



Kapur, S., & Munafo, M. (2019). Small Sample Sizes and a False Economy for Psychiatric Clinical Trials. *JAMA Psychiatry*, 76(7), 676-677. <https://doi.org/10.1001/jamapsychiatry.2019.0095>

Publisher's PDF, also known as Version of record

Link to published version (if available):
[10.1001/jamapsychiatry.2019.0095](https://doi.org/10.1001/jamapsychiatry.2019.0095)

[Link to publication record on the Bristol Research Portal](#)
PDF-document

This is the final published version of the article (version of record). It first appeared online via AMA at <https://jamanetwork.com/journals/jamapsychiatry/fullarticle/2729439> . Please refer to any applicable terms of use of the publisher.

University of Bristol – Bristol Research Portal

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available: <http://www.bristol.ac.uk/red/research-policy/pure/user-guides/brp-terms/>

Small Sample Sizes and a False Economy for Psychiatric Clinical Trials

Shitij Kapur, MBBS, PhD, FRCPC; Marcus Munafò, PhD

In 2013, *JAMA Psychiatry* published an exciting new finding by Hallak et al¹: patients with schizophrenia who were treated with a single infusion of the antihypertensive agent sodium nitroprusside showed a dramatic, instantaneous, and sustained improvement in psychotic and negative symptoms. The study by Hallak



Related article

et al¹ was not only double-blind, randomized, and placebo-controlled, but the authors had also presented a prospective power analysis, ensured good interrater reliability, and reported all the relevant aspects of the trial. Understandably, this study generated tremendous excitement and several attempts were made to replicate the finding. In this issue of *JAMA Psychiatry*, Brown et al² report an equally systematic study that fails to replicate the original finding. In fact, the findings by Brown et al² join those of 2 other prior studies^{3,4} that also failed to replicate the original finding. A careful analysis of the competing studies fails to reveal an obvious explanation for this discrepancy—except that the original finding was based on a rather small sample of just 10 patients per treatment arm.

Failure to reproduce such dramatic findings is neither new, nor limited to psychiatry. As provocatively claimed by Ioannidis⁵ in 2005, perhaps most published studies are indeed false. Alert to this possibility, the Academy of Medical Sciences in the United Kingdom published a report on the reproducibility and reliability of biomedical research, and identified the many causes for the low replicability of published studies, from deficiencies in the conduct and reporting of studies to incentive structures that promote publication over rigour.^{6,7} However, in the matter of the sodium nitroprusside studies the methods of the studies seem to be strong and the reporting seems to be thorough. But rigor of conduct and reporting cannot overcome the challenge of starting with a small number of patients.

A review of 83 psychiatric intervention studies indicated that only half had been subject to any attempt at replication.⁸ Of those that had been subject to an attempt at replication, only 37% replicated the original effect size, while the others either contradicted the original finding or observed a much smaller effect. The review clearly showed that, the larger the original study, the more likely it was to be replicated.

It is a statistical inevitability that (all other things being equal) smaller studies will provide more imprecise estimates of any true effect. Therefore, if a number of small studies are performed, there will be a wider variation of findings (sometimes described as *vibration of effects*). If many small studies

are conducted, only those that generate a large effect size (and therefore reach a *P* value below the conventional 5% threshold) will be published and these findings from studies with a small sample size are likely to represent inflated effect size estimates, and at worst, false positives.

Based on this finding it would be easy to recommend larger sample sizes in general, given evidence that studies in the biomedical sciences are usually too small to detect credible or likely effect sizes.⁹ But any such general commandment would be too simplistic. Ultimately, studies should be large enough to give a sufficiently precise estimate, but not so large as to be wasteful. Funders and journals increasingly require a justification of sample size. However, calculation of sample size is not a perfect science and requires an estimation of the expected outcomes in the drug and placebo groups and an estimation of the expected variance. Getting these estimates right and realistic is at the heart of the matter.

Hallak et al¹ predicted an effect size of $d = 1.5$. How does this size compare with effect sizes in this field? We provide not 2, but 3 comparators. A study of clinical interventions in medicine and psychiatry across a range of domains indicates a median observed an effect size of 0.37 for general medicine and 0.41 for psychiatry.¹⁰ A review of more than 7000 patients treated with antipsychotics in randomized clinical trials showed that the effect size of the newer atypical antipsychotics, compared with placebo, was 0.48.¹¹ Finally, studies of schizophrenia clinicians show that they consider an improvement from baseline (which includes placebo response plus drug response) equivalent to an approximate effect size of $d = 1.0$ to be clinically significant.¹² Thus, seen from these 3 perspectives, an expectation of $d = 1.5$ vs placebo seems rather optimistic. The temptation is understandable. One would like to discover the big findings, and such an assumption comes with the payoff that it decreases the number of patients required by sample size calculators. However, such over-optimism comes with 2 wasteful consequences: a higher than appropriate chance of a false negative and, in the event of a positive finding, a higher chance that the positive finding might be a false positive with an inflated effect.

So, what is the right effect size to aim for? There is not, and cannot be, a single answer. What seems prudent is that trials of any new treatment should assume the median observed in the field (which usually in the range of 0.3–0.5), and those who hope for a much larger effect size should be required to provide a strong justification for such optimism. Because, over-optimism that leads to too small a sample may just be a false economy.

ARTICLE INFORMATION

Author Affiliations: Faculty of Medicine, Dentistry, and Health Sciences, University of Melbourne, Melbourne, Australia (Kapur); School of Psychological Science, University of Bristol, Bristol, United Kingdom (Munafò); Medical Research Council of UK Integrative Epidemiology Unit, University of Bristol, Bristol, United Kingdom (Munafò).

Corresponding Author: Shitij Kapur, MBBS, PhD, FRCP, Faculty of Medicine, Dentistry, and Health Sciences, University of Melbourne, Melbourne 3010, Australia (shitij.kapur@unimelb.edu.au).

Published Online: March 27, 2019.
doi:10.1001/jamapsychiatry.2019.0095

Conflict of Interest Disclosures: Dr Kapur reported being coauthor on one of the previously published studies that failed to replicate the sodium nitroprusside effect. No other disclosures were reported.

REFERENCES

- Hallak JE, Maia-de-Oliveira JP, Abrao J, et al. Rapid improvement of acute schizophrenia symptoms after intravenous sodium nitroprusside: a randomized, double-blind, placebo-controlled trial. *JAMA Psychiatry*. 2013;70(7):668-676.
- Brown HE, Freudenreich O, Fan X, et al. Efficacy and tolerability of adjunctive intravenous sodium nitroprusside treatment for outpatients with schizophrenia: a randomized clinical trial [published online March 27, 2019]. *JAMA Psychiatry*. doi:10.1001/jamapsychiatry.2019.0151
- Stone JM, Morrison PD, Koychev I, et al. The effect of sodium nitroprusside on psychotic symptoms and spatial working memory in patients with schizophrenia: a randomized, double-blind, placebo-controlled trial. *Psychol Med*. 2016;46(16):3443-3450. doi:10.1017/S0033291716002245
- Wang X, Zhao J, Hu Y, et al. Sodium nitroprusside treatment for psychotic symptoms and cognitive deficits of schizophrenia: a randomized, double-blind, placebo-controlled trial. *Psychiatry Res*. 2018;269:271-277. doi:10.1016/j.psychres.2018.08.079
- Ioannidis JP. Why most published research findings are false. *PLoS Med*. 2005;2(8):e124.
- Academy of Medical Sciences. *Reproducibility and Reliability of Biomedical Research: Improving Research Practice*. London, UK: Academy of Medical Sciences; 2015.
- Higginson AD, Munafò MR. Current incentives for scientists lead to underpowered studies with erroneous conclusions. *PLoS Biol*. 2016;14(11):e2000995. doi:10.1371/journal.pbio.2000995
- Tajika A, Ogawa Y, Takeshima N, Hayasaka Y, Furukawa TA. Replication and contradiction of highly cited research papers in psychiatry: 10-year follow-up. *Br J Psychiatry*. 2015;207(4):357-362. doi:10.1192/bjp.bp.113.143701
- Dumas-Mallet E, Button KS, Boraud T, Gonon F, Munafò MR. Low statistical power in biomedical science: a review of three human research domains. *R Soc Open Sci*. 2017;4(2):160254. doi:10.1098/rsos.160254
- Leucht S, Hierl S, Kissling W, Dold M, Davis JM. Putting the efficacy of psychiatric and general medicine medication into perspective: review of meta-analyses. *Br J Psychiatry*. 2012;200(2):97-106. doi:10.1192/bjp.bp.111.096594
- Leucht S, Arbter D, Engel RR, Kissling W, Davis JM. How effective are second-generation antipsychotic drugs? a meta-analysis of placebo-controlled trials. *Mol Psychiatry*. 2009;14(4):429-447. doi:10.1038/sj.mp.4002136
- Hermes ED, Sokoloff D, Stroup TS, Rosenheck RA. Minimum clinically important difference in the Positive and Negative Syndrome Scale with data from the Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE). *J Clin Psychiatry*. 2012;73(4):526-532. doi:10.4088/JCP.11m07162