Elements

WRESTLING WITH REPRODUCIBILITY IN RESEARCH

Trish Dove and Skips Reaction at home in Virginia

With a mounting sense of alarm, the life sciences community has noted an increasing number of formal corrections to scientific papers. This trend, combined with a highly publicized report stating that only 25% of preclinical studies could be validated to the extent that projects could continue (Begley and Ellis 2012), has motivated discussions about how to improve reproducibility in research. Insights from recent workshops to examine the reproducibility of cancer and genetic research suggest that the "lessons learned" provide a timely opportunity for all scientific areas to review and update best practices.

At the center of the discussions is the issue of "irreproducibility"—a catch-all term that includes many things. Although the aspect that gets the most media attention is scientific fraud, studies show this is actually quite rare (Claxton 2005). It is no surprise that the vast majority of investigators are working to do the right thing.

...a good place to begin is a thoughtful assessment of whether a research question should be tackled with an exploratory or hypothesis-based approach.

More typically, irreproducibility is the unintentional result of poor experimental design, laziness, and sloppiness. Examples include studies that lack (or fail to report) appropriate controls, lack sufficient repetitions, have an accidental data selection bias, include statistical analysis that is inaccurate or absent, selectively report only a portion of the full data set, and suffer from publication bias—that is, peer pressure to conform to a certain result. More broadly, attendees of a workshop hosted by the US National Institutes of Health on this topic agreed that poor methods reporting, poor experimental design, and overstated findings often go hand-in-hand (Landis et al. 2012).

These observations raise the point that it is critical to provide students and postdocs with proper training in scientific data collection and analysis. For example, too few biologists receive adequate training in statistics and other quantitative aspects of their subject (MacArthur 2012; Russell 2013). Can the same be said of some Earth scientists? Are statistical methods covered in our curriculum and correctly applied in our own studies?

Earlier this year, Donald Berry remarked that this is a discipline-specific issue where educators—the senior scientists—can be the problem (PCAST 2014a). They set the stage, through training, for the habits and quality of next-generation research. Message to students: Alert! Ask the questions! Does your research have sufficient replication and error analysis? Do you need independent confirmation using orthogonal approaches?

Ironically, a new and doubly unintentional type of irreproducibility is emerging with the increasing sophistication of instrumentation. This is the situation where a single datum can be a challenge to collect or where researchers are simply unable to obtain the right training. Here, technology may be advancing faster than "best practices" can be established. This is particularly critical for deciphering the nuances of and artifacts from complex instruments. Another example is data analysis. If a particular instrumentation makes a big advance every five years but statistical training occurs only once per lifetime (every 30 years), a significant disconnect between modern statistical approaches and one's training becomes possible.

Computational studies do not face the same issues of reproducibility in the narrowest sense because the same code will always give the same result. However, there are significant needs in validation and verification. For example, rooting out and estimating errors can be even more elusive. One challenge is finding the correct boundaries of the model. Another involves the assumptions and modeling errors within the code itself. They can be so deeply buried and sophisticated that comparisons between researchers are limited (PCAST 2014a).

One implication of these insights for the Earth science community is to recognize that *reproducibility* is a spectrum of endeavor that is, to some extent, discipline-specific. Ioannidis and Khoury (2011) propose that the spectrum ranges from the minimum standard of *repeatability* to the gold standard of *replicability*.

Achieving *repeatability* means that another group can access the data, analyze them using the same methodology, and obtain the same result. Here, a collaborative spirit of freely comparing data across laboratories or research groups is a powerful means to building a community-based consensus.

Amicable cooperation is key to identifying differences in methodologies that can lead to different results and misinterpretation (Bissell 2013). Marcia McNutt notes this is critical for experimental studies because authors often have tacit knowledge that is not considered important to report, yet leads to an apparent irreproducibility (PCAST 2014b). By identifying these factors and encouraging greater communication among groups, apparent inconsistencies may melt away as differences are reconciled, and new insights may emerge.

At the other end of the reproducibility spectrum is *replication*. In this case, a study can be repeated from start to finish in an effort to obtain the same result by collecting new data, using new analysis, and employing fresh materials and reagents. This is typically the domain of biological studies that are designed to measure the responses of hundreds to thousands of mice, or cells in culture, etc.

Cont'd on page 324

THIS ISSUE

This issue on cosmogenic nuclides illustrates a frontier area that is fast moving thanks to improvements in analytical instrumentation. It is interesting to see that theoretical developments in particle physics have found applications in our effort to understand today's landscapes. And I can only salute the technological prowess needed to measure these rare nuclides—one atom in a million billion—as laid out in the Toolkit article. Guest editors Friedhelm von Blanckenburg and Jane Willenbring, together with the cast of authors they assembled, chose to illustrate how cosmogenic nuclides can help us understand Earth-surface processes.

2013 IMPACT FACTOR = 4.5

*Element*s' impact factor climbed from 3.156 in 2012 to 4.5 in 2013. As a reminder the 2013 impact factor measures the number of citations articles published in 2011 and 2012 received during 2013. Thus, *Elements*' articles are well cited, oftentimes in journals outside the mineralogy–geochemistry –petrology community. The 335 articles published in *Elements* up to the end of 2013 received 1555 citations in 2013, giving an average of 4.6 citations per article.

The issues that have been most cited since the time of publication are:

V3n1 – Zircon (534 citations up to June 2014)
v2n2 – Arsenic (368)
V4n5 – CO_2 Sequestration (275)
V1n5 – Large Igneous Provinces (270)
V2n6 – The Nuclear Fuel Cycle (240)

ERRATUM

In the August issue of *Elements*, in the article by Larter and Head [Larter SR, Head IM (2014) Oil sands and heavy oil: Origin and exploitation, Elements 10: 277-284], the cost of the Manhattan Project should have read US$22 billion instead of US$22 million (page 282, right column, line 14). This error has been corrected in the online version.

In the same article, some symbols in Figure 1a did not transfer when the files were transmitted to the printer. Please refer to the online version of the article at www.elementsmagazine.org.

ELEMENTS AT IMA

Kevin Murphy (right) with BSc honours geology students from the University of Limpopo. From left to right, Mr M. M. Langa, Ms K. A. Ramakadi, Mr L. F. Ledwaba, Ms P. M. Masapa, Mr M. E. Mogale. PHOTO CREDIT: JOHN DUNLEVEY

Many thanks go to Kevin Murphy who oversaw *Elements*' presence at the recent IMA meeting in Johannesburg, South Africa, and to everyone who spent time at the booth. Kevin reports the following from his experience:

"The *Elements* societies' booth at IMA was, as the name suggests, a joint affair and staffed by a range of representatives from several of the societies involved. One of the good things about the joint effort was that there was much of interest to attract delegates to the booth. A large selection of back issues of *Elements* on the table prompted lots of comments, from "oooh, that's my favourite one" to "I'd really like to get this magazine for myself." Also on display were sample copies of *Geochemical Perspectives*, *Mineralogical Magazine*, *Clay Minerals*, *Clays and Clay Minerals*, the CMS Workshop Lecture Series, the EMU Notes in Mineralogy series and the Reviews in Mineralogy & Geochemistry series. The booth was busy throughout the five days and especially so at break times. *Elements* representatives visited all of the other display stands to talk about possible advertising in *Elements* and to thank those who already support our magazine by doing so. At the end of the conference, left-over display material was donated to the relatively new University of Limpopo – they have gone to a good home."

Pierrette Tremblay, Executive Editor

EDITORIAL *Cont'd from page 323*

For Earth scientists, particularly those investigating questions in field settings, modern or ancient events can never be identically repeated, much less replicated. Whether one researches long-term processes or discrete events, such as storms, earthquakes, and eruptions, the Earth does not comply! Thus, true replication cannot be achieved. Our approach must be to collect data using well-designed protocols that include sufficient replications for meaningful statistical analysis and a detailed log of accompanying observations.

While there is no single way to conduct an experiment, a good place to begin is a thoughtful assessment of whether a research question should be tackled with an exploratory or hypothesis-based approach. Earth environments and the systems that we study are so complex that there always seem to be yet-unrecognized variables. Thus, I have come to think that much of what we do in the Earth sciences is exploratory, despite our best intentions to strictly test hypotheses. By accepting the limitations that each type of research can yield, we begin with a stronger framework for interpreting and predicting without overstatement.

At the end of the day, this must be a community-wide endeavor. The responsibility is shared governance with other researchers, reviewers, funding agencies, editors, and publishers. This is already underway. For example, journal editors are contributing to reproducibility by allowing more space for authors to describe their methods, by providing online repositories, and by selecting articles that provide an unusually excellent treatment of the data (Russell 2013).

The benefits are many. First, we must invest limited time and research money with utmost care. Second, the public nature of the reproducibility discussion affirms the strength of our scientific system because it is there for all to see. It upholds the validity and legitimacy of the scientific method. This is particularly important during an era when scientific findings and recommendations are sometimes treated with cynicism. Iron-clad data collection, replication, error analysis, and documentation are critical. We must be mindful when making claims of findings and their implications. In the long run, our diligence will build, not erode, public confidence.

Patricia M. Dove, Virginia Tech
Principal Editor

REFERENCES

Begley CG, Ellis LM (2012) Drug development: Raise standards for preclinical cancer research. Nature 497: 433-434

Bissell M (2013) The risks of the replication drive. Nature 503: 333-334

Claxton, LD (2005) Scientific authorship Part 1. A window into scientific fraud? Mutation Research 589: 17-30

Ioannidis JPA, Khoury MJ (2011) Improving validation practices in "Omics" research. Science 334: 1230-1232

Landis SC and 7 coauthors (2012) A call for transparent reporting to optimize the predictive value of preclinical research. Nature 490: 187-190

MacArthur D (2012) Face up to false positives. Nature 487: 427-428

Russell JF (2013) Reducing our irreproducibility. Nature 496: 398

PCAST (2014a) Improving Scientific Reproducibility in an Age of International Competition and Big Data: Researchers. Presidential Council of Advisors on Science Technology, January 31, 2014, www.tvworldwide.com/events/pcast/140131/globe_show/default_go_archive.cfm?gsid=2434&type=flv&test=0&live=0

PCAST (2014b) Improving Scientific Reproducibility in an Age of International Competition and Big Data: Editors. Presidential Council of Advisors on Science Technology, January 31, 2014, www.tvworldwide.com/events/pcast/140131/globe_show/default_go_archive.cfm?gsid=2435&type=flv&test=0&live=0