

## COMMENTARY

### Rigor, Reproducibility, and Responsibility: A Quantum of Solace



Lack of reproducibility in biomedical science is a serious and growing issue. Two publications, in 2011<sup>1</sup> and 2012,<sup>2</sup> along with other analyses,<sup>3</sup> documented failures to replicate key findings and other fundamental flaws in high-visibility research articles. This triggered action among funding bodies, journals, and other change agents. Here, I examine well-recognized and underrecognized factors that contribute to experimental failure and suggest individual and community approaches that can be used to attack these factors and eschew the SPECTRE of irreproducibility.

#### Shaken, Not Stirred

In some cases, irreproducibility may reflect methodological variation. The legendary secret agent James Bond appreciated that, just as in experimentation, technical details can be critical, even for tasks as apparently routine as preparing mixed drinks. Bond instructed the bartender that his martini should be shaken, not stirred. It's not clear why this was his preference, but 007 may have known that shaking increases the antioxidant content of a martini.<sup>4</sup> Alternatively, he may have simply wanted the release of small ice chips during shaking to ensure that the martini was very cold. In the lab similar small details, such as choice of pipette, can also be critical. For example, after several failed attempts to isolate lymphocytes by density gradient centrifugation, my lab realized that it had been impossible to maintain a sharp interface and avoid mixing solution layers of different densities using a serological pipette. Simply exchanging the serological pipette for a Pasteur pipette allowed successful implementation of a standard technique. Although we never doubted that it was possible to isolate lymphocytes using density gradients, extrapolation of this experience to non-standard procedures indicates that irreproducibility can reflect inattention to small, seemingly insignificant, experimental details. In this case, it was not the protocol, but rather the nuances, or art, of experimentation<sup>5</sup> that shifted the outcome. Perhaps owing to increased automation and computerization, there appears to be less emphasis on these subtle, but sometimes critical, procedural details than there was in the past. This source of error can be attenuated by teaching the importance of these skills, particularly when working with those who lack experience relevant to the task, and by documenting vital details in experimental protocols.

Inadequate documentation can also compromise reproducibility. My anecdotal observation is that protocols have become shorter and less informative. For example, I no longer see protocols specifying "dropwise" addition of one solution to another, as was once common. Method sections of published articles have become abbreviated. The strict

word limits enforced by many journals are one cause of this format shift, as shortening or eliminating the methods has become standard practice when words must be cut. Fortunately, many journals, including *Cellular and Molecular Gastroenterology and Hepatology (CMGH)* (which has never included methods in the word count), are working to reverse this trend by requiring detailed methods.

Inadequate analytical approaches are yet another underrecognized source of irreproducibility. These include inappropriate use of statistics,<sup>6</sup> failure to report key results, and insufficient biological replicates. In most cases this is not intentional subversion, but flawed decision making. For instance, how many independent samples are needed for RNA sequencing? Is it necessary to repeat RNAseq analyses in independent experiments? High cost, generation of massive datasets, and assumed repeatability of technical procedures are commonly used to excuse the absence of replicate studies. Previous over confidence in the reliability of microarray tools may be a cautionary tale in this case.<sup>7</sup>

To combat statistical misadventures, *CMGH* and other journals increasingly ask that each figure legend document the statistical test used, the number of biological replicates in the experiment shown, the number of independent times the experiment was performed, and whether all of the experimental replicates yielded similar results. *CMGH* also strongly encourages presentation of individual data points rather than simple bar graphs. While uncommon, editorial requests for implementation of these policies have occasionally revealed major deficiencies that called the reported results into question. In response to these and other concerns, the National Institutes of Health now requires training in the responsible conduct of research as part of all fellowships and career development programs. Inclusion of similar requirements is becoming prevalent in graduate, postgraduate, and ongoing faculty certification curricula. Biologically oriented statistics courses and texts have also become more common. Nevertheless, the quality of education in statistics, responsible conduct, and related topics is variable and, more importantly, the extent to which these experiences will impact scientists' behaviors is not yet known.

#### Specific Ingredients

Variation in reagent quality can also be a source of irreproducibility. To continue the metaphor, Bond's martini recipe included 3 measures of Gordon's, 1 of vodka, and half a measure of Kina Lillet. Had planned travel to malaria endemic areas led to 007's request for Kina Lillet, which contains the antimalarial remedy quinine? It is also curious that, despite naming the brand of gin, Bond provided no guidance in choice of vodka or even whether he preferred a corn- or potato-based variety. Did he simply expect the bartender to know which vodka should be used? Was this an oversight? Similar lapses can have serious consequences in the lab. For example, one could prepare a buffer using a

weak base with a mixture of the acid and  $\text{Na}^+$  salt, only the  $\text{Na}^+$  salt, or only with the acid. If the pH were adjusted correctly, all 3 solutions would have the same buffering capacity. However, if only the  $\text{Na}^+$  salt was used, addition of HCl to lower the pH would result in a solution with  $\text{Cl}^-$  and greater osmolality (reflecting increased  $\text{Na}^+$  and  $\text{Cl}^-$ ) than the other 2 preparations. Whether this would affect results depends on the planned use of the buffer. Similarly, a protocol specifying “MOPS buffer” could refer to inclusion of MOPS salts to buffer the solution or, alternatively, to the MOPS buffer used in electrophoresis. However, EDTA in the electrophoresis buffer would block  $\text{Ca}^{2+}$ -dependent processes. Thus, if used, for example, in an assay measuring activity of a  $\text{Ca}^{2+}$ -dependent kinase, inadvertent use of the EDTA-containing buffer would lead to false negative results. Failure to specify reagents in detail and the resulting use of different reagents in replicate experiments is, therefore, another potential source of irreproducibility.

Beyond reagent preparation within one's lab, it is also critical that biological reagents are reliable. The National Institutes of Health is attempting to address this by adding a section entitled “Authentication of Key Biological and/or Chemical Resources” to grant applications. Regarding cell lines, the instructions advise applicants to “describe the method they plan to use to verify the identity and purity of the lines, which might include short tandem repeat profiling and mycoplasma testing.” Is this too much, too little, just right? So much more is possible,<sup>8,9</sup> but is it necessary? The National Institutes of Health instructions also state that “If key resources have been purchased or obtained from an outside source... the investigator is still expected to provide their own authentication plans.” Despite catastrophic examples, including article retractions and failure of commercial entities,<sup>10</sup> poor practices persist. I have, for example, seen grant applications stating that, “we will only buy antibodies from reputable sources.” However vendors commonly sell one another's products, making it impossible to know who produced the antibody. Even if this approach was adequate, there are no standards for purification or validation. Incredibly, many polyclonal antibodies are sold as protein A purified immunoglobulins, which are no more specific than whole serum. Antibodies purified by antigen affinity chromatography should be a minimum standard, as this assures specific reactivity with the intended target. Cross-reactivity with other targets must, however, also be considered. Fortunately, validation data using protein arrays, knockout cells, or other approaches to exclude cross-reactivity are becoming increasingly prevalent. Finally, it is important to recognize that separate lots of polyclonal antibodies can have distinct characteristics due to interanimal variation in immune responses. Some companies acknowledge this by discontinuing old products when new sera are generated and marketing the new lots under new catalog numbers. Unfortunately, most manufacturers do not follow this approach, perhaps because of fear of losing consumers who repeatedly order the same item. This can cause great confusion, as the Research Resource Identifier code is tied to the catalog, rather than lot, number. Lack of transparency,

which, in effect, misleads the research community, may therefore be one more contributor to irreproducibility.

## The Right Glass

Finally, James Bond specified that his martini be served in a deep champagne goblet. Why? Perhaps he preferred it because it created a longer path for the small bubbles created by the shaking, thereby extending the interaction between gas and liquid. Alternatively, the smaller surface area of a champagne goblet, relative to a martini glass, may have had other effects. Would using a standard martini glass have significantly affected the drink? It is hard to know, but in the lab subtle changes in the quality of materials not usually thought of as ingredients, like Bond's glass, can be crucial. This applies, for example, to tissue culture-treated glass or plastic and other cellular growth surfaces. Several members of my lab recently noted a change in the behavior of cells grown on surfaces within a product that we had used for years. Further investigation revealed that one of the manufacturer's suppliers had gone out of business and that the manufacturer had found a new source for the material provided by that supplier. The manufacturer went on to produce and sell a final product that was visually indistinguishable from the original. Nevertheless, this nearly invisible change markedly affected experimental results. Because the manufacturer made no effort to inform the research community of the material change, the cause of irreproducibility in this case would not have been discovered if it weren't for some careful investigation. To be reliable, manufacturers must share this type of information. Failure to do so represents a breach of trust and yet another difficult-to-identify source of experimental irreproducibility.

## License to Kill

James Bond had license, or permission, to kill. As investigators and consumers, we have license, or authority, to effect change. We can emphasize procedural details, particularly with those performing procedures for the first time. We can insist on seeing all of the data, even those which do not support our hypotheses. We can make reagent validation standard practice in our own laboratories. We can also promote ethical behavior among our suppliers by favoring companies that, for example, stand behind their products by providing both meaningful validation and refunds for products that do not perform as advertised. Conversely, shunning vendors whose products fail frequently will send a strong message and, perhaps, eliminate sale of inferior or compromised materials. Finally, the scientific community must define best practices and insist on their adoption by ourselves, our colleagues, and our suppliers. By maintaining rigorous standards and taking responsibility for small details that can have enormous impact we can exercise our license to “kill” irreproducibility and, like 007, make the world (of science) a safer place.

JERROLD R. TURNER

Editor-in-Chief, *CMGH*

Brigham and Women's Hospital and Harvard Medical School

## References

1. Prinz F, Schlange T, Asadullah K. Believe it or not: how much can we rely on published data on potential drug targets? *Nat Rev Drug Discov* 2011;10:712.
2. Begley CG, Ellis LM. Drug development: raise standards for preclinical cancer research. *Nature* 2012; 483:531–533.
3. Ioannidis JP. Why most published research findings are false. *PLoS Med* 2005;2:e124.
4. Trevithick CC, Chartrand MM, Wahlman J, Rahman F, Hirst M, Trevithick JR. Shaken, not stirred: bioanalytical study of the antioxidant activities of martinis. *BMJ* 1999; 319:1600–1602.
5. Kneebone R, Schlegel C, Spivey A. Science in hand: how art and craft can boost reproducibility. *Nature* 2018; 564:188–189.
6. Chavalarias D, Wallach JD, Li AH, Ioannidis JP. Evolution of reporting P values in the biomedical literature, 1990–2015. *JAMA* 2016;315:1141–1148.
7. Ioannidis JP, Allison DB, Ball CA, Coulibaly I, Cui X, Culhane AC, Falchi M, Furlanello C, Game L, Jurman G, Mangion J, Mehta T, Nitzberg M, Page GP, Petretto E, van Noort V. Repeatability of published microarray gene expression analyses. *Nat Genet* 2009;41:149–155.
8. Lorsch JR, Collins FS, Lippincott-Schwartz J. Cell Biology. Fixing problems with cell lines. *Science* 2014; 346:1452–1453.
9. Yu M, Selvaraj SK, Liang-Chu MM, et al. A resource for cell line authentication, annotation and quality control. *Nature* 2015;520:307–311.
10. Baker M. Reproducibility crisis: Blame it on the antibodies. *Nature* 2015;521:274–276.

---

### Correspondence

Address correspondence to: Jerrold R. Turner, MD, PhD, AGAF, 77 Avenue Louis Pasteur, HNRB 730B, Boston, Massachusetts 02115. e-mail: [jrtturner@bwh.harvard.edu](mailto:jrtturner@bwh.harvard.edu).

### Conflicts of interest

The author discloses no conflicts.

### Most current article

© 2019 The Author. Published by Elsevier Inc. on behalf of the AGA Institute. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

2352-345X

<https://doi.org/10.1016/j.jcmgh.2019.03.006>