RESEARCH · RESEARCH INTEGRITY

# Research on Research (Meta-Research) for Improving Knowledge Generation Practices

What meta-research is, the five areas it studies, what it has revealed about methods, reporting, reproducibility, evaluation and incentives, and how its findings are used to improve the way knowledge is produced.

ModQu research article · 1,109 words · 12 references (APA 7th edition)

Researchers study almost everything, including, increasingly, research itself. **Meta-research**, or research on research, uses the methods of science to examine how research is designed, conducted, reported, evaluated and rewarded. Its aim is practical: to find out where research goes wrong and to test ways of improving it. This article explains what meta-research covers, what it has found and how its findings are changing practice.

## What meta-research is

Ioannidis et al. (2015) defined meta-research as the study of research itself: its methods, reporting, reproducibility, evaluation and incentives. Ioannidis (2018) argued that it matters because the reliability and efficiency of all other research depend on the practices it examines. A single flawed method used across thousands of studies can do more harm than any one bad study.

Meta-research draws on statistics, epidemiology, psychology, sociology, economics and information science. Its studies take many forms: surveys of researchers, analyses of large samples of published papers, comparisons of protocols with publications, replication projects and randomised trials of interventions such as new review procedures.

## Five areas of study

The figure below sets out the five areas identified by Ioannidis et al. (2015), with examples of what meta-research has found in each.

### Five areas of meta-research

What each studies, with an example finding

| Area | Question | Example finding |
|---|---|---|
| **Methods** | How is research designed? | Many studies have low power |
| **Reporting** | Is it reported fully? | Outcomes switched after protocol |
| **Reproducibility** | Does it hold up? | Replication effects about half |
| **Evaluation** | How is it judged? | Reviewers often disagree |
| **Incentives** | What is rewarded? | Output rewarded over rigour |

Areas from Ioannidis et al. (2015) ModQu.com

Figure 1. The five areas of meta-research, with example findings. Areas from Ioannidis et al. (2015); table by ModQu.

## Methods

Meta-research examines how studies are designed and analysed. Button et al. (2013) showed that many studies in neuroscience had very low statistical power, which means that real effects are often missed and that significant findings are likely to overstate effects. Simmons et al. (2011) showed how flexible analysis can produce false positives.

## Reporting

Studies of reporting compare what researchers planned with what they published. Chan et al. (2004) found that outcomes in many randomised trials had been changed, added or omitted between protocol and publication. Fanelli (2010) found that the share of papers reporting positive results was implausibly high in some fields.

## Reproducibility

Large replication projects test whether published findings hold up. The Open Science Collaboration (2015) repeated 100 psychology studies and found that replication effects were, on average, about half the size of the original effects.

## Evaluation

Meta-research studies how research is judged, through peer review, metrics and funding decisions, and whether those judgements are reliable and fair. Studies of this kind have found, for example, that reviewers of the same manuscript or grant application often disagree, and that citation-based metrics reward some kinds of work over others regardless of quality.

### Incentives

Smaldino and McElreath (2016) modelled how rewards for publication output can favour weak methods over time, even when no one intends to cut corners. Studies of incentives ask what researchers are rewarded for and how this shapes what they do.

## How meta-research is done

Meta-research uses several distinctive designs.

**Cohorts of studies.** Researchers follow a defined group of studies, such as all trials approved by an ethics committee or registered in a given year, from start to publication. Comparing what was planned with what was published reveals selective reporting and non-publication.

**Large-scale literature analysis.** Thousands of published papers are analysed, by hand or with text-mining tools, to measure features such as sample sizes, reported p-values, statistical errors or the share of positive results.

**Replication projects.** Teams repeat published studies, often in coordinated multi-site efforts, to estimate how often findings hold up.

**Surveys and interviews.** Researchers are asked about their practices, pressures and attitudes, sometimes using techniques that encourage honest answers about sensitive behaviour.

**Experiments on the research process.** Interventions such as new peer review formats, reporting checklists or data-sharing policies are tested, ideally with randomised designs, to see whether they improve outcomes.

**Modelling.** Simulations explore how incentives and practices interact over time, as in the work of Smaldino and McElreath (2016).

## The scale of the problem

Chalmers and Glasziou (2009) combined estimates of losses at each stage of biomedical research: questions that did not matter to patients or clinicians, designs that could not answer the question, studies never published, and reports too incomplete to use. They estimated that as much as 85 per cent of investment in biomedical research was wasted. The figure has been debated, but it drew attention to how much value is lost through avoidable problems.

## From diagnosis to cure

Meta-research does not only identify problems; it tests solutions. Turner et al. (2008) used regulatory records to show how selective publication distorted the evidence on antidepressants, strengthening the case for trial registration. Kaplan and Irvin (2015) found that the share of large trials reporting a benefit fell sharply after advance registration of outcomes became mandatory, providing evidence that registration curbs flexible reporting. Scheel et al. (2021) compared Registered Reports with standard articles and found far fewer positive results, consistent with reduced publication bias.

These studies show the cycle that meta-research makes possible: measure a problem, introduce a reform, then measure again to see whether it worked.

## Who uses the findings

The findings of meta-research reach practice through several routes. Journals use them to justify new requirements, such as data availability statements and reporting checklists. Funders use them to design policies on registration, open access and data sharing. Universities use them in training and in revising how they assess researchers. Individual researchers use them to choose sample sizes, plan analyses and decide which published findings to trust. The field has also produced a growing set of tools, such as automated checks for statistical inconsistencies, that make some problems easier to detect.

## Meta-research beyond academia

The same approach can improve research in organisations. A market research team can audit its own past projects: how often did sample sizes allow reliable conclusions, how often were findings acted on, how often did follow-up studies confirm earlier results? Such internal meta-research reveals systematic weaknesses that no single project review would catch, and allows the team to test whether changes to its methods actually improve results.

## Limits

Meta-research is itself research and has the same weaknesses. Samples of studies may be unrepresentative, measures of quality are imperfect, and replication projects make their own choices that can be disputed. Meta-researchers have therefore increasingly applied open practices, such as preregistration and data sharing, to their own work.

## Conclusion

Meta-research turns the tools of science on science itself. It has shown that low power, flexible analysis, selective reporting and misaligned incentives are common, and that a large share of research effort is wasted. It has also shown that reforms such as registration and Registered Reports can work. As a discipline devoted to improving how knowledge is produced, it offers every field, including business research, a way to check and strengthen its own practices.

## References

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, *14*(5), 365–376. https://doi.org/10.1038/nrn3475

Chalmers, I., & Glasziou, P. (2009). Avoidable waste in the production and reporting of research evidence. *The Lancet*, *374*(9683), 86–89. https://doi.org/10.1016/S0140-6736(09)60329-9

Chan, A.-W., Hróbjartsson, A., Haahr, M. T., Gøtzsche, P. C., & Altman, D. G. (2004). Empirical evidence for selective reporting of outcomes in randomized trials: Comparison of protocols to published articles. *JAMA, 291*(20), 2457–2465. https://doi.org/10.1001/jama.291.20.2457

Fanelli, D. (2010). "Positive" results increase down the hierarchy of the sciences. *PLoS ONE, 5*(4), Article e10068. https://doi.org/10.1371/journal.pone.0010068

Ioannidis, J. P. A. (2018). Meta-research: Why research on research matters. *PLoS Biology, 16*(3), Article e2005468. https://doi.org/10.1371/journal.pbio.2005468

Ioannidis, J. P. A., Fanelli, D., Dunne, D. D., & Goodman, S. N. (2015). Meta-research: Evaluation and improvement of research methods and practices. *PLoS Biology, 13*(10), Article e1002264. https://doi.org/10.1371/journal.pbio.1002264

Kaplan, R. M., & Irvin, V. L. (2015). Likelihood of null effects of large NHLBI clinical trials has increased over time. *PLOS ONE, 10*(8), Article e0132382. https://doi.org/10.1371/journal.pone.0132382

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349*(6251), Article aac4716. https://doi.org/10.1126/science.aac4716

Scheel, A. M., Schijen, M. R. M. J., & Lakens, D. (2021). An excess of positive results: Comparing the standard psychology literature with Registered Reports. *Advances in Methods and Practices in Psychological Science, 4*(2). https://doi.org/10.1177/25152459211007467

Simmons et al. (2011) had already shown how undisclosed flexibility in collecting and analysing data could make almost any hypothesis appear to be supported. Their proposed remedy was disclosure: if researchers reported every decision, measure and exclusion, readers could judge how much flexibility had been used. Open materials, code and preregistration extend this logic by making the decisions visible as a matter of routine. There is early evidence that such measures work. Scheel et al. (2021) found that Registered Reports in psychology reported support for their first hypothesis far less often than standard articles, which suggests that publishing studies regardless of outcome gives a more accurate picture of the evidence.

Smaldino, P. E., & McElreath, R. (2016). The natural selection of bad science. *Royal Society Open Science, 3*(9), Article 160384. https://doi.org/10.1098/rsos.160384

Turner, E. H., Matthews, A. M., Linardatos, E., Tell, R. A., & Rosenthal, R. (2008). Selective publication of antidepressant trials and its influence on apparent efficacy. *New England Journal of Medicine, 358*(3), 252–260. https://doi.org/10.1056/NEJMsa065779