Open Science in Mental Health Research

Vaughan Bell

Division of Psychiatry, University College London
South London and Maudsley NHS Foundation Trust

We hear a lot of talk about parity for mental health research but sometimes, as researchers, we need to take the lead. One area where clinical science is failing is in the quality of research, in terms of the reliability and transparency of published studies. Evidence shows that psychiatry is slightly ahead in some respects. A recent study by Estelle Dumas-Mallet and colleagues on "Low statistical power in biomedical science" found that psychiatry was the best powered research field among a sample of medical disciplines but, overall, the picture left much to be desired. Median statistical power in psychiatry research was still only 20%. Effectively, we're a nose ahead in a race where everyone is performing poorly.

Concerns about the poor replicability of studies in psychology have led cognitive scientists to re-examine how research is planned, conducted, and reported. And it is now becoming clear that this is not a problem restricted to psychology. Early studies show that the reliability of results in biomedicine may be as bad or worse. This has led to a vibrant open science movement that puts transparency, reliability, and openness at its core, aided by online tools such as the *Open Science Framework*. Mental health researchers, however, have been slow to catch up to the advances made by their psychology colleagues. There are some existing good examples—genetics stands out and neuroimaging is making considerable strides—but to make our field a leader in conducting the most reliable, translatable research in medicine, we need to make open methods standard rather than a research sub-culture.

Open science is an umbrella term for making more of the scientific process transparent, from hypothesising and collecting data, to analysis and negotiating the final report. Importantly, no radical changes are needed from standard scientific practice and there are several tools researchers can use right now. First, there is the pre-registration of hypotheses and analysis plans. One of the surest ways to get statistically significant results is to run lots of tests, in different ways, on different versions of the dataset. This is not a problem unless you fail to mention the other non-significant tests and additional variables: a questionable research practice known as "p-hacking". You can assure others that this has not happened in your research by pre-registering your predictions and analysis plans before you collect data. It's the concept of trial pre-registration extended to all hypothesis-based research. Pre-registration provides an added degree of confidence that the significant results you publish are not the result of unreported decisions. Around 50 journals now allow formal registered reports, including publications by *Nature* and *The Royal Society*, where you can submit the introduction and methods sections to be peer reviewed before you collect data with a

conditional accept for publication of the final report. But you can also just register your analysis and predictions on the web using a brief form. Any changes to your registered analysis plan should be reported in the paper, and you can still do non-registered or exploratory research.

[SIDEBAR LINKS: For analysis and prediction registration sites, see https://aspredicted.org and https://osf.io].

Second, you can make materials available. This won't work for everything, because some materials may be copyrighted or won't travel well over the internet (antibodies, for example, are notoriously difficult to download). But most research involves materials that can be put online and used by others. This improves the ability of other researchers to replicate your study, and allows them to build on your methods.

Third, you can publish your analysis code. Most modern research involves multi-stage data processing and statistical analysis written as code, syntax, or scripts. Having your code available to all means others can see exactly what you did, with every decision clearly apparent, and can learn from your good work. As any software developer will tell you, bugs in code are common and allowing others to find and fix these will make your analysis and future analyses more reliable.

Fourth, you can make your data available—so-called open data. The intrinsic value of data increases as it becomes more widely available. Nevertheless, making data public seems to make researchers nervous. It shouldn't, because we typically declare when we publish that we will make anonymised data available to other researchers if they request it. Studies show, however, that compliance rates are very poor and making data available on the basis of a single email contact on a published article is bound to be unreliable. To publish data online, there are now numerous free or low cost data archiving sites. Some data archives also provide a digital object identifier or DOI, meaning you can be cited when your data is reused, and there are now journals dedicated to publishing data sets.

[SIDEBAR LINK: For data archiving sites, see https://figshare.com , https://osf.io , http://datadryad.org , https://zenodo.org , and http://neurovault.org . For a directory of research data repositories, see http://www.re3data.org].

One common concern is that publishing data from studies will violate ethical approval or consent. It is good practice to mention the possibility of anonymous publication on study information but if data is sufficiently anonymised to protect against de-identification, it typically does not require specific consent to publish it—and this includes clinical data. I confirmed the above with the Health Research Authority in the UK, but check with your relevant ethics body. There are really very few barriers to publishing adequately anonymised data online for others to use, even in retrospect.

Finally, consider preprints. Physicists have been publishing pre-publication versions of their articles for decades, but it has only recently started to become common in biomedicine. Pre-prints are not intended to replace peer-review but they do make your provisional results available early, and they mean you can put a mark in the sand if getting scooped is a concern. The *Medical Research Council* and the *Wellcome Trust* now accept pre-prints in grant applications and the majority of journals have agreed to accept pre-printed papers. Indeed, some journals are actively soliciting submissions from preprint servers. DOIs are provided by these servers, so pre-prints can be cited.

[SIDEBAR LINK: For preprint servers, see http://biorxiv.org for biomedical work, https://ssrn.com for social sciences and https://osf.io/preprints/psyarxiv for psychology.]

If it's so easy to make mental health science more open, why haven't more people done it? There are a few worries that are commonly expressed by researchers. First, there is the concern that open science creates additional work. It's true, it does. But not much more, and it will make your work more valuable to other researchers. And from a career perspective, it will increase the number of citable research outputs you produce, potentially increasing your impact. It may be additional work but it's also necessary work: the Transparency and Openness Promotion Guidelines, which list open science standards, have now been signed by over 700 journals and 60 scientific organisations, including learned societies and funding bodies. Open science is the future, and it's worth being ahead of the curve.

[SIDEBAR LINK: For the Guidelines, see https://cos.io/our-services/top-guidelines]

Second, there is the understandable fear of being wrong in public. Science is about being usefully wrong, but some still believe a stereotype that being wrong in public affects your reputation. Evidence, however, is not on their side. A recent study entitled *The Reputational Consequences of Failed Replications and Wrongness Admission among Scientists* found that the stereotype was common but incorrect. Admitting being wrong actually had a positive effect on reputation; it was not admitting that was harmful. Yes, more errors will be found in your work if more is available to the world, but more errors will get fixed, and how you respond will signal your quality to others.

Third, and related to this, is the concern about so-called 'hostile reanalysis'. Science can be adversarial, and psychiatry is a controversial area. Some researchers fear not an honest re-evaluation of their work, but bad faith attempts to prove them wrong. Open science isn't going to stop this but it helps protect against it. It makes everything open from the point of publication, making accusations of obscuring data and analysis methods fall flat, and motivated reanalyses mostly have the status of exploratory research if you have pre-registered your hypotheses.

It would be naïve to think that open science is the only solution to improving research quality. A persistent problem is low statistical power and many mental health journals are still filled with small N studies that are well designed and administratively difficult, but of

low statistical value in terms of being likely to reflect a true effect. This needs change on multiple levels beyond the lab. Negative stereotypes about internet-conducted research need to be discarded in light of evidence to show that performance on experimental psychology tasks is typically equal to lab studies and how well participants represent the wider population is usually only bettered by formal sampling. Grant bodies need to be prepared to fund studies to collect datasets that are statistically powered to miss only irrelevant effects, and journals need to be stricter on how smaller studies are selected and interpreted – too often technical complexity propels studies to the pages of high impact journals accompanied by evidential claims that are only weakly supported by the data. Career incentive structures need to move away from primarily valuing high status publications and grant income to valuing sound scientific questions and methods, regardless of results. We also need to change dusty research methods training to reflect new scientific practices. Students need to be introduced to the value of replication, and the value of facilitating replication through transparency, as the bedrock of scientific practice, in addition to the traditional focus on sound research methods to test new ideas.

I'm aware that calls for improvement can sometimes feel like hectoring for people already doing good work, and adapting to new methods can feel like a slowdown. But this is an investment of time that will increase the value of your work in the scientific community and, most important, make it more useful to the millions of people worldwide who need and deserve substantial advances in mental health science.