Title: Hidden Publication Bias in Epidemiologic Secondary Data Analysis

Authors: Rebecca C. Stebbins¹ and Stuart J. Ritchie¹

Affiliations:

1. Social, Genetic, & Developmental Psychiatry Centre; Institute of Psychiatry, Psychology, and Neuroscience; King's College London.

Word Count: 2,986 (max 3,000)

Abstract

We highlight a particular type of publication bias unique to secondary data analysis, and particularly common in epidemiologic research. We begin by setting a reminder of the scientific method of inquiry, and—by analogy with the movement for full transparency in clinical trials—present arguments for reporting all results of secondary data analysis. We then describe the ways in which data dredging—a subtle form a *p*-hacking—can lead to a distorted scientific literature; we highlight prior research that has empirically demonstrated this. We conclude by arguing that in order to combat this bias, epidemiologists should move toward preregistering analyses, and epidemiologic journals should encourage this through the implementation of Registered Reports. Finally, we respond to some common criticisms of preregistration.

Word count: 114 (unstructured, 150 max)

Reporting *all* the results

We all learn the scientific method in primary school: ask a question, formulate an hypothesis, then design an experiment to test it; analyse the data; draw conclusions; and present the results. This method, though simple at its core, is the foundation of all modern scientific inquiry and the bedrock on which science bases its claims of impartiality and integrity. However, in the modern era of academic journal impact factors, tenure metrics, click-bait headlines, and generally perverse incentives, this method has in some ways run aground. Today, publication, an important metric for promotion and prestige in academia, can often depend on finding flashy or “significant” results. This can often lead to cherry-picking of results to publish, in search of a “high impact” article.¹ This phenomenon, which substantially distorts the research we see published in the scientific literature, has been described at length elsewhere,²⁻⁴ but the application of these incentives to epidemiologic research—particularly “secondary” research on pre-existing datasets—remains underexplored.

In 2017, Brassington et al. presented arguments for reporting ALL results from clinical trials.⁵ The arguments are as follows:

1. *Scientific integrity*: a fully transparent literature should report all results, regardless of outcome.
2. *Patient welfare*: if null results are not reported in the literature, or if positive results are exaggerated, clinicians may be misled with regard to treatments.
3. *Avoiding waste*: publishing all results will diminish waste in cases where new trials are needlessly run on the same question.
4. *Ethics*: patients in clinical research have made a sacrifice (at least of time, but in many cases involving pain or discomfort), thinking they’re contributing to

the scientific literature. If all results are not published, we are not respecting that sacrifice.

We contend that each of these arguments applies to “secondary” data analysis in epidemiology research.

1. *Scientific integrity*: Regardless of whether the results support the initial hypothesis, the answer to a well-defined question is of interest. This applies no less strongly whether the data are pre-existing or collected specifically for a new project.
2. *Patient welfare*: Epidemiologic research ultimately aims to influence policy and have impacts on health outcomes via medical interventions, social welfare programs, and the like. Publishing all results is essential to making these decisions without bias.
3. *Avoiding waste*: Each year, time, effort, and millions of dollars in grant funding goes to research projects for which “the literature has not yet come to a consensus”. But decisions on which research questions to prioritise are often made based on a biased literature.
4. *Ethics*: Even though epidemiologists often use previously-collected data, participants have almost always agreed to store data/samples for future research, and that research should be published.

Although the reasons may not appear as strong or as pressing for research where participants are not put at immediate risk (i.e. they are not involved in a drug trial with potential side effects) or are not related to particular medical treatments or interventions, they still apply: the scientific literature is left incomplete if all results are not reported, with all the negative consequences that implies. Not only that, but with publicly-available datasets the

problem of failing to report analyses may be far worse than in analyses of clinical trial data.

In the next section, we explain why.

“Mucking about” in big datasets

There are several ways in which the scientific research that is ultimately published is a biased representation of analyses conducted.⁶ These include, but are not limited to, the following:

- Journals are less likely to publish null or ambiguous results;
- Scientists internalise the preferences of journals, and do not submit their null or ambiguous results for publication;
- Publication bias occurs within a single analysis (i): scientists can “*p*-hack”,⁷ that is, leave out non-significant results out or use multiple methods or analyses and report only the results that reach statistical significance;
- Publication bias occurs within a single analysis (ii): scientists can data-dredge, testing many relations within a dataset without a clear hypothesis until a statistically-significant result is found.

The last of these—data-dredging, or informally “mucking about” in the data—is a practice that is unfortunately easy to slip into even unconsciously and is perhaps the most insidious in epidemiologic research. The ready accessibility of data, sometimes available with a simple download of public datasets from the internet, or by writing a short and often vague project proposal to demonstrate one’s credentials as a ‘*bona fide* researcher’, gives epidemiologists the opportunity to conduct an untold number of exploratory analyses in search of an interesting ‘story’ for a research article. Substantial proportions of researchers in fields such as psychology,⁸ economics,⁹ and biomedical statistics¹⁰ admit to using such

practices; we have little reason to suspect epidemiology is different. Indeed, given the existence of many large observational studies with hundreds or thousands of variables, it may be even more tempting for epidemiologists to engage in data dredging, exploring a dataset in multiple ways to find the “right story”. This type of “Questionable Research Practice” is not only questionable due to the lack of scientific rigor, but also bias-producing in that multiple analyses, usually producing null results, are run informally by researchers, not necessarily as part of a pre-specified analysis, and not reported in research articles. They thus never become part of the “existing literature” that we use to justify our future research priorities. It is easy to imagine how this issue has been further amplified by the introduction of ‘big data’ to healthcare research.¹¹

Further biasing the published results is the issue of the “garden of forking paths” wherein researchers may select the analysis specific to the data they have.¹² In the field of epidemiologic research, though, the issue is not simply that of a different test being performed given different data; rather, given various assumptions, expertise, and preferences, a researcher makes hundreds of analytic decisions that could lead to a slightly (or drastically) different answer to the question.^{13,14} Often, these decisions aren’t made explicit prior to analysis, and are even changed after initial analyses in order to get certain results. These decisions can be as minute as selecting the functional form of an exposure variable in a regression model based on which form yields statistically significant results. A 2018 study demonstrated how these seemingly minute decisions can lead to a range of estimates to answer the same question in the same dataset.¹⁵ Though the results converge on a similar range of answers, $\frac{1}{3}$ of the results were not “statistically significant” and thus may not have been published if it were a “real” project. Fortunately, a re-analysis of this work found that by reanalysing the data with a clearer research question and precise definitions, results varied within a much more narrow range.¹⁶

It is not uncommon, in our experience, for a researcher to review preliminary (and often null) results and recommend as next steps to “look at it a different way”, implying that they are looking for a specific answer, and have no qualms about altering our analytic methods to find it. However, this practice is essentially a version of *p*-hacking: it is a method of asking one question in several ways—thus increasing the likelihood of false-positive results—and only publishing the one that provides a statistically significant result.

This method of scientific inquiry has harmful consequences for everyone: researchers, funders, and consumers of scientific output. From the beginning of our training, scientists are taught to “find the gaps in the literature” as the first step to motivate research. Students in particular spend countless hours reading and synthesizing the existing literature in their specific area in order to understand the state of the field and identify potential areas of further inquiry. However, when null findings are excluded from publication, they are excluded from students’ (and all scientists’) reading and synthesis – the literature appears as if it is full of significant findings but has glaring (or not-so-glaring) holes that the student could fill.⁶ This skewed synthesis is then at the core of all of the harmful consequences mentioned above: it can lead health care providers to give the general public inaccurate guidance on interventions, behaviours, etc; it can lead funders to provide funds to investigate research questions that have already been asked but the results never published; and it ultimately causes a great deal of waste.

Preregistration in Epidemiology

Despite the problems presented by these hidden biases, there are several potential solutions that epidemiologists and epidemiologic journals can and should move toward implementing. The first and simplest is pre-registration of the research question and planned methods before any analysis takes place. The level of detail in a pre-registration can vary

from highly general (e.g. merely stating the research question) to as specific as writing the entire background and methods section of the manuscript before the analysis, or even writing the analytic code, testing it on a simulated, “dummy” dataset, and posting it online before moving on to the real thing. Specifying and publicly posting these or other details prior to delving into the data can serve as a commitment device, holding researchers to their original aims regardless of the results. In epidemiology specifically, pre-registration of both the analytic plans as well as the conceptual models used to determine adjustment sets, e.g. DAGs, can force researchers to be more open and explicit about assumptions and ensure they stick to decisions made based on the most appropriate methods and theories, regardless of the results. In fact, a recent preprint by Barnard-Mayers argues for the publication of DAGs alongside research and suggests a template for such.¹⁷ We would encourage researchers to consider including these as part of a pre-registration.

Specifically for epidemiologic academic journals, pre-registration can take the form of Registered Reports.¹⁸ Registered Reports are a form of publication in which an initial peer review occurs before results are known and written, to focus specifically on the importance of the research question and appropriateness of the methodology.¹⁹ A journal may then agree to publish the paper based on this question, using the specified methods, without consideration to the ‘significance’ or ‘impact’ of the results, though still subject to peer review after completion. This is different from the pre-registration route in that the question and methods are submitted to a journal and adjudicated by peer-reviewers prior to analyses, rather than simply stated and made public, and that the journal agrees to publish the article. To our knowledge and at the time of writing, none of the major epidemiology journals offer Registered Reports as a publishing option. The list of journals that offer registered reports, maintained by the Centre for Open Science (<https://www.cos.io/initiatives/registered-reports>), does not include any of our field’s journals. Though we acknowledge that epidemiology

research is often published in substantive-area journals (e.g., research on cancer epidemiology is often published in cancer journals), and those areas would do well to consider Registered Reports as well, the research published in generalist epidemiologic journals would particularly benefit from this opportunity.

There are some aspects of Registered Reports that may be particularly beneficial to epidemiologic research. Most epidemiologic projects take advantage of pre-existing data, which makes the issue of looking through the data to see “what is there” is even more prevalent in our field. Having an analytic plan with pre-specified exposure and outcome relationships, covariates, and statistical methods would diminish that practice. Furthermore, there is sometimes a disconnect between epidemiologic methods research and development and the application of the most cutting-edge methods to substantive research. Registered reports could help bridge that gap by providing an opportunity for peer review on the analytic plan before the analysis has been run. Reviewers may suggest less-familiar but more appropriate methods to researchers who are less familiar or comfortable with new methods. A recent article by Van Calster, et al., suggests just that; the authors support Registered Reports as one component of a ‘top-down’ approach to solving the current lack of methodological rigor in medical research.²⁰ We agree wholeheartedly with their concerns regarding prioritization of metrics over methodology in the current medical research system, and indeed believe that the top-down approach these authors suggest—one led by journals and funding institutions—would be the most effective way to improve research quality quickly.

The clearest of the Registered Reports benefits, though not limited to epidemiology, is that journals will be committed to publishing the answer to the research question, regardless of the answer to the question (i.e. they will publish results even if the answer is null, or the results are ambiguous or confusing). When the journal agrees that a research question is interesting and the analytic plan is sound, the answer to that question remains of interest

regardless of the result. Registered reports in epidemiology journals would thus be enormously beneficial to the research community, solving one major driver of publication bias.

Criticisms of pre-registration

A 2020 Letter to the Editor of *Epidemiology* offered some criticism of registered reports and preregistration more generally.²¹ The authors noted some important considerations for the implementation of registration: first, that researchers might not follow their pre-registered analysis plan (evidence of this demonstrated in Claesen, 2021²²) or it may not be specific enough to combat the issues described here (see Kahan, 2021²³); second, that preregistration may in future become a mere box-ticking exercise, without the ability to shape researchers' behaviours (see also Buck, 2021²⁴); third, that, even in the Registered Report format, decisions on what to accept or reject may still be based on irrational criteria, such as the prestige of individual researchers.

These are all valid points, and should be dealt with by systems in place at the journal level: templates for pre-registrations, for example, could help in the process of matching plans to analyses that are eventually carried out, as could more widespread sharing of analytic code. However, in our view none of these issues come close to negating the positive aspects of preregistration and registered reports - the fundamental idea of registering an analysis (which is itself potentially peer-reviewed) is a clear positive step beyond the status quo, where no analysis plans are shared at all. In many cases, such plans are never even made in the first place.

Other objections to preregistrations centre around the additional time and effort required to preregister analytic plans (especially, perhaps, if those plans undergo peer review as part of a Registered Report). It is undeniable that producing a plan takes additional time and effort compared to an unplanned dive into a dataset. But we believe that many

researchers will see the benefits of increased transparency and openness as more than making up for the extra labour involved.

A final concern about preregistration concerns a possible adverse effect on scientific creativity: with our hands tied by a pre-registered analytic plan, is there still room for serendipitous findings - the lucky, unplanned accidents that have often led to advances in various scientific fields?

There are two responses to this objection. First, preregistration does not preclude further analysis: it merely requires that in the final writeup, scientists label which analyses were pre-registered and which were not. These could be separate subsections in the paper, or otherwise be clearly marked. Second, some journals have come up with innovative ways of publishing that allow researchers explicitly to explore a dataset as much as they see fit: the journal *Cortex* now accepts “Exploratory Reports”, specifically for hypothesis-generating, inductive research and which rely on full transparency of the methods and results presented in the final paper.²⁵ We would encourage epidemiology journals to consider formats of this kind - though they should bear in mind that exploration should be done with some pre-arranged constraints (e.g. on which variables are involved), otherwise it risks reproducing the same problems we describe above.

At base is the fundamental issue that many analyses in (and far beyond) epidemiology are performed using exploration but written as if they were entirely confirmatory – more transparency on which is which would allow readers of epidemiologic research to more accurately assess the quality of the evidence any given study provides.

Conclusion

Scientists value the ability to produce research quickly, and the world—particularly in times of health-related crises, like the coronavirus pandemic—is crying out for more research

to be done. But without robust quality control, extra speed dangerously distorts the scientific literature.

One controversial solution might be to advise scientists to *do less research* – spend more time on fewer projects, taking each analysis more seriously, and prioritizing quality over quantity.²⁶ There is a growing movement supporting more thoughtful application of causal methods and an increased use of descriptive studies in epidemiology^{27,28}—a “back to basics” approach, to better ground our etiologic hypotheses (relatedly, in psychology research, one radical proposal has been to *test fewer hypotheses*²⁹).

Realistically, though, fast research will always be in demand. There is, however, a set of tools that can increase the reliability and robustness of epidemiological “secondary analysis” without necessarily stopping scientists from testing as many hypotheses as their investigation requires. That tool is pre-registration. Whether organised by a journal or self-directed, a renewed focus on transparency—on reporting *all* the tests we run on our datasets—could restore a great degree of trust in the epidemiological literature.

References

1. Simmons JP, Nelson LD, Simonsohn U. False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological science*. 2011;22(11):1359-1366.
2. Easterbrook PJ, Gopalan R, Berlin J, Matthews DR. Publication bias in clinical research. *The Lancet*. 1991;337(8746):867-872.
3. Franco A, Malhotra N, Simonovits G. Publication bias in the social sciences: Unlocking the file drawer. *Science*. 2014;345(6203):1502-1505.
4. Edwards MA, Roy S. Academic research in the 21st century: Maintaining scientific integrity in a climate of perverse incentives and hypercompetition. *Environmental engineering science*. 2017;34(1):51-61.
5. Brassington I. The ethics of reporting all the results of clinical trials. *British medical bulletin*. 2017;121(1):19-29.
6. De Vries Y, Roest A, de Jonge P, Cuijpers P, Munafò M, Bastiaansen J. The cumulative effect of reporting and citation biases on the apparent efficacy of treatments: the case of depression. *Psychological medicine*. 2018;48(15):2453-2455.
7. Nelson LD, Simmons J, Simonsohn U. Psychology's renaissance. *Annual review of psychology*. 2018;69:511-534.
8. John LK, Loewenstein G, Prelec D. Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological science*. 2012;23(5):524-532.
9. Necker S. Scientific misbehavior in economics. *Research Policy*. 2014;43(10):1747-1759.
10. Wang MQ, Yan AF, Katz RV. Researcher requests for inappropriate analysis and reporting: A US survey of consulting biostatisticians. *Annals of internal medicine*. 2018;169(8):554-558.
11. Calude CS, Longo G. The deluge of spurious correlations in big data. *Foundations of science*. 2017;22(3):595-612.
12. Gelman A, Loken E. The garden of forking paths: Why multiple comparisons can be a problem, even when there is no “fishing expedition” or “p-hacking” and the research hypothesis was posited ahead of time. *Department of Statistics, Columbia University*. 2013;348.
13. Hoffmann S, Schönbrodt F, Elsas R, Wilson R, Strasser U, Boulesteix A-L. The multiplicity of analysis strategies jeopardizes replicability: lessons learned across disciplines. *Royal Society Open Science*. 2021;8(4):201925.
14. Huntington-Klein N, Arenas A, Beam E, et al. The influence of hidden researcher decisions in applied microeconomics. *Economic Inquiry*. 2021.
15. Silberzahn R, Uhlmann EL, Martin DP, et al. Many analysts, one data set: Making transparent how variations in analytic choices affect results. *Advances in Methods and Practices in Psychological Science*. 2018;1(3):337-356.
16. Auspurg K, Brüderl J. Has the Credibility of the Social Sciences Been Credibly Destroyed? Reanalyzing the “Many Analysts, One Data Set” Project. *Socius*. 2021;7:23780231211024421.
17. Barnard-Mayers R, Kouser H, Cohen JA, et al. A case study and proposal for publishing directed acyclic graphs: The effectiveness of the quadrivalent HPV vaccine in perinatally HIV exposed girls. 2021.
18. Registered Reports offer recognition for rigour. *Nature Communications*. 2020;11(1):3443.
19. Chambers CD, Tzavella L. The past, present and future of Registered Reports. *Nature Human Behaviour*. 2021:1-14.
20. Van Calster B, Wynants L, Riley RD, van Smeden M, Collins GS. Methodology over metrics: Current scientific standards are a disservice to patients and society. *Journal of Clinical Epidemiology*. 2021.

21. Bian J, Min JS, Prosperi M, Wang M. Are preregistration and registered reports vulnerable to hacking? *Epidemiology (Cambridge, Mass)*. 2020;31(3):e32.
22. Claesen A, Gomes SLBT. Preregistration: Comparing dream to reality. 2019.
23. Kahan BC, Morris TP, White IR, Carpenter J, Cro S. Estimands in published protocols of randomised trials: urgent improvement needed. *Trials*. 2021;22(1):1-10.
24. Buck S. Beware performative reproducibility. *Nature*. 2021;595(7866):151-151.
25. McIntosh RD. Exploratory reports: A new article type for Cortex. *Cortex*. 2017;96:A1-A4.
26. Stengers I. *Another science is possible: A manifesto for slow science*. John Wiley & Sons; 2018.
27. Lesko CR, Keil AP, Edwards JK. The Epidemiologic Toolbox: Identifying, Honing, and Using the Right Tools for the Job. *American Journal of Epidemiology*. 2020;189(6):511-517.
28. Conroy S, Murray EJ. Let the question determine the methods: descriptive epidemiology done right. *British Journal of Cancer*. 2020;123(9):1351-1352.
29. Scheel AM, Tiokhin L, Isager PM, Lakens D. Why hypothesis testers should spend less time testing hypotheses. *Perspectives on Psychological Science*. 2021;16(4):744-755.