The problem with data

Lonnie W. Aarssen

Lonnie W. Aarssen (aarssenl@queensu.ca), Department of Biology, Queen’s University, Kingston ON, Canada K7L 3N6

The progress of science requires inspiration. Some researchers find this only from data. "Show me the evidence", they say. Many peer-reviewed publications in science, however, have no data. They involve a different kind of inspiration: proposals for original ideas or new hypothesis development. These are found within the 'Forum', 'Perspectives', 'Opinion' and 'Commentary' sections of many journals, and in some journals, like IEE, devoted entirely to new ideas and commentary.

I have always been particularly drawn to the honesty and beauty in this creative brand of enquiry. And so I am puzzled to hear it often dismissed out of hand with pejorative labeling, like ‘hand-waving’ and ‘just-so-stories’. Many—especially among the elites and self-appointed guardians of established theory—would have us believe that only ‘evidence-based’ practice and products can be taken seriously as legitimate sources of inspiration and discovery. This is plainly arrogant and wrongheaded. The scientific method means doing whatever is necessary to get good answers to questions worth asking. And so data collection that is not guided by interesting, novel, and important ideas is usually boring at best. At worst, it is a waste of research grant funds.

But published data are plagued with an even more serious problem: we never know how much to trust them. A few minutes of Google searching under the terms "research bias", "scientific misconduct", "publication bias", and “retractions” shows that the follies of faith in published data have come sharply and painfully into the public spotlight in recent years. The latest bad news is particularly troubling: most published studies are not reproducible (Baker 2015, Bartlett 2015, Begley et al 2015, Jump 2015). The statistical implication from this is unavoidable: it means that the results of at least half of all empirical research that has ever been published, probably in all fields of study, are inconclusive at best. They may be reliable and useful, but maybe not. Mounting evidence in fact leans toward the latter (Ioannidis 2005; Lehrer 2010; Hayden 2013).

Moreover, these inconclusive reports, I suspect, are likely to involve mostly those that had been regarded as especially promising contributions—lauded as particularly novel and ground-breaking. In contrast, the smaller group that passed the reproducibility test is likely to involve mostly esoteric research that few people care about, or so-called ‘safe research’: studies that report merely confirmatory results, designed to generate data that were already categorically expected, i.e. studies that aimed to provide just another example of support for well-established theory—or if not the latter, support for something that was already an obvious bet or easily believable anyway, even without data collection (or theory). A study that anticipates only positive results in advance is pointless. There is no reason for doing the science in the first place; it just confirms what one already knows must be true. This probably accounts for why the majority of published research remains uncited in the literature—or virtually so, attracting only a small handful of citations, many (or most) of which are self-citations (Bauerlein et al. 2010).

Are there any remedies for this reproducibility problem? Undoubtedly some, and researchers are scrambling, ramping up efforts to identify them [see Nature Special (2015) on Challenges in Irreproducible Research, http://www.nature.com/news/reproducibility-1.17552]. Addressing them effectively (if it is possible at all) will require nothing short of a complete restructuring of the culture of science, with new and revised manuals of ‘best practice’ (e.g. see Nosek et al.
Some of the reasons for irreproducibility, however, will not go away easily. In addition to outright fraud, there are at least six more—some more unscrupulous than others:

1. **Page space restrictions of some journals.** For some studies, results cannot be reproduced because the authors were required to limit the length of the paper. Hence, important details required for repeating the study are missing.

2. **Sloppy record keeping/data storage/accessibility.** Researchers are not all equally meticulous by nature. In some cases, methodological details are missing inadvertently because the authors simply forgot to include them, or the raw data were not stored or backed up with sufficient care.

3. **Practical limitations that prevent ‘controls’ for everything that might matter.** For many study systems, there are variables that simply cannot be controlled. In some cases, the authors are aware of these, and acknowledge them (and hence also the inconclusive nature of their results). But in other cases, there were important variables that could have been controlled but were innocently overlooked, and in still other cases there were variables that simply could not have been known or even imagined. The impact of these ‘ghost variables’ can severely limit the chances of reproducing the results of the earlier study.

4. **Pressure to publish a lot of papers quickly.** Success in academia is measured by counting papers. Researchers are often anxious, therefore, to publish a ‘minimum publishable unit’ (MPU), and as quickly as possible, without first repeating the study to bolster confidence that the results can be replicated and were not just a fluke. Inevitably of course, some (perhaps a lot) of the time, results (especially MPUs) will be a fluke, but it is generally better for one’s career not to take the time and effort to find out (time and effort taken away from cranking out more papers). When others do however take the time and effort to check, more incidences of irreproducible results make news headlines—news that would be a lot less common if the culture of academia encouraged researchers to replicate their own studies before publishing them.

5. **Using secrecy (omissions) to retain a competitive edge.** As Collins and Tabak (2014) note: “…some scientists reputedly use a ‘secret sauce’ to make their experiments work—and withhold details from publication or describe them only vaguely to retain a competitive edge.”

6. **Pressure to publish in ‘high end’ journals.** Successful careers in academia are measured not just by counting papers, but especially by counting papers in ‘high end’ journals—those that generate high Impact Factors because of their mission to publish only the most exciting findings, and disinterest in publishing negative findings. Researchers are thus addicted to chasing Impact Factor (IF) as a status symbol within a culture that breeds elitism—and the high end journals feed that addiction (many of them while cashing in on it). The traditional argument for defending the value of ‘high-end branding’ for journals (supposedly measured by high IF) is that it provides a convenient filtering mechanism allowing one to quickly find and read the most significant research contributions within a field of study. In fact, however, the IF of a ‘high end’ journal says very little to nothing about the eventual relative impact (citation rates) for the vast majority of papers published in it (Leimu and Koricheva 2005). A high journal IF, in most cases, is driven by publication of only a small handful of ‘superstar’ articles (or articles by a few famous authors). Journal ‘brand’ (IF) therefore has only marginal value at best as a filtering mechanism for readers and researchers.

Moreover, addiction to chasing Impact Factor, despite not delivering what its gate-keepers proclaim, is ironically at the heart of the irreproducibility problem—for at least two reasons. First, it fuels incentives for researchers to be biased in the selection of study material (e.g. using a certain species) that they already have reason to suspect, in advance, is particularly likely to provide support for the ‘favoured hypothesis’—the ‘exciting story’. Any data collected for different study material that fail to support the ‘exciting story’ must of course be shelved—the so-called ‘file drawer’ problem—because high end journals won’t publish them. Second, addiction to chasing IF can motivate researchers to report their findings selectively, excluding certain data or failing to mention the results of certain statistical analyses that do not fit neatly with the ‘exciting story’. This may, for example, include ‘p-hacking’—searching for and choosing to emphasize only analyses that produce small p-values. And obviously there is no incentive here to repeat one’s experiment, ‘just to be sure’; self-replication would run the risk that the ‘exciting story’ might go away.

All of this means that the research community and the general public are commonly duped—led to believe that support for the ‘exciting story’ is stronger than it really is. And this is revealed when later research attempts, unsuccessfully, to replicate the effect sizes of earlier supporting studies. Negative findings in this
context then, ironically, become more ‘publishable’ (including for study material that was already used earlier but ended up in a file drawer somewhere). Hence, empirical support for an exciting new idea commonly accelerates rapidly at first, but eventually starts to fall off (‘regression to the mean’) as more replications are conducted—the so called ‘decline effect’ (Lehrer 2010).

The progress of science happens when research results reject a null hypothesis, thus supporting the existence of a relationship between two measured variables, or a difference among groups—i.e. a ‘positive result’. But progress is also supposed to happen when research produces a ‘negative result’—i.e. results that fail to reject a null hypothesis, thus failing to find a relationship or difference. Science done properly then, with precision and good design but without bias, should commonly produce negative results, even perhaps as much as half of the time. But negative results are mostly missing from published literature. Instead, they are hidden in file drawers, destroyed altogether, or they never have a chance of being discovered. Because positive results are more exciting to authors, and especially journal editors, researchers commonly rig their study designs and analyses to maximize the chances of reporting positive results.

The absurdity of this contemporary culture of science is now being unveiled by the growing evidence of failure to reproduce the results of most published research. The results of new science submitted for publication today, in the vast majority of cases, conforms to researchers’ preconceived expectations—and virtually always so for those published in high end journals. This is good reason to suspect a widespread misappropriation of science.

Lots of fixing needed

Well-established researchers have always been privately aware of the above limitations of published data, and many have had direct knowledge of the scale of the problem—usually while showing a brave face for the public, intimating that the scientific establishment is virtuous, and it’s integrity (save for a few ‘isolated cases’), rock solid. This would be amusing if it weren’t so egregious and pathetic. Good data will of course always contribute importantly to the progress of science, and in a perfect world, ‘evidence-based’ would be the gold standard. But unfortunately, proclamations of evidence are rarely free of suspicion. Both researchers and the general public are well-advised to be wary of the potential for false confidence, questionable inspiration, and misguided recommendations.

Remedies for the data problem remain an important challenge for researchers. In the meantime, however, a rich and reliable source of inspiration—thankfully—remains strong in supporting the progress of science: published ideas and stories. Write them, publish them, read them, challenge them, revise them, and be inspired!

References

http://www.nature.com/news/reproducibility-will-not-cure-what-ails-science-1.18339?WT.ec_id=NEWS-20150910&spMailingID=49515108&spUserID=Mjk1NjczMjc0MgS2&spJobID=761320940&spReportId=NzYxMzIwOTQwS0