How to reduce scientific irreproducibility

Clare Fiala[1] and Eleftherios P. Diamandis[1 2,3]

[1] Department of Pathology and Laboratory Medicine, Mount Sinai Hospital, Toronto, Ontario, Canada

[2] Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Ontario, Canada

[3] Department of Clinical Biochemistry, University Health Network, Toronto, Ontario, Canada

*To whom the correspondence should be addressed:
Eleftherios P. Diamandis, MD, PhD, FRCP(C), FRSC
Head of Clinical Biochemistry,
Mount Sinai Hospital and University Health Network
60 Murray St. Box 32, Floor 6, Rm L6-201.
Toronto, ON, MST 3L9,
Canada Tel. (416) 586-8443
E-mail: Eleftherios.diamandis@sinaihealthsystem.ca

Word count: 1,247

It seems that these days everybody is talking about irreproducibility in science (1). But is this a new problem? Certainly not! There is now much awareness to it, since the number of irreproducible papers published in high-impact journals seems to be on the rise. In 2006 one of us wrote a mentorship paper in this journal addressing this problem (2). Also, John Ioannidis and others published repeatetly on the same issue, including one with the catchy title "Why most published research findings are false" (3). Can this menace of science be eliminated? Likely not, but it could be significantly reduced with some measures (see below).

There are at least 6 types of irreproducible results or papers in science (our definitions). Type 1 is due to fraud, which is supposed to be rare. You can Google the names Jan Hendrick Schön, Hwang-Woo-Suk, Haruko Obokata and Jon Sudbø for some of the most infamous frauds in science.

Type 2 irreproducible papers are usually written by prominent authorities, such as Nobel Prize winners, in an area that is unrelated to their Nobel-winning specialty. Some of these authors may suffer from a disease that I coined recently as "Nobelitis", which is related to the well-known "Hubris syndrome" (4, 5). In such cases, these authors overestimate their abilities due to their fame and publish papers that are not well-founded. A classic example of this is the case of the brilliant Chemist Linus Pauling (a double Nobelist) who claimed in the 1970s that mega-doses of vitamin C can cure cancer. Specialists in this field worked for years and succeeded in discrediting this notion but Pauling never accepted the verdict.

Type 3 irreproducibility is false discovery due to bias (pre-analytical/analytical/post-analytical) (6). Bias can lead to statistically significant differences between the comparison groups (e.g.

control subjects and cancer patients) that are due to something else other than the disease at hand. Ideally, scientists should control for every factor in the comparison groups but this is easier said than done. There are numerous examples of papers published in top-tier journals where there was obvious bias between the comparison groups. Here is one: comparing a test for prostate cancer diagnosis between a control group of 20-30 year old men vs prostate cancer patients (who are usually 60-70 years old). The difference between the groups could be due to cancer or to different age. Although this appears to be an obvious bias, we previously identified papers in top-tier journals which describe such inappropriate comparisons. (7). Other biases could be very subtle. For example, consider the case of a company called "Atairgin" which was formed in year 2000 with millions of starting investment, to commercialize a test for early ovarian cancer diagnosis. The company was folded a few years later after realizing that the observed differences between controls and ovarian cancer patients with their test were due to differences in the centrifugation speed of blood samples collected at two different locations. A very subtle bias that was very difficult to spot.

Irreproducibility type 4 is usually due to technical deficiencies of those who do the experiments and especially, unawareness of the limitations of the techniques used. We have plenty of experience with our own graduate students working with mass spectrometry instrumentation. They are familiar with the routine operation of the machines but generally lack in-depth knowledge of the technology. For example, some used the wrong parameters to search their results in public databases and others were using the wrong Zip-Tips (a solid-phase extraction device) for sample preparation. These small technical details could bring about big mistakes. As we usually tell our students "a small hole can sink a big ship".

Irreproducibility type 5, which is on the increase these days, is due to fragmented science. Mega projects are now executed in pieces in various laboratories and then the results are knitted together, usually by a single principal investigator. Since it is nearly impossible to know exactly what is happening in other sites, the PI accepts external results based on faith. This practice could lead to disasters, as was the case with the Haruko Obokata scandal (8). One of the co-investigators and Nature paper co-author, Teruhiko Wakayama, is an internationally recognized expert in stem cells. But it escaped to him that the Obokata results may be suspect. As fragmented science is expanding, this problem will likely become more acute.

Irreproducibility type 6 is associated with the recent trend towards big data generation and analysis by bioinformatics. As a Nature editorial puts it, "it is next to impossible to be absolutely sure that the bioinformatic tools are analyzing the huge data with great robustness and that small glitches in the bioinformatic tools may produce false interpretations". A classic example of this type of error is the case of crystallographer Geoffrey Chang who in 2007 retracted five high-impact papers (3 in Science, 1 in PNAS and 1 in J Mol Biol) due to a very small computer program glitch (9).

Is there a radical solution to this irreproducibility issues in science? Not really, but some recent efforts to minimize the problem are worth mentioning.

The reproducibility project was launched in 2013 with the goal of attempting to reproduce data from 50 cancer papers published in high-impact journals (10). Due to financial constraints the project was subsequently downsized and data on 5 papers are now published (11). The results were not clear-cut, precluding firm conclusions. In these author's opinion, the reproducibility

issue could not be solved with this project for four major reasons: high cost, inconclusive results, technical demands and slow process.

More recently, Mogil and MaCleod proposed inclusion of very strong preclinical data before publishing any full publication with perceived impact (12). I doubt that this suggestion will be effective since generation of such preclinical data will be expensive, time consuming and will likely delay publication of critical information in the scientific literature.

We here propose an alternative solution to address the issue of irreproducibility of papers with translational or commercial importance (seemingly "breakthroughs"). We suggest that authors of such papers are invited to provide a 5 (and maybe a 10 year) reflection on their papers. With this system, authors will be obliged to sign a declaration form (similar to the commonly used conflict of interest forms) in which they will agree to write a 5 year reflection. This should include information if the invention was translated or commercialized, or on what is the state of its development and any other hints (including identification of errors, misinterpretations or other roadblocks). We proposed this idea to some authors of such "breakthrough" papers a few years back, but none of them agreed to write such a reflection. With the advent of electronic publishing, it is now possible for journals to publish such a 5-10 year reflection, which could be an addendum to the original manuscript. This no cost procedure will allow readers to asses very quickly what happened to these "discoveries".

I believe that the obligation of authors to write a reflection on their high-impact papers will alert them to be more careful when doing the work and more conservative, thus avoiding overselling of their results.

We hope that some high-impact journals, especially in the Clinical Sciences, will adopt this proposal immediately, as a mandatory requirement for papers with seemingly large translational potential. The results of this experiment could be evaluated in the future, to see if this cost-free process contributes to reduction of irreproducible papers.

References

1. Schulz JB, Cookson MR, Hausmann L. The impact of fraudulent and irreproducible data to the translational research crisis - solutions and implementation. J Neurochem. 2016;139:253-270.

2. Diamandis EP. Quality of the scientific literature: all that glitters is not gold. Clin Biochem. 2006;39:1109-11.

3. Ioannidis JP. Why most published research findings are false. PLoS Med. 2005;2:e124.

4. Diamandis EP. Nobelitis: a common disease among Nobel laureates? Clin Chem Lab Med. 2013;51:1573-4.

5. Owen D, Davidson J. Hubris syndrome: an acquired personality disorder? A study of US Presidents and UK Prime Ministers over the last 100 years. Brain. 2009;132(Pt 5):1396-406.

6. Ransohoff DF. Bias as a threat to the validity of cancer molecular-marker research. Nat Rev Cancer. 2005;5:142-9.

7. Villanueva J, Shaffer DR, Philip J, Chaparro CA, Erdjument-Bromage H, et al. Differential exoprotease activities confer tumor-specific serum peptidome patterns. J Clin Invest. 2006;116:271-84.

8. Sugawara Y, Tanimoto T, Miyagawa S, Murakami M, Tsuya A, Tanaka A, Kami M, Narimatsu H. Scientific Misconduct and Social Media: Role of Twitter in the Stimulus Triggered Acquisition of Pluripotency Cells Scandal. J Med Internet Res. 2017;19:e57.

9. Miller G. Scientific publishing. A scientist's nightmare: software problem leads to five retractions. Science. 2006;314:1856-7.

10. Baker M, Dolgin E. Cancer reproducibility project releases first results. Nature. 2017;541:269-270.

11. Anonymous: The challenges of replication. Elife. 2017;6. pii: e23693. doi: 10.7554/eLife.23693.

12. Mogil JS, Macleod MR. No publication without confirmation. Nature. 2017;542:409-411