



It is illegal to post this copyrighted PDF on any website.

HARKing, Cherry-Picking, P-Hacking, Fishing Expeditions, and Data Dredging and Mining as Questionable Research Practices

Chittaranjan Andrade, MD



Each month in his online column, Dr Andrade considers theoretical and practical ideas in clinical psychopharmacology with a view to update the knowledge and skills of medical practitioners who treat patients with psychiatric conditions.

Department of Clinical Psychopharmacology and Neurotoxicology, National Institute of Mental Health and Neurosciences, Bangalore, India (candrade@psychiatrist.com).

ABSTRACT

Questionable research practices (QRPs) in the statistical analysis of data and in the presentation of the results in research papers include HARKing, cherry-picking, P-hacking, fishing, and data dredging or mining. HARKing (Hypothesizing After the Results are Known) is the presentation of a post hoc hypothesis as an a priori hypothesis. Cherry-picking is the presentation of favorable evidence with the concealment of unfavorable evidence. P-hacking is the relentless analysis of data with an intent to obtain a statistically significant result, usually to support the researcher's hypothesis. A fishing expedition is the indiscriminate testing of associations between different combinations of variables not with specific hypotheses in mind but with the hope of finding something that is statistically significant in the data. Data dredging and data mining describe the extensive testing of relationships between a large number of variables for which data are available, usually in a database. This article explains what these QRPs are and why they are QRPs. This knowledge must become widespread so that researchers and readers understand what approaches to statistical analysis and reporting amount to scientific misconduct.

J Clin Psychiatry 2021;82(1):20f13804

To cite: Andrade C. HARKing, cherry-picking, P-hacking, fishing expeditions, and data dredging and mining as questionable research practices. *J Clin Psychiatry*. 2021;82(1):20f13804.

To share: <https://doi.org/10.4088/JCP.20f13804>

© Copyright 2021 Physicians Postgraduate Press, Inc.

In psychiatry and in psychopharmacology, as in other branches of medicine, researchers may knowingly or unknowingly violate scientific principles when analyzing their data and, afterward, when communicating their results in their research papers to the scientific community. This article uses examples to briefly present and explain certain of such questionable research practices (QRPs).

HARKing

HARKing means “Hypothesizing After the Results are Known.”^{1–3} HARKing is a QRP wherein a researcher analyzes data, observes a (not necessarily expected) statistically significant result, constructs a hypothesis based on that result, *and then presents the result and the hypothesis as though the study had been designed, conducted, and analyzed or at least oriented to test that hypothesis.* The italicized part of the definition constitutes the QRP in HARKing. This is because there is nothing wrong with constructing hypotheses after examining results; after all, many important discoveries in science were serendipitously made after, not before, generating and inspecting data. What is important, therefore, is that when hypotheses are made after the results have been examined, the analysis should be acknowledged as having been exploratory or hypothesis generating, and the post hoc hypothesis should be acknowledged as requiring to be confirmed in future research.

Why is HARKing problematic? If we choose a primary outcome to test a stated hypothesis *before* conducting a study and then find that the outcome is statistically significant, we can be fairly confident that the result is true for the population; this is a principle on which medical research is based.⁴ However, if we analyze data without preset hypotheses and *afterward* find something unusual or unexpected, that finding could be a chance finding.

Consider a hypothetical example. We compare patients who did and did not respond to antidepressant treatment. We find that the responders and nonresponders do not differ in age, sex, education, socioeconomic status, baseline severity of depression, presence of melancholia, and a clutch of other variables. However, we do find that the value of one variable, body mass index (BMI), is higher in the nonresponders. We now sit back and construct an elaborate explanation for how gut microbiota, inflammatory, insulin signaling, and other mechanisms related to overweight and obesity may influence neuronal functioning and reduce the likelihood of response to antidepressant treatment. Everything in our paper that reports the findings, from title and introduction to discussion and conclusion, focuses on this hypothesis that was constructed after the results were known.

Why is this wrong? When we compare a large number of variables between groups, by the laws of probability, one or more variables may, by chance, be statistically significantly different between the groups. This is a false positive or type I statistical error.^{4–6} In the context of

You are prohibited from making this PDF publicly available.

the example provided, we may have found, instead, that BMI was lower in the nonresponders. In such an event, we might have constructed an elaborate explanation for how lower BMI is a proxy for a diet that is deficient in micronutrients that are important for neuronal health and neurotransmission, and prepare our paper accordingly. If some other variable had been found to differ significantly between responders and nonresponders, we would have found other explanations. This is not to say that the finding is spurious. It merely means that when a finding may be a false positive finding, discovered in an exploratory analysis, this possibility needs to be explicitly communicated to readers.

The biggest danger of HARKing is that it may result in type I errors gaining lasting presence in mainstream scientific literature. The difference between HARKing and responsible reporting is that the latter will acknowledge that the analyses were exploratory and that the post hoc hypothesis, presented briefly in the discussion, requires formal examination in future research.

Cherry-Picking

People who pick cherries will not select fruit that appear unpalatable; so, researchers who cherry-pick are those who select and report only whatever supports their hypothesis.³ As an example, a researcher may find that, in an antidepressant trial, the study drug was superior to placebo on one but not another depression rating scale and for global improvement but not improvement in quality of life. The researcher cherry-picks only the significant outcomes for the paper that presents the findings; the nonsignificant outcomes are omitted as though those outcomes had not been studied. Or, when discussing the findings of their study, authors may cherry-pick for consideration research that favors their viewpoint and may criticize or even neglect to cite studies that do not support their arguments. Cherry-picking is a QRP because the reader is deceived into seeing a picture that is more favorable than it truly is.

In an interesting example using real data, Mayo-Wilson et al⁷ showed how strikingly different the results of a meta-analysis could be if authors wished to cherry-pick individual study outcomes to support or discredit hypotheses about the efficacy of gabapentin for neuropathic pain and the efficacy of quetiapine for bipolar depression.

P-Hacking

P-hacking is a QRP wherein a researcher persistently analyzes the data, in different ways, until a statistically significant outcome is obtained; the purpose is not to test a hypothesis but to obtain a significant result. Thus, the researcher may experiment with different statistical approaches to test a hypothesis; or may include or exclude covariates; or may experiment with different cutoff values; or may split groups or combine groups; or may study different subgroups; and the analysis stops either when a significant result is obtained or when the researcher runs out of options. The researcher then reports only the

approach that led to the desired result.^{3,8} P-hacking is very obviously a QRP because the researchers have decided in advance what the data should show.

Fishing Expeditions, Data Mining, and Data Dredging

The term *fishing expedition* is used to describe what researchers do when they indiscriminately examine associations between different combinations of variables not with the intention of testing a priori hypotheses but with the hope of finding something that is statistically significant in the data. As an example, a researcher may test every possible sociodemographic, clinical, radiologic, and biochemical variable with every outcome variable available to identify possible predictors of antidepressant response. Very obviously, because of the large number of statistical tests involved, such an exercise would be associated with a high risk of false positive findings. Fishing expeditions may be followed by HARKing.

Data dredging or mining are fishing expeditions that are carried to an extreme. *Data dredging* and *data mining* are synonymous terms that describe the extensive testing of relationships between variables for which data are available in a study or database. The difference between P-hacking and data dredging is that whereas P-hacking usually refers to the dragging of statistical significance out of data related to one or more hypotheses of interest, data dredging is the extensive search for significant relationships in a dataset without necessarily having specific hypotheses in mind. As with fishing expeditions, with data dredging the probability of false positive findings is very high because of the very large number of statistical tests conducted.⁹

Data mining may be ethically done in health care and other fields, such as when searching for new leads in anticancer drug development and when studying “big data”; the data mining approach is clearly acknowledged as such.¹⁰⁻¹² In this specific context, therefore, the term *data mining* is not pejorative. The term *data dredging*, however, is restricted to the context of QRP.^{13,14}

Comments

Many researchers who HARK do not realize that presenting their post hoc formulations as a priori formulations amounts to scientific misconduct. Cherry-picking favorable over unfavorable evidence is more obviously dishonest because researchers cannot fail to recognize that they are misrepresenting the evidence. P-hacking comes the closest to deliberate cheating because researchers are forcing the data into a conclusion that they have already drawn. Fishing expeditions and data dredging are QRPs because they build on patterns that may not exist in other datasets; the exercises may, however, be mitigated by due acknowledgment of the process. Whether data mining is a QRP or not depends on the context and on how it is acknowledged.

Researchers who are involved in these QRPs may not necessarily realize that their actions are scientific misconduct; this is especially true of researchers who are

It is illegal to post this copyrighted PDF on any website.

young, inexperienced, and inadequately mentored. It is therefore important for knowledge about these QRPs to become widespread and for readers and researchers to know what approaches to statistical analysis and reporting amount to scientific misconduct.

The QRPs explained in this article can be hard to detect (unless the study was registered in a database that is accessible to reviewers and readers; if so, efforts will need to be made to locate the study and check whether the objectives and plan of analysis agree between what was registered and what was published). This is why it is important for authors to honestly describe what they have done, such as when they HARK, or not to do it at all, such as by *P*-hacking.

Readers may note that, sometimes, observation of unusual findings in data may be just that: unusual, but due to chance. This is because, for example, in normally distributed data, extreme values may be observed by chance on about 5% of occasions. At other times, observation of anomalies in the data can lead to unexpected discoveries. Post hoc

observations should therefore be considered hypothesis generating until confirmed in subsequent studies.

Parting Notes

The QRPs discussed in this article are not necessarily defined and explained in the same way in all articles on the subject. This is because there is some overlap in concepts across some of the QRPs. What's important, therefore, is for readers to understand the principles involved.

Regulatory clinical trials are required to clearly state all primary and secondary outcomes as well as the detailed plan of analysis, and the complete protocol needs to be submitted to the regulatory authorities before the trials receive approval to begin. In an ideal world, all research protocols should conform to these norms and be made available in publicly accessible online registries before or by the time of study commencement. This would help reduce the temptation to indulge in QRPs and make QRPs possible to identify.

Published online: February 18, 2021.

REFERENCES

1. Kerr NL. HARKing: hypothesizing after the results are known. *Pers Soc Psychol Rev*. 1998;2(3):196–217.
2. Prosperi M, Bian J, Buchan IE, et al. Raiders of the lost HARK: a reproducible inference framework for big data science. *Palgrave Commun*. 2019;5(1):125.
3. Büttner F, Toomey E, McClean S, et al. Are questionable research practices facilitating new discoveries in sport and exercise medicine? the proportion of supported hypotheses is implausibly high. *Br J Sports Med*. 2020;54(22):1365–1371.
4. Andrade C. The primary outcome measure and its importance in clinical trials. *J Clin Psychiatry*. 2015;76(10):e1320–e1323.
5. Andrade C. Multiple testing and protection against a type 1 (false positive) error using the Bonferroni and Hochberg corrections. *Indian J Psychol Med*. 2019;41(1):99–100.
6. Andrade C. The P value and statistical significance: misunderstandings, explanations, challenges, and alternatives. *Indian J Psychol Med*. 2019;41(3):210–215.
7. Mayo-Wilson E, Li T, Fusco N, et al. Cherry-picking by trialists and meta-analysts can drive conclusions about intervention efficacy. *J Clin Epidemiol*. 2017;91:95–110.
8. Head ML, Holman L, Lanfear R, et al. The extent and consequences of p-hacking in science. *PLoS Biol*. 2015;13(3):e1002106.
9. Smith GD, Ebrahim S. Data dredging, bias, or confounding. *BMJ*. 2002;325(7378):1437–1438.
10. Yang J, Li Y, Liu Q, et al. Brief introduction of medical database and data mining technology in big data era. *J Evid Based Med*. 2020;13(1):57–69.
11. Momeni Z, Hassanzadeh E, Saniee Abadeh M, et al. A survey on single and multi omics data mining methods in cancer data classification. *J Biomed Inform*. 2020;107:103466.
12. Zdrzil B, Richter L, Brown N, et al. Moving targets in drug discovery. *Sci Rep*. 2020;10(1):20213.
13. Norman G. Data dredging, salami-slicing, and other successful strategies to ensure rejection: twelve tips on how to not get your paper published. *Adv Health Sci Educ Theory Pract*. 2014;19(1):1–5.
14. Elston DM. Data dredging and false discovery. *J Am Acad Dermatol*. 2020;82(6):1301–1302.