Hypotheses in the Life Sciences Volume 2 Issue 1 pp 1-3

What we know, what we don't

William Bains

SENS Lab., Chemical Engineering and Biotechnology, Cambridge University, Cambridge CB2 1QT, UK, and MIT (EAPS), Cambridge, MA 02139, USA

Abstract

Hypotheses have to be based on the real world. But what if the reported knowledge of the real world is itself wrong? Recent studies have suggested that a substantial fraction, maybe a majority, of the biomedical literature is at best flawed, at worst completely wrong. This means that scientists trying to make sense of the literature need to have a deep understanding of the subject, must look at the whole literature and not just a few selected papers for tests of a hypothesis, and if they build a hypothesis on a small number of outlier results, they must realise how weak its foundations may be.

Keywords: Hypotheses, life sciences, fraud, replication rate, journal

Correspondence Address: Dr William Bains, Rufus Scientific Ltd, 37 The Moor, Melbourn, Royston, Herts SG8 6ED, UK. EAPS, MIT, 77 Massachusetts Avenue, Cambridge, MA, 02139, USA. *E-mail:* editor@hy-ls.org

---

This journal asks that hypotheses are based on facts about the real world. But what is a fact? We can assume that if a scientist says that they ran a gel and saw a band, nearly all the times they are right. Outright fraud in the life sciences is mercifully rare: fraud involving biotechnology, for example, is more likely to involve finance than science, despite there being far more scientists than financiers in biotechnology institutes and companies [1]. If we judge by the number of papers that are retracted (Figure 1), the biomedical literature is an almost perfect record of accurate science, and the large majority of the retractions in the last decade have been for plagiarism or multiple publication[1], which, while unethical, do not invalidate the findings reported.

So does a new hypothesis about the life sciences have to take any published result as "true"? Recent studies suggest that it does not. Academic science is rarely repeated in the lab.. Journals will not accept a paper that says you have repeated someone's work exactly and got the same result. But pharmaceutical research frequently *does* exactly repeat prior work. A company has to be sure of the correctness of academic research before it bases an investment of upward of half a billion dollars [2] on that work in developing a medicine. Recently two pharmaceutical companies – Amgen [3] and Bayer [4] – have published statistics on how often they can replicate key academic findings. They find the replication rate is around 1 in 4. Around 75% of the leading studies they examined in detail could not be exactly replicated as described. Sometimes the reasons were minor, small details of experimental technique. Sometimes they were major, and the main conclusions of the study were just not replicable. Their finding was that a significant fraction of high impact scientific papers were wrong.

This was not a case of stupid industrial scientists failing to replicate the work of brilliant academics. Amgen and Bayer routinely publish in top journals, and the managers of the replication programmes always tried to contact the academic authors of the original papers to iron out discrepancies – differences in methods, undocumented protocols and so on. Nor is it atypical: my informal conversations with several other senior pharmaceutical company researchers suggest that Amgen's and Bayer's experience is entirely typical.

So is 75% of the biomedical literature wrong? That is an extreme interpretation. Another recent case study suggests that things are not that bad, but it is not good. Academics *do* on occasion try to replicate important or controversial results. When several groups working on the sirtuin pathway came up with conflicting results, some dug into the reasons for the discrepancies in more depth. Sirtuins were claimed to be central control proteins in the extension of lifespan by caloric restriction [5], and the path of action of resveratrol. What was found in the sirtuin case, and

[1] See the Retraction Watch web site for more on retractions and their causes http://retractionwatch.wordpress.com/

what I suspect is the cause of many other cases of non-replicability, is not that the results cannot be replicated, but that they are specific to a specific experimental system, which itself may be specific to a single lab.. These specific experiments were interpreted as having wide explanatory power, whereas in fact they were limited to particular strains of yeasts, *c. elegans* mutants, and the experimental conditions that only pertained in the research group that generated them [6]. To extend the example above, a specific band on a gel was extrapolated to apply to many bands on many gels – and it did not.

Experienced bench scientists know this for their own field. Any discussion with someone who has worked at the bench for a decade will sooner or later come to the comment "Well, we tried that, but it did not work" or "We do not believe that result because …". They are not saying that the scientists involved are deliberately lying. They are saying, as in the recent case of the "Arsenic Bacterium" GFAJ-1, that experimental design was poor, controls were not done, and as a result the conclusions made were not justified (see [7-11] on GFAJ-1.)

The finding of such a high rate of mis-interpretation and over-generalization is important for this journal. We receive many papers (and reject nearly all of them) that are based on a small number of oddball results from the literature, on which is built an inverted pyramid of hypothesis about how large sections of the life science are wrong. The problem is not that the papers are poorly argued (although they may be). The problem is that with enough ingenuity and time one can find a set of papers to support almost any hypothesis, providing you ignore everything else. HIV does not cause AIDS, genes are not coded in DNA, the sky is not blue – all of these will find "experimental" support somewhere. The chances are very high, though, that the "experimental" papers' conclusions are wrong.

Of course, it is possible that the oddball result is right and the majority of the literature is wrong. But if you argue that, you must make a strong case. You may be right. But to be effective as well as right you must convince at least a small number of readers. Just saying "They are wrong and I am right" is not really enough.

In this regard, I urge authors to follow Freeman Dysons' wise advice to young scientists [12]. The first thing you should do is look, really look hard, for evidence that disproves your hypothesis. A good scientist does not look for confirming evidence, but for tests that *disprove* her ideas. Only when he or she has exhausted every test they can think of do they consider the idea worth publishing. This needs deep understanding of not just what a paper's summary says, but of the methods used, the strengths and flaws of the experimental system, and of course of the prior and the citing papers that may point to different experimental systems that cast doubt on the experimenter's interpretation.

Hypotheses that are wrong, incomplete, even ones that are contradicted by known results, can still have value (13, 14). After all, the consensus may be wrong. But admit where the weakness lies, admit to the reports that flatly contradicts your idea, and explain why you chose the supporting data you use. In short, admit fallibility to reflect the fallibility of the literature. If you do not, your editorial team will do it for you, probably with a rejection.

References

[1]. Bains, W. (2006) Fraud and Scandal in Biotech. *Nature Biotechnology* 24: (7), 475-476

[2]. Bains, W. (2004). Failure rates in drug discovery and development: will we ever get any better? *Drug Discovery World* 2004(Fall): 9 - 18.

[3]. Begley, C. G. and Ellis, L. M. (2012) Drug development: Raise standards for preclinical cancer research. Nature 483: 531 – 533

[4]. Prinz, F., Schlange, T., Asadullah, Khusru (2011) Believe it or not: how much can we rely on published data on potential drug targets? Nature Reviews Drug Discovery 10: 712 - 713

[5]. Wood, J. G, Rogina, B., Lavu, S., Howitz, K., Helfand, S. L., Tatar, M., Sinclair, D. (2004_) Sirtuin activators mimic caloric restriction and delay ageing in metazoans. *Nature* 430: 686 - 689

[6]. Couzin-Frankel, J. (2011) Ageing genes: the sirtuin story unravels. *Science* 334: 1198 -1198

[7]. Schoepp-Cothenet, Nitschke1,B., Barge, L. M., Ponce, A., Russell, M. J., Tsapin. A. I. (2011) Comment on "A Bacterium That Can Grow by Using Arsenic Instead of Phosphorus" *Science* 332: DOI: 10.1126/science.1201438

[8]. Redfield, R. (2011) Comment on "A Bacterium That Can Grow by Using Arsenic Instead of Phosphorus" *Science* 332: DOI: 10.1126/science.1201482

[9]. Benner, S. (2011) Comment on "A Bacterium That Can Grow by Using Arsenic Instead of Phosphorus" *Science* 332 DOI:10.1126/science.1201304

[10]. Oehler, S. (2011) Comment on "A Bacterium That Can Grow by Using Arsenic Instead of Phosphorus" *Science* 332 DOI:10.1126/science.1201381

[11]. Reaves M.L., Sinha S., Rabinowitz J.D., Kruglyak L., Redfield R.J. Absence of arsenate in DNA from arsenate-grown GFAJ-1 cells. *Science* 2012 *in press*

[12]. Dyson, F. (1974) Cargo Cult Science. *in* Surely you're joking, Mr. Feynman Bantam Books (1985), pp308 - 317

[13]. Bains, W. (2009) Hypotheses and humility: ideas to not have to be right to be useful. *Bioscience Hypotheses*. 2(1): 1 – 2

[14]. Steinhauser, G. *and 192 other authors* (2012) Peer review vs. editorial review and their role in innovative science. *Medical Hypotheses in press*

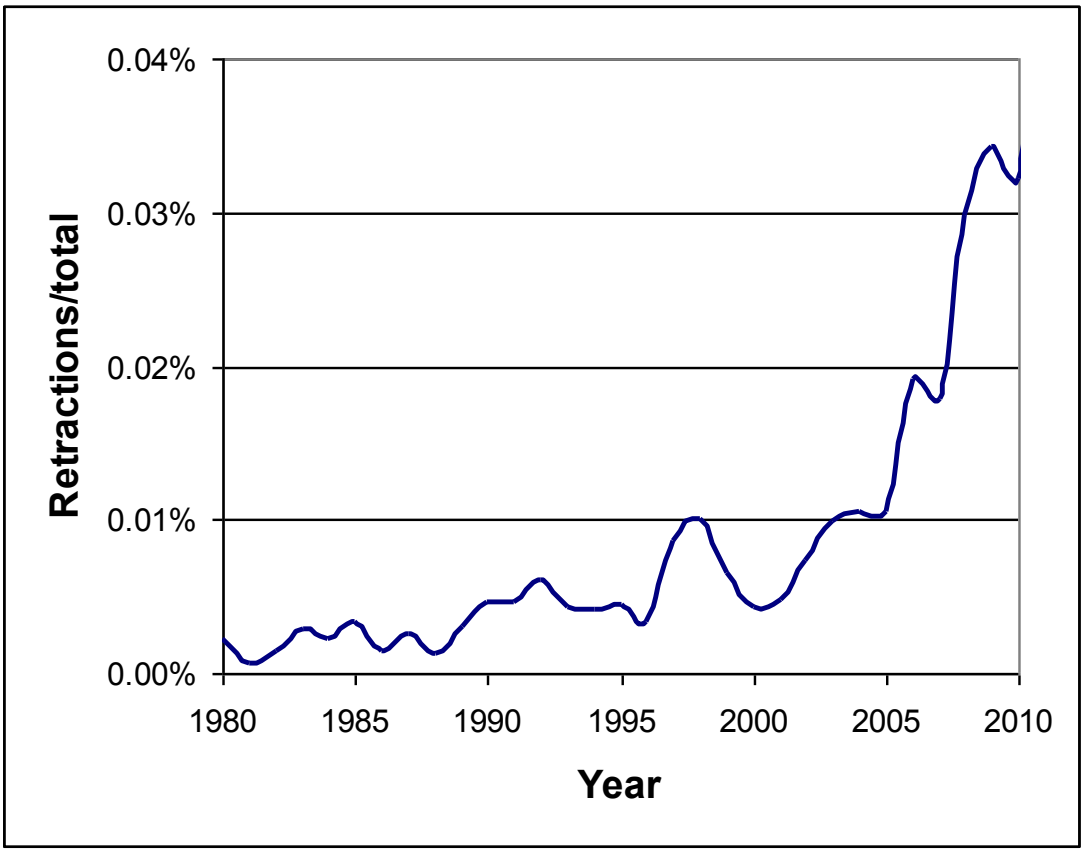

Figure 1. Fraction of papers indexed in Medline each year that are flagged as 'retractions', 1980 – 2010. Data extracted from PubMed using http://dan.corlan.net/medline-trend.html