

Is withholding your data simply bad science, or should it fall under scientific misconduct?

 blogs.lse.ac.uk/impactofsocialsciences/2015/07/03/data-secrecy-bad-science-or-scientific-misconduct/

7/3/2015

A recent study sent data requests to 200 authors of economics articles where it was stated 'data available upon request'. Most of the authors refused. What does the scientific community think about those withholding their data? Are they guilty of scientific misconduct? [Nicole Janz](#) argues that if you don't share your data, you are breaking professional standards in research, and are thus committing scientific misconduct. Classifying data secrecy as misconduct may be a harsh, but it is a necessary step.



I recently read a blog [post](#) by statistician Andrew Gelman, in which he commented on authors unwilling to share their data: "I'm not accusing [them] of scientific misconduct in not sharing their data." I immediately remembered how I [said](#) to a group of grad students and post-docs at [Berkeley](#) that not sharing your data is not really misconduct, because they are not plagiarizing or committing fraud.

But was I right in saying that? Is withholding your data simply bad science, or does it – should it – fall under *scientific misconduct*? This question is crucial because we need to find new ways to fight data secrecy. A study by [Krawczyk and Reuben](#) published in 2015 sent data requests to 200 authors of articles in economic journals, and to authors of working papers. Only 44% provided the data on request. We are not talking about data that cannot be shared due to confidentiality or privacy concerns – obviously it is fine not to make these data public. In fact, the study had not addressed authors that did not promise to publish their data. Only those who stated that 'data are available on request' were targeted. If we can punish data secrecy – and breaking promises – by labelling it misconduct, this could send a strong signal to the community.

Definition of scientific misconduct

What is scientific misconduct? Most definitions talk about the extreme cases of data fabrication, manipulation, and plagiarism, e.g. the National Science Foundation:

Research misconduct means fabrication, falsification, or plagiarism ... Research misconduct does not include honest error or differences of opinion. ([National Science Foundation](#))

The [National Institutes of Health](#) and the [American Psychological Association](#) use a very similar definition. And it makes sense to list the worst possible cases first and foremost when talking about misconduct. Fabrication means making up data or results. Falsification means manipulating your materials. Plagiarism means using ideas from others without credit. This is straightforward. However, there are cases when data secrecy should be added to the list of scientific misconduct examples.

Case 1: What if you try to cover up misconduct by hiding your data – is that misconduct in itself?

The UK's "[Concordat to support research integrity](#)" (which is signed by the UK Government, funders and universities) states that misconduct includes:

improper dealing with allegations of misconduct: failing to address possible infringements such as

Therefore, if you fail to provide information that shows you did not hide fabrication or falsification of your results, you are guilty of misconduct. For example, in the case of [LaCour](#)'s study on gay marriage that recently fell apart, data were manipulated, and in order to prevent anyone from finding out, the main author [deleted](#) his raw data. Most articles on the scandal saw all his actions as misconduct. If you cover up data manipulation or fabrication by 'withholding' your data, no one would doubt that this is part of the overall misconduct.

But what if you do not try to cover up any misconduct, but you simply don't want to share your data? Reasons for withholding data can include valid concerns such as patient privacy, confidentiality and copyright issues. [Savage and Vickers](#) found out in a survey among researchers that some authors withhold their data because they want to publish more articles with the data. Data collection can be expensive and time-consuming – and some simply want to keep the data exclusively to themselves for that reason. Unfortunately this means that no one can cross-check or replicate their results.

So should we see that as misconduct? Are these authors doing some form of harm to the advancement of knowledge out of self-interest, or are they simply being practical? Again, it depends on how you define misconduct.

Case 2: What if you break professional standards in your field – is that misconduct?

Yes! Some institutions state that it is scientific misconduct when you don't comply with your field's professional standards. For example, the National Institutes of Health [website](#) lists, after the usual fabrication, falsification and plagiarism problems, another requirement for "making a finding of research misconduct":

| *[If] there be a significant departure from accepted practices of the relevant research community"*
(National Institutes of Health)

Similarly, the UK's "[Concordat to support research integrity](#)" states that research misconduct is the "failure to meet ethical, legal and professional obligations" which includes "behaviour or actions that fall short of the standards of ethics, research and scholarship required to ensure that the integrity of research is upheld."

Based on such wider definitions that look beyond the usual extreme cases, it would not be far-fetched to say that when you withhold your data you don't meet professional obligations as a researcher. Of course, this would imply that your research community's professional standards include transparency and data sharing. And this is exactly the case.

[Professional guidelines](#) for political science state that "researchers have an ethical obligation to facilitate the evaluation of their evidence based knowledge claims through data access, production transparency, and analytic transparency." The American Psychological Association [affirmed](#) the principle that sharing data "promotes scientific progress" and "encourages a culture of openness and accountability in scientific research." Similar guidelines apply in economics, where one of the top journals [states](#):

| *It is the policy of the American Economic Review to publish papers only if the data used in the analysis are clearly and precisely documented and are readily available.*

If a researcher departs from these professional standards – according to the wider definitions I presented, scientific misconduct has occurred.

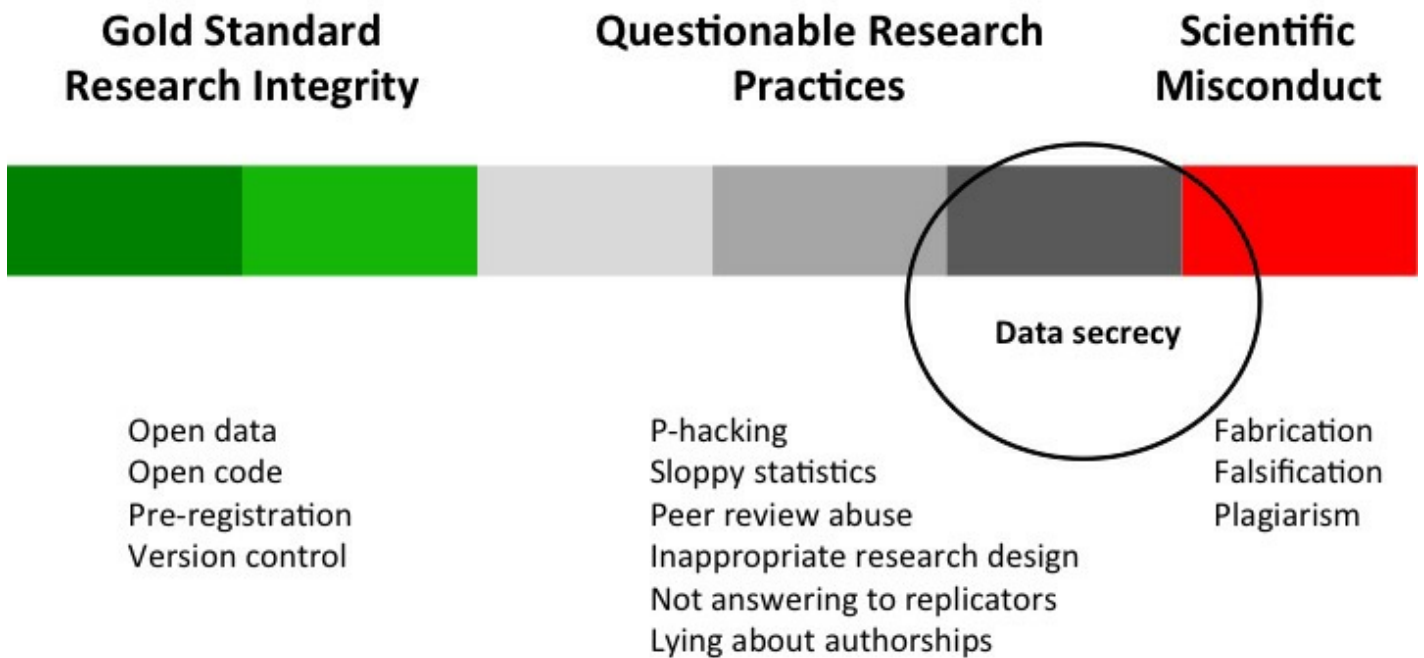


Figure: Representing good science versus misconduct, with a large grey area of Questionable Research Practices (QRP). Data sharing could be QRP or misconduct. Figure adapted from René Custers, VIB (2013).

My figure shows the scenario proposed here. On the left you can see features of good science, with authors providing their data and software code, and in the best cases even using pre-registration of their study and version control for maximum transparency. The grey area in the middle shows questionable research practices, which can include p-hacking, sloppy statistics, peer review abuse etc. On the right side and marked 'red' is scientific misconduct as commonly defined (falsification, fabrication, plagiarism). Between the grey and red are is data secrecy.

Some may argue that it is not actually misconduct, while I have argued that in some cases one could say that it is indeed misconduct: (1) when trying to cover up misconduct; (2) when deviating significantly from professional standards in your field.

In times where only [few authors](#) provide their data on request, classifying data secrecy as misconduct may be a harsh, but necessary step.

Note: This article gives the views of the author, and not the position of the Impact of Social Science blog, nor of the London School of Economics. Please review our [Comments Policy](#) if you have any concerns on posting a comment below.

About the Author

Nicole Janz is a political scientist at Cambridge University and teaches research methods, including a Replication Workshop. She [blogs](#) and [tweets](#) at [@polsci replicate](#).

- Copyright 2015 LSE Impact of Social Sciences - Unless otherwise stated, this work is licensed under a Creative Commons Attribution Unported 3.0 License.

20 Comments



Heidi Laine July 3, 2015 at 12:08 pm - Reply

Good post! I personally wouldn't go as far as calling data secrecy misconduct, even though I agree that it is ethically questionable. Research misconduct is pretty well defined as a concept. For example the Finnish Advisory Board on Research Integrity, which is responsible for one of the longest standing guidelines on the issue, divides research misconduct into fabrication, falsification, plagiarism and misappropriation. In Finland a suspected misconduct always requires an investigation and a "verdict" (the process is not legally binding, but has a lot of authority, since the entire Finnish research community has voluntarily signed into the procedure). Since data secrecy is currently pretty much the norm, it would be an administrative nightmare in Finland to start calling it misconduct. In my opinion it falls better under the heading of disregard for responsible conduct of research. The following is mentioned among the types of disregard listed in the Finnish guidelines: "inadequate record-keeping and storage of results and research data". The full text can be found here (it's in English): <http://www.tenk.fi/en/responsible-conduct-research-guidelines>



Buck Field July 3, 2015 at 5:28 pm - Reply

Thanks for linking to the guidelines Heidi – a great resource!

Clearly, data secrecy itself is not inherently misconduct and may be obligatory for privacy, etc.

It is true that the 4 most common species of misconduct listed wouldn't seem to apply even to deliberate, outright lying about data availability. However, there seems no obvious reason to assume the list is exhaustive, and it seems we ought to be able to apply the general scope of misconduct which precedes the subcategories. Presenting false (perhaps fraudulent) claims that source data is available on

request' seems to meet the criteria of "misleading the research community" and/or "misleading decision-makers", as well as "presenting false data or results", if we grant metadata can be fairly called a type of data.

That's the puritanical view, but there is the pragmatic concern you raise which would seem the strongest objection to Nicole's recommendation: the "administrative nightmare" consequence. That seems like a sufficiently strong claim as to warrant evidence, especially given the prima facie argument that if the verdict "has a lot of authority", researchers would be strongly motivated not to make false claims.



Buck Field July 3, 2015 at 12:15 pm - [Reply](#)

Falsification applies not only to data, but claims about the data and the author. Claims by authors that they will provide data when in fact they will not (unless under threat) is falsification – and in many professional contexts, certainly in the corporate/commercial world, there are very good laws against it.

It is not merely misconduct, it is a crime.

It is my sense that science should at least be able to rise to levels enforced on Wall Street bankers.



Fredrick Welfare July 3, 2015 at 7:01 pm - [Reply](#)

Data should be replicable, so one issue is why someone wants to look at another's data. It could be that researchers hold their data so that it cannot be misappropriated or misinterpreted by others. There is treachery in paradigm competition. The statement, "data provided upon request," might mean that data is provided to certain other persons after a vetting process. There are two important sources of information missing: editors of journals and books which expect authors to follow through on data provision should be interviewed, and leading researchers should weigh in on the reasons why data is provided or not to those requesting.



David Gibbens July 4, 2015 at 11:20 am - Reply

The original post is very interesting and raises some critical issues, but I think it is important to distinguish between two different aspects that are being slightly blurred. One is integrity: the other is competence or quality. "Misconduct" must refer to integrity; competence or quality is another matter.

So in the middle category of "questionable research practices" there is a big difference between sloppy statistics and "inappropriate research design" (quality) and "lying about authorships" (integrity). P-hacking, admittedly, does seem to lie in an interesting grey area.

I support what Buck Field has said: if someone has material published on the basis that they are going to share data and then they don't do so, their original statement was a lie: that's misconduct.

In general terms, it seems to me that the power lies with the publishers. Publishing research studies without ensuring that relevant data is also published is not exactly "misconduct" – but it can be construed as poor quality publishing. From time immemorial (OK the 17th century) presenting research findings in a way that allows others to replicate the processes leading to those findings has been a fundamental of scientific publishing. With research studies that present their results in the form of statistical analysis of data, the only way that you can replicate the analysis of the data is by actually having the data. Anyone who has worked with data and statistics can surely confirm the numerous ways in which error can creep in (deliberately or not). So it should be a sine qua non of normal research processes to make the data available.



Nate Breznau July 4, 2015 at 3:11 pm - Reply

Many associations, organizations and journals are on board with standardizing practices, let us hope things improve: <http://centerforopenscience.org/top/>. I am grateful that we this topic, although some of the extreme cases of falsifying data are helping raise

awareness ultimately in our favor (i.e. the favor of open science) (see a record of such cases at: <http://retractionwatch.com/>). I think it is not only bad for science but it is also an ethical issue. For example, the American Sociological Association's Code of Ethics (<http://www.asanet.org/images/asa/docs/pdf/CodeofEthics.pdf>; section 13.05) clearly states that Sociologists share data. Not that they 'should' share data but that they do. Acceptable reasons for not sharing data can be considerable cost burdens of the researchers in collecting the data which they can ask to partially recover, or if sharing the data compromises the anonymity of the subjects. I think these 200 refusing sharers would not be considered inline with these ethics, nor most others. Of course there are no consequences for these un-ethical practices, other than a reduction in human knowledge and scientific progress. Or, the occasional strong motivation of someone whose request is snubbed to then publish research that debunks others' findings as I experienced during my dissertation (Breznau forthcoming in Sociological Science and you can read more about the experience in a forthcoming blogpost at <https://politicalsciencereplication.wordpress.com>)



Peter Kenny July 4, 2015 at 3:23 pm - Reply

A number of articles in drug discovery research focus on models for prediction of biological activity and what are termed ADMET behavior (e.g. metabolic stability, permeability etc) which determines how readily a drug gets to its target. Data sharing appears to be the exception rather than the rule in the published drug discovery modeling literature and I have made the point in blog posts that certain articles would have packed a heavier punch had the authors shared their data.

A related problem is that it is not only the data which is not shared. Sometimes the authors do not disclose the models themselves although that doesn't inhibit them from making comparisons between different models. I have linked a post from my blog that illustrates the issue as the URL for this comment.

BibSonomy :: url :: [Impact of Social Sciences – Is withholding your data simply bad science, or should it fall under scientific misconduct?](#) July 4, 2015 at 7:32 pm - Reply

[...] [1 Impact of Social Sciences – Is withholding your data simply bad science, or should it fall under s...](#) [...]



Klaas van Dijk July 5, 2015 at 11:12 am - Reply

Good posting. I would like to make some additional comments.

.
(1). The "ESF-ALLEA European Code of Conduct for Research Integrity" (http://www.esf.org/fileadmin/Public_documents/Publications/Code_Conduct_ResearchIntegrity.pdf) also lists such an obligation to share raw research data.

.
* page 6. "1.4 Good Research Practices. (...). 1 Data. All primary and secondary data should be stored in secure and accessible form, documented and archived for a substantial period. It should be placed at the disposal of colleagues."

.
* page 10/11. "2.2.3 Integrity in science and scholarship. Principles. (...). These are principles that all scientific and scholarly researchers and practitioners should observe individually, among each other and toward the outside world. These principles include the following: (...) open communication, in discussing the work with other scientists (...). This openness presupposes a proper storage and availability of data, and accessibility for interested colleagues."

(2). The obligation to share raw research data is mandatory for any researcher affiliated to any of the Dutch universities and/or affiliated to any of the Dutch research institutes which are endorsing the 'The Netherlands Code of Conduct for Academic Practice' (<http://www.rug.nl/about-us/organization/rules-and-regulations/algemeen/gedragcodes-nederlandse-universiteiten/code-wetenschapsbeoefening-14-en.pdf>).

* page 8. "Principle 3 Verifiability. Presented information is verifiable. Whenever research results are published, it is made clear (..) how they can be verified. (..). 3.3. Raw research data are stored for at least ten years. These data are made available to other academic practitioners upon request, unless legal provisions dictate otherwise."

(3). There is a recent case at the University of Groningen (<http://www.rug.nl>) where two researchers have been found guilty of violating the rules of research integrity because they were refusing to share raw research data.

This case is covered by science journalist Frank van Kolfschooten in a recent article in the Dutch newspaper NRC (<http://www.nrc.nl/nieuws/2015/07/01/universiteit-integriteit-in-geding-bij-taalfoutonderzoek/>).

The researchers in question, Dr Anouk van Eerden and Dr Mik van Es, were unwilling to share raw research data from their PhD thesis to other Dutch researchers (Peter-Arno Coppen of Radboud University in Nijmegen, Carel Jansen of RUG and Marc van Oostendorp of the University of Leiden). These three researchers have filed a complaint to RUG when Dr Anouk van Eerden and Dr Mik van Es continued with their refusal to share these data. A Committee of RUG has decided that the allegations were founded. Dr Anouk van Eerden and Dr Mik van Es are not anymore affiliated to RUG, so RUG was unable to punish them. Please note that both Dr Anouk van Eerden and Dr Mik van Es had promised in public, during the defence of their thesis, that they should 'act in accordance with the Netherlands Code of Conduct for Scientific Practice'.



PG Bhat July 8, 2015 at 4:40 am - [Reply](#)

Thank you for this thoughtful article. In addition to research data, other public data are also hidden or manipulated to avoid public scrutiny.

Much public data held by Governments are hidden from the public. Thus, often they avoid being answerable to their misdeeds. Breaking professional standards in their work on data processing too is common in several government bodies.

Making these agencies accountable is not an easy task.

Reblog: Is withholding your data simply bad science, or should it fall under scientific misconduct? | Political Science Replication July 8, 2015 at 8:49 pm - [Reply](#)

[...] the LSE Impact Blog, I argue that if you don't share your data, you are breaking professional standards in research, [...]

Links 7/11/15 | Mike the Mad Biologist July 11, 2015 at 9:45 pm - [Reply](#)

[...] Is withholding your data simply bad science, or should it fall under scientific misconduct? NIH sees reversal of fortune with proposed funding boosts (while many scientists dislike Collins, he's pretty good at dealing with people on the Hill) Why Can't We Fall Asleep? Optically Clear Aluminum Provides Bulletproof Protection Texas Scientist With a Thing for Longhorns [...]

What is research misconduct and why it matters for open science | THE HONEST BROKER July 13, 2015 at 1:56 pm - [Reply](#)

[...] have recently come across some discussion concerning the concept of research misconduct and whether withholding data falls into the category. [...]

Potpourri: Statistik #23 | Erik Gahner Larsen July 22, 2015 at 10:41 am - [Reply](#)

[Potpourri: Statistik #23 | Erik Gahner Larsen](#) July 22, 2015 at 10:41 am - [Reply](#)

[...] How Can I Get Started Using R? – Is withholding your data simply bad science, or should it fall under scientific misconduct? – Data poverty makes it harder to fix real poverty. That’s why the UN should push countries [...]

[A Reblog of a Reblog from Political Science Replication | The Replication Network](#) July 25, 2015 at 8:44 am - [Reply](#)

[...] FROM THE ORIGINAL BLOG: “A recent study sent data requests to 200 authors of economics articles where it was stated ‘data available upon request’. Most of the authors refused.” Is this scientific misconduct? If so, what should be done about it? To read more, click here. [...]

[Horizon 2020 and the social sciences: an overview | Open Science](#) October 1, 2015 at 1:10 pm - [Reply](#)

[...] RDM is increasingly considered to be a part of good research practice, and no longer as a “nice-to-have” optional extra, and the impetus for this comes from both top-down (i.e. the introduction of funder and institutional data policies and mandates) and bottom-up (e.g. disciplinary norms which are dictated by fellow scholars. In fact, Nicola Janz, a political scientist at the University of Cambridge, has a recent blog post in which she questions whether withholding data unnecessarily is just bad practic.... [...]

[The Great Enrichment | Skeptical Swedish Scientists](#) October 31, 2015 at 3:24 pm - [Reply](#)

[...] rocks. A bit of critical thinking might help, or maybe not How about wayward scientists? Well, Nicole Janz is whipping them into shape: Hiding data = breaking professional standards = scientific [...]

Faut-il partager ses données? | Data Sciences Sociales November 20, 2015 at 6:13 pm - Reply

[...] qui refusent de donner à voir les coulisses du travail scientifique sont vite dénoncés comme corporatistes ou archaïques face aux modernes de la science ouverte. La meilleure preuve de cette modernité n'est-elle [...]

Is withholding your data simply bad science, or should it fall under scientific misconduct? | Nader Ale Ebrahim December 16, 2015 at 3:11 pm - Reply

[...] Sourced through Scoop.it from: blogs.lse.ac.uk [...]



KC Wheeler June 16, 2016 at 4:32 am - Reply

The history and political back-drop to the fight against the sharing of PACE trial research data, by researchers who appear to have misrepresented their results, is explained in this report: <http://www.centreforwelfarereform.org/news/misleading-mability-cuts/00270.html>

Unusual goings on with what was described as PACE's sister trial, FINE, have also occurred. Anonymised data had been made available as a part of PLoS's data sharing requirements, then this data was removed, and now the data has been put back in place: <https://forbetterscience.wordpress.com/2016/05/20/plos-correction-removes-previously-available-anonymised-patient-clinical-trial-data/>

A statement explaining the return of the FINE data would seem to undermine the PACE trial researcher's argument that anonymised data cannot be released without explicit consent from participants, yet they continue to fight against making available the data that would allow for the calculation of results for the trial's unreleased pre-specified outcome measures.

Considering the number of influential figures likely to be embarrassed by the release of the data from the PACE trial, it seems highly unlikely that those fighting against its release have any reason to fear disciplinary action.