

Education Brief Report

Cite this article: Kalichman M and Mills PJ.
(2020) Workshop on reproducibility in research.
Journal of Clinical and Translational Science 4:
562–564. doi: [10.1017/cts.2020.496](https://doi.org/10.1017/cts.2020.496)

Received: 17 April 2020
Revised: 13 May 2020
Accepted: 9 June 2020
First published online: 19 June 2020

Keywords:

Reproducibility; education; research ethics;
responsible conduct of research; good
practices of research

Address for correspondence:

M. Kalichman, PhD, Research Ethics Program,
University of California San Diego, 9500 Gilman
Dr. M/C 0612, La Jolla, CA 92093-0612, USA.
Email: mkalichman@ucsd.edu

© The Association for Clinical and Translational
Science 2020. This is an Open Access article,
distributed under the terms of the Creative
Commons Attribution-NonCommercial-
ShareAlike licence (<http://creativecommons.org/licenses/by-nc-sa/4.0/>), which permits non-
commercial re-use, distribution, and
reproduction in any medium, provided the
same Creative Commons licence is included
and the original work is properly cited. The
written permission of Cambridge University
Press must be obtained for commercial re-use.



Workshop on reproducibility in research

Michael Kalichman¹  and Paul J. Mills²

¹Research Ethics Program, University of California San Diego, La Jolla, CA, USA and ²Family Medicine and Public Health, University of California San Diego, La Jolla, CA, USA

Abstract

In recent years, concern about research reproducibility has increased dramatically for scientists, funders of research, and the general public. With a view to explicitly address what is often called a reproducibility crisis and putting the focus on research being done by individual trainees, a two-hour workshop was developed and introduced into six courses at UC San Diego. Participation in the workshop resulted in a statistically significant increase in the number of different types of strategies identified by the trainees for fostering reproducibility. The findings are consistent with having increased awareness of strategies to promote reproducibility.

Introduction

In recent years, the research and academic communities have been increasingly aware of and concerned about the apparent lack of reproducibility of scientific research, including published findings [1–4]. While the perceived crisis may in part be a product of increased attention to work that is not reproducible [5], it is plausible that the majority of the irreproducibility is simply due to a failure to pursue good practices in the design, conduct, and reporting of research. Many genuine methodological reasons have been documented [3], including a well-publicized rash of poor quality antibodies and reagents [6,7]. These issues can theoretically be addressed in mentoring, or in courses with a focus on methods, ethics, or good practices of research. However, the premise of this project is that these ad hoc efforts have been insufficient and that it might be beneficial to design an approach that more explicitly addresses the reproducibility crisis and puts the focus on the research being done by the trainees.

The proposed plan is to develop a curriculum for a 2-hour module on rigor and reproducibility, pilot test it as part of an existing course for an audience of graduate students and/or post-doctoral researchers, and use a pre-/post-test design to assess impact of the module on trainees' understanding of scientific reproducibility. A description, instructor's guide, and resources for the resulting module will then be shared widely for use and modification by others.

The two underlying goals set for the proposed intervention are that trainees would:

1. Better appreciate the wide range of ways in which reproducibility of research might be compromised.
2. Be able to identify a number of practices that might protect the reproducibility of their own research.

Methods

This project was reviewed and approved as exempt by the UC San Diego Institutional Review Board (#190336XX). Based on a review of the literature as well as past workshops run by the authors, a curriculum and materials were developed for a 2-hour workshop on the topic of rigor and reproducibility. The goal was to produce a program with the following features: (1) highly interactive; (2) pre- and post-testings to monitor impact; (3) independent of the particular area of research for the participants; (4) emphasis on the many different ways in which diverse stakeholders could easily improve the reproducibility of research; and (5) portability such that others could readily implement the workshop.

The project began with a pilot version for participants in a seminar for an NIH training grant, followed by a test phase in six sections of a course titled “Scientific Ethics.” Participants in the courses were largely graduate students, but also included some postdoctoral researchers. The disciplines most heavily represented were biological sciences, biomedical sciences, and engineering. All or nearly all participants were in these courses because of requirements for responsible conduct of research training set by their graduate programs, National Institutes of Health (NIH), and/or National Science Foundation (NSF).

Pilot Workshop

In advance of the 2-hour pilot workshop, prospective participants were asked to identify recent experimental papers in their disciplines. Two of the five proposed manuscripts were selected by

the instructors and, just before the formal start of the workshop, copies of one was distributed to half of the participants and the other to the remaining participants. All individuals were given a worksheet with the following prompt regarding the manuscript they received: "Please note below distinct issues you see that would increase the risk that the published findings would not be reproducible." The exercise was repeated at the end of the workshop, but after switching the manuscripts between the participants. The workshop itself consisted of an opening explanation of the plans and an introductory lecture on the topic of reproducibility, a brainstorming discussion to identify the stakeholders who have roles in whether research will be reproducible, a small groups' assignment to identify strategies and approaches that might increase research reproducibility, and a plenary discussion of what each group had identified.

Test Workshops

Based on experience with the pilot workshop, a 2-hour test workshop was then repeated in each of six sections of the UC San Diego Scientific Ethics course in the Spring of 2019. Participants received the pre- and post-worksheets with the same prompt as the pilot, but were asked to consider the question in general rather than for a specific experimental paper. It became clear from the Pilot Workshop that choosing one or two papers in advance of a workshop could severely narrow the range of issues identified and would be problematic with an audience representing diverse research disciplines. The test workshop agenda was otherwise largely the same as that used previously.

Scoring

Following completion of the six workshops, the co-authors (MK and PM) each scored the pre- and post-worksheets for half of the workshops (three each). Scoring consisted of assigning each item listed by the participant into one of the nine categories of factors that might be important to the reproducibility of research (Table 1). Two dependent variables were defined. First, the primary endpoint was number of categories represented by the items identified by the participants. The hypothesis was that post-workshop participants would identify a wider variety of categories of factors that might contribute to reproducibility, thus an increase in the number of categories identified in the post-workshop worksheet compared to pre-workshop. A secondary endpoint was the total number of items listed by the participants. The hypothesis was that participants would list more items in the post-workshop worksheet compared to pre-workshop. The hypotheses were tested by conducting paired sample t-tests for the pre- and post-workshop categories and factors, respectively, for each of the six workshops.

Materials from the workshops, including an Instructor's Guide, will be linked from the UC San Diego Altman Clinical and Translational Research Institute (ACTRI) Website at <http://actri.ucsd.edu> and posted on the UC San Diego Research Ethics Program Website at <https://ethics.ucsd.edu/resources/instructor-resources>. These include

1. Instructor's guide
2. PowerPoint slides for 10-minute introductory presentation
3. Video of introductory presentation
4. Schedule: Overview of workshop with notes
5. Worksheet: For participants to list factors relevant to reproducibility of their research

Table 1. Categories of factors that might have an impact on the reproducibility of research

- | |
|---|
| <ol style="list-style-type: none"> 1. Incentives (e.g., funding, promotion) 2. Training (e.g., scientific methods, statistics) 3. Research tools (e.g., standardization, validation) 4. Experimental design (e.g., blinding, controls) 5. Data management (e.g., record keeping, sharing) 6. Reporting (e.g., sample size, management of outliers, inclusion/exclusion criteria) 7. Openness (e.g., study pre-registration, sharing of methods, materials, and data) 8. Peer review (e.g., items expected for reporting, ongoing open peer review) 9. Other (e.g., expectation of reproducing the study prior to publication, handling of misconduct allegations, handling of retractions, options to publish negative findings) |
|---|

6. Scoring rubric: For participants and/or instructor to assign factors to different categories relevant to the reproducibility of research

7. Resources: List of resources for more information on reproducibility

Results

Results for each of the six workshops are summarized in Table 2. In each case, the average number of categories identified by the participants significantly increased from baseline and the total number of items identified also increased.

Discussion

Based on the pre-/post-exercise, the impact of these workshops was substantial (Table 2). Over six different workshops, participants increased the number of categories of reproducibility improving interventions by no less than 35% and up to 105%. Similarly, the number of items identified increased by 16–77%. These findings are consistent with the argument that this format of workshop at least acutely increases conscious awareness of the wide variety of ways in which the rigor and therefore reproducibility of research can be increased.

Four limitations of this project are worth noting. First, it is plausible that participants did not learn something new, but simply were prompted to think about additional elements post-workshop because they had just been discussed. This leaves open the possibility that participants will similarly fail to remember and think more broadly about these issues later. On the other hand, it is fair to say that at least acutely participants were able to see a wider range of ways that reproducibility might be improved. To the extent that any such intervention might have an impact, this is certainly a desirable outcome.

A second limitation is related to the highly interactive nature of this style of workshop. Thus, different instructors might have different levels of success. Anecdotally, we know of at least one instance in which a faculty member who sat through this workshop went on to try it successfully with her own class. This is consistent with the portability of the approach, but as with any possible intervention, results are likely to vary with instructors.

The third limitation noted is that assessing outcomes in this study is to some degree subjective. For example, one person might consider a reference to a student's flawed experimental design as

Table 2. Results for average number of categories and items identified before and after each of the six test workshops

Workshop	#	Categ. (pre)	Categ. (post)	% Δ	Items (pre)	Items (post)	% Δ	Scored by
1	25	3.0	4.8	64*	6.2	9.4	52*	PM
2	25	2.9	6.0	105*	6.7	11.9	77*	MK
3	17	3.2	5.5	72*	6.4	10.7	69*	PM
4	14	3.8	5.3	40*	8.1	11.1	37*	MK
5	23	3.7	4.9	35*	7.3	10.2	40*	PM
6	20	4.0	5.6	42*	8.5	9.9	16 [^]	MK

* $P < 0.05$, [^]n.s.

Columns represent workshop number, number (#) of participants, average number of categories (Categ.) identified before (pre) and after (post) each workshop, % increase (% Δ) in categories identified, average number of items identified before (pre) and after (post) each workshop, % increase (% Δ) in items identified, and scoring of categories and items by PM or MK. All changes except the number of items identified in workshop 6 were statistically significant ($P < 0.05$).

evidence falling in the category of training or mentoring, while someone else would assign the item to the category of experimental design. For this reason, we thought it useful to see if the broad strokes of the findings for this study held up despite scoring of different classes by different people. As noted above, although three of the workshops were scored by one of us (PM) and the other three by the other (MK), the pattern of findings was for all practical purposes the same. This is despite the fact that there were examples of differences of opinion about how individual items should be scored.

The fourth and final limitation to be noted is that this module was tested in audiences primarily limited to graduate students in biology, biomedical sciences, physical sciences, and engineering. However, the audiences also included at least some postdoctoral researchers, faculty, and staff, and the disciplines represented have occasionally extended to social sciences. By emphasizing principles of student-centered learning, it is hoped that the approach can translate to most if not all experimental disciplines.

The National Institutes of Health continues to emphasize the importance of properly training the next generation of scientists to improve rigor and reproducibility [8]. In a 2018 anonymous national survey to trainees, a significant percentage of trainees expressed that they had not received optimal mentoring in research practice, integrity, and how pressures to publish influence reporting practices [9]. This brief workshop approach to help trainees in improving the rigor and reproducibility of scientific research is promising and remains needed.

Acknowledgments. This research was partially supported by the National Institutes of Health Grant UL1TR001442 (MK, PJM) and the National

Science Foundation Grant 1835029 (MK). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH or NSF.

Disclosures. The authors have no conflicts of interest to declare.

References

1. Ioannidis JPA. Why most published research findings are false. *PLOS Med* 2005; 2: e124. doi: [10.1371/journal.pmed.0020124](https://doi.org/10.1371/journal.pmed.0020124)
2. Begley CG, Ellis LM. Drug development: raise standards for preclinical cancer research. *Nature* 2012; 483: 531–533. <http://www.nature.com/nature/journal/v483/n7391/full/483531a.html>
3. Freedman LP, Cockburn IM, Simcoe TS. The economics of reproducibility in preclinical research. *PLOS Biol* 2015; 13: e1002165. doi: [10.1371/journal.pbio.1002165](https://doi.org/10.1371/journal.pbio.1002165)
4. Baker M. Is there a reproducibility crisis? *Nature* 2016; 533: 452–454. <https://www.nature.com/news/1-500-scientists-lift-the-lid-on-reproducibility-1.19970>
5. Fanelli D. Is science really facing a reproducibility crisis, and do we need it to? *Proc Natl Acad Sci* 2018; 115: 2628–2631. doi: [10.1073/pnas.1708272114](https://doi.org/10.1073/pnas.1708272114)
6. Voskuil JL. The challenges with the validation of research antibodies. *F1000Res*. 2017; 6: 161. doi: [10.12688/f1000research.10851.1](https://doi.org/10.12688/f1000research.10851.1)
7. Williams M. Reagent validation to facilitate experimental reproducibility. *Curr Protocols Pharmacol* 2018; 81: e40. doi: [10.1002/cpph.40](https://doi.org/10.1002/cpph.40)
8. NIH and AHRQ. NIH and AHRQ Announce Upcoming Changes to Policies, Instructions and Forms for Research Training Grant, Fellowship, and Career Development Award Applications. 2019; Notice Number: NOT-OD-20-033. (<https://grants.nih.gov/grants/guide/notice-files/NOT-OD-20-033.html>)
9. Boulbes DR, Costello T, Baggerly K, et al. A survey on data reproducibility and the effect of publication process on the ethical reporting of laboratory research. *Clin Cancer Res* 2018; 24: 3447–3455. doi: [10.1158/1078-0432.CCR-18-0227](https://doi.org/10.1158/1078-0432.CCR-18-0227)