Biological Psychiatry: CNNI

Commentary

Special Issue on Reliable Mechanisms for Translational Applications

Quentin J.M. Huys and Martin Paulus

The reliability of neuroscientific measurements is critical to the translation of neuroscientific advances into clinical applications. In recent years, researchers have uncovered substantial limitations in the reliability of many neuroscience, social science, and psychology findings. Shortcomings in the reliability of scientific research have been prominently featured in both scientific publications and the wider press (1,2). Unreliable scientific findings can be identified by further research, but this involves diverting efforts away from the truly promising research directions. Unreliable or incorrect research results can have long-lasting and pernicious effects on the state of knowledge (3). These findings have subsequently galvanized efforts into understanding the source of unreliable findings and into addressing them, resulting in important changes to scientific procedures that aim to ensure improved replicability.

This special issue of *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging* presents the state of the art in a series of articles covering advances in our understanding of reliability relevant to mental health neuroscience research. The topics range from novel methodologies to a focus on analytical and specific issues in particular settings. First, Botvinik-Nezer and Wager (4) focus on reproducibility of neuroimaging, but their insights and lessons apply much more broadly to the field.

They describe novel tools and practices to improve reproducibility, i.e., the ability to identify the same set of results using the same analysis methods on the same data. The fact that this is frequently not possible is a major reminder of the challenges ahead and points to the necessity of improving reporting standards, code and data sharing practices, management of computing environments, and analytic flexibility. They point to a novel form of "doing" science, the Psychological Science Accelerator, whereby a global network of laboratories coordinates data collection for democratically selected studies (5), and a novel form of "doing" analyses involving diverse analytical approaches involving a multiverse of analyses, i.e., examining whether a particular finding holds up across multiple different analytical approaches.

Haines *et al.* (6) address a core issue in translational and computational psychiatry: the challenges in establishing reliable correlates between mechanistic measurements of brain and behavior, and measurements of symptoms or psychopathology (7). Such associations are fundamental determinants of the clinical validity of mechanistic assessments. The authors provide a thorough yet delightfully concise overview of the theoretical foundations of how noise (i.e., unreliability) in measurements affects translation. They identify a key culprit that is rarely appreciated—uncertainty in psychopathological measurements (as opposed to just the mechanistic/brain measurement)—and describe in intuitive detail how hierarchical and structural equation approaches correctly, and therefore effectively, address such issues. Zorowitz and Niv (8) complement the Haines *et al.* (6) discussion by identifying specific ways in which cognitive tasks can be made more reliable. They provide easy-to-understand reasons why some results might not be reliable. They also offer useful tips for experiments and data analyses that future researchers will find helpful, drawing on their vast experience. Importantly, and reflecting many of the suggestions by Botvinik-Nezer and Wager (4), both articles developed an online R notebook for readers to experiment with.

Kwon *et al.* (9) extend this theoretical framework and introduce the notion of Bayesian optimal adaptive design. In short, the concept is about adjusting the experiment as it is happening. For example, if we study how a person responds to a sad face and we see halfway through the experiment that the person classifies most faces as sad, we should explore faces with stronger expressions of sadness. This adjustment is based on the real-time data acquired during the experiment. In principle, this adaptation can be done such that the maximal information is extracted about the underlying quantity on every trial. This is particularly important for translational settings in which the long durations typical in preclinical laboratories are not feasible.

Parmigiani *et al.* (10) discuss the combination of transcranial magnetic stimulation with electroencephalography. This combination is particularly promising because it potentially allows for causal assessment of brain connectivity. Such a causal assessment could confer substantial additional validity, but the authors summarize important issues around reliable noise and unreliable signals and provide an in-depth discussion of how to address these issues.

Finally, Pezzoli *et al.* (11) consider the issue of reliability specifically in developmental settings. Many similar issues occur in treatment settings, where we study how brain functions change over time and between different groups. Two big challenges are distinguishing between effects due to practice or those due to actual cognitive development, and dealing with noise or random variations in the data. As outlined above, the latter is critical for reliability, but in developmental settings this can itself be expected to change and be a marker of cognitive function, be it over short or longer timescales.

In conclusion, this special issue provides a comprehensive exploration of the various aspects of ensuring reliable and replicable results in applied neuroscience research. Although the advances are numerous, multifaceted, and impressive,

  https://doi.org/10.1016/j.bpsc.2023.06.004
Biological Psychiatry: Cognitive Neuroscience and Neuroimaging August 2023; 8:778–779 www.sobp.org/BPCNNI ISSN: 2451-9022

much work remains to be done. Indeed, the demands of both time and effort in ensuring reliability are challenges for individual researchers facing additional work and for funders and policymakers who need to recognize and support this: Good science comes at a cost, but robust results will yield better outcomes sooner—it's worth it.

Acknowledgments and Disclosures

QJMH acknowledges support by the University College London Hospitals NHS Foundation Trust National Institute for Health and Care Research Biomedical Research Centre (UCLH NIHR BRC) and ongoing funding from Wellcome Trust Grant No. 221826/Z/20/Z, Carigest S.A., and Koa Health.

QJMH has obtained fees and options for consultancies for Aya Technologies and Alto Neuroscience. MP has no relevant biomedical financial interests or potential conflicts of interest.

Article Information

From the Applied Computational Psychiatry Laboratory, Mental Health Neuroscience Department, Division of Psychiatry and Max Planck UCL Centre for Computational Psychiatry and Ageing Research, Queen Square Institute of Neurology, University College London, London, United Kingdom (QJMH), and the Laureate Institute for Brain Research, Tulsa, Oklahoma (MP).

Address correspondence to Quentin J.M. Huys, M.D., Ph.D., at qhuys@cantab.net.

Received Jun 7, 2023; accepted Jun 12, 2023.

References

1. Ioannidis JPA (2005): Why most published research findings are false. PLoS Med 2:e124.
2. Open Science Collaboration (2015): Psychology. Estimating the reproducibility of psychological science. Science 349:aac4716.
3. Nissen SB, Magidson T, Gross K, Bergstrom CT (2016): Publication bias and the canonization of false facts. Elife 5:e21451.
4. Botvinik-Nezer R, Wager TD (2023): Reproducibility in neuroimaging analysis: Challenges and solutions. Biol Psychiatry Cogn Neurosci Neuroimaging 8:780–788.
5. Moshontz H, Campbell L, Ebersole CR, IJzerman H, Urry HL, Forscher PS, *et al.* (2018): The psychological science accelerator: Advancing psychology through a distributed collaborative network. Adv Methods Pract Psychol Sci 1:501–515.
6. Haines N, Sullivan-Toole H, Olino T (2023): From classical methods to generative models: Tackling the unreliability of neuroscientific measures in mental health research. Biol Psychiatry Cogn Neurosci Neuroimaging 8:822–831.
7. Eisenberg IW, Bissett PG, Enkavi AZ, Li J, MacKinnon DP, Marsch LA, *et al.* (2019): Uncovering the structure of self-regulation through data-driven ontology discovery. Nat Comm 10:2319.
8. Zorowitz S, Niv Y (2023): Improving the reliability of cognitive task measures: A narrative review. Biol Psychiatry Cogn Neurosci Neuroimaging 8:789–797.
9. Kwon M, Lee SH, Ahn W-Y (2023): Adaptive design optimization as a promising tool for reliable and efficient computational fingerprinting. Biol Psychiatry Cogn Neurosci Neuroimaging 8:798–804.
10. Parmigiani S, Ross JM, Cline CC, Minasi CB, Gogulski J, Keller CJ (2023): Reliability and validity of transcranial magnetic stimulation–electroencephalography biomarkers. Biol Psychiatry Cogn Neurosci Neuroimaging 8:805–814.
11. Pezzoli P, Parsons S, Kievit RA, Astle DE, Huys QJM, Steinbeis N, Viding E (2023): Challenges and solutions to the measurement of neurocognitive mechanisms in developmental settings. Biol Psychiatry Cogn Neurosci Neuroimaging 8:815–821.