Published in final edited form as:

Oral Dis. 2017 October ; 23(7): 813–816. doi:10.1111/odi.12588.

Reproducibility and transparency in biomedical sciences

Gallya Gannot^{#*}, Mary A. Cutting[#], Dena J. Fischer[#], and Laura J. Hsu[#]

Center for Clinical Research, National Institute of Dental and Craniofacial Research, National Institutes of Health, Bethesda, MD, 6701 Democracy Blvd, Bethesda, MD, 20892-4878, Office phone: 301-451-5096 Mobile: 301-648-8224

[#] These authors contributed equally to this work.

The biomedical research sciences are currently facing a challenge highlighted in several recent publications: concerns about the rigor and reproducibility of studies published in the scientific literature.

Research progress is strongly dependent on published work. Basic science researchers build on their own prior work and the published findings of other researchers. This work becomes the foundation for preclinical and clinical research aimed at developing innovative new diagnostic tools and disease therapies. At each of the stages of research, scientific rigor and reproducibility are critical, and the financial and ethical stakes rise as drug development research moves through these stages.

Recent reports from leading pharmaceutical companies indicate that industry scientists are cautious about accepting published results from basic science studies. A report on in-house target validation from one company described their inability to reproduce published data in several research fields; they indicated that their in-house results were consistent with published results for only 20-25% of 67 target validation projects that were analyzed (Prinz et al., 2011). Similarly, when scientists from a biotechnology company tried to confirm oncology-related findings from publications that they identified as ‘landmark’ papers, they found that only 11% of these studies (6 of the 53 assessed) had scientifically reproducible data, although all studies had been cited numerous times (Begley and Ellis, 2012). The authors of these reports suggested several possible factors contributing to the lack of reproducibility, including poor experimental design, inappropriate statistical analyses, inadequate sample sizes, poor control of experimental conditions, lack of repetition of experiments, insufficient reporting on materials and methods, and failure to report negative results (Begley and Ellis, 2012, Prinz et al., 2011). A recent publication reported that the success rates for phase II trials of new drug candidates in projects from 16 leading pharmaceutical companies that represent more than 60% of global research and development spending decreased from 28% in 2006-2007 to 18% in 2008-2009. Insufficient efficacy was the most frequently cited reason for failure (Arrowsmith, 2011). Based on their own concerns with reproducibility described above, Prinz et al. (2011) suggest that the efficacy concerns in these trials may be due to the scientific limitations noted above and problems with reproducibility in the preclinical work that preceded the clinical trials.

* Corresponding author; GALLYA.GANNOT@NIH.GOV.

Pharmaceutical companies are not alone in reporting difficulty in reproducing published data. A 2015 editorial in the journal *GigaScience* reported that reproduction of results in a computational biology paper required at least 280 hours of work (Kenall et al., 2015), even when using the same data sets. Further, a *Journal of Cell Biology* editorial from 2015 highlighted a study that assessed agreement among more than 200 publications addressing an important research question of determining the cellular source of matrix-degrading proteases in four types of human tumors. The study results indicated major discrepancies between the conclusions of the 200 studies, even among those that appeared to use appropriate controls (Madsen and Bugge, 2015).

Animal model research is another field with reported poor reproducibility. Animal models are an important tool to evaluate basic research findings in a complex living organism. However, recent publications have reported that many model-organism studies were not reproducible (Steward et al., 2012, van der Worp and Macleod, 2011). At a workshop convened by the National Institute of Neurological Disorders and Stroke, stakeholders suggested that the problems with reproducibility of animal model studies may be due to inadequate study design and incomplete descriptions of the experimental design in publications. The workshop participants expressed the concern that these deficiencies could result in the scientific community being unable to identify problems in the experimental design and data analysis, limiting the potential to derive benefit from the findings (Landis et al., 2012). As an example of this, a review that examined 100 animal-model papers published in *Cancer Research* found that randomization was reported for only 28% of the studies for which it would have been feasible, and only 2 papers indicated that the examiners were blinded to the treatment groups. Moreover, none of the examined papers described the methods used to determine the number of animals per treatment group (Hess, 2011).

Human subjects research and specifically clinical trials are regulated by various laws and policies that require rigorous study designs and independent oversight. Beginning in July 2005, the International Committee of Medical Journal Editors (ICMJE) made prospective registration of clinical trials a requirement for publishing results in member journals (Dechartres et al., 2016). The US Food and Drug Administration (FDA) Modernization Act of 1997 and the Food and Drug Administration Amendments Act of 2007 required “applicable clinical trials” of FDA-regulated drugs, biologics, or devices to register at [ClinicalTrials.gov](https://clinicaltrials.gov). This was intended to benefit patients, clinicians, and the research community by ensuring that information on study design and study results would be publicly available (<https://clinicaltrials.gov/ct2/manage-recs/fdaaa>). However, a study published in 2009 found that nearly 28% of a sample of randomized controlled trials published in general medical journals and specialty journals with the highest impact factors were not registered. Approximately 14% of the trials included in the sample registered only after completion of the study, and another 13% were registered, but did not have a clear description of the primary outcome (Mathieu et al., 2009). A 2016 study assessed 67 Cochrane review meta-analyses to determine if there was a difference in treatment effect estimates between prospectively and retrospectively registered/unregistered trials. The results suggested that retrospectively-registered or unregistered trials had larger treatment effect estimates than trials registered prospectively (Dechartres et al., 2016). Another systematic analysis was performed in 2015 on 74 oncology-related randomized studies for which both study

protocols and study results were published in leading journals. Approximately 38% of the papers reported endpoints that were not planned in the protocol, and 80% of these were not identified in the publication as unplanned endpoints. The abstracts for these papers reported positive unplanned end points ($P=0.002$) and unplanned analyses ($P=0.007$) more frequently than negative outcomes. The authors suggested that, because statistically significant results are more likely to be published, clinical investigators may tend to suppress the non-significant outcomes and inappropriately emphasize only the significant outcomes from their studies (Raghav et al., 2015).

Concerns with reproducibility and transparency exist across the scientific research spectrum, from the basic sciences, to preclinical research, to clinical studies and trials. Journal editors and others commenting on this issue have indicated that the lack of scientific rigor and reproducibility is not likely to be the result of scientific misconduct. Rather, it has been suggested that investigators make errors in designing and performing their research, in selecting statistical tests, and in reporting results (Yamada and Hall, 2015, Steward, 2016). Another contributing factor may be the tendency of scientists to be influenced by their own biases. According to a recent article in *Nature*, this “self-deception” may include searching only for evidence that supports rather than refutes a hypothesis, assuming that a random pattern in data represents an interesting finding, and only checking unexpected data rather than all data (Nuzzo, 2015). As noted by Collins and Tabak many of the problems with reproducibility of studies may be explained by simple factors: “different animal strains, different lab environments or subtle changes in protocol.” Even when investigators repeating an experiment obtain the same animal strain from the same vendor, and carefully maintain the experimental animals under the same conditions as used in previous studies, there may be hidden variables that can confound study results (Collins and Tabak, 2014). In a recent article, Servick described an investigator’s discovery that variations in the gut microbiomes of mice might account for her inability to replicate her initial results in a study testing the effect of a drug on bone density (Servick, 2016). Regardless of the reasons, the lack of reproducibility in the biomedical sciences needs to be addressed, especially since faulty conclusions impede the advancement of knowledge and may lead to further faulty conclusions.

Addressing Rigor and Reproducibility Issues

Approaches from the research community

To help address current weaknesses in published literature, Moher and Altman suggested the concept of publications officers at universities and research centers, who would guide researchers through the process of preparing manuscripts, and the development of core competencies and training for editors, peer reviewers and authors (Moher and Altman, 2015). In line with this recommendation, the Transparency and Openness Promotion (TOP) Committee, made up of leaders from various disciplines, journal editors, and funding agency representatives, developed and published a set of standards to improve journal policies for acceptance of papers.

The TOP Committee offered journals a total of eight standards for publications that include citation standards, data transparency, analytical methods transparency, research materials

transparency, design and analysis transparency, preregistration of studies, preregistration of analysis plans, and replication (Nosek et al., 2015).

Publications in several journals have begun to address specific issues in statistical analysis to minimize bias in research and publications. For example, an editorial by Motulsky recommended that investigators refrain from reanalyzing data, express the actual size of an observed effect rather than emphasizing only the P values, and rely less on standard errors (Motulsky, 2015).

Other journals have recently offered tutorials for the research community in basic principles of preclinical research, including defining the research question, considering aspects of experimental design to minimize bias, completing pre-experimental power calculations, defining starting points and end points, using random assignment to groups and blinding, establishing a statistical analysis plan, observing best practices and transparency regarding decisions on data inclusion and exclusion, and maintaining rigorous standards for recording and storing data (Steward and Balice-Gordon, 2014, Groves, 2010). With the goal of improving reproducibility in laboratory research by improving study design, study conduct, and data analysis, the Statistics Group of the United Kingdom National Institute for Health Research published a framework to support early-stage discussions between scientists and statisticians (Masca et al., 2015).

Approaches from the National Institutes of Health

In January 2014, Drs. Francis Collins and Lawrence Tabak published the proposed plan of the National Institutes of Health (NIH) to enhance reproducibility of biomedical research. The commentary highlighted preclinical research, especially work using animal models, as an area to address, since it is particularly susceptible to reproducibility issues. Drs. Collins and Tabak outlined several proposals to address the problem, and requested involvement of the broader scientific community in this initiative (Collins and Tabak, 2014).

In June 2014, NIH held a joint workshop with the Nature Publishing Group and Science on the issue of reproducibility and rigor of research findings, with journal editors representing more than 30 basic/preclinical science journals. The journal editors came to consensus on a set of principles in the scientific publishing arena to support research that is reproducible, robust, and transparent (<https://www.nih.gov/research-training/rigor-reproducibility/principles-guidelines-reporting-preclinical-research>). There was general agreement on the importance of providing detailed descriptions of methods, to permit transparency and replication by other laboratories (Yamada and Hall, 2015, Kenall et al., 2015, Loew et al., 2015, Kuhar, 2014, Waller and Miller, 2016). Some journals developed a policy to accept refutations after a thorough peer review (Daugherty et al., 2016, Bandrowski and Martone, 2016). Suggestions were also made to improve reproducibility by introducing resource identifiers to standardize research materials, such as antibodies and cell lines (Bandrowski and Martone, 2016). Other journals introduced checklists for authors to use before submitting a paper for review (Kenall et al., 2015).

In October 2015, the NIH published notices informing the biomedical research community of updates to grant application instructions and review criteria to enhance reproducibility of research findings through increased scientific rigor and transparency (see: <http://grants.nih.gov/grants/guide/notice-files/NOT-OD-16-011.html> and <http://grants.nih.gov/grants/guide/notice-files/NOT-OD-16-012.html>). These updates focus on four areas deemed important for enhancing rigor and transparency: 1) the scientific premise forming the basis of the proposed research, 2) rigorous experimental design for robust and unbiased results, 3) consideration of relevant biological variables, and 4) authentication of key biological and/or chemical resources. The basic principles of rigor and transparency and the four areas of focus apply to the full spectrum of research, from basic to clinical. Investigators will need to consider how all four areas apply to their proposed research. Likewise, reviewers will assess whether these areas have been appropriately addressed by the applicant through revised language defining the peer review criteria. Expectations for grant applications include the following:

Scientific premise: Discussion of the strengths and weaknesses of prior research crucial to support the application, including attention to rigor of published research, methodology analysis and interpretation, relevant biological variables, authentication of key resources.

Scientific rigor: Strict application of scientific method to ensure robust and unbiased experimental design, methodology, analysis, interpretation and reporting of results. Applicants are expected to describe experimental design and methods proposed and how they will achieve robust and unbiased results, as appropriate for the work proposed.

Consideration of relevant biological variables such as sex: Explanation of how relevant biological variables, such as sex, are factored into research designs, analyses, and reporting in vertebrate animal and human studies. Applicants proposing to use only one sex are expected to provide strong justification from the scientific literature or preliminary data to support this decision.

Authentication of key biological and/or chemical resources: Key biological or chemical resources are resources that may differ from laboratory to laboratory, over time; have qualities or qualifications that could influence results; are integral to the proposed research. Examples include cell lines, specialty chemicals, antibodies, and other biologics. Applicants are expected to briefly describe methods to ensure the identity and validity of key biological and/or chemical resources used in the proposed studies.

Scientific rigor and transparency in the conduct of clinical trials was specifically addressed by the release in September 2016 of a Department of Health and Human Services (HHS) final rule on clinical trials registration and results information submission, by an accompanying NIH policy on the Dissemination of NIH-Funded Clinical Trial Information, and by additional NIH policies related to clinical trials. The HHS regulation and the complementary NIH policy on trial registration and results reporting include financial consequences for failure to comply; it is expected that these provisions will improve the rates of trial registration and data sharing. Another new NIH policy will require that all NIH-

funded investigators and staff complete training in Good Clinical Practice (GCP). In addition, an NIH policy will require that all grant applications that include clinical trials be submitted only in response to a Funding Opportunity Announcement that is specifically designed for clinical trials; this is intended to ensure that key data about a proposed trial will be included in the application and be evaluated by review groups that include individuals with clinical trial expertise. These policies and other initiatives are summarized in a recent publication authored by NIH leaders (Hudson et al., 2016), and the following notices

<http://grants.nih.gov/grants/guide/notice-files/NOT-OD-16-147.html>;

<http://grants.nih.gov/grants/guide/notice-files/NOT-OD-16-148.html>;

<http://grants.nih.gov/grants/guide/notice-files/NOT-OD-16-149.html>

In an effort to enhance scientific reproducibility, the National Institute of Dental and Craniofacial Research (NIDCR) has included elements of scientific rigor in instructions to grant applicants and reviewers. One example of many is the NIDCR Funding Opportunity Announcement titled, “Tailoring dental treatment for individuals with systemic diseases that compromise oral health (R01),” which aims to stimulate research addressing gaps in our knowledge of how best to treat oral diseases of patients with systemic diseases or conditions known to compromise oral health, to identify factors predictive of treatment outcomes within patient groups, and to generate evidence for more precise dental treatment guidelines tailored to patient needs. To ensure scientific rigor, the instructions request that the applicant provide evidence from past research that the systemic disease or condition compromises oral health, and evidence that the population proposed for study is representative of the medical condition/disease and is of sufficient size to obtain clinically meaningful oral health outcomes, that the application include descriptions of the selection criteria for study inclusion and the method of measuring oral health outcomes, and that the application contain the analysis plan.

Since 2008, NIDCR has required that investigators submit applications in response to clinical trial-specific Funding Opportunity Announcements for the planning (R34) and implementation (U01) of such studies. The NIDCR Policy for Data and Safety Monitoring of Clinical Research describes NIDCR’s system for appropriate oversight and monitoring of the conduct of NIDCR-supported clinical research, which is intended to ensure the safety of participants, the validity and integrity of the data, the appropriate conduct of the study, and the availability of data in a timely manner. The [NIDCR website](#) also contains many tools that were developed to help investigators design and implement rigorous clinical studies.

Conclusion

The recognition that rigor and reproducibility standards for research need to improve is an important first step toward better quality science. Adhering to the recommendations of experts from academic communities, industry, and funding agencies should enhance reproducibility in future scientific studies. A recent editorial by a National Library of Medicine scientist advocated a strong post-publication culture as another means to improve research quality. The author suggested that developing an encouraging climate for

communicating errors and weaknesses is critical, and that the ability of the scientific community to ask questions and conduct discussions about publications is essential to assess the research and obtain information for other studies (Bastian, 2014). Reproducibility and transparency in science can be achieved only by working as a community of multidisciplinary scientists, reviewers, editors and readers.

References

- Arrowsmith J. Trial watch: Phase II failures: 2008-2010. *Nat Rev Drug Discov*. 2011; 10:328–9.
- Bandrowski AE, Martone ME. RRIDs: A Simple Step toward Improving Reproducibility through Rigor and Transparency of Experimental Methods. *Neuron*. 2016; 90:434–6. [PubMed: 27151636]
- Bastian H. A stronger post-publication culture is needed for better science. *PLoS Med*. 2014; 11:e1001772. [PubMed: 25548904]
- Begley CG, Ellis LM. Drug development: Raise standards for preclinical cancer research. *Nature*. 2012; 483:531–3. [PubMed: 22460880]
- Collins FS, Tabak LA. Policy: NIH plans to enhance reproducibility. *Nature*. 2014; 505:612–3. [PubMed: 24482835]
- Daugherty A, Hegele RA, Mackman N, Rader DJ, Schmidt AM, Weber C. Complying With the National Institutes of Health Guidelines and Principles for Rigor and Reproducibility: Refutations. *Arterioscler Thromb Vasc Biol*. 2016; 36:1303–4. [PubMed: 27335467]
- Dechartres A, Ravaud P, Atal I, Riveros C, Boutron I. Association between trial registration and treatment effect estimates: a meta-epidemiological study. *BMC Med*. 2016; 14:100. [PubMed: 27377062]
- Groves T. What makes a high quality clinical research paper? *Oral Dis*. 2010; 16:313–5. [PubMed: 20233312]
- Hess KR. Statistical design considerations in animal studies published recently in cancer research. *Cancer Res*. 2011; 71:625. [PubMed: 21239476]
- Hudson KL, Lauer MS, Collins FS. Toward a New Era of Trust and Transparency in Clinical Trials. *JAMA*. 2016
- Kenall A, Edmunds S, Goodman L, Bal L, Flintoft L, Shanahan DR, Shipley T. Better reporting for better research: a checklist for reproducibility. *Gigascience*. 2015; 4:32. [PubMed: 26207176]
- Kuhar MJ. Letter from the Editor-in-Chief: Irreproducible Results. *J Drug Alcohol Res*. 2014; 3
- Landis SC, Amara SG, Asadullah K, Austin CP, Blumenstein R, Bradley EW, Crystal RG, Darnell RB, Ferrante RJ, Fillit H, Finkelstein R, Fisher M, Gendelman HE, Golub RM, Goudreau JL, Gross RA, Gubitza AK, Hesterlee SE, Howells DW, Huguenard J, Kelner K, Koroshetz W, Krainc D, Lazic SE, Levine MS, Macleod MR, McCall JM, Moxley RT 3rd, Narasimhan K, Noble LJ, Perrin S, Porter JD, Steward O, Unger E, Utz U, Silberberg SD. A call for transparent reporting to optimize the predictive value of preclinical research. *Nature*. 2012; 490:187–91. [PubMed: 23060188]
- Loew L, Beckett D, Egelman EH, Scarlata S. Reproducibility of research in biophysics. *Biophys J*. 2015; 108:E1. [PubMed: 25863074]
- Madsen DH, Bugge TH. The source of matrix-degrading enzymes in human cancer: Problems of research reproducibility and possible solutions. *J Cell Biol*. 2015; 209:195–8. [PubMed: 25918222]
- Masca NG, Hensor EM, Cornelius VR, Buffa FM, Marriott HM, Eales JM, Messenger MP, Anderson AE, Boot C, Bunce C, Goldin RD, Harris J, Hinchliffe RF, Junaid H, Kingston S, Martin-Ruiz C, Nelson CP, Peacock J, Seed PT, Shinkins B, Staples KJ, Toombs J, Wright AK, Teare MD. RIPOSTE: a framework for improving the design and analysis of laboratory-based research. *Elife*. 2015; 4
- Mathieu S, Boutron I, Moher D, Altman DG, Ravaud P. Comparison of registered and published primary outcomes in randomized controlled trials. *JAMA*. 2009; 302:977–84. [PubMed: 19724045]

- Moher D, Altman DG. Four Proposals to Help Improve the Medical Research Literature. *PLoS Med*. 2015; 12:e1001864. [PubMed: 26393914]
- Motulsky HJ. Common misconceptions about data analysis and statistics. *Pharmacol Res Perspect*. 2015; 3:e00093. [PubMed: 25692012]
- 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. SCIENTIFIC STANDARDS. Promoting an open research culture. *Science*. 2015; 348:1422–5. [PubMed: 26113702]
- Nuzzo R. How scientists fool themselves - and how they can stop. *Nature*. 2015; 526:182–5. [PubMed: 26450039]
- Prinz F, Schlange T, Asadullah K. Believe it or not: how much can we rely on published data on potential drug targets? *Nat Rev Drug Discov*. 2011; 10:712. [PubMed: 21892149]
- Raghav KP, Mahajan S, Yao JC, Hobbs BP, Berry DA, Pentz RD, Tam A, Hong WK, Ellis LM, Abbruzzese J, Overman MJ. From Protocols to Publications: A Study in Selective Reporting of Outcomes in Randomized Trials in Oncology. *J Clin Oncol*. 2015; 33:3583–90. [PubMed: 26304898]
- Servick K. Of mice and microbes. *Science*. 2016; 353:741–3. [PubMed: 27540148]
- Steward O. A Rhumba of "R26's": Replication, Reproducibility, Rigor, Robustness: What Does a Failure to Replicate Mean? *eNeuro*. 2016; 3
- Steward O, Balice-Gordon R. Rigor or mortis: best practices for preclinical research in neuroscience. *Neuron*. 2014; 84:572–81. [PubMed: 25442936]
- Steward O, Popovich PG, Dietrich WD, Kleitman N. Replication and reproducibility in spinal cord injury research. *Exp Neurol*. 2012; 233:597–605. [PubMed: 22078756]
- van der Worp HB, Macleod MR. Preclinical studies of human disease: time to take methodological quality seriously. *J Mol Cell Cardiol*. 2011; 51:449–50. [PubMed: 21549125]
- Waller LA, Miller GW. More than Manuscripts: Reproducibility, Rigor, and Research Productivity in the Big Data Era. *Toxicol Sci*. 2016; 149:275–6. [PubMed: 26811418]
- Yamada KM, Hall A. Reproducibility and cell biology. *J Cell Biol*. 2015; 209:191–3. [PubMed: 25918221]