

Reproducibility crisis timeline—milestones in tackling research reliability

December 6 2016, by Hilda Bastian

It's not a new story, although "the reproducibility crisis" may seem to be. For life sciences, I think it started in the late 1950s. Problems caused in clinical research burst into the open in a very public way then.

But before we get to that, what is "research reproducibility"? It's a euphemism for unreliable research or research reporting. Steve Goodman and colleagues (2016) say 3 dimensions of [science](#) that affect reliability are at play:

- Methods reproducibility – enough detail available to enable a study to be repeated;
- Results reproducibility – the findings are replicated by others;
- Inferential reproducibility – similar conclusions are drawn about results, which brings statistics and interpretation squarely into the mix.

There is a lot of history behind each of those. Here are some of the milestones in awareness and proposed solutions that stick out for me.

1950s

[Estes Kefauver](#) was a U.S. Senator and Adlai Stevenson's presidential running mate. He had become famous at the start of the decade with hearings into [organized crime](#). More than 30 million people watched some of those hearings, on television or free in movie theaters.

He turned his attention to the pharmaceutical industry at the end of the decade, holding [hearings](#) into drug prices in 1958 and 1959. One of the issues that emerged was the "sorry state of science supporting drug effectiveness". This outcry about research reproducibility led to a major change in research requirements for FDA approval in 1962.

Part of the problem's solution had a major milestone in [1959](#), too. [Austin Bradford Hill](#) led a meeting in Vienna that codified the methodology for controlled clinical trials.

Earlier in the 1950s, though, there was a development that stands as a milestone in fueling science's reproducibility crisis by inadvertently introducing a perverse incentive: Eugene Garfield proposed the journal impact factor in [1955](#).

1960s

In [1962](#), the Kefauver-Harris amendments required "adequate and well-controlled clinical investigations" for FDA approval: adequate was at least 2 studies – a major step in expecting replication of results.

A new problem in biological research was revealed in [1966](#) at a conference in Bethesda. Stanley Gartler was the first to report contamination in cell lines in cancer research (reported in Nature, [1968](#)). Still not adequately dealt with, that's grown into a juggernaut of unreproducibility, making thousands of studies unreliable.

That year also saw what may be the first example of the science of systematically studying science's methods – what John Ioannidis and colleagues have dubbed [meta-research](#). It was a study of statistics and methods of evaluation in medical journal publications (discussed [here](#)).

1970s

The 70s brought us 2 key methods: systematically reviewing evidence ([1971](#)) and meta-analysis: a way to analyze the results of several studies at once ([1976](#)) (explainer [here](#)). Those methods didn't just help make sense of bodies of evidence. They also propelled meta-research.

And [1977](#) surfaced a problem in scientists' behavior that allows unreliable scientific findings to take root and thrive. Michael Mahoney showed confirmation bias thrived in peer review:

Confirmatory bias is the tendency to emphasize and believe experiences which support one's views and to ignore or discredit those which do not...[R]eviewers were strongly biased against manuscripts which reported results contrary to their theoretical perspective.

1980s

Analyzing the reliability of studies took a formal step forward in [1981](#), with the publication of a method for assessing the quality of clinical trials.

Doug Altman and colleagues took on the issue of problems with the use and interpretation of statistics, with [1983](#) statistical guidelines for medical journals.

And a way forward was proposed for the problem of unpublished research results. The unpublished research results are often the disappointing ones. That leaves us with a deceptive public record of science.

John Simes called for an international registry of clinical trials in [1986](#).

Registering the details of all research before it is actually done would mean we could at least identify gaps in the published research record down the line.

1990s

According to Goodman & co ([2016](#)), the term "reproducible research" was coined in 1992. Computer scientist Jon Claerbout used it in the sense of methods reproducibility – enough published details for someone else to be able to reproduce the steps of a study.

Formal standardized guidelines for reporting research – one of the key strategies for trying to improve research reproducibility – arrived in clinical research [in the 1990s](#). The CONSORT statement for clinical trials in 1996 is the milestone here.

2000s

The registration of clinical trials at inception took a giant leap forward in [2004](#) when the International Committee of Medical Journal Editors (ICMJE) announced they would not publish a trial that had not been registered.

"Why most research findings are false" landed like a bomb on science's landscape in [2005](#). John Ioannidis' paper has now been viewed more than 1.8 million times, and definitely counts as a milestone in awareness of science's reproducibility problems. (I wrote about the article and controversies around it [here](#).)

In [2007](#), FDAA, the FDA amendments act, added public results reporting to an earlier requirement that drug clinical trials be registered at inception. Final Rules and NIH policies that give those requirements

teeth go into effect in January [2017](#).

2010s

Pre-clinical research got a major jolt in [2012](#), when Begley and Ellis at the biotech company Amgen set out to confirm cancer research findings:

Fifty-three papers were deemed 'landmark' studies...[S]cientific findings were confirmed in only 6 (11%) cases. Even knowing the limitations of preclinical research, this was a shocking result.

In [2013](#), registering studies at inception took a leap at the journal *Cortex* with registered reports. [Registered reports](#) are peer reviewed detailed protocols, with the journal's commitment to publishing the results after the study is completed. (A tally of journals following suit is kept [here](#).)

In [2014](#), the NIH began reporting on its strategies for addressing reproducibility, including [reporting guidelines](#) for preclinical research, and PubMed Commons, which had arrived in [2013](#). That's the commenting system in PubMed, the largest biomedical literature database.

Psychology's big jolt came in [2015](#), when the Open Science Collaboration reported that they replicated between a third and a half of 100 experiments and correlation studies.

And in 2016, the American Statistical Association (ASA) issued its first-ever guidance, trying to stem the tide of misuse and misinterpretation of p values. (Explainer [here](#).)

[2016](#) is ending, though, with a potential roll-back in the [clinical research](#) rigor and transparency required for FDA approval. Lesser levels of evidence than adequately controlled [clinical trials](#) might be back, and full

raw data might not be necessary, either.

We have a "reproducibility crisis" because we need to do more to try to prevent bias in research design, conduct, interpretation, and reporting. And after that we need others to interrogate what's found and replicate it in different contexts. As Christie Aschwanden [points out](#), that's just plain hard to do. But "science isn't really scientific without all of that, is it?"

This story is republished courtesy of PLOS Blogs: blogs.plos.org.

Provided by Public Library of Science

Citation: Reproducibility crisis timeline—milestones in tackling research reliability (2016, December 6) retrieved 18 April 2024 from <https://phys.org/news/2016-12-crisis-timelinemilestones-tackling-reliability.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.