








## OPINION ARTICLE

# Improving open and rigorous science: ten key future research opportunities related to rigor, reproducibility, and transparency in scientific research [version 1; peer review: 3 approved]

Danny Valdez <sup>1</sup>, Colby J. Vorland <sup>1</sup>, Andrew W. Brown <sup>1</sup>,  
Evan Mayo-Wilson <sup>1</sup>, Justin Otten<sup>1</sup>, Richard Ball<sup>2</sup>, Sean Grant<sup>3</sup>, Rachel Levy<sup>4</sup>,  
Dubravka Svetina Valdivia <sup>5</sup>, David B. Allison<sup>1</sup>

<sup>1</sup>Indiana University School of Public Health, Bloomington, IN, 47403, USA

<sup>2</sup>Project TIER, Haverford College, Haverford, Pennsylvania, 19041, USA

<sup>3</sup>Indiana University Purdue University Indianapolis Fairbanks School of Public Health, Indianapolis, IN, 46223, USA

<sup>4</sup>Rachel Levy, Mathematical Association of America, 1529 18th St. NW, Washington, DC, 20036, USA

<sup>5</sup>Indiana University School of Education, Bloomington, IN, 47401, USA

**V1** First published: 14 Oct 2020, 9:1235  
<https://doi.org/10.12688/f1000research.26594.1>  
Latest published: 14 Oct 2020, 9:1235  
<https://doi.org/10.12688/f1000research.26594.1>

## Abstract

**Background:** As part of a coordinated effort to expand research activity around rigor, reproducibility, and transparency (RRT) across scientific disciplines, a team of investigators at the Indiana University School of Public Health-Bloomington hosted a workshop in October 2019 with international leaders to discuss key opportunities for RRT research.

**Objective:** The workshop aimed to identify research priorities and opportunities related to RRT.

**Design:** Over two-days, workshop attendees gave presentations and participated in three working groups: (1) Improving Education & Training in RRT, (2) Reducing Statistical Errors and Increasing Analytic Transparency, and (3) Looking Outward: Increasing Truthfulness and Accuracy of Research Communications. Following small-group discussions, the working groups presented their findings, and participants discussed the research opportunities identified. The investigators compiled a list of research priorities, which were circulated to all participants for feedback.



**Results:** Participants identified the following priority research questions: (1) Can RRT-focused statistics and mathematical modeling courses improve statistics practice?; (2) Can specialized training in

## Open Peer Review

Reviewer Status 

Invited Reviewers

	1	2	3
<b>version 1</b>			
14 Oct 2020	report	report	report

1. **Judith A Hewitt** , Office of Biodefense, Research Resources and Translational Research, Division of Microbiology and Infectious Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, USA
2. **Christopher A Mebane** , Idaho Water Science Center, US Geological Survey, Boise, USA
3. **Sheenah M Mische**, New York University (NYU) Langone Medical Center, New York,

scientific writing improve transparency?; (3) Does modality (e.g. face to face, online) affect the efficacy RRT-related education?; (4) How can automated programs help identify errors more efficiently?; (5) What is the prevalence and impact of errors in scientific publications (e.g., analytic inconsistencies, statistical errors, and other objective errors)?; (6) Do error prevention workflows reduce errors?; (7) How do we encourage post-publication error correction?; (8) How does 'spin' in research communication affect stakeholder understanding and use of research evidence?; (9) Do tools to aid writing research reports increase comprehensiveness and clarity of research reports?; and (10) Is it possible to inculcate scientific values and norms related to truthful, rigorous, accurate, and comprehensive scientific reporting?

**Conclusion:** Participants identified important and relatively unexplored questions related to improving RRT. This list may be useful to the scientific community and investigators seeking to advance meta-science (i.e. research on research).

### Keywords

Meta-Science; Science of Science; Rigor Reproducibility and Transparency (RRT); Workshop;

USA

Any reports and responses or comments on the article can be found at the end of the article.



This article is included in the **Future of Research (FoR)** collection.

**Corresponding author:** Justin Otten ([jotten@indiana.edu](mailto:jotten@indiana.edu))

**Author roles:** **Valdez D:** Conceptualization, Writing – Original Draft Preparation; **Vorland CJ:** Writing – Original Draft Preparation, Writing – Review & Editing; **Brown AW:** Funding Acquisition, Writing – Review & Editing; **Mayo-Wilson E:** Conceptualization, Funding Acquisition, Writing – Review & Editing; **Otten J:** Funding Acquisition, Project Administration; **Ball R:** Writing – Review & Editing; **Grant S:** Writing – Review & Editing; **Levy R:** Writing – Review & Editing; **Svetina Valdivia D:** Writing – Review & Editing; **Allison DB:** Conceptualization, Funding Acquisition, Writing – Review & Editing

**Competing interests:** No competing interests were disclosed.

**Grant information:** This work was funded by the Alfred P. Sloan Foundation (G-2019-11438) and awarded to David B. Allison. *The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.*

**Copyright:** © 2020 Valdez D *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**How to cite this article:** Valdez D, Vorland CJ, Brown AW *et al.* **Improving open and rigorous science: ten key future research opportunities related to rigor, reproducibility, and transparency in scientific research [version 1; peer review: 3 approved]** F1000Research 2020, 9:1235 <https://doi.org/10.12688/f1000research.26594.1>

**First published:** 14 Oct 2020, 9:1235 <https://doi.org/10.12688/f1000research.26594.1>

## Introduction

Rigor, reproducibility, and transparency (RRT) are scientific cornerstones that promote truthful, accurate, and objective science (McNutt, 2014). In the context of scientific research, rigor is defined as a thorough, careful approach that enhances the veracity of findings (Casadevall & Fang, 2012). There are several types of reproducibility, which include the ability to evaluate and follow the same procedures as previous studies, obtain comparable results, and draw similar inferences (Goodman *et al.*, 2016; National Academies of Sciences, 2019). Transparency is a process by which methodology, experimental design, coding, and data analysis tools are reported clearly and openly shared (Nosek *et al.*, 2015; Prager *et al.*, 2019). Together, these scientific norms represent the best means of obtaining objective knowledge of the world (Anderson *et al.*, 2010; Allison *et al.*, 2016). The science concerning these norms is a specific branch of meta-science, or “research on research”, led by scientists who promote these values through the education of early career scientists, identifying areas of concern for scientific validity, and postulating paths toward stronger, more credible science (Ioannidis *et al.*, 2015).

Several factors compete with the pursuit of rigorous, reproducible, and transparent research. For example, the rate of scientific publication has risen dramatically in the last two decades. Although this is indicative of many important scientific breakthroughs (Van Noorden, 2014), the rate of manuscript retractions due to either researcher error or malfeasance has also increased (Steen *et al.*, 2013). A survey found between 40% and 70% of scientists agreed that factors including fraud, selective reporting, and pressure to publish contribute to the irreproducibility of scientific findings (Fanelli, 2018). These concerns also have the potential to decrease public trust in science, although research on this question is needed (National Academies of Sciences, 2017).

Basic and applied science are undermined when scientists fail to uphold high standards of conduct (Prager *et al.*, 2019). Given that many authors have identified issues or concerns in science, the emerging challenge for scholars in this area is to find workable solutions to improve RRT, rather than simply continuing to illustrate problems related to RRT (Allen & Mehler, 2019). To this end, in October 2019, Indiana University School of Public Health-Bloomington hosted a multidisciplinary meeting of leading scholars to discuss ongoing RRT-related challenges. The purpose of the meeting, which was funded by the Alfred P. Sloan Foundation, was to identify new opportunities to advance sound scientific practice, from the early stages of planning a study, through to execution and the communication of findings. This paper presents findings from that meeting.

## Methods

The meeting was structured around three areas:

- (1) Improving education & training in RRT.
- (2) Reducing statistical errors and increasing analytic transparency.

- (3) Looking outward: increasing truthfulness and accuracy of research communications.

## Participants

We invited participants based on prior contributions to RRT research. Participants included representatives from several leading organizations and Indiana University (IU) faculty, staff, and graduate students who were invited to participate in the meeting and proceedings (Table 1). For their participation in the meeting, invited guests who were not federal employees or IU employees received a \$1,000 honorarium.

## Meeting format

The two-day meeting was comprised of nine prepared research talks, moderated panel discussions, and small-group open-forum style sessions related to each of the three previously stated goals.

**Day one.** On the first day, participants presented 10–12 minute research talks, each of which was followed by a moderated question-and-answer period. Participants discussed questions pertaining to RRT and sought to identify emerging areas of research including novel approaches, testable outcomes, and potential limitations. During the afternoon session, participants were divided into three small-groups to discuss potential research opportunities, moderated by an IU faculty representative charged with compiling notes for record keeping and dissemination.

**Day two.** On the second day, one representative from each group summarized major points through a brief presentation, which was followed by a question-and-answer session with all participants. This dialogue was intended to clarify ideas raised and to identify fundable research opportunities. The meeting concluded with a call to action by the Dean of the School of Public Health-Bloomington and Co-Principal Investigator of the project (DA), to continue promoting interdisciplinary RRT Science.

## Results

### Subgroup 1: improving education & training in RRT

We asked the first subgroup to discuss research opportunities related to implementing and testing RRT-guided academic curricula. The group identified elements of current undergraduate and graduate education that contribute to problematic data practices, including possible underlying causes and potential solutions (see Table 2). Three primary education-related questions guided the discussion:

- (1) Can RRT-focused statistics and mathematical modeling courses improve statistical practice?
- (2) Can specialized training in scientific writing improve transparency?
- (3) Does modality affect the efficacy of RRT-related education?

**Table 1. List of invited participants and other attendees.**

Name	Affiliation	Subgroup
<b>Invited participants</b>		
Richard Ball, Ph.D.	Project TIER [Teaching Integrity in Empirical Research]	Improving Education & Training in rigor, reproducibility, and transparency [RRT]
Rachel Levy, Ph.D.	Mathematical Association of America	Improving Education & Training in RRT
Keith Baggerly, Ph.D.	University of Texas	Reducing Statistical Errors and Increasing Analytic Transparency
John Ioannidis, M.D., DSc	METRICS [Meta-Research Innovation Center at Stanford]	Reducing Statistical Errors and Increasing Analytic Transparency
Brian Nosek, Ph.D.	Center for Open Science	Reducing Statistical Errors and Increasing Analytic Transparency
Phillipe Ravaud, M.D., Ph.D.	Paris Descartes University	Looking Outward: Increasing Truthfulness and Accuracy of Research Communications
Machell Town, Ph.D.	Centers for Disease Control and Prevention	Looking Outward: Increasing Truthfulness and Accuracy of Research Communications
Matt Vassar, MBA, Ph.D.	Oklahoma State University	Looking Outward: Increasing Truthfulness and Accuracy of Research Communications
<b>Indiana University faculty &amp; staff</b>		
David B. Allison, Ph.D.	Dean of the School of Public Health	Improving Education & Training
Dubravka Svetina, Ph.D.	School of Education	Improving Education & Training
Elizabeth Housworth, Ph.D.	Mathematics	Improving Education & Training
Emily Meanwell, Ph.D.	Social Science Research Commons	Improving Education & Training
Roger Zoh, MS, Ph.D.	School of Public Health	Improving Education & Training
Andrew W. Brown, Ph.D.	School of Public Health	Reducing Statistical Errors and Increasing Analytic Transparency
Stephanie Dickinson, MS	School of Public Health	Reducing Statistical Errors and Increasing Analytic Transparency
Mandy Mejia, Ph.D.	Statistics	Reducing Statistical Errors and Increasing Analytic Transparency
Carmen Tekwe, MS, Ph.D.	School of Public Health	Reducing Statistical Errors and Increasing Analytic Transparency
Evan Mayo-Wilson, DPhil	School of Public Health	Looking Outward: Increasing Truthfulness and Accuracy of Research Communications
Ana Bento, Ph.D.	School of Public Health	Looking Outward: Increasing Truthfulness and Accuracy of Research Communications
Jutta Schickore, Ph.D.	History and Philosophy of Science and Medicine	Looking Outward: Increasing Truthfulness and Accuracy of Research Communications
Jamie Wittenberg, M.B.S, MSLIS	Library System	Looking Outward: Increasing Truthfulness and Accuracy of Research Communications
<b>Non-presenting attendees</b>		
Lilian Golzarri Arroyo, MS (IU School of Public Health); Chris Bogert, Ph.D. (IU Applied Pharmacology & Toxicology); Sean Grant, DPhil (IUPUI School of Public Health); Stasa Milojevic, Ph.D. (IU Informatics); Luis Mestre, MS (IU School of Public Health); Justin Otten, Ph.D. (IU School of Public Health); Danny Valdez, Ph.D. (IU School of Public Health); Colby J. Vorland, Ph.D. (IU School of Public Health)		

**Table 2. Summary of group discussion.**

Group	Question	Challenges <sup>1</sup>
<p><b>Subgroup 1:</b> improving education &amp; training in rigor, reproducibility, and transparency (RRT)</p>	<p>1. Can RRT-focused statistics and mathematical modeling courses improve statistical practice?</p>	<p>1. It would be difficult to isolate and to evaluate the effects of changes to existing curricula.</p> <p>2. Proximal measures related to technical skills might not translate into improved research practices.</p>
	<p>2. Can specialized training in scientific writing improve transparency?</p>	<p>1. Writing is an abstract science, which would make measuring outcomes challenging.</p> <p>2. There are currently limited existing graduate level curricula that pertain exclusively to writing.</p>
	<p>3. Does modality affect the efficacy of RRT-related education?</p>	<p>1. Feasibility concerns including, cost, time, and other additional resources needed to facilitate an intervention.</p> <p>2. Examining heterogeneity requires large and diverse populations, and is practically difficult.</p>
<p><b>Subgroup 2:</b> reducing statistical errors and increasing analytic transparency</p>	<p>4. Can automation help identify errors more efficiently?</p>	<p>1. Automation may be technically possible for only certain types of errors.</p> <p>2. New programs intended to automate error correction require a certain level of computer programming expertise.</p>
	<p>5. What is the prevalence and impact of errors?</p>	<p>1. It would be difficult to generalize the prevalence of errors, because many common errors have field-specific names.</p> <p>2. Assessing the impact of errors is largely subjective, unless strict guidelines are agreed upon and adopted.</p>
	<p>6. Do error prevention workflows reduce errors?</p>	<p>1. It would be difficult to determine if workflows are entirely responsible for reduced error and improved research practice.</p> <p>2. It may be challenging to identify generalizable workflows that logically function across disciplines.</p>
	<p>7. How do we encourage post-publication error correction?</p>	<p>1. It would be difficult to implement standard post-publication error correction guidelines that function effectively across disciplines.</p> <p>2. There is a hesitancy to embrace error correction as a normal component of the editorial process.</p>
	<p>8. How does 'spin' in research communication affect stakeholders' understanding and use of research evidence?</p>	<p>1. The effects of spin in controlled research settings might not generalize to real-world decisions.</p> <p>2. Reviewing and categorizing text is both subjective and time consuming.</p>
	<p>9. Do tools to aid writing research reports increase the comprehensiveness and clarity of research reports?</p>	<p>1. Although tools could be developed for testing, implementation challenges could mitigate their effectiveness in practice.</p> <p>2. Previous guidelines have had minimal impact on reporting quality.</p>
	<p>10. Is it possible to inculcate scientific values and norms related to truthful, rigorous, accurate, and comprehensive scientific reporting?</p>	<p>1. There are few model interventions to form self-identity.</p> <p>2. There may be limited opportunities and enthusiasm to integrate values-based education in classes that focus on technical skills.</p>

<sup>1</sup>We present here only two of the most salient challenges.

With respect to each question the existing and entrenched practices, feasibility of change, and proper audience for interventions were discussed.

### 1. Can RRT-focused statistics and mathematical modeling courses improve statistical practice?

Incorrect analyses are some of the most common, preventable errors in science (Resnik, 2012). Scholars attribute mistakes to gaps in statistics education (Thompson, 2006). With the rise in data science as a component of scientific exploration, students need more exposure to evidence-based pedagogical approaches to statistics and mathematical modeling (GAISE, 2016; GAIMME, 2016, and NASEM, 2018). Many introductory data science courses include topics from statistics (e.g., contingency tables [chi-square tests], multiple regression, analysis of variance, and the broader general linear model) (Gorsuch, 2015), as well as mathematical modeling approaches and computational algorithms. These topics can be reframed through an RRT lens as modules/domains within existing mathematics or data-science courses or structured as new data-driven courses entirely.

Indeed, participants noted that to improve RRT practices, there are opportunities to design new courses with a direct RRT focus at the undergraduate, graduate and postdoctoral levels (Willig *et al.*, 2018). Courses could include modules related to the identification of errors in published research, proposing solutions to these errors, addressing real-world contexts and demonstrating the importance of careful methodological decision-making (Peng, 2015). Specific assignments could test for and reinforce RRT principles, such as research compendia (i.e. sharable electronic folders containing code, and other exploratory information to validate reported results) (Ball & Medeiros, 2012; King, 1995; Stodden *et al.*, 2015), workflows, which are described later in this paper, and other research projects related to communication and computational reproducibility. The learning practices could be assessed to ensure that students appropriately apply concepts rather than demonstrate rote-formula memorization (Thompson, 2002; Ware *et al.*, 2013). Integrating learning into stages of education where students are concurrently engaged in research can help improve both retention and transfer of the RRT ideas to future scientific settings.

### 2. Can specialized training in scientific writing improve transparency?

Clear scientific writing is necessary to reproduce and build on research findings. To facilitate better writing, scholars have developed curricula to help academics improve writing practice and quality (e.g., Goodson, 2016). However, many academic writing programs focus on personal habit building and development of linguistic mechanics to craft more powerful prose (Elbow, 1998; Kellogg & Whiteford, 2009). In such courses, RRT-related dimensions of writing (such as writing transparently or minimizing ‘spin’) may not be emphasized. Thus, the subgroup discussed how existing writing curricula could incorporate RRT principles, what new writing courses guided by RRT would entail, and research opportunities to test the efficacy of new writing curricula.

Participants identified several RRT-specific writing principles and discussed how a deeper understanding of the extent to which writing and research are intertwined may increase transparency. Examples included learning about methodological reporting guidelines, writing compelling post-publication peer reviews, and other transparent writing practices. The group also discussed how courses could be developed or redesigned specifically to center on RRT principles. One theme of the discussion was the need for rigorous testing of student learning outcomes associated with novel writing content. However, a primary concern was the identification of the appropriate outcome measures for writing-specific interventions (Barnes *et al.*, 2015) given the subjective and nebulous nature of constructs like writing quality, individual improvement, and writing-related self-efficacy.

### 3. Does modality affect the efficacy of RRT-related education?

Another research opportunity discussed by the subgroup related to instructional modality, which refers to the manner in which a curriculum or intervention is experienced by the learner (Perry & Pilati, 2011). These may include traditional face-to-face instruction, synchronous or asynchronous online meetings/trainings, and various hybrid formats (Beall *et al.*, 2014). Understanding the relative benefits of each modality is important in choosing an appropriate intervention. Indeed, educational needs vary among learner groups; for example, what is most effective for undergraduate students may not be effective or feasible for post-doctoral researchers with full-time professional commitments. Broad research questions identified by the group included:

- What modalities exist beyond face-to-face, online, or hybrid instruction?;
- How can technology push modality beyond online courses and other Massive Open Online Course formats?
- Which modality is most effective and among which audiences?

In the context of previously discussed coursework in statistics and writing, participants explored the strengths and weaknesses of various modalities and how interventions could be conducted to test them empirically. There are logistical considerations, such as cost, space, and faculty time, that further complicate the feasibility of these interventions. For example, a face-to-face intervention may offer more tailored instruction to individual learners, while an online intervention may better deliver content to a wider audience. Thus, the subgroup identified several areas for future research, including comparisons of student learning across modalities, strategies for scaling educational content to institutional constraints, and the moderating effects of learner demographics on intervention efficacy.

#### Subgroup 2: reducing statistical errors and increasing analytical transparency

Errors are “actions or conclusions that are demonstrably and unequivocally incorrect from a logical or epistemological point of view” (Brown *et al.*, 2018). Despite the adage that science

is self-correcting, uncorrected errors are prevalent in the scientific literature (Brown *et al.*, 2018; Ioannidis, 2012). Subgroup 2 discussed questions related to reducing and mitigating such errors, including:

- (4) Can automation help identify errors more efficiently?;
- (5) What is the impact of errors within disciplines?;
- (6) Do standardized procedures (i.e., workflows) prevent errors?;
- (7) How do we encourage post-publication error correction?;

The costs and benefits associated with each question were also discussed (see Table 2).

#### 4. Can automation help identify errors more efficiently?

Various automated and manual methods have been developed and applied to assess analytic inconsistencies, statistical errors and improbabilities, and other errors (e.g., Anaya, 2016; Baggerly & Coombes, 2009; Brown & Heathers, 2017; Georgescu & Wren, 2018; Labbé *et al.*, 2019; Monsarrat & Vergnes, 2018). An increase in automation (i.e., producing more user-friendly tools and algorithms) has the potential for surveilling the prevalence, prevention, and correction of errors. However, more work is needed to determine the most efficient use of such tools, including their collective abilities to detect field-specific issues that require subject matter expertise (Lakens & Debruine, 2020). For example, the automatic recomputation of some p-values is possible using the program ‘Statcheck’, but only for articles that utilize the American Psychological Association’s (APA) in-text citation style for statistical reporting (Nuijten *et al.*, 2017). Other examples require statistical ratios (Georgescu & Wren, 2018), or integer-based data and sample sizes (e.g., Brown & Heathers, 2017; Heathers *et al.*, 2018), which are both challenging to automate and not recurrent across all fields.

Automated error detection is currently limited to a narrow range of errors. Other types of errors might be detected by careful readers, such as the ignoring of clustering in cluster-randomized trials (Brown *et al.*, 2015; Heo *et al.*, 2018), misinterpretation of differences in nominal significance, and post-hoc fallacies (Brown *et al.*, 2019; George *et al.*, 2016). The subgroup discussed opportunities to define, and possibly automate, diagnostic checklists, advanced natural language processing, or other computational informatics approaches that would facilitate the detection of these errors. These novel automated measures could be tested empirically for effectiveness.

#### 5. What is the prevalence and impact of errors?

Different errors will have varying impacts on study conclusions. While some errors can be easily corrected and reported, some fundamentally invalidate study conclusions. Some general statistical errors have occurred repeatedly across disciplines for decades (e.g., mistaken differences due to “regression to the mean” since at least 1886 [Thomas *et al.*, 2020] and “differences

in nominal significance” for decades [Altman, 2002; Thompson, 2002]). Automated methods, such as those outlined above, have been used almost exclusively to illuminate problems but not necessarily correct them (Georgescu & Wren, 2018; Monsarrat & Vergnes, 2018; Nuijten *et al.*, 2017).

To achieve the goal of error reduction, one must first know how pervasive errors are. Yet, it remains challenging to generalize the detection and correction of scientific errors across disciplines because of field specificity (i.e. the unique nuances and methodological specificities inherent to a specific field of study) (Lohse *et al.*, 2020), the various terminologies used for describing the same models (e.g. ‘Hierarchical Linear’ models vs ‘Multilevel’ models), as well as the seeming need to repackage the same problem as new disciplines arise (e.g. ongoing multiple comparison issues raised anew with the advent of genome-wide association studies, microarray, microbiome, and functional magnetic resonance imaging methods). Thus, this subgroup discussed the value of longitudinal, discipline-specific error surveillance and error frequency estimation to collect empirical evidence about error rate differences among disciplines. Other issues discussed were the identification of better prevalence estimates across fields, and how simulation studies can modify our confidence in the understanding of the prevalence of errors and their generalizability across disciplines.

#### 6. Do error prevention workflows reduce errors?

Workflows are the various approaches for accomplishing scientific objectives, usually expressed as tasks and dependencies (Ludäscher *et al.*, 2009). The implementation of clear, logical workflows can potentially prevent errors and improve research transparency. Workflows may be of value to catch errors at various stages of the research process, from planning, to data collection and handling procedures, and reporting/manuscript screening (Cohen-Boulakia *et al.*, 2017). Error detection processes within scientific workflows may serve as mechanisms to prevent errors before publication, akin to how text duplication software (e.g. iThenticate) is used prophylactically to catch inadvertent plagiarism. Separately, some research groups implement workflows that require two independent scientists to verify data, analyses, and statistical reporting prior to manuscript publication, with at least one of those individuals being a professional statistician (George *et al.*, 2016). A similar workflow is to establish “red teams”, consisting of methodologists, statisticians, and subject-matter experts, to critique the study design and analysis for errors, offering incentives akin to “bug bounty” programs in computer software development (Lakens, 2020).

The development and dissemination of research workflows could be modeled after those outlined above, or in other ways such as the use of checklists to complete work systematically. Registrations, reporting guidelines, and other workflow approaches essentially serve as checklists of the plan for a study and what should be reported. Although this subgroup agreed about the importance of preventive versus post-publication workflows

and integration of automated methods to detect errors, questions regarding their efficacy remained. For example, how might workflows be generalized across academic disciplines? At what level should standardizing data collection and handling be taught to scientists to maintain data provenance (e.g. Long, 2009)? And can workflows be tested empirically? What is the cost of automated versus manual workflows, versus none at all, at detecting and preventing errors? How do workflows impact productivity?

### 7. How do we encourage post-publication error correction?

Science cannot self-correct without processes that facilitate correction (Firestein, 2012). Unfortunately, errors in science may be tied with perceived reputation costs, yet it is unclear that correcting errors actually harms a researcher's reputation (Azoulay *et al.*, 2015). Thus, destigmatizing error correction and likewise embracing the importance of scientific failures may be of value for individual scientists and editors overseeing content through the peer-review process (Teixeira da Silva & Al-Khatib, 2019). Journals and their editors, as gatekeepers of science, are key stakeholders in this culture shift. They may also require practical guidelines to facilitate judgement-free corrections that would be acceptable to editors and reviewers.

Error correction should be done in a fair and efficient manner (e.g., Vorland *et al.*, 2020). Although there are several existing standards for publication ethics and norms (e.g., Committee on Publication Ethics [COPE], and the International Committee of Medical Journal Editors [ICMJE]), few have been tested empirically. The subgroup debated how journals and their editors could be part of empirically tested trials on best approaches to facilitate correction and minimize the incurring of additional costs. For example, based on our experiences, journals have few procedures for handling errors separate from typical scholarly dialogue. We believe it is important to examine which procedures are more efficient and fair to authors, whether such procedures can be standardized to enable editors to handle different types of errors consistently and transparently, whether correction mechanisms are sufficient or require additional innovation (e.g. retraction and republication is sufficient or versioning), and how authors can be supported and encouraged in the process. Three such costs that require further study include the actual cost of post-publication error correction across all parties involved (e.g. page charges, salary), how those costs to the scientific enterprise compare to implementing prevention strategies, and the cost-benefit of salvaging a publication containing an error depending on the quality of the collected data versus simply retracting.

### Subgroup 3 - looking outward: increasing truthfulness and accuracy of research communications

The third working group discussed opportunities for research related to research reporting and dissemination, primarily highlighting the importance of accuracy and truthfulness when communicating research findings (see Table 2). Specifically, this group identified research opportunities tied to the following questions:

- (8) How does 'spin' in research communication affect stakeholders' understanding and use of research evidence?

- (9) Do tools to aid writing research reports increase the comprehensiveness and clarity of research reports?
- (10) Is it possible to inculcate scientific values and norms related to truthful, rigorous, accurate, and comprehensive scientific reporting?

### 8. How does "spin" in research communication affect stakeholders' understanding and use of research evidence?

In addition to conducting research rigorously, investigators should describe their research comprehensively and interpret their findings by balancing the strengths and limitations of their methods and results (Brown *et al.*, 2017). By contrast, researchers might 'spin' their results through misleading reporting, misleading interpretation, and inappropriate extrapolation (Fletcher & Black, 2007; Yavchitz *et al.*, 2016). Some evidence suggests that spin is common in reports of clinical trials and meta-analyses (Boutron *et al.*, 2019; Lazarus *et al.*, 2015) and that authors in a variety of research disciplines often draw inappropriate causal inferences (Bleske-Rechek *et al.*, 2015; Casazza *et al.*, 2013; Chiu *et al.*, 2017; Knight *et al.*, 1996; Ochodo *et al.*, 2013). Moreover, spin in popular media (e.g., newspapers) appears to stem from spin in scientific reports (e.g., journal articles) and associated press releases (de Semir *et al.*, 1998; Schwartz *et al.*, 2012; Schwitzer, 2008).

Spin is unscientific, and could have implications for policy and practice (Adams *et al.*, 2016; Boutron *et al.*, 2019; Matthews *et al.*, 2016). Workshop participants discussed the need for more evidence to determine whether and how spin in scientific reports affects other stakeholders such as healthcare and social service providers, service users, policymakers, and payers. Evidence concerning the ways in which stakeholders use and interpret research evidence could inform future efforts to improve research communication (Boutron *et al.*, 2019; Lazarus *et al.*, 2015).

### 9. Do tools to aid writing research reports increase the comprehensiveness and clarity of research reports?

Research reports (e.g., journal articles) should describe what was done and what was found (von Elm *et al.*, 2007). Stakeholders need comprehensive and accurate information about research methods and results to assess risk of bias, interpret the generalizability of study results, and reproduce the conditions (e.g., interventions) described (Moher *et al.*, 2011). Reporting guidelines describe the minimum information that should be included in reports of different types of research, yet much evidence suggests that scientific reports do not include this information (e.g., Grant *et al.*, 2013). Some tools to help authors write better reports have been developed, such as the consort-based web tool (COBWEB) (Barnes *et al.*, 2015); some preliminary evaluations suggest that these tools could help authors write better reports.

Workshop participants identified a need for research to develop and to test tools that could help authors write reports that adhere to existing guidelines. Some tools could be used when writing scientific manuscripts (Turner *et al.*, 2012) while other tools could be used in graduate education (e.g. class assignments, dissertation writing) or continuing education. Guidelines



designed to increase authors' and reviewers' knowledge of reporting requirements are not commonly adhered to and, thus, have minimal impact on reporting quality (Capers *et al.*, 2015). Participants emphasized the need for new interventions and implementation research that promote guideline adherence.

### 10. Is it possible to inculcate scientific values and norms related to truthful, rigorous, accurate, and comprehensive scientific reporting?

In the 1940s, Robert Merton proposed that communism/communality, universalism, disinterestedness, and organized skepticism constitute the ethos of modern science (Merton, 1942). As the National Research Council stated in their report "Scientific Research in Education", these fundamental principles are enforced by the community of researchers that shape scientific understanding (Shavelson & Towne, 2003). Evidence suggests that most scientists endorse these positive values and norms, but fewer scientists believe that their colleagues behave in accordance with these positive norms (Anderson *et al.*, 2007). Better incentives (Begley *et al.*, 2017; Fanelli, 2010; Nosek *et al.*, 2012) and better methods for detecting scientific errors, might improve scientific practice and communication; yet fundamentally, we will always have to place some trust in the veracity of our fellow scientists (Jamieson *et al.*, 2017).

Participants agreed that ethics and responsibility are vital across scientific disciplines, yet graduate research often neglects the philosophy of science and the formation of professional identity as a scientist. Instead, training tends to focus on the

technical skills needed to conduct experiments and analyze data in specific disciplines (Bosch, 2018; Bosch & Casadevall, 2017). Technical skills are essential to produce good science; to apply them ethically and responsibly, however, it is paramount that scientists also endorse scientific values and norms. Participants identified a need for research to determine how these scientific values could be inculcated in scientists and how scientists should be taught to enact those values in their research.

### Conclusion

Scientists slow the pursuit of truth when research is not rigorous, reproducible, or transparent (Collins & Tabak, 2014). To improve the state of science, RRT leaders have long raised concerns about many of the current challenges the scientific enterprise faces by identifying novel strategies intended to uphold and improve scientific validity. Discussions among RRT leaders at Indiana University Bloomington reinforce the value and importance of promoting accurate, objective, and truthful science. The proposal, execution, and evaluation of the ideas presented herein showcases how the collective and interdisciplinary efforts of those investing in the future of science can solve problems in unique and exciting ways.

### Data availability

No data are associated with this article.

All participants have provided their permission to be named in this article.

## References

- Adams SA, Choi SK, Eberth JM, *et al.*: **Adams *et al.* Respond.** *Am J Public Health.* 2016; **106**(6): e8–9.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Allen C, Mehler DMA: **Open science challenges, benefits and tips in early career and beyond.** *PLoS Biol.* 2019; **17**(5): e3000246.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Allison DB, Brown AW, George BJ, *et al.*: **Reproducibility: A tragedy of errors.** *Nature.* 2016; **530**(7588): 27–9.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Altman DG: **Poor-Quality Medical Research: What Can Journals Do?** *JAMA.* 2002; **287**(21): 2765–2767.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Anaya J: **The GRIMMER test: A method for testing the validity of reported measures of variability (e2400v1).** *PeerJ Inc.* 2016.
- Anderson MS, Martinson BC, De Vries R, *et al.*: **Normative Dissonance in Science: Results from a National Survey of U.S. Scientists.** *J Empir Res Hum Res Ethics.* 2007; **2**(4): 3–14.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Anderson MS, Ronning EA, DeVries R, *et al.*: **Extending the Mertonian Norms: Scientists' Subscription to Norms of Research.** *J Higher Educ.* 2010; **81**(3): 366–393.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Azoulay P, Bonatti A, Krieger JL: **The Career Effects of Scandal: Evidence from Scientific Retractions.** (Working Paper No. 21146). National Bureau of Economic Research. 2015.  
[Publisher Full Text](#)
- Baggerly KA, Coombes KR: **Deriving chemosensitivity from cell lines: Forensic bioinformatics and reproducible research in high-throughput biology.** *The Annals of Applied Statistics.* 2009; **3**(4): 1309–1334.  
[Publisher Full Text](#)
- Ball R, Medeiros N: **Teaching Integrity in Empirical Research: A Protocol for Documenting Data Management and Analysis.** *The Journal of Economic Education.* 2012; **43**(2): 182–189.  
[Publisher Full Text](#)
- Barnes C, Boutron I, Giraudeau B, *et al.*: **Impact of an online writing aid tool for writing a randomized trial report: The COBWEB (Consort-based WEB tool) randomized controlled trial.** *BMC Med.* 2015; **13**(1): 221.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Beall RF, Baskerville N, Golfam M, *et al.*: **Modes of delivery in preventive intervention studies: A rapid review.** *Eur J Clin Invest.* 2014; **44**(7): 688–696.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Begley, EB, Ware JM, Hexem SA, *et al.*: **Personally Identifiable Information in State Laws: Use Release, and Collaboration at Health Departments.** *Am J Public Health.* 2017; **107**(8): 1272–1276.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Bleske-Rechek A, Morrison KM, Heidtke LD: **Causal Inference from Descriptions of Experimental and Non-Experimental Research: Public Understanding of Correlation-Versus-Causation.** *J Gen Psychol.* 2015; **142**(1): 48–70.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Bosch G: **Train PhD students to be thinkers not just specialists.** *Nature.* 2018; **554**(7692): 277.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Bosch G, Casadevall A: **Graduate Biomedical Science Education Needs a New**

- Philosophy.** *mBio.* 2017; **8**(6): e01539–17.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Boutron I, Haneef R, Yavchitz A, *et al.*: **Three randomized controlled trials evaluating the impact of “spin” in health news stories reporting studies of pharmacologic treatments on patients’/caregivers’ interpretation of treatment benefit.** *BMC Med.* 2019; **17**(1): 105.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Brown AW, Altman DG, Baranowski T, *et al.*: **Childhood obesity intervention studies: A narrative review and guide for investigators, authors, editors, reviewers, journalists, and readers to guard against exaggerated effectiveness claims.** *Obes Rev.* 2019; **20**(11): 1523–1541.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Brown AW, Kaiser KA, Allison DB: **Issues with data and analyses: Errors, underlying themes, and potential solutions.** *Proc Natl Acad Sci U S A.* 2018; **115**(11): 2563–2570.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Brown NJL, Heathers JA: **The GRIM Test: A Simple Technique Detects Numerous Anomalies in the Reporting of Results in Psychology.** *Social Psychological and Personality Science.* 2017; **8**(4): 363–369.  
[Publisher Full Text](#)
- Brown AW, Li P, Bohan MMB, *et al.*: **Best (but oft-forgotten) practices: designing, analyzing, and reporting cluster randomized controlled trials.** *Am J Clin Nutr.* 2015; **102**(2): 241–248.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Brown AW, Mehta TS, Allison DB: **Publication bias in science: what is it why is it problematic, and how can it be addressed?** *The Oxford Handbook of the Science of Science Communication.* 2017; 93–101.  
[Publisher Full Text](#)
- Capers PL, Brown AW, Dawson JA, *et al.*: **Double sampling with multiple imputation to answer large sample meta-research questions: introduction and illustration by evaluating adherence to two simple CONSORT guidelines.** *Front Nutr.* 2015; **2**: 6.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Carver R, Everson M, Gabrosek J, *et al.*: **Guidelines for assessment and instruction in statistics education (GAISE) college report 2016.** 2016.  
[Reference Source](#)
- Casadevall A, Fang FC: **Reforming Science: Methodological and Cultural Reforms.** *Infect Immun.* 2012; **80**(3): 891–896.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Casazza K, Fontaine KR, Astrup A, *et al.*: **Myths, Presumptions, and Facts about Obesity.** *N Engl J Med.* 2013; **368**(5): 446–454.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Chiu K, Grundy Q, Bero L: **‘Spin’ in published biomedical literature: A methodological systematic review.** *PLoS Biol.* 2017; **15**(9): e2002173.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Cohen-Boulakia S, Belhajjame K, Collin O, *et al.*: **Scientific workflows for computational reproducibility in the life sciences: Status, challenges and opportunities.** *Future Generation Computer Systems.* 2017; **75**: 284–298.  
[Publisher Full Text](#)
- Collins FS, Tabak LA: **Policy: NIH plans to enhance reproducibility.** *Nature.* 2014; **505**(7485): 612–613.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Consortium for Mathematics and its Applications & Society for Industrial and Applied Mathematics: **Guidelines for assessment and instruction in mathematical modeling education.** 2016.  
[Reference Source](#)
- de Semir V, Ribas C, Revuelta G: **Press releases of science journal articles and subsequent newspaper stories on the same topic.** *JAMA.* 1998; **280**(3): 294–295.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Elbow P: **Writing With Power: Techniques for Mastering the Writing Process.** Oxford University Press, 1998.  
[Reference Source](#)
- Fanelli D: **“Positive” Results Increase Down the Hierarchy of the Sciences.** *PLoS One.* 2010; **5**(4): e10068.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Fanelli D: **Opinion: Is science really facing a reproducibility crisis, and do we need it to?** *Proc Natl Acad Sci U S A.* 2018; **115**(11): 2628–2631.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Firestein S: **Ignorance: How It Drives Science.** (1 edition), Oxford University Press, 2012.  
[Reference Source](#)
- Fletcher RH, Black B: **“Spin” in Scientific Writing: Scientific Mischief and Legal Jeopardy.** *Med Law.* 2007; **26**(3): 511–525.  
[PubMed Abstract](#)
- George BJ, Beasley TM, Brown AW, *et al.*: **Common scientific and statistical errors in obesity research.** *Obesity (Silver Spring).* 2016; **24**(4): 781–790.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Georgescu C, Wren JD: **Algorithmic identification of discrepancies between published ratios and their reported confidence intervals and P-values.** *Bioinformatics.* 2018; **34**(10): 1758–1766.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Goodman SN, Fanelli D, Ioannidis JPA: **What does research reproducibility mean?** *Sci Transl Med.* 2016; **8**(341): 341ps12.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Goodson P: **Becoming an Academic Writer: 50 Exercises for Paced, Productive, and Powerful Writing.** SAGE Publications, 2016.  
[Reference Source](#)
- Gorsuch RL: **Enhancing the Teaching of Statistics by Use of the Full GLM.** *Journal of Methods and Measurement in the Social Sciences.* 2015; **6**(2): 60–69.  
[Reference Source](#)
- Grant SP, Mayo-Wilson E, Melendez-Torres GJ, *et al.*: **Reporting Quality of Social and Psychological Intervention Trials: A Systematic Review of Reporting Guidelines and Trial Publications.** *PLoS One.* 2013; **8**(5): e65442.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Heathers JA, Anaya J, Zee T, *et al.*: **Recovering data from summary statistics: Sample Parameter Reconstruction via Iterative TEchniques (SPRITE).** PeerJ Inc., 2018; e26968v1.  
[Publisher Full Text](#)
- Heo M, Nair SR, Wylie-Rosett J, *et al.*: **Trial characteristics and appropriateness of statistical methods applied for design and analysis of randomized school-based studies addressing weight-related issues: a literature review.** *J Obes.* 2018; **2018**: 8767315.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Ioannidis JPA: **Why Science Is Not Necessarily Self-Correcting.** *Perspect Psychol Sci.* 2012; **7**(6): 645–64.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Ioannidis JPA, Fanelli D, Dunne DD, *et al.*: **Meta-research: evaluation and improvement of research methods and practices.** *PLoS Biol.* 2015; **13**(10): e1002264.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Jamieson KH, Kahan DM, Scheufele DA: **The Oxford Handbook of the Science of Science Communication.** Oxford University Press, 2017.  
[Publisher Full Text](#)
- Kellogg RT, Whiteford AP: **Training advanced writing skills: The case for deliberate practice.** *Educational Psychologist.* 2009; **44**(4): 250–266.  
[Publisher Full Text](#)
- King G: **Replication, Replication.** *PS: Political Science and Politics.* 1995; **28**(3): 444–452.  
[Publisher Full Text](#)
- Knight GP, Fabes RA, Higgins DA: **Concerns about drawing causal inferences from meta-analyses: An example in the study of gender differences in aggression.** *Psychol Bull.* 1996; **119**(3): 410–421.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Labbé C, Grima N, Gautier T, *et al.*: **Semi-automated fact-checking of nucleotide sequence reagents in biomedical research publications: The Seek & Blastn tool.** *PLoS One.* 2019; **14**(3): e0213266.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Lakens D: **Pandemic researchers - recruit your own best critics.** *Nature.* 2020; **581**(7807): 121.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Lakens D, DeBruine L: **Improving transparency, falsifiability, and rigour by making hypothesis tests machine readable.** 2020.  
[Publisher Full Text](#)
- Lazarus C, Haneef R, Ravaud P, *et al.*: **Classification and prevalence of spin in abstracts of non-randomized studies evaluating an intervention.** *BMC Med Res Methodol.* 2015; **15**: 85.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Lohse K, Sainani K, Taylor JA, *et al.*: **Systematic Review of the use of “Magnitude-Based Inference” in Sports Science and Medicine.** [Preprint]. SportRxiv, 2020.  
[Publisher Full Text](#)
- Long JS: **The workflow of data analysis using Stata.** Stata Press, 2009.  
[Reference Source](#)
- Ludäscher B, Weske M, McPhillips T, *et al.*: **Scientific workflows: Business as usual? In: International Conference on Business Process Management.** Springer, Berlin, Heidelberg, 2009; **5701**: 31–47.  
[Publisher Full Text](#)
- Matthews DD, Smith JC, Brown AL, *et al.*: **Reconciling Epidemiology and Social Justice in the Public Health Discourse Around the Sexual Networks of Black Men Who Have Sex With Men.** *Am J Public Health.* 2016; **106**(5): 808–814.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- McNutt M: **Reproducibility.** *Science.* 2014; **343**(6168): 229.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Merton RK: **A Note on Science and Democracy.** *Journal of Legal and Political Sociology.* 1942; **1**: 115.  
[Reference Source](#)
- Moher D, Weeks L, Ocampo M, *et al.*: **Describing reporting guidelines for health research: a systematic review.** *J Clin Epidemiol.* 2011; **64**(7): 718–742.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Monsarrat P, Vergnes JN: **Data mining of effect sizes from PubMed abstracts: a cross-study conceptual replication.** *Bioinformatics.* 2018; **34**(15): 2698–2700.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- National Academies of Sciences, Engineering, and Medicine: **Data science for**

**undergraduates: Opportunities and options.** National Academies Press. 2018.  
[Publisher Full Text](#)

National Academies of Sciences, Engineering, and Medicine, Policy and Global Affairs, Government-University-Industry Research Roundtable: **Examining the Mistrust of Science: Proceedings of a Workshop—in Brief.** 2017.  
[PubMed Abstract](#) | [Publisher Full Text](#)

National Academies of Sciences, Engineering, and Medicine: **Reproducibility and Replicability in Science.** 2019.  
[Publisher Full Text](#)

Nosek BA, Spies JR, Motyl M: **Scientific Utopia: II. Restructuring incentives and practices to promote truth over publishability.** *ArXiv: 1205.4251 [Physics]*. 2012.

[Reference Source](#)

Nuijten MB, Assen MAL, van Hartgerink CHJ, et al.: **The Validity of the Tool “statcheck” in Discovering Statistical Reporting Inconsistencies.** *PsyArXiv*. 2017.  
[Publisher Full Text](#)

Ochodo EA, de Haan MC, Reitsma JB, et al.: **Overinterpretation and misreporting of diagnostic accuracy studies: evidence of “spin”.** *Radiology*. 2013; **267**(2): 581–588.  
[PubMed Abstract](#) | [Publisher Full Text](#)

Nosek et al.: Open Science Collaboration: **PSYCHOLOGY. Estimating the reproducibility of psychological science.** *Science*. 2015; **349**(6251): aac4716.  
[PubMed Abstract](#) | [Publisher Full Text](#)

Peng R: **The reproducibility crisis in science: A statistical counterattack.** *Significance*. 2015; **12**(3): 30–32.  
[Publisher Full Text](#)

Perry EH, Pilati ML: **Online learning.** *New Directions for Teaching and Learning*. 2011; **128**: 95–104.  
[Reference Source](#)

Prager EM, Chambers KE, Plotkin JL, et al.: **Improving transparency and scientific rigor in academic publishing.** *J Neurosci Res*. 2019; **97**(4): 377–390.  
[PubMed Abstract](#) | [Publisher Full Text](#)

Resnik DB: **Ethical virtues in scientific research.** *Account Res*. 2012; **19**(6): 329–343.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Schwartz LM, Woloshin S, Andrews A, et al.: **Influence of medical journal press releases on the quality of associated newspaper coverage: retrospective cohort study.** *BMJ*. 2012; **344**: d8164.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Schwitzer G: **How do US journalists cover treatments, tests, products, and procedures? An evaluation of 500 stories.** *PLoS Med*. 2008; **5**(5): e95.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Shavelson RJ, Towne L: **Committee on Scientific Principles for Education Research.** 2003; 204.

Steen RG, Casadevall A, Fang FC: **Why has the number of scientific retractions increased?** *PLoS One*. 2013; **8**(7): e68397.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Stodden V, Miguez S, Seiler J: **ResearchCompendia.org: Cyberinfrastructure for reproducibility and collaboration in computational science.** *Computing in Science and Engineering*. 2015; **17**(1): 12–19.  
[Publisher Full Text](#)

Teixeira da Silva JA, Al-Khatib A: **Ending the retraction stigma: Encouraging the reporting of errors in the biomedical record.** *Research Ethics*. 2019.  
[Publisher Full Text](#)

Thomas DM, Clark N, Turner D, et al.: **Best (but oft-forgotten) practices: identifying and accounting for regression to the mean in nutrition and obesity research.** *Am J Clin Nutr*. 2020; **111**(2): 256–265.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Thompson B: **“Statistical,” “Practical,” and “Clinical”: How Many Kinds of Significance Do Counselors Need to Consider?** *Journal of Counseling Development*. 2002; **80**(1): 64–71.  
[Publisher Full Text](#)

Thompson B: **Foundations of Behavioral Statistics: An Insight-Based Approach.** Guilford Press. 2006.  
[Reference Source](#)

Turner L, Shamseer L, Altman DG, et al.: **Does use of the CONSORT Statement impact the completeness of reporting of randomised controlled trials published in medical journals? A Cochrane review.** *Syst Rev*. 2012; **1**(1): 60.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Van Noorden R: **Online collaboration: Scientists and the social network.** *Nature*. 2014; **512**(7513): 126–9.  
[PubMed Abstract](#) | [Publisher Full Text](#)

von Elm E, Altman DG, Egger M, et al.: **Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies.** *BMJ*. 2007; **335**(7624): 806–808.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Vorland CJ, Brown AW, Ejima K, et al.: **Toward fulfilling the aspirational goal of science as self-correcting: A call for editorial courage and diligence for error correction.** *Eur J Clin Invest*. 2020; **50**(2): e13190.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Ware WB, Ferron JM, Miller BM: **Introductory Statistics: A Conceptual Approach Using R.** Routledge. 2013.  
[Reference Source](#)

Willig J, Croker J, Wallace B, et al.: **2440 Teaching rigor, reproducibility, and transparency using gamification.** *J Clin Transl Sci*. 2018; **2**(S1): 61.  
[Publisher Full Text](#) | [Free Full Text](#)

Yavchitz A, Ravaud P, Altman DG, et al.: **A new classification of spin in systematic reviews and meta-analyses was developed and ranked according to the severity.** *J Clin Epidemiol*. 2016; **75**: 56–65.  
[PubMed Abstract](#) | [Publisher Full Text](#)

# Open Peer Review

Current Peer Review Status:   

---

Version 1

Reviewer Report 19 February 2021

<https://doi.org/10.5256/f1000research.29358.r78363>

© 2021 Mische S. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



## Sheenah M Mische

Department of Pathology, New York University (NYU) Langone Medical Center, New York, NY, USA

This manuscript is a concise summary of a two day workshop held at Indiana University School of Public Health - Bloomington on identifying key opportunities for rigor, reproducibility & transparency (RRT) in research. This is not a research report, rather a report on the status of scientific research. The meeting attendance was invitation only and IU faculty staff and graduate students were joined by invited participants with recognized expertise in RRT, and reflected in the extensive references. Opportunities were focused in three key areas: 1) education and training, 2) reducing statistical errors while increasing analytical transparency and 3) improving transparency (truthfulness) and accuracy of research communications to promote accurate, objective, and truthful science. The article reads well with a focus on biomedical research.

### Specific Comments:

This manuscript provided an excellent summary of numerous and important challenges facing the research enterprise. There were no applicable outcomes or solutions. Of particular note:

1. Education and training: instructional modality, and understanding the relative benefits of various hybrid formats: there is no debate on the importance of RRT education and training. Both formal and informal forums are critical to research-integrity issues. Bringing scientific integrity issues into the open provides practical guidance for everyone from graduate students to faculty members. Faced with the pandemic, we all have adapted modalities to virtual platforms, emphasizing the importance of instruction regardless of format. Furthermore, formal instruction in rigorous experimental design and transparency is now required for NIH training, career development, and fellowship applications.
2. Reducing statistical errors while increasing analytical transparency and 3) improving transparency (truthfulness) and accuracy of research communications: sharing knowledge is what drives scientific progress—each new advance or innovation in biomedical research builds on previous observations. Experimental reports must have sufficient information to validate the original results and be verified by other researchers to be broadly accepted as credible by the scientific community. While statistics is a necessary for data interpretation by clinical researchers, psychologists, and epidemiologists whose conclusions depend

wholly on statistics, the interpretation of data in papers published in the biological sciences does not always require sophisticated statistical analyses; rather, diligent data reporting and transparency is essential.

**Conclusion:**

The authors summarize with “proposal, execution, and evaluation of the ideas presented herein showcases how the collective and interdisciplinary efforts of those investing in the future of science can solve problems in unique and exciting ways”. While appreciating this forward looking statement, the message is clear: the issue of reproducibility in science is complex and will continue to be debated and discussed in workshops such as this manuscript describes in the coming years. In response to well-publicized allegations of the inability to reproduce published biomedical research there have been numerous declarations of the components of trustworthy research and research integrity such as the Singapore Statement in 2010, the Montreal Statement in 2013, the Hong Kong Principles in 2019 and the European Code of Conduct for Research Integrity in 2017, and U.S. NIH and NSF Federal RRT policies. Ultimately we are all responsible for careful assessment of the rigor of the prior research, rigorous experimental design for robust and unbiased results by application of the scientific method; consideration of relevant biological variables and authentication of key biological and/or chemical resources used to conduct research and the use of numerical identifiers and the RRID syntax to improve communication of critical experimental details within the research community and to the public.

**Is the topic of the opinion article discussed accurately in the context of the current literature?**

Yes

**Are all factual statements correct and adequately supported by citations?**

Yes

**Are arguments sufficiently supported by evidence from the published literature?**

Yes

**Are the conclusions drawn balanced and justified on the basis of the presented arguments?**

Yes

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** Pathology, Technology, Shared Research Resource

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

Reviewer Report 18 February 2021

<https://doi.org/10.5256/f1000research.29358.r78351>

© 2021 Mebane C. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**Christopher A Mebane** 

Idaho Water Science Center, US Geological Survey, Boise, Idaho, USA

The article "Improving open and rigorous science...." is a report out on a workshop intended to make recommendations on improving rigor, reproducibility, and transparency (RRT) in interdisciplinary science. The idea of peer reviewing a workshop report is a bit of a curious assignment. What's a reviewer to say? No, those weren't the best topics to debate at your workshop, please reconvene and discuss something else? Raise questions about whether the article faithfully reports the workshop deliberations and consensus, when the reviewer wasn't there? As such this review is rather limited. The article reads well and has clearly been well vetted by the authors. The workshop and paper are interdisciplinary, although the focus is strongly slanted toward biomedical research and the health sciences.

**Not all errors are mistakes**

My only criticism of substance is the use of the term "statistical errors." Consider replacing it with "statistical mistakes" throughout the manuscript. In many fields, including mine (environmental science), the word "error" could refer to variability in the data, such as "the standard error of the mean." In other contexts, the word error is often used to describe the limits of precision. DNA and cells replicate with small errors, which over time lead to aging and senescence. In analytical chemistry, deviations from instrument values for calibration or quality control samples may be termed measurement error. Measurement error might refer to the inherent limits of a sensor in the instrument or the combined errors of the method. For example, in a bathymetric survey, errors accrue from inherent limits in the measuring distance as a function of sound through water, temperature changes in the water introduce error, a breeze adding motion to the boat introduces error, plants growing on the bottom muddy the signal increasing error, imprecision in the Earth's spheroid and canyon walls interfere with the GPS, and on and on. The hydrologist tries to reflect the accumulated error with a margin of error statement on overall accuracy. Those are examples of error – something the scientist always seeks to reduce and to accurately report the uncertainties associated with measurements, modeling, etc., but the presence of error is unavoidable. A mistake on the other hand is a blunder. Attaching the bathymetric sensor backwards, entering the wrong units into the calculations, using a long-wave, deep ocean sensor in shallow water, using the wrong datum, using a poorly suited method, neglecting calibrations, .... Just as with statistical mistakes, the topic of the argument, while there are often different appropriate methods of measurement for just about any scientific setting, some controversial or debatable methods, and some that are just plain wrong. The focus of the authors is on the latter – helping scientists avoid statistical blunders that are just plain wrong. I strongly urge you to call these "statistical mistakes" which is less ambiguous than "errors." There are supposed to be interdisciplinary RRT recommendations.

**Minor suggestions**

p7., in subsection titled "5. What is the prevalence and impact of errors," I thought the second paragraph was particularly dense and probably impenetrable to those not already in the know:

:  
*"Thus, [Subgroup 2] discussed the value of longitudinal, discipline-specific error surveillance and error frequency estimation to collect empirical evidence about error rate differences among disciplines. Other issues discussed were the identification of better prevalence estimates across fields, and how simulation studies can modify our confidence in the understanding of the prevalence of errors and their generalizability across disciplines."*

I think if you could expand on these points with some examples or examples with citations, readers might have better understanding of what is being recommended.

That's all. This was a tightly written report out of the workshop. Thank you for considering my rant about mistakes versus errors, where depending on the field and context, the latter is often a neutral descriptor of uncertainty.

**Is the topic of the opinion article discussed accurately in the context of the current literature?**

Yes

**Are all factual statements correct and adequately supported by citations?**

Yes

**Are arguments sufficiently supported by evidence from the published literature?**

Yes

**Are the conclusions drawn balanced and justified on the basis of the presented arguments?**

Yes

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** I am an environmental scientist who has published on related topics of research rigor, bias, and transparency in the environmental sciences.

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

Reviewer Report 28 January 2021

<https://doi.org/10.5256/f1000research.29358.r77096>

© 2021 Hewitt J. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The author(s) is/are employees of the US Government and therefore domestic copyright protection in USA does not apply to this work. The work may be protected under the copyright laws of other jurisdictions when used in those jurisdictions.



**Judith A Hewitt** 

Office of Biodefense, Research Resources and Translational Research, Division of Microbiology and Infectious Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD, USA

This is a concise and well written summary of a meeting of ~30 people on the vitally important topic of rigor, reproducibility & transparency. The meeting discussion questions were very well

formulated, though with the small size of the meeting and the limited number of invited participants outside of the university host, it is difficult to say whether the discussions, presented in a very succinct format of key challenges, is representative of all of the issues or viewpoints on the topic. Nevertheless, this appears to have been a good discussion that raised significant challenges. I would have preferred to see a bit more focus on solutions, as the challenges raised are all daunting.

### **Specific Comments:**

#### **Introduction:**

Regarding the statement that 40-70% of scientists agreed on factors contributing to irreproducibility, the original citation be used, (Baker, 2016; added<sup>1</sup>). Also, reference to the funder for the meeting is very much appreciated - but it is "Alfred P Sloan" not "Afred". In the last sentence, ""through to execution" is unwieldy - either "through" or "to" works but no need for both.

#### **Methods:**

I very much appreciate the list of participants and acknowledgement of honorariums - kudos on the transparency! I also appreciate knowing who participated in the small groups, but it would have been nice to see the agenda or titles of the Day One research presentations. Were those research or meta-research presentations? Also, "small-groups" should not be hyphenated, in fact you could just say three groups and let the reader come to their own conclusion about size; "breakout" is another useful term.

#### **Results:**

Subgroup 1, first paragraph: the following wording could be more precise by changing "three primary education-related questions" (where primary modifies education and not questions) to "three primary questions, education-related," or something similar. Precision of language is one of the articulated goals of training and communication in this article!

Q5, 2nd paragraph: I disagree with the first sentence, "To achieve the goal of error reduction, one must first know how pervasive errors are." I think any reduction in errors is a win, even without understanding the entire landscape, and needing to fully understand the landscape before attempting solutions is just kicking the can down the road. It's the "measurement" of error reduction or assessing progress toward a particular goal (which is not articulated) that requires knowing the pervasiveness first, and I agree that is extremely difficult to measure.

Q7, 2nd paragraph, last sentence: I question whether understanding "salary" costs of error correction is a valid pursuit, whether it's a case of pay now or pay later; page charges are a different matter.

#### **Conclusion:**

Since the Methods section stated that the meeting ended with a "call to action" to continue promoting interdisciplinary RRT science, I wonder if that call to action is accurately summarized? I found a great summary of the discussion but didn't walk away with a clearly articulated call to action in the very brief conclusion.

### **General Comments:**



I tend to agree that the challenges are many and difficult, though the small group discussions are distilled down to two challenges per question. They are mainly framed in negative terms, which is hard to read as a "call to action" without more detail. Nonetheless, the challenges raised are important and should be addressed, I'm just left scratching my head on what the next step is for many of these, given how they are stated.

I note that many of the references are from participants at the meeting, which may reflect the meeting content (difficult to judge without seeing the agenda), but does not necessarily instill in others an unbiased approach; this is perhaps a limitation of a small-meeting-by-invitation and could be formally recognized in the paper. This is not a value judgement on the references, indeed there is some balance, but it is a selected view that focuses on the meeting participants.

### References

1. Baker M: 1,500 scientists lift the lid on reproducibility. *Nature*. 2016; **533** (7604): 452-454  
[Publisher Full Text](#)

### Is the topic of the opinion article discussed accurately in the context of the current literature?

Yes

### Are all factual statements correct and adequately supported by citations?

Yes

### Are arguments sufficiently supported by evidence from the published literature?

Partly

### Are the conclusions drawn balanced and justified on the basis of the presented arguments?

Partly

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** infectious diseases; animal models of infectious diseases; translational research; scientific rigor; reproducibility

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

---

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact [research@f1000.com](mailto:research@f1000.com)

**F1000Research**