Open Science Practices to Support Addiction Research

January 27, 2020

Editor's note: This op-ed was written by David Mellor, PhD, of the Center for Open Science, as part of this month's Special Series on Open Science Practices in addiction research.

Being able to replicate a colleague's reported empirical observations is a central premise of how scientific discoveries are expected to be disseminated. This ideal has been central to the expectations of scientific processes for centuries and separates credible findings from incredible observations. The motto of the Royal Society, "Nullius in verba" or "On the word of no one," exemplifies this ideal by specifying that *demonstrating* a finding is more important than *claiming* a finding. That demonstration can of course entail replication, but it also assumes that researchers will transparently report the entirety of the evidence- the data and the precise methods by which they established and analyzed their questions. This transparency, broadly described as the Open Science movement, is essential for science to work as expected, as a self-correcting process by which explanations are proposed, evaluated, and winnowed into a more accurate representation of how the world works. BASIS's series on Open Science practices is covering a broad swath of these behaviors and the reasons behind them. At their core, these practices will make the scientific enterprise more efficient, more credible, and more democratic.

Robert Merton espoused these ideas as communal ownership of scientific goods, universally valid scientific processes, disinterested pursuit of evidence, and organized skepticism of methods and conduct (Merton, 1942; 1973). Scientists almost universally endorse such norms and widely self-report engaging in such practices, while at the same time there is widespread belief that they are not universally followed (Anderson, Martinson, & De Vries, 2007).

This belief presents a particular challenge in a movement that aims to increase transparency into scientific practice: When faced with such a seemingly toxic environment, how can one be expected to be more open to critique than others? Further, the current environment doesn't necessarily reflect a conscious decision to be opaque but can simply be a natural continuation of the status quo, an unawareness that particular practices can be problematic, and the reality that we are all too busy to pick up new skills. However, questionable research behaviors, such as cherry picking evidence or gathering data until a desired result is achieved, occur by a large majority of researchers across several disciplines (Agnoli et al., 2017; Fraser et al., 2018; John et al., 2012; Makel et al., 2019), because of the incentives to obtain desired findings for career success. Overcoming these challenges is necessary if we wish to reach a better scientific culture in which credibility and transparency is recognized as more important than primacy or incredibility of findings (Nosek, Spies, & Motyl, 2012). Achieving that cultural change requires both top-down and grassroots efforts to recognize, reward, and require the types of open practices we need to see. Both are underway.

Policy makers are beginning to recognize that open science practices are necessary for scientific advancement and credibility. Dozens of publishers and funders, and thousands of journals of scientific research, have endorsed standards that lay out a roadmap for improving scientific practice: the Transparency and Openness Promotion Guidelines, TOP. Over 200 journals have implemented Registered Reports, a format that emphasize the importance of the research questions and methodology over the surprisingness of the results. However, individual researchers also are taking steps to be the change that they wish to see in their communities. Grassroots networks are forming in departments and universities to advocate for improved practices and share experiences and lessons with colleagues. And hundreds of thousands are using tools to collaborate, register studies, share data, and quickly disseminate findings via preprints (see table).

This reformation in scientific practice is taking place because we are finally beginning to systematically gather evidence on an empirical question that has, to date, largely been the subject of hushed discussions outside of conference center symposia: How replicable are published findings in the scientific literature? These systematic attempts (e.g., Begley & Ellis, 2012; Board of Governors of the Federal Reserve System, Chang, & Li, 2015; Border et al., 2019; C. F. Camerer et al., 2016; Colin F. Camerer et al., 2018; Collaboration, 2015; Cova et al., 2018; Ioannidis et al., 2009) have convinced the majority of the research community that there is a crisis in reproducibility (Baker, 2016). For a different perspective on the importance of replicability and the importance of these issues, a National Academies report on the matter pointed to the importance of generalizability through methods other than replication (National Academies of Sciences, 2019), but still recommended that funders and journals take clear steps to improve the reproducibility and replicability of scientific outputs. Fixing the crisis in reproducibility requires transparency into the collected evidence and into the practice of science itself.

Transparency into the *evidence* of science requires clear and comprehensive reporting of what happened over the course of a study: documentation of research materials, data gathered, and analytical code generated. Use of clear reporting guidelines, such as those curated at the Equator Network (https://www.equator-network.org/) can ensure that all important details are reported. What is gained from this transparency is a more complete record of the research process that can be used to evaluate the credibility of reported results.

Transparency into the *practice* of science requires new habits be formed. Preregistration is the act of specifying in advance hypotheses and how a study will be conducted, and data analyzed (Nosek, Ebersole, DeHaven, & Mellor, 2018). It is particularly important for hypothesis testing research, which requires that the data used to test a hypothesis are not the same data used to generate that hypothesis. When that occurs, we fool ourselves by overfitting models or describing a hypothesis after results are known (i.e. HARKing, see Kerr, 1998), which invalidates hypo-deductive model of statistical inference.[1] Likewise, the unreported flexibility in data analytical decisions, such as choosing the covariates or exclusion criteria that lead to a "statistically significant" finding diminishes the diagnostic value of p-values, known as p-hacking (Simmons, Nelson, & Simonsohn, 2011).

What is gained from this type of transparency is a research method that is less biased by implicit or explicit biases. By making a clear research plan ahead of time, with specific, testable hypotheses and a precise statistical model to test each pre-specified model, we can generate a purely confirmatory research plan. Subjecting the data to that preregistered model will create a clear hypothesis test with meaningful results. Of course, there is a chance that the results will be non-significant, but by specifying the tests ahead of time we will not be motivated to torture the data until it confesses. Doing so in the pursuit of finding an unexpected trend or difference between sub-groups is perfectly acceptable in the pursuit of discovery, but this exploration must be transparently reported as such. The results of this exploration will be a testable hypothesis that deserves to be put to a fair test on a new dataset that was not used to generate it.

The time-stamped preregistration creates ancillary benefits beyond facilitating the clear distinction between confirmation and exploration. By submitting a research plan to a registry, the work becomes citable and discoverable (perhaps after an embargo period), which can make it easier for researchers to receive credit for an original research idea. Furthermore, the act of pre-planning can improve the design and analysis plan early enough for researchers to develop improvements. If a research submits the research plan to a journal as part of a Registered Report (cos.io/rr) (Chambers, et.al, 2014), they can incorporate suggestions through the peer review process and the journal can grant the project an "in-principle acceptance", or a promise to publish the findings regardless of outcome.

There are challenges to preregistering some research methods. The use of existing datasets, for example, can raise the possibility that knowledge of the data biases the generation of hypothesis tests. However, there are solutions to this problem. One particularly useful method, used in machine learning for many years now, is the use of "hold-off" datasets (Anderson & Magruder, 2017; Dwork et al., 2015; Fafchamps & Labonne, 2016). Researchers hold off a random section of the dataset in a separate folder or physical location, away from any analysts. Researchers use the other portion (half, a fifth, or any other randomly generated sub-section) to test model assumptions, look for promising trends in the data, or otherwise explore for relevant discoveries. When ready, the researchers pre-register a plan and uses the unanalyzed data in confirmatory hypothesis testing.

What does all of this mean for addiction research? This is a particularly challenging field of inquiry. We cannot create experiments where we randomly make half of the participants addicted to harmful substances, doing so would be wildly unethical. What we can do is expect the highest form of evidence, given the constraints that will always exist. Large, shared datasets (with identifying information removed, or curated by professionals who evaluate access to the data based on reviews of ethical standards), preregistration of analyses before accessing existing datasets that would otherwise be subject to data-dredging, and advocating to policy makers to implement transparency standards in publication or funding decisions will improve research outcomes.

Our credibility as scientists requires that we acknowledge the incentives that drive our behavior and the biases that cloud our judgement. Transparency into the complete process of science is a necessary condition for obtaining and preserving that credibility. This transparency does not guarantee that perfectly rigorous methods will follow, but it does provide a more direct incentive for this level of rigor and it does allow for an accurate assessment of rigor to take place. This transparency is new to most scientists and we owe it to the community to reward it whenever we see it.

Literature Cited

Agnoli, F., Wicherts, J. M., Veldkamp, C. L. S., Albiero, P., & Cubelli, R. (2017). Questionable research practices among italian research psychologists. PLOS ONE, 12(3), e0172792. https://doi.org/10.1371/journal.pone.0172792

Anderson, M., & Magruder, J. (2017). Split-Sample Strategies for Avoiding False Discoveries (No. w23544). https://doi.org/10.3386/w23544

Anderson, M. S., Martinson, B. C., & De Vries, R. (2007). Normative Dissonance in Science: Results from a National Survey of U.S. Scientists. Journal of Empirical Research on Human Research Ethics: An International Journal, 2(4), 3–14. https://doi.org/10/cfs74f

Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. Nature, 533(7604), 452–454. https://doi.org/10/gdgzjx

Begley, C. G., & Ellis, L. M. (2012). Drug development: Raise standards for preclinical cancer research. Nature, 483(7391), 531–533. https://doi.org/10/gd3xdh

Board of Governors of the Federal Reserve System, Chang, A. C., & Li, P. (2015). Is Economics Research Replicable? Sixty Published Papers from Thirteen Journals Say "Usually Not." Finance and Economics Discussion Series, 2015(83), 1–26. https://doi.org/10/gfgv79

Border, R., Johnson, E. C., Evans, L. M., Smolen, A., Berley, N., Sullivan, P. F., & Keller, M. C. (2019). No Support for Historical Candidate Gene or Candidate Gene-by-Interaction Hypotheses for Major Depression Across Multiple Large Samples. The American Journal of Psychiatry, 176(5), 376–387. https://doi.org/10/gfwnhp

Camerer, C. F., Dreber, A., Forsell, E., Ho, T.-H., Huber, J., Johannesson, M., … Wu, H. (2016). Evaluating replicability of laboratory experiments in economics. Science, 351(6280), 1433–1436. https://doi.org/10/bdps

Camerer, Colin F., Dreber, A., Holzmeister, F., Ho, T.-H., Huber, J., Johannesson, M., … Wu, H. (2018). Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015. Nature Human Behaviour, 1. https://doi.org/10/gd3v2n

Chambers, C., Feredoes, E., D. Muthukumaraswamy, S., J. Etchells, P., & 1 Cardiff University Brain Research Imaging Centre, School of Psychology, Cardiff University; (2014). Instead of "playing the game" it is time to change the rules: Registered Reports at AIMS Neuroscience and beyond. AIMS Neuroscience, 1(1), 4–17. https://doi.org/10/gdnbt7

Collaboration, O. S. (2015). Estimating the reproducibility of psychological science. Science, 349(6251), aac4716. https://doi.org/10/68c

Cova, F., Strickland, B., Abatista, A., Allard, A., Andow, J., Attie, M., … Zhou, X. (2018). Estimating the Reproducibility of Experimental Philosophy. Review of Philosophy and Psychology. https://doi.org/10/gf28qh

Dwork, C., Feldman, V., Hardt, M., Pitassi, T., Reingold, O., & Roth, A. (2015). The reusable holdout: Preserving validity in adaptive data analysis. Science, 349(6248), 636–638. https://doi.org/10/f7nccv

Fafchamps, M., & Labonne, J. (2016). Using Split Samples to Improve Inference about Causal Effects (No. w21842). https://doi.org/10.3386/w21842

Fraser, H., Parker, T., Nakagawa, S., Barnett, A., & Fidler, F. (2018). Questionable research practices in ecology and evolution. PLOS ONE, 13(7), e0200303. https://doi.org/10/gdtmg2

Ioannidis, J. P. A., Allison, D. B., Ball, C. A., Coulibaly, I., Cui, X., Culhane, A. C., … van Noort, V. (2009). Repeatability of published microarray gene expression analyses. Nature Genetics, 41(2), 149–155. https://doi.org/10/bcn2tr

John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the Prevalence of Questionable Research Practices With Incentives for Truth Telling. Psychological Science, 23(5), 524–532. https://doi.org/10/f33h6z

Kerr, N. L. (1998). HARKing: Hypothesizing After the Results are Known. Personality and Social Psychology Review, 2(3), 196–217. https://doi.org/10/dnqm8w

Makel, M. C., Hodges, J., Cook, B. G., & Plucker, J. (2019). Questionable and Open Research Practices in Education Research [Preprint]. EdArXiv. https://doi.org/10.35542/osf.io/f7srb

Merton, Robert K. (1973). The Normative Structure of Science. In The Sociology of Science: Theoretical and Empirical Investigations (pp. 267–278). University of Chicago Press.

National Academies of Sciences, E. (2019). Reproducibility and Replicability in Science. https://doi.org/10.17226/25303

Nosek, B. A., Spies, J. R., & Motyl, M. (2012). Scientific Utopia: II. Restructuring Incentives and Practices to Promote Truth Over Publishability. Perspectives on Psychological Science, 7(6), 615–631. https://doi.org/10/f4fc2k

Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. Proceedings of the National Academy of Sciences, 115(11), 2600–2606. https://doi.org/10/gc6xk8

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant. Psychological Science, 22(11), 1359–1366. https://doi.org/10/bxbw3c

Author Bio

David Mellor is the Director of Policy Initiatives at the Center for Open Science (https://cos.io), a non-profit company whose mission is to increase credibility in scientific research by increasing transparency. David's background is in citizen science and behavioral ecology. He completed his PhD at Rutgers University studying the behavior of cichlid fishes and a post-doc at Virginia Tech working with citizen scientists conducting authentic scientific inquiry in water quality and invasive plant control.
Find David online at https://orcid.org/0000-0002-3125-5888

Acknowledgments

The Center for Open Science is a 501(c)3 non-profit organization and is funded by private and government funders to support its mission to increase trust and replicability of scientific research. You can learn more about COS's funders at https://cos.io/about/our-sponsors/. COS builds and maintains the open source OSF (https://osf.io) to enable the activities described in this paper.

---

1. Of course, the process of generating theories or data exploration when no hypotheses are reported and no inferences made to wider populations cannot suffer from these particular problems.