REPLICATION CRISIS [1]

The Replication Crisis is the term that describes the realisation that a very large part of the literature on science, perhaps half, is flawed. Professor John Ioannidis, a Stanford University mathematician who specialises in statistics of health and medical research, has a strong claim to have first recognised the Replication Crisis.

Ioannidis published a now classic paper that has been cited by thousands of other scientists with the striking title, "*Why Most Published Research Findings are False*" [2]. In this paper Ioannidis stated:

"There is increasing concern that in modern research, false findings may be the majority or even the vast majority of published research claims. However, this should not be surprising. It can be proven that most claimed research findings are false."

A remarkable finding, it is certainly not going unnoticed. It has been addressed by the big-name science journals for more than a decade and, within some, but certainly not the majority, of science institutions. It is remarkable that the media, with their interest in controversy, are largely unconscious of the Replication Crisis despite its being well known [3]. Fifty percent failure rate of reported science results would be better termed a replication scandal. Does any other profession have such a huge failure rate?

Replication is where a scientist tries to repeat the work of another scientist. It is fundamental to science. If research cannot be repeated by another scientist, and produce equivalent results, there is a problem. When multiple attempts to replicate fail, the original work, the original findings, must be regarded as wrong. In contrast, if multiple attempts to replicate succeed, a solid basis for relying on the work has been laid. The work becomes genuine "science".

Some fields of science are massively replicated and are utterly reliable. For example, the world relies upon Newton's laws of motion and gravitation every day when people travel in a car, walk across a bridge, or fly in a plane. Einstein's laws of relativity are tested every time the GPS on a phone is activated and when aircraft navigate the globe. These basic laws of physics are totally reliable within an uncertainty margin of a tiny fraction of a percent. Importantly for this well-replicated science, this uncertainty margin is accurately known so that the boundary between reliability and unreliability is established.

But most science is not massively replicated. Most science reported in specialist journals, and often reported in the media, is not replicated at all, and ***is thus not strictly science at all***.

Professor John Ioannidis should be given Nobel prizes for Physics, Chemistry, and Medicine, because the implications of his work will ultimately clean up science research and restore reliability. Compared with most Nobel prizes, Ioannidis's work is of far greater importance because it focuses on quality assurance

systems of ***all*** science.

Ioannidis highlighted the inadequate understanding of statistical methods that are often used in the health and medical sciences. Of these, possibly the best known is called P-hacking or data-dredging. With enough data and different parameters, it is almost always possible to find some correlation between two of the parameters.

For example, to find relationships regarding human health, hundreds of different types of measurements, such as blood pressure, ethnic background, weight, hair colour, length of the thigh, dozens of chemical tests of the blood, social data, and economic data can be used. Almost inevitably, by pure luck, there will be some correlation between two of them.

Wikipedia gives a telling example of this type of spurious correlation that has recently become common because we now have so much data about almost everything. Apparently, the number of letters in an American national spelling competition's winning word is correlated with the number of people in the United States killed by venomous spiders. [4]. So, it could be concluded that if this year's winning word was the 13-letter word "conscientious", then the number of people who would die this year of spider bite would be 13. This ridiculous result is as believable as 50 percent of biomedical research - it is just plain wrong. The result came from a spurious correlation that occurred entirely by chance.

In one study, a correlation between river discharge and crown of thorns starfish plagues that is almost certainly spurious was demonstrated. It is a classic example of the type of the P-hacking problem that Ioannidis warns about. Unlike the Wikipedia example, the correlation between river discharge and crown-of-thorns starfish plagues is being used as a basis for legislation that is affecting every farmer in north-east Queensland.

The financial costs of irreproducible biomedical research are large. In the United States alone it has been estimated [5] that the cumulative prevalence of irreproducible preclinical research exceeds 50 percent, and results in approximately US$28 billion per annum spent on research that is not reproducible. A similar analysis of Great Barrier Reef research has not been performed.

The Replication Crisis started to become widely known when Prinz et al. (2011), [6] of Bayer, the German drug company, writing in *Nature Reviews Drug Discovery,* claimed that 75 percent of the literature used for potential drug discovery targets is not reliable. For a drug company to take a promising scientific finding, perhaps made at a university laboratory, to a commercial drug will cost around $2 billion. An essential first step is to check the original finding. *The Economist* (19/10/2013) commented:

> *"A rule of thumb among biotechnology venture-capitalists is that half of published research cannot be replicated. Even that may be optimistic. Last year researchers at one biotech firm, Amgen, found they could reproduce just*

six of 53 "landmark" studies in cancer research."

In a sense this high failure rate of replication is not a problem for the drug companies. They check the work, find it is wrong and move on. But if science was working professionally, they could spend a few hundred thousand dollars in their replication tests and saved a couple of billion. But are rigorous replication tests happening in all areas of science - particularly in the environmental sciences? The answer is "No".

A concern over reproducibility is shared by editors of major journals. Marcia Angell, a former editor of *The New England Journal of Medicine,* lamented:

"It is simply no longer possible to believe much of the clinical research that is published, or to rely on the judgment of trusted physicians or authoritative medical guidelines. I take no pleasure in this conclusion, which I reached slowly and reluctantly over my two decades as an editor of The New England Journal of Medicine." **[7]**

A few years later, the editor of The *Lancet,* another prestigious medical journal, acknowledged that -

"The case against science is straightforward: much of the scientific literature, perhaps half, may simply be untrue. Afflicted by studies with small sample sizes, tiny effects, invalid exploratory analyses, and flagrant conflicts of interest, together with an obsession for pursuing fashionable trends of dubious importance, science has taken a turn towards darkness." [8]

Another observation in the same year reported there is:

"A growing concern regarding the replicability of findings in psychology, including a mounting number of prominent findings that have failed to replicate via high-powered independent replication attempts." **[9]**

In introducing a special edition on "Replicability in Psychological Science: A Crisis of Confidence', the editors, H. Pashler and E.J. VVagenmakers, asked:

"Is there currently a crisis of confidence in psychological science reflecting an unprecedented level of doubt among practitioners about the reliability of research findings in the field?" **[10]**

They answered themselves in the affirmative, warning that –

"Research findings that do not replicate are worse than fairy tales; with fairy tales the reader is at least aware that the work is fictional."

The problem has also recently been recognised in environmental science, with a call for "organised scepticism" to improve the reliability of the environmental marine sciences by Duarte et al. (2015) **[11]** and Browman (2016) **[12]**. In particular, Duarte et al. (2015) argue that some of the major threats to ocean ecosystems may not be as severe as is portrayed in some accounts, and that -

“The scientific community concerned with problems in the marine ecosystem [should] undertake a rigorous and systematic audit of ocean calamities. ... Such an audit of ocean calamities would involve a large contingent of scientists co-ordinated by a global program.”

Problems of replication are thus a big, well-known problem, but its true implications have not been fully examined, especially in the environmental sciences. Indeed, many environmental scientists appear ignorant of the existence of the Replication Crisis.

An example of this was demonstrated on July 2020, by the CEO of the Australian Institute of Marine Sciences (AIMS) giving evidence at a Senate inquiry [13] delving into possible deficiencies of scientific evidence being used to formulate regulations for the farming industry. When asked by Senator Roberts:

“Are you aware of the replication crisis, where it is regularly found that a large fraction of the peer reviewed literature - maybe 50 per cent in some estimates around the world - is in error?”

The response of the AIMS CEO was: *"Absolutely not.”*

Another example of Replication Crisis denial was demonstrated when Piers Larcombe and I wrote a paper explaining why the Replication Crisis needs to be addressed in the context of the Great Barrier Reef (GBR). The response from a large group of senior Great Barrier Reef scientists was that although there may be replication problems in the medical and biomedical sciences, GBR science was just fine, and -

“The quality control procedures that are applied to the GBR science are appropriate and fit for purpose." [14]

As the potentially erroneous science of the GBR affects every major industry in north-eastern Australia, such complacency is difficult to fathom. Indeed, it is astonishing that such a strenuous argument was made with the purpose of ***not*** doing a little more checking. It is a puzzle that there should be such resistance to the checking and replication so fundamental to the scientific method.

A key problem is that, until recently, it has been very difficult to get funding for replication studies as checking old work is not discovering "new" science. On one occasion, an application to the Australian Research Council to do replication studies on some GBR science was refused because, according to the funding rules, checking past work was not "new" and therefore not fundable.

The Australian Research Council, the major science funding agency in Australia with a budget of close to a billion dollars, is effectively prohibited from funding replication although it is essential to the scientific process. Any more graphic evidence of a failure in the system is not easy to imagine.

Considerable effort is now being made to improve the quality of science literature, especially in the biomedical area. Ioannidis has followed his 2005 paper on "*Why Most Published Research Findings are False*" with another entitled, "*How to*

Make More Published Research True": [15]

National research granting bodies, such as the Dutch and US equivalents of the ARC (the Royal Netherlands Academy of Arts and Sciences, and the National Science Foundation), have both recently instituted policies to fund replication studies. [16 & 17]. However, at the last time of checking (January 2020), nothing has changed at the Australian Research Council.

Notes:

1. **This reading is a precis of what was written on the topic by Ridd, Peter, "*Reef Heresy*", pages 160-166, Connor curt, Redland Bay QLD, 2020.**
2. **Ioannidis, J.P.A. (2005) "*Why Most Published Research Findings Are False",* PLoS Medicine, 2(8), p.e. 124.**
3. **Finkel, A. (2019). *"There is a problem":* Australia's top scientist Alan Finkel pushes to eradicate bad science. [online] The Conversation. Available at: https://theconversation.com/there- is-a - problem – Australia's- top- scientist -Alan – Finkel pushes-to-eradicate-bad-science-123374 [Accessed 12 Nov. 2019].**
4. **Data dredging. (2020). *Wikipedia.* [online] Available at: https:// en.wikipedia.org/wiki/Data dredging [Accessed 1 Jun. 2020].**
5. **Freedman, L.P., Cockburn, LM. and Simcoe, T.S. (2015). The Economics of Reproducibility in Preclinical Research. *PLOS Biology,* 13(6), p.el 002165.**
6. **Ioannidis, J.P.A. (2005) "*Why Most Published Research Findings Are False",* PLoS Medicine, 2(8), p.e. 124.**
7. **Angell, M. (2009). Drug Companies & Doctors: A Story of Corruption. *The New York Review of Books,* 56(1).**
8. **Horton, R. (2015). Offline: What is medicine's 5 sigma? *The Lancet,* 385(9976), p.1380.**
9. **LeBel, E.P. (2015). A New Replication Norm for Psychology. *Collabra,* 1(1), pp. 1-13.**
10. **Pashler, H. and Wagenmakers, E. (2012). Editors' Introduction to the Special Section on Replicability in Psychological Science. *Perspectives on Psychological Science,* 7(6), pp. 528-530.**
11. **Duarte,C.M., Fulweiler, R.W. Lovelock, C.E., Martinetto,P., Saunders, M.I. Pandolfi, J.M.,Gelcich, s., and Nixon,s.w. (2014) "*Reconsidering Ocean Calamaties",* Bioscience, 652(2), pp130-139.**
12. **Browman, H., (2016). "*Applying organised scepticism to ocean acidification research*." *ICES Journal of Marine Science,* 73(3), pp. 529-536.**
13. **Senate Standing Committees on Rural and Regional Affairs and Transport (2019). *Identification of leading practices in ensuring evidence-based regulation of farm practices that impact water quality outcomes in the Great Barrier Reef Terms of Reference.* [online] Parliament of Australia. Available at: https:/ /www.aph.gov.au/Parliamentary _Business/Committees/Senate/Rural_ and_Regional_Affairs_and_ Transport/ GreatBarrier Reef/Terms_ of Reference [Accessed 1 Nov. 2019].**
14. **Schaffelke, B., Fabricius, K., Kroon, E, Brodie, J., Death, G., Shaw, R., Tarte, D., Warne, M. and Thorburn, P. (2018). Support for improved quality control but misplaced criticism of GBR science. Reply to viewpoint "The need for a formalised system of Quality Control for environmental policy-science" by P. Larcombe and P. Ridd *(Marine Pollution Bulletin,* 126: 449-461, 2018). *Marine Pollution Bulletin,* 129(1), pp. 357-363.**
15. **Ioannidis, J.P.A. (2014). How to Make More Published Research True. *PLoS Medicine,* 11(10), p.e. lOOI747.**
16. **KNAW (2018). *Replication studies - Improving reproducibility in the empirical sciences.* Amsterdam: KNAW**

17. **Cook, EL. (201S). Dear Colleague Letter: Achieving New Insights through Replicability and Reproducibility. *National Science Foundation.* [online] 9 Mar. Available at: https://www.nsf.gov/pubs/201S/nsflS053/nsflS053.jsp [Accessed 1 May 2020].**