society for Neuroscience Rigor and Reproducibility Training Webinars	SfN has partnered with NIH and leading neuroscientists who are experts in the field of scientific rigor to offer the **webinar series** Promoting Awareness and Knowledge to Enhance Scientific Rigor in Neuroscience.	http://neuronline.sfn.org/TMEDR
Experimental Design	ILAR Roundtable: Reproducibility Issues in Research with Animals and Animal Models	Public workshop to discuss fundamental aspects of experimental design of research using animals and animal models, aimed at improving reproducibility. **Interactive summary, presentations, and video recordings of workshop** available online.	http://nas-sites.org/ilar-roundtable/roundtable-activities/reproducibility/
	ILAR Journal Issues on Experimental Design and Statistics in Biomedical Research	ILAR **Journal issues** dedicated to experimental design and statistics, published in 2002 and 2014	http://ilarjournal.oxfordjournals.org/content/55/3.toc http://ilarjournal.oxfordjournals.org/content/43/4.toc
	LabRoots Laboratory Animal Science Virtual Conference: Optimizing Design, Conduct and Reproducibility of Animal Studies	**Conference presentations** from the following three tracks available online for viewing: optimizing animal study designs; optimizing the conduct and implementation of animal studies; and optimizing reproducibility of animal studies.	http://www.labroots.com/virtual-event/laboratory-animal-sciences-2016
	3Rs- Reduction.co.uk	Interactive **short course** on experimental design for research scientists working with laboratory animals.	http://www.3rs-reduction.co.uk/
	Research Randomizer	**Resource** that provides a quick way of generating random numbers or assigning participants to experimental conditions.	https://www.randomizer.org/
	NC3Rs Experimental Design Assistant	The NC3Rs developed an **Experimental Design Assistant** (EDA) that can be used by researchers to generate experimental plans and diagrams that can help address potential bias at critical points in their experiments. The EDA can output suggestions for including randomization, concealment, and blinding, as well as	https://eda.nc3rs.org.uk/

Continued

Table 2 Continued

Topic	Name of resource	Description	Website*
		performing power calculations to determine and adequate number of animals for any experiment.	
	Center for Open Science	The Center for Open Science is a non-profit organization dedicated to increasing the openness, integrity, and reproducibility of scientific research. They have **tools** available, including free training, statistical consulting, webinars, and workshops.	https://cos.io/stats_consulting/
	STAR Methods	STAR Methods promote rigor and robustness with an intuitive, consistent **framework** that integrates seamlessly into the scientific information flow, making reporting easier for the author and replication easier for the reader.	http://www.cell.com/star-methods
	Stu Hunter Teaches Statistics	**Video series** on various topics related to statistics and experimental design.	https://www.youtube.com/playlist?list=PL335F9F2DE78A358B
	Assay Guidance Manual	This **manual** provides guidance for the design, development, and statistical validation of in vivo assays residing in flow schemes of discovery projects. It provides statistical methodology for prestudy, cross-study (lab-to-lab transfers and protocol changes), and in-study (quality control monitoring) validation.	https://www.ncbi.nlm.nih.gov/books/NBK92013/pdf/Bookshelf_NBK92013.pdf
Sex as a Biological Variable	NIH Office of Research on Women's Health: The Science of Sex & Gender in Human Health Online Course Series	**Online series of courses** that provides a foundation for sex and gender accountability in medical research and treatment.	https://www.sexandgendercourse.org/
	NHLBI Working Group **Executive Summary**: Sex Bias In Cardiovascular Research	The National Heart, Lung, and Blood Institute held a working group meeting to examine the topic of sex bias in cardiovascular research on September 22, 2014, in Bethesda, MD. The working group gathered leading scientists in the field, who discussed the current knowledge and identified scientific gaps and challenges related to sex differences in nonclinical and clinical research in cardiovascular diseases. Representatives from the NIH Office of Research on Women's Health, Office of Extramural Research, Center for Scientific Review, and the Food and Drug Administration also participated in the working group's deliberations.	http://www.nhlbi.nih.gov/research/reports/sex-bias-cardiovascular-research
	Gendered Innovations in Science, Health and Medicine, Engineering, and Environment	The peer-reviewed Gendered Innovations project develops **practical methods** of sex and gender analysis for scientists and engineers; and provides **case studies** as concrete illustrations of how sex and gender analysis leads to innovation.	http://genderedinnovations.stanford.edu/
	Organization for the Study of Sex Differences (OSSD)	**Resources** from the OSSD 2015 Workshop entitled "How to Study Sex Differences"	http://www.ossd.wildapricot.org/teaching-materials
	NIH Office of Research on Women's Health (ORWH)	**Methods and techniques** for integrating sex into research	http://orwh.od.nih.gov/research/sex-gender/methods-and-techniques/
Reporting Guidelines	United States National Library of Medicine—Research Reporting Guidelines and Initiatives: By Organization	**Summary listing** of the major biomedical research reporting guidelines that provide advice for reporting research methods and findings	https://www.nlm.nih.gov/services/research_report_guide.html
	Equator Network: Enhancing the QUAlity and Transparency Of health Research	Online library that contains a comprehensive searchable **database** of reporting guidelines and also links to other resources relevant to research reporting	http://www.equator-network.org/
	Principles and Guidelines for Reporting Preclinical Research	NIH held a joint workshop in June 2014 with the Nature Publishing Group and Science on the issue of reproducibility and rigor of research findings, with journal editors representing over 30 basic/preclinical science journals in which NIH-funded investigators have most often published. The workshop focused on identifying the common opportunities in the scientific publishing arena to enhance rigor and further support research that is reproducible, robust, and transparent. The journal editors came to consensus on a set of **principles** to facilitate these goals, which a considerable number of journals have agreed to endorse.	https://www.nih.gov/research-training/rigor-reproducibility/principles-guidelines-reporting-preclinical-research https://www.nih.gov/sites/default/files/research-training/initiatives/reproducibility/rigor-reproducibility-endorsements.pdf

	National Centre for the Replacement, Refinement and Reductions of Animals in Research (NC3Rs) ARRIVE Guidelines and Checklist	ARRIVE (Animal Research: Reporting of *In Vivo* Experiments) **guidelines** are intended to improve the reporting of research using animals, maximizing information published and minimizing unnecessary studies. The ARRIVE guidelines, originally published in PLOS Biology, were developed in consultation with the scientific community as part of an NC3Rs initiative to improve the standard of reporting of research using animals.	http://www.nc3rs.org.uk/arrive-guidelines
Data Sharing	Collaborative Approach to Meta-Analysis and Review of Animal Data from Experimental Studies (CAMARADES)	Collaboration that provides a supporting **framework** for groups involved in the **systematic review and meta-analysis** of data from experimental animal studies	http://www.dcn.ed.ac.uk/camarades/default.htm
	NIH Sharing Policies and Related Guidance on NIH-Funded Research Resources	Selected NIH **policies and related guidance** on sharing of research resources developed with NIH funding	https://grants.nih.gov/policy/sharing.htm
	biomedical and HealthCAre Data Discovery Index Ecosystem (bioCADDIE)	The bioCADDIE team will develop a **data discovery index** (DDI) prototype, which will index data that are stored elsewhere. The DDI will play an important role in promoting data integration through the adoption of content standards and alignment to common data elements and high-level schema.	https://biocaddie.org/
	Preclinical Reproducibility and Robustness Channel	The Preclinical Reproducibility and Robustness channel is a **platform for open and transparent publication** of confirmatory and nonconfirmatory studies in biomedical research. The channel is open to all scientists from both academia and industry and provides a centralized space for researchers to start an open dialogue, thereby helping to improve the reproducibility of studies.	http://f1000research.com/channels/PRR
Examples of NIH Data Archives	National Center for Biotechnology Information	The National Center for Biotechnology Information advances science and health by providing **access to biomedical and genomic information**.	http://www.ncbi.nlm.nih.gov/
	Bioinformatics Resource Centers (BRCs)	The BRCs for Infectious Diseases program collects, archives, updates, and integrates a variety of research data and provides such information through user friendly **interfaces and computational analysis tools** made freely available to the scientific community.	https://www.niaid.nih.gov/research/bioinformatics-resource-centers
	Immunology Database and Analysis Portal (ImmPort)	ImmPort is a long-term, sustainable **data warehouse** for the purpose of promoting re-use of immunological data generated by NIAID DAIT and DMID-funded investigators.	http://immport.niaid.nih.gov

*Active and accessible on 10/1/2016.

properly authenticated? Answering these questions should help investigators identify the strengths and weaknesses of the prior research in order to either address them prior to submitting a grant application or to incorporate those strategies in the application, depending on the nature of the risk to the proposed research.

Evaluation of scientific premise is intended to prompt both applicants and reviewers to consider the quality of the research foundation for the proposed project. This is critical for maximizing the benefits of research dollars as well as providing a strong ethical foundation for the use of both animal and human subjects. To illustrate the value of premise, we present an example. The non-profit Amyotrophic Lateral Sclerosis (ALS) Therapy Development Institute (TDI) published the results of 5 years of rigorous studies exploring the effects of more than 70 drugs in a mouse model of ALS (Perrin 2014; Scott et al. 2008). None of the drugs was found to have a significant effect on the survival of the mice, including several drugs that had previously been reported to extend the lifespan of the animals. The authors concluded that the previously reported apparent effects of the various drugs were likely due to chance arising from the use of small sample sizes. Based on an elegant analysis, they recommended the use of 24 litter-matched mice per group, half male and half female, and to exclude deaths not related to ALS, or mice with low copy number of the transgene, to achieve the lowest level of noise and therefore maximize the ability to detect true drug effects. Notably, one of the drugs they tested was minocycline, which played an important role in the decision of NINDS to address the issue of scientific premise. In 2002 a number of studies reported that minocycline extends the lifespan of the SOD1 mouse model (Kriz et al. 2002; Van Den Bosch et al. 2002; Zhu et al. 2002). This prompted NINDS to fund a randomized, placebo-controlled clinical trial in which 412 patients were treated with minocycline for 9 months. The results of the trial, published 1 year before the report by ALS-TDI, found no beneficial effect of minocycline on patients with ALS (Gordon et al. 2007). Reexamination of the three preclinical publications cited by the minocycline clinical trial revealed that one study used just 10 animals per group without reporting on litter matching, sex of the animals, exclusion of non-ALS deaths, or exclusion of low copy transgenics. Furthermore, there was no mention of whether the animals were appropriately randomized to comparison groups and whether experiments were conducted blind (Zhu et al. 2002). A second study used just 7 females per group without reporting on litter-matching or exclusion of non-ALS deaths and low copy transgenics (Van Den Bosch et al. 2002), and the third study used 12 to 17 littermates per group and did not report on sex, exclusion of low copy transgenics, or non-ALS deaths (Kriz et al. 2002). Thus, in hindsight, examination of the scientific premise for the clinical trial suggests that more rigorous preclinical data should have been obtained before investing in a Phase III trial.

The cornerstones of laboratory animal welfare are the 3Rs, or replacement, reduction, and refinement (Russell and Burch 1959). Reduction in animal numbers potentially contributes to decreased rigor if the number of animals is insufficient to minimize chance observations or to increase the likelihood of observing a true effect (Button et al. 2013; Scott et al. 2008). Reduced purchasing power and an increasingly competitive environment have likely also contributed to inadequate numbers of animals being used. An alternative set of 3Rs was recently proposed: relevance, robustness, and reproducibility (Everitt 2015). Focusing on the scientific aspects articulated by these alternative 3Rs may increase the number of animals used in the near term but is also likely to reduce the number of animals used in the long term, as a focus on quality improves research overall. Institutional animal care and use committees as well as researchers, peer reviewers, institutions, journals, scientific societies, and funders should support this culture shift.

A frequently raised concern is that requiring rigor in both previous and proposed work might stifle innovative or high-risk research. It is important to keep in mind that all types of research fall somewhere on the exploratory—confirmatory continuum and one should acknowledge the nature of the research in applications and publications (Jaeger and Halliday 1998; Kimmelman et al. 2014; Landis et al. 2012). *Exploratory* research minimizes type II error (false negatives) and is used to generate hypotheses and models for testing, and is therefore not conclusive. Exploratory research may be able to accommodate less rigor than confirmatory research. *Confirmatory* research tests these hypotheses or models and therefore minimizes type I error (false positives). Done properly, confirmatory research should be adequately powered, controlled for possible confounding variables by appropriately randomizing subjects to comparison groups, and designed to minimize potential bias by blinding subject allocation and outcome analysis. Indeed, it is possible that part of the current concerns about low reproducibility result from a lack of distinction between these dissimilar research objectives. Innovative or high-risk research may involve a greater level of uncertainty because of the novelty of the research questions, yet it must clearly be defined as exploratory or confirmatory and in the latter case be carried out in a more scientifically rigorous manner.

Authentication of Key Resources

Problems with the identity and purity of cell lines have been known for decades (Chatterjee 2007; Hughes et al. 2007). An NIH Guide Notice released in 2007 *promoting* the authentication of cultured cell lines (NIH 2007) appears to have had little impact on authentication practices. NIH leadership subsequently published a commentary on the issues of cell line misidentification (Lorsch et al. 2014) and the potential benefits of *requiring* authentication. During development of the policy, similar issues surrounding antibodies were considered as well, resulting in a general definition of key biological and/or chemical resources that allows investigators to decide which of their resources should be authenticated. Key biological and/or chemical resources are broadly defined as resources that may or may not be generated with NIH funds and: (1) may differ from laboratory to laboratory or over time; (2) may have qualities and/or qualifications that could influence the research data; and (3) are integral to the proposed research. These include, but are not limited to, cell lines, specialty chemicals, antibodies, and other biologics. Standard laboratory reagents not expected to vary do not need to be included in the plan; examples are buffers and common salts. Genetically modified animals, while not specifically called out by the policy, clearly meet the definition of a key biological resource and as such, should be authenticated (Lloyd et al. 2015). For example, the Mutant Mouse Resource and Research Centers (MMRRC) document, verify, and authenticate each detail on archived mouse models as part of the MMRRC's reproducibility assurances (MMRRC 2016).

NIH encourages the scientific community to develop standards for authenticating various types of resources. A recent publication describes a framework for human cell line authentication, annotation, and quality control using short tandem repeat (STR) profiling as well as single nucleotide polymorphisms (Yu et al. 2015). Mouse cells of certain species may be

authenticated using STR profiles (Almeida et al. 2014) while strain differences may require single nucleotide polymorphism profiling (Didion et al. 2014). The process of development and adoption of standards has been well described (Almeida et al. 2016). The National Institute for General Medical Sciences (NIGMS) at NIH held a workshop in September 2015 to highlight issues surrounding cell line authentication (NIH 2015b), and recommendations from that workshop will be published. NIH also supports the development of new technologies that will assist researchers in authenticating various reagents, for example, PA-16-186, Tools for Cell Line Identification (NIH 2016c). The National Center for Biotechnology Information (NCBI) is supporting this effort through the BioSample database (NCBI 2012), which includes STR profiles and misidentified cell lines. Another aspect of key resources is the variation in nomenclature, which could be addressed by the use of unique identifiers for all key resources (Bandrowski and Martone 2016).

Sex and Other Relevant Biological Variables

NIH expects that sex will be factored into research design, analyses, and reporting in vertebrate animal and human studies; this may take different forms, depending upon the field of science and the information available (Clayton 2016). Strong justification must be provided for applications proposing to study only one sex, and that justification should come from the scientific literature, preliminary data, or other relevant considerations. Understanding terminology is important: sex is a biological variable determined by the presence of sex chromosomes and organs, whereas gender is a psychosocial construct of male vs female identity (Torgrimson and Minson 2005). Cell lines, primary cells, tissues, and animals have a sex but do not have a gender.

Prior to the policy, it was common for investigators to use only female laboratory animals to avoid male aggression or to use male animals to avoid variability due to the estrus cycle in females. It is not the intent of the policy that investigators surgically manipulate females to alter the estrus cycle, nor to tolerate male aggression, as either approach may have an undesired impact on results. Indeed, some have demonstrated that the estrus cycle does not create more variability in female mice than is seen in males and both sexes show less variability when singly housed compared to group housed (Prendergast et al. 2014). Similar data are available for rats (Becker et al. 2016). Meta-analyses of microarray data have also shown that inter-individual variability of gene expression in female mice and humans is not more variable than males (Itoh and Arnold 2015). Experimental design should consider sex and data should always be analyzed by sex even if differences are not expected; indeed, it is possible to see opposite effects in males and females such that aggregated data do not show a difference (McCullough et al. 2005). However, not all experiments need to be powered to detect sex differences, since powering studies requires that one first understands the magnitude of any difference. If the published literature does not include data in both males and females, then it would be appropriate to begin to collect that data.

While sex **must** be considered in studies of vertebrate animals and humans, it is up to the investigator to determine if any other relevant biological variables will be considered and clearly state that. Other relevant biological variables may include age, weight, and underlying health conditions, as these are often critical factors affecting health or disease and experimental outcomes in research. Overlooking relevant biological variables in designing experiments and not reporting these

variables may lead to differences when researchers try to reproduce experimental findings from other laboratories. Determining which biological variables should be considered in the design and analyses of proposed studies and how they should be appropriately controlled will depend on the scientific discipline(s) involved and the research question(s) being examined. Clearly, it is more difficult to consider continuous variables, compared to the dichotomous variable of sex, and researchers should consider the context of the research question and field in deciding which factors to observe or control. For example, studies of stroke may take into account relevant risk factors such as age, hypertension, diabetes, and/or obesity, perhaps by studying older, hypertensive animals. The principles of observing, reporting, analyzing, and, if appropriate, studying differences, apply to these biological variables in addition to sex (Clayton 2016).

A growing body of literature demonstrates that variables in conducting animal studies, previously assumed to be minor, may in fact have profound effects on the experimental results. For example, restricting the pathogens in animal facilities produces an immune cell repertoire similar to human neonates, while modifying the environment of laboratory mice to more closely mimic pet store or field mice environments (i.e., dirty mice) induces different populations of immune cells that more closely resemble adult human physiology (Beura et al. 2016). To control for the effects of pathogen status of animals in research studies, current guidelines and recommended best practices within the biomedical research and laboratory animal science communities promote standardization of health monitoring (i.e., FELASA 2015) and advocate more detailed reporting of the health status of research animals in manuscripts and applications (i.e., Animal Research: Reporting of *In Vivo* Experiments guidelines; NC3Rs 2010).

Similarly, the notion that any given genotype-phenotype relationship can be extrapolated from one genetic background to another has recently come under question (Sittig et al. 2016). The phenotypic effects of two different null alleles were examined on 30 genetic backgrounds and there were significant differences due to genetic background and in a few cases, the phenotypic effects were opposite. The authors call for the research community to broaden their focus and leverage the genetic diversity among inbred strains to unravel the genetic basis of disease traits and develop new therapeutics. Researchers should give careful thought to the generalizability of their findings.

In summary, these underappreciated differences in potentially biologically relevant variables contribute to the reproducibility issue, and investigators should give them consideration in planning and/or reporting experimental results.

Resources and Examples

There are numerous resources available for each area of the policy, and we present some of them here; see the full list in Table 2. This compilation is not intended to be exhaustive or even vetted, but rather is intended to provide a starting point for addressing some of the more common issues.

Training Materials

The NIH Office of Extramural Research created a general policy overview on the issue of rigor and transparency that is publicly available (NIH 2015a). This 30-minute narrated slide presentation presents the rationale behind the policy and is a good starting point for anyone trying to gain familiarity with the

issues. NIH has supported grants to develop training modules ([NIH 2014](#)), and once completed and tested, those products will be made available at an NIGMS clearinghouse ([NIH 2017](#)). NIH has published advance notice that funding for training grants and other training activities will require an emphasis on formal instruction in rigor and transparency as early as 2017 (NOT-OD-16-034 in Table [1](#)).

Qualification of Animal Models

As discussed above, in addition to advancing fundamental knowledge in biomedical research, animal models serve as a platform to evaluate the likely benefit and risk of diagnostic tests and therapeutic interventions. Therefore, deliberate efforts to improve the predictive accuracy of preclinical animal models are needed to increase the translatability of the data and consequently increase confidence that developers are selecting the most promising interventions, drugs, and/or biomarkers. It is critical for the scientific community to collaborate in this effort. Qualification of animal models can be performed by developing consensus among multiple stakeholders about key characteristics and endpoints and through greater use of publicly funded animal repositories such as mouse biobanks, which verify and authenticate mouse strains prior to preservation or distribution ([Lloyd et al. 2015](#)). Furthermore, qualified animal models should also be accompanied by thorough and transparent methods that include all the necessary details to ensure that investigators fully understand the possibilities and limitations of each animal model. For example, the National Institute of Allergy and Infectious Diseases (NIAID), in collaboration with the Critical Path to Tuberculosis drug Regimens consortium, held a workshop to evaluate the status of preclinical models for drug development in tuberculosis ([Nuermberger et al. 2016](#)). The workshop participants proposed the development of a battery of qualified animal models that can be used in a modular fashion to evaluate newly developed drugs and therapies. In this roadmap, candidate drugs are evaluated in a variety of preclinical assays and animal models, each of which has been qualified for specific characteristics and/or endpoints, and therefore in aggregate provide confidence in the evidence of the efficacy of any new candidate.

Avoiding Bias

Numerous forms of experimental bias have been identified and described in biomedical research ([Chavalarias and Ioannidis 2010](#); [Sackett 1979](#)), and thus there is no single definition of experimental bias nor a single measure that can be taken to limit its influence. Because biases are unintentional and unconscious, the best way to minimize their impact is to assure that comparison groups are treated equally in the design, conduct, and interpretation of an experiment, with the exception of the actual parameters being compared ([Ransohoff 2005](#)). For example, when the effect of a biological substance is being tested on cells in culture and the medium containing the biological substance needs to be filtered through a 0.2-μM syringe filter, then the control medium should also be filtered in a similar fashion. Otherwise, one could not exclude the possibility that chemicals eluting from disposable plastic syringe and syringe filter account for observed differences been the comparison groups ([Lee et al. 2015](#)). Similarly, if behavioral studies are conducted on animals, the individuals conducting the experiment and/or analyzing the results should be blinded to the identity of the animals to avoid expectation bias ([Bello et al. 2014](#); [Rosenthal and Fode 1963](#)).

There is increasing attention to the need for randomization, allocation concealment, and blinded assessment of outcomes in animal studies as means to reduce bias, which tends to magnify the significance of results. For example, analyses of studies testing drugs have found that lack of bias-reducing measures may overexaggerate the effect size by as much as 30–45% ([Hirst et al. 2014](#)). A meta-analysis of animal studies showed that as many as 70–75% do not report any type of randomization or blinding ([Macleod et al. 2015](#)). Consequently, some researchers have argued that these exaggerated effect sizes in animal studies might be in part responsible for the lack of translatability from animal models into human clinical trials ([Ioannidis 2006](#); [Macleod 2010](#)).

Researchers must, therefore, proactively attempt to mitigate bias. The NIH policy on rigor and transparency formalizes the expectation that grant applications will explain in detail how this will be accomplished in the proposed work and that reviewers will assess the suitability of such plans. A number of resources are available for researchers, journal editors, and reviewers to reduce the risk of bias in their animal studies. For one, the National Centre for the Replacement, Refinement and Reduction of Animals in Research (NC3Rs) developed an Experimental Design Assistant (EDA) ([NC3Rs 2015](#)), which can be used by researchers to generate experimental plans and diagrams that can help address potential bias at critical points in their experiments. The EDA can output suggestions for including randomization, concealment, and blinding, as well as performing power calculations to determine the adequate number of animals for any experiment.

Overemphasis on Statistical Significance

The vast majority of published biomedical research relies on the concept of "statistical significance," generally assigned by p values; however, there is a considerable amount of misunderstanding and misuse of p values ([Button et al. 2013](#); [Colquhoun 2014](#); [Head et al. 2015](#); [Nuzzo 2014](#); [Vaux 2012](#)). For instance, p values only become reliable when statistical power to find a true effect is >90% ([Halsey et al. 2015](#)); when experiments have lower statistical power, the p values themselves can vary widely from one experiment to another. As described above, <2% of animal studies reported performing a power analysis ([Kilkenny et al. 2009](#)) and yet continue to rely on p values, as though these were an absolute index of the truth, automatically assuming that a p value ≤ 0.05 is sufficient evidence that the phenomenon is real ([Halsey et al. 2015](#)).

There have been many calls for the scientific community to adopt better-suited statistical measures, including a recent statement from the American Statistical Association ([Wasserstein and Lazar 2016](#)). Among these measures are calculation of effect sizes and confidence intervals ([Nakagawa and Cuthill 2007](#); [Sullivan and Feinn 2012](#)), which could alleviate the issue of overestimation of the significance of experimental results. Some have questioned the appropriateness of the near-universal use of null hypothesis significance testing ([Gigerenzer 2004](#)). Researchers should consult with biostatisticians during the design of their experiments and become better informed about the statistical methods they apply. As described above, the EDA provides some statistical support for animal studies. As with bias mitigation, the NIH rigor and transparency policy addresses this issue by expecting that applications will include details on the design and analysis, to include power calculations and sample size selection when appropriate. Reviewers are directed to assess the rigor of the design as part of their evaluation criteria.

Reporting Guidelines

Numerous guidelines have been put forward on transparently reporting information; some are generally applicable, while others pertain to specific fields or types of data. An example of a generally applicable reporting guideline is the ARRIVE or Animals in Research: Reporting *In Vivo* Experiments (NC3Rs 2010), which provides a list of elements that should be included when reporting animal studies (Kilkenny et al. 2010). An example of more specific reporting guidelines can be found in the Minimal Information About series, such as Minimal Information for Mouse Phenotyping Procedures or Minimal Information about T-Cell Assays (biosharing 2017). With the large number of guidelines available, there are sites devoted to curation, such as the Equator Network, Enhancing the Quality and Transparency of Health Research (EQUATOR 2017).

These reporting guidelines are generally created by the community in response to differences noted in the literature and use is voluntary. Journals have endorsed the use of reporting guidelines, though adoption has been variable. Active use of the ARRIVE guidelines, even by journals that have endorsed them, has been poor (Baker et al. 2014). Researchers should become familiar with the reporting guidelines that pertain to their area of research, since the use of such guidelines will help build consensus on standards and best practices.

Data Sharing and Transparency

An additional component of transparency beyond reporting all the necessary experimental and analytical details is sharing of the original data. Making data publically available ensures transparency with respect to the analysis and conclusions. However, data sharing offers many additional benefits, including promoting higher quality experimentation and more rigorous results, encouraging collaborations, fostering meta-analysis and repurposing of large scale data sets, reducing redundancy and stagnation, and opening possibilities for researchers that lack access to expensive or unique technologies, among others. The numerous NIH policies on data sharing highlight the importance of these activities (NIH 2016a). With the advent of large-scale technologies, such as genomics (including genome-wide association studies), transcriptomics, proteomics, etc., the NIH has addressed sharing of novel data types and most recently has also instituted a policy for sharing data from clinical trials. Furthermore, in recognition of the importance of data sharing, the NIH has made significant investments into data archives including those housed by the NCBI (NCBI 2017) as well as IC-specific portals (e.g., the NIAID Bioinformatics Resource Centers (NIAID 2016) and Immunology Database and Analysis Portal (NIAID 2017). Most recently, the NIH Big Data Initiative is aimed at the development of a Data Discovery Index as a way to measure the impact and use of public use of primary data that is independent and unrelated to journal publications (bioCADDIE 2017).

A number of grassroots initiatives have propelled the concept of broad sharing of pre- and postpublication data, including unprocessed data. Some examples include the publication of standards and guidelines for Transparency and Openness Promotion (Nosek et al. 2015) that can be used by researchers, journal editors, and reviewers during preparation of research articles. Prepublication release of manuscripts has gained traction in repositories such as bioRxiv (CSHL 2017). Importantly, repositories also serve the important function of housing so-called unpublishable results (i.e., negative, or nonstatistically significant). Recently, researchers responding to public health emergencies such as the Ebola and Zika outbreaks have opted for an open notebook approach to share up-to-date protocols and data analysis that enable more rapid response all around (LabKey 2017; virological.org 2016). Importantly, the NIAID and National Center for Advancing Translational Science have collaborated with the WHO to establish a repository where researchers can share all experiments, including negative results, in Ebola research in order to avoid unnecessary redundancy while developing novel therapeutics (WHO 2017). Adoption of such practices across the board is unlikely unless fundamental changes are made to current parameters used to evaluate the scientific contributions of individual scientists.

Conclusion

Science is a process of exploration leading to knowledge, and higher quality experimentation will lead more efficiently and directly to the acquisition of that knowledge. Scientists can control the methods and the questions asked but not the actual answers. In this era of more complex science, it is especially important to focus on the quality of research. The use of animals demands an ethical focus on rigor and transparency as well. Although the near term may require the use of more animals in a single study, over the long term, fewer animals may be used if researchers are able to reach go/no-go decisions sooner. Science has become more of a community effort, and it is imperative that researchers serve the community by sharing information and developing consensus as well as asking tough questions equally of their own research as that of their peers. The published literature will always include results that may later be corrected as science evolves, and that should not be viewed as a failure of reproducibility but rather a success of the scientific endeavor, as long as rigor and transparency are not in doubt. Therefore, individual scientists participating in community efforts to establish standards or best practices will serve the scientific enterprise in the long run. Emphasis on rigor and transparency is a fundamental aspect of good science that will lead to more informative models, a necessary but not sufficient feature required for translatability.

Acknowledgments

References

Almeida JL, Cole KD, Plant AL. 2016. Standards for cell line authentication and beyond. PLoS Biol 14(6):e1002476.

Almeida JL, Hill CR, Cole KD. 2014. Mouse cell line authentication. Cytotechnology 66(1):133–147.

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.

Baker M. 2016. 1,500 scientists lift the lid on reproducibility. Nature 533(7604):452–454.

Bandrowski AE, Martone ME. 2016. RRIDs: A simple step toward improving reproducibility through rigor and transparency of experimental methods. Neuron 90(3):434–436.

Becker JB, Prendergast BJ, Liang JW. 2016. Female rats are not more variable than male rats: A meta-analysis of neuroscience studies. Biol Sex Differ 7:34.

Begley CG, Ellis LM. 2012. Drug development: Raise standards for preclinical cancer research. Nature 483(7391):531–533.

Bello S, Krogsboll LT, Gruber J, Zhao ZJ, Fischer D, Hrobjartsson A. 2014. Lack of blinding of outcome assessors in animal model experiments implies risk of observer bias. J Clin Epidemiol 67(9):973–983.

Beura LK, Hamilton SE, Bi K, Schenkel JM, Odumade OA, Casey KA, Thompson EA, Fraser KA, Rosato PC, Filali-Mouhim A, Sekaly RP, Jenkins MK, Vezys V, Haining WN, Jameson SC, Masopust D. 2016. Normalizing the environment recapitulates adult human immune traits in laboratory mice. Nature 532(7600):512–516.

BioCADDIE (biomedical and healthCAre Data Discovery Index Ecosystem). 2017. Available online (https://biocaddie.org/), accessed on March 6, 2017.

BioSharing.org. 2017. A curated, informative and educational resource on inter-related data standards, databases, and policies in the life, environmental and biomedical sciences. Available online (https://biosharing.org), accessed on March 6, 2017.

Button KS, Ioannidis JP, Mokrysz C, Nosek BA, Flint J, Robinson ES, Munafo MR. 2013. Power failure: Why small sample size undermines the reliability of neuroscience. Nat Rev Neurosci 14(5):365–376.

Chatterjee R. 2007. Cell biology. Cases of mistaken identity. Science 315(5814):928–931.

Chavalarias D, Ioannidis JP. 2010. Science mapping analysis characterizes 235 biases in biomedical research. J Clin Epidemiol 63(11):1205–1215.

Clayton JA. 2016. Studying both sexes: A guiding principle for biomedicine. FASEB J 30(2):519–524.

Collins FS, Tabak LA. 2014. Policy: NIH plans to enhance reproducibility. Nature 505(7485):612–613.

Colquhoun D. 2014. An investigation of the false discovery rate and the misinterpretation of p-values. R Soc Open Sci 1(3): 140216.

CSHL (Cold Spring Harbor Laboratory). 2017. bioRxiv beta: The Preprint Server for Biology. Available online (https://biocaddie.org), accessed on March 6, 2017.

Didion JP, Buus RJ, Naghashfar Z, Threadgill DW, Morse HC 3rd, de Villena FP. 2014. SNP array profiling of mouse cell lines identifies their strains of origin and reveals cross-contamination and widespread aneuploidy. BMC Genomics 15:847.

EQUATOR Network (Enhancing the QUAlity and Transparency Of health Research Network). 2017. Available online (http://www.equator-network.org/), accessed on March 6, 2017.

Everitt JI. 2015. The future of preclinical animal models in pharmaceutical discovery and development: a need to bring in cerebro to the in vivo discussions. Toxicol Pathol 43(1):70–77.

FELASA (Federation of European Laboratory Animal Science Associations). 2015. Recommendations. Available online (http://www.felasa.eu/recommendations), accessed on March 6, 2017.

Glasziou P, Meats E, Heneghan C, Shepperd S. 2008. What is missing from descriptions of treatment in trials and reviews? BMJ 336(7659):1472–1474.

Gigerenzer G. 2004. Mindless statistics. The Journal of Socio-Economics 33(2004):587–606.

Gordon PH, Moore DH, Miller RG, Florence JM, Verheijde JL, Doorish C, Hilton JF, Spitalny GM, MacArthur RB, Mitsumoto H, Neville HE, Boylan K, Mozaffar T, Belsh JM, Ravits J,

Bedlack RS, Graves MC, McCluskey LF, Barohn RJ, Tandan R. 2007. Efficacy of minocycline in patients with amyotrophic lateral sclerosis: A phase III randomised trial. Lancet Neurol 6(12):1045–1053.

Hackam DG, Redelmeier DA. 2006. Translation of research evidence from animals to humans. JAMA 296(14): 1731–1732.

Halsey LG, Curran-Everett D, Vowler SL, Drummond GB. 2015. The fickle P value generates irreproducible results. Nat Methods 12(3):179–185.

Hartshorne JK, Schachner A. 2012. Tracking replicability as a method of post-publication open evaluation. Front Comput Neurosci 6:8.

Hay M, Thomas DW, Craighead JL, Economides C, Rosenthal J. 2014. Clinical development success rates for investigational drugs. Nat Biotechnol 32(1):40–51.

Head ML, Holman L, Lanfear R, Kahn AT, Jennions MD. 2015. The extent and consequences of p-hacking in science. PLoS Biol 13(3):e1002106.

Hirst JA, Howick J, Aronson JK, Roberts N, Perera R, Koshiaris C, Heneghan C. 2014. The need for randomization in animal trials: An overview of systematic reviews. PLoS One 9(6): e98856.

Howells DW, Macleod MR. 2013. Evidence-based translational medicine. Stroke 44(5):1466–1471.

Hughes P, Marshall D, Reid Y, Parkes H, Gelber C. 2007. The costs of using unauthenticated, over-passaged cell lines: How much more data do we need? Biotechniques 43(5):575, 577–8, 581–2 passim.

Ioannidis JP. 2006. Evolution and translation of research findings: From bench to where? PLoS Clin Trials 1(7):e36.

Itoh Y, Arnold AP. 2015. Are females more variable than males in gene expression? Meta-analysis of microarray datasets. Biol Sex Differ 6:18.

Jaeger RG, Halliday TR. 1998. On confirmatory versus exploratory research. Herpetologica 54(Suppl.):S64–S66.

[No authors listed]. 2014. Journals unite for reproducibility. Nature 515(7525):7.

Kilkenny C, Browne WJ, Cuthill IC, Emerson M, Altman DG. 2010. Improving bioscience research reporting: The ARRIVE guidelines for reporting animal research. PLoS Biol 8(6): e1000412.

Kilkenny C, Parsons N, Kadyszewski E, Festing MF, Cuthill IC, Fry D, Hutton J, Altman DG. 2009. Survey of the quality of experimental design, statistical analysis and reporting of research using animals. PLoS One 4(11):e7824.

Kimmelman J, Mogil JS, Dirnagl U. 2014. Distinguishing between exploratory and confirmatory preclinical research will improve translation. PLoS Biol 12(5):e1001863.

Kriz J, Nguyen MD, Julien JP. 2002. Minocycline slows disease progression in a mouse model of amyotrophic lateral sclerosis. Neurobiol Dis 10(3):268–278.

Labkey. 2017. Zika Open-Research Portal. Available online (https://zika.labkey.com), accessed on March 6, 2017.

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, Gubitz AK, Hesterlee SE, Howells DW, Huguenard J, Kelner K, Koroshetz W, Krainc D, Lazic SE, Levine MS, Macleod MR, McCall JM, Moxley Iii RT, Narasimhan K, Noble LJ, Perrin S, Porter JD, Steward O, Unger E, Utz U, Silberberg SD. 2012. A call for transparent reporting to optimize the predictive value of preclinical research. Nature 490(7419):187–191.

Lee TW, Tumanov S, Villas-Boas SG, Montgomery JM, Birch NP. 2015. Chemicals eluting from disposable plastic syringes and syringe filters alter neurite growth, axogenesis and the microtubule cytoskeleton in cultured hippocampal neurons. J Neurochem 133(1):53–65.

Lloyd K, Franklin C, Lutz C, Magnuson T. 2015. Reproducibility: Use mouse biobanks or lose them. Nature 522(7555):151–153.

Lorsch JR, Collins FS, Lippincott-Schwartz J. 2014. Cell biology. Fixing problems with cell lines. Science 346(6216):1452–1453.

Macleod M. 2010. How to avoid bumping into the translational roadblock. In: Dirnagl U, ed. Rodent Models of Stroke. New York: Humana Press, Springer Science+Business Media. p 7–15.

Macleod MR, Lawson McLean A, Kyriakopoulou A, Serghiou S, de Wilde A, Sherratt N, Hirst T, Hemblade R, Bahor Z, Nunes-Fonseca C, Potluru A, Thomson A, Baginskaite J, Egan K, Vesterinen H, Currie GL, Churilov L, Howells DW, Sena ES. 2015. Risk of bias in reports of in vivo research: A focus for improvement. PLoS Biol 13(10):e1002273.

Macleod MR, O'Collins T, Horky LL, Howells DW, Donnan GA. 2005. Systematic review and metaanalysis of the efficacy of FK506 in experimental stroke. J Cereb Blood Flow Metab 25(6):713–721.

Macleod MR, van der Worp HB, Sena ES, Howells DW, Dirnagl U, Donnan GA. 2008. Evidence for the efficacy of NXY-059 in experimental focal cerebral ischaemia is confounded by study quality. Stroke 39(10):2824–2829.

McCullough LD, Zeng Z, Blizzard KK, Debchoudhury I, Hurn PD. 2005. Ischemic nitric oxide and poly (ADP-ribose) polymerase-1 in cerebral ischemia: Male toxicity, female protection. J Cereb Blood Flow Metab 25(4):502–512.

McNutt M. 2014. Journals unite for reproducibility. Science 346 (6210):679.

MMRRC (Mutant Mouse Resource and Research Centers). 2016. MMRRC reproducibility & rigor. Available online (https://www.mmrrc.org/about/reproducibility.php), accessed on March 6, 2017.

Nakagawa S, Cuthill IC. 2007. Effect size, confidence interval and statistical significance: A practical guide for biologists. Biol Rev Camb Philos Soc 82(4):591–605.

NCBI (National Center for Biotechnology Information). 2012. BioSample. Available online (http://www.ncbi.nlm.nih.gov/biosample/), accessed on March 6, 2017.

NCBI (National Center for Biotechnology Information). 2017. Welcome to NCBI. Available online (https://www.ncbi.nlm.nih.gov/), accessed on March 6, 2017.

NC3Rs (National Centre for the Replacement, Refinement and Reduction of Animals in Research). 2010. ARRIVE Guidelines. Available online (http://www.nc3rs.org.uk/arrive-guidelines), accessed on March 6, 2017.

NC3Rs (National Centre for the Replacement, Refinement and Reduction of Animals in Research). 2015. Experimental Design Assistant version 1.0. Available online (https://eda.nc3rs.org.uk/), accessed on March 6, 2017.

NIAID (National Institute of Allergy and Infectious Diseases. 2016. Bioinformatics Resource Centers. Available online (https://www.niaid.nih.gov/research/bioinformatics-resource-centers), acceed on March 6, 2017.

NIAID (National Institute of Allergy and Infectious Diseases). 2017. IMMPORT: bioinformatics for the future of immunology. Available online (http://www.immport.org/immport-open/public/home/home), accessed on March 6, 2017.

NIH (National Institutes of Health). 2007. Notice regarding authentication of cultured cell lines. Available online (https://grants.nih.gov/grants/guide/notice-files/NOT-OD-08-017.html), accessed on March 6, 2017.

NIH (National Institutes of Health). 2014. Training modules to enhance data reproducibility. Available online (http://grants.nih.gov/grants/guide/rfa-files/RFA-GM-15-006.html), accessed on March 6, 2017.

NIH (National Institutes of Health). 2015a. NIH policy: Rigor and transparency—module 1. Available online (http://grants.nih.gov/reproducibility/module_1/presentation.html), accessed on March 6, 2017.

NIH (National Institutes of Health). 2015b. NIH workshop on reproducibility in cell culture studies. Available online (https://loop.nigms.nih.gov/2015/09/nih-workshop-on-reproducibility-in-cell-culture-studies/), accessed on March 6, 2017.

NIH (National Institutes of Health). 2016a. NIH sharing policies and related guidance on NIH-funded research resources. Available online (https://grants.nih.gov/policy/sharing.htm), accessed on March 6, 2017.

NIH (National Institutes of Health). 2016b. Principles and guidelines for reporting preclinical research. Available online (https://www.nih.gov/research-training/rigor-reproducibility/principles-guidelines-reporting-preclinical-research), accessed on March 6, 2017.

NIH (National Institutes of Health). 2016c. Tools for cell line identification. Available online (http://grants.nih.gov/grants/guide/pa-files/PA-16-186.html), accessed on March 6, 2017.

NIH (National Institutes of Health). 2017. Clearinghouse for training modules to enhance data reproducibility. Available online (https://www.nigms.nih.gov/training/pages/clearinghouse-for-training-modules-to-enhance-data-reproducibility.aspx), accessed on March 6, 2017.

Nosek BA, Alter G, Banks GC, Borsboom D, Bowman SD, Breckler SJ, Buck S, Chambers CD, Chin G, Christensen G, Contestabile M, Dafoe A, Eich E, Freese J, Glennerster R, Goroff D, Green DP, Hesse B, Humphreys M, Ishiyama J, Karlan D, Kraut A, Lupia A, Mabry P, Madon TA, Malhotra N, Mayo-Wilson E, McNutt M, Miguel E, Paluck EL, Simonsohn U, Soderberg C, Spellman BA, Turitto J, VandenBos G, Vazire S, Wagenmakers EJ, Wilson R, Yarkoni T. 2015. Scientific standards. Promoting an open research culture. Science 348(6242):1422–1425.

Nuermberger E, Sizemore C, Romero K, Hanna D. 2016. Toward an evidence-based nonclinical road map for evaluating the efficacy of new tuberculosis (TB) drug regimens: Proceedings of a critical path to TB drug regimens-National Institute of Allergy and Infectious Diseases In Vivo Pharmacology Workshop for TB Drug Development. Antimicrob Agents Chemother 60(3):1177–1182.

Nuzzo R. 2014. Scientific method: Statistical errors. Nature 506 (7487):150–152.

Open Science Collaborative. 2015. Estimating the reproducibility of psychological science. Science 349(6251):943.

Perel P, Roberts I, Sena E, Wheble P, Briscoe C, Sandercock P, Macleod M, Mignini LE, Jayaram P, Khan KS. 2007. Comparison of treatment effects between animal experiments and clinical trials: Systematic review. BMJ 334(7586):197.

Perrin S. 2014. Preclinical research: Make mouse studies work. Nature 507(7493):423–425.

Pouwels KB, Widyakusuma NN, Groenwold RH, Hak E. 2016. Quality of reporting of confounding remained suboptimal after the STROBE guideline. J Clin Epidemiol 69:217–224.

Prendergast BJ, Onishi KG, Zucker I. 2014. Female mice liberated for inclusion in neuroscience and biomedical research. Neurosci Biobehav Rev 40:1–5.

Prinz F, Schlange T, Asadullah K. 2011. Believe it or not: How much can we rely on published data on potential drug targets? Nat Rev Drug Discov 10(9):712.

Ransohoff DF. 2005. Bias as a threat to the validity of cancer molecular-marker research. Nat Rev Cancer 5(2): 142–149.

Rosenthal R, Fode KL. 1963. The effect of experimenter bias on the performance of the albino rat. Syst Res Behav Sci 8(3): 183–189.

Russell W, Burch R. 1959. The Principles of Humane Experimental Technique. London: Methuen.

Sackett DL. 1979. Bias in analytic research. J Chronic Dis 32(1–2): 51–63.

Scott S, Kranz JE, Cole J, Lincecum JM, Thompson K, Kelly N, Bostrom A, Theodoss J, Al-Nakhala BM, Vieira FG, Ramasubbu J, Heywood JA. 2008. Design, power, and interpretation of studies in the standard murine model of ALS. Amyotroph Lateral Scler 9(1):4–15.

Sena ES, Currie GL, McCann SK, Macleod MR, Howells DW. 2014. Systematic reviews and meta-analysis of preclinical studies: why perform them and how to appraise them critically. J Cereb Blood Flow Metab 34(5):737–742.

Sittig LJ, Carbonetto P, Engel KA, Krauss KS, Barrios-Camacho CM, Palmer AA. 2016. Genetic background limits generalizability of genotype-phenotype relationships. Neuron 91(6): 1253–1259.

STAIR. 1999. Recommendations for standards regarding preclinical neuroprotective and restorative drug development. Stroke 30(12):2752–2758.

Sullivan GM, Feinn R. 2012. Using effect size-or why the P value is not enough. J Grad Med Educ 4(3):279–282.

Torgrimson BN, Minson CT. 2005. Sex and gender: What is the difference? J Appl Physiol 99(3):785–787.

Van Den Bosch L, Tilkin P, Lemmens G, Robberecht W. 2002. Minocycline delays disease onset and mortality in a transgenic model of ALS. Neuroreport 13(8):1067–1070.

Vasilevsky NA, Brush MH, Paddock H, Ponting L, Tripathy SJ, Larocca GM, Haendel MA. 2013. On the reproducibility of science: Unique identification of research resources in the biomedical literature. Peer J 1:e148.

Vaux DL. 2012. Research methods: Know when your numbers are significant. Nature 492(7428):180–181.

virological.org. 2016. virological.org—discussion forum for molecular evolution and epidemiology of viruses. Available online (http://virological.org), accessed on March 6, 2017.

Wasserstein RL, Lazar NA. 2016. The ASA's statement on p-values: Context, process, and purpose. Am Stat 70(2):129–133.

WHO (World Health Organization). 2017. Ebola drug test database. Available online (http://www.who.int/medicines/ebola-treatment/test-database/en/), accessed on March 6, 2017.

Yu M, Selvaraj SK, Liang-Chu MM, Aghajani S, Busse M, Yuan J, Lee G, Peale F, Klijn C, Bourgon R, Kaminker JS, Neve RM. 2015. A resource for cell line authentication, annotation and quality control. Nature 520(7547):307–311.

Zhu S, Stavrovskaya IG, Drozda M, Kim BY, Ona V, Li M, Sarang S, Liu AS, Hartley DM, Wu DC, Gullans S, Ferrante RJ, Przedborski S, Kristal BS, Friedlander RM. 2002. Minocycline inhibits cytochrome c release and delays progression of amyotrophic lateral sclerosis in mice. Nature 417(6884):74–78.