Biological Psychiatry: GOS

Review

Open Science Practices in Psychiatric Genetics: A Primer

Adrianna P. Kępińska, Jessica S. Johnson, and Laura M. Huckins

ABSTRACT

Open science ensures that research is transparently reported and freely accessible for all to assess and collaboratively build on. Psychiatric genetics has led among the health sciences in implementing some open science practices in common study designs, such as replication as part of genome-wide association studies. However, thorough open science implementation guidelines are limited and largely not specific to data, privacy, and research conduct challenges in psychiatric genetics. Here, we present a primer of open science practices, including selection of a research topic with patients/nonacademic collaborators, equitable authorship and citation practices, design of replicable, reproducible studies, preregistrations, open data, and privacy issues. We provide tips for informative figures and inclusive, precise reporting. We discuss considerations in working with nonacademic collaborators and distributing research through preprints, blogs, social media, and accessible lecture materials. Finally, we provide extra resources to support every step of the research process.

https://doi.org/10.1016/j.bpsgos.2023.08.007

Open science is an approach where research and resulting publications are transparent and freely accessible. Consequently, findings may be replicated, and errors prevented or corrected (1). The core of open science is a commitment to sharing research with everyone (2). Thus, several definitions of open science also address collaboration, inclusion, and diversity to improve participation in science (3,4). Consistent with this multifaceted view, we discuss open science practices related to transparency, integrity, diversity, inclusivity, equity, reproducibility, rigor, and accessibility (see Box S1 for definitions).

Psychiatric genetics has been praised as an early adopter of open science, championing genome analyses with embedded replication, initially through genome-wide association studies (GWASs) (5,6). Multiple consortia now exist to improve data pooling and provide free bioinformatic methods, e.g., the Psychiatric Genomics Consortium (PGC) (6), the Genotype-Tissue Expression Project (7), and the Global Biobank Meta-analysis Initiative (8). Geneticists frequently use preprint servers (where manuscripts get freely deposited before peer review), e.g., bioRxiv (9). All these efforts are aimed at remedying years of unreplicable research (10,11) due to small samples that were underpowered to detect associations between psychiatric disorders (12), singular gene variants, and environmental factors (13); questionable statistical practices generating false positives (14,15); and inconsistent testing of different variants on the same genes (16).

However, GWASs may be losing their status as a "paragon of open science" (17). GWASs have been criticized for being increasingly difficult to reproduce owing to inconsistent, un(der)reported methods and increasing use of private company data to enlarge samples. Commercial releases of partial data hinder replication. Finally, data come from unrepresentative samples (individuals who are predominantly affluent, of European ancestry, better educated, and/or healthy volunteers).

For science to remain open, existing research standards must be consistently implemented and constantly refined (1). Recent literature has addressed open science in psychiatry (18–20) or GWASs (21,22), but not specifically psychiatric GWASs. When reviews have addressed psychiatric genetics, they have prioritized replicability and reproducibility (23,24). These otherwise crucial dimensions are not the only facets of open scholarship. Here, we provide a primer on implementing open science practices in psychiatric genetics at multiple research steps (Figure 1): topic choice and authorship order, study design and preregistration, data access, open code, informative figures, inclusive language, reporting standards, preprints, nonacademic (citizen scientist) involvement, and research distribution. We identify common issues, suggest solutions, and include resources for every step.

RESEARCH TOPIC AND AUTHORSHIP

Several open science values–integrity, transparency, equity, and accessibility–are intertwined with study conceptualization.

Choice of Research Topic

Patient experts and researchers increasingly recognize that patients and community representatives should be successfully involved in every step of the research process (25). Such inclusion ensures that research addresses nonstigmatizing topics that are of the highest benefit to community-specific health needs (26).

ISSN: 2667-1743

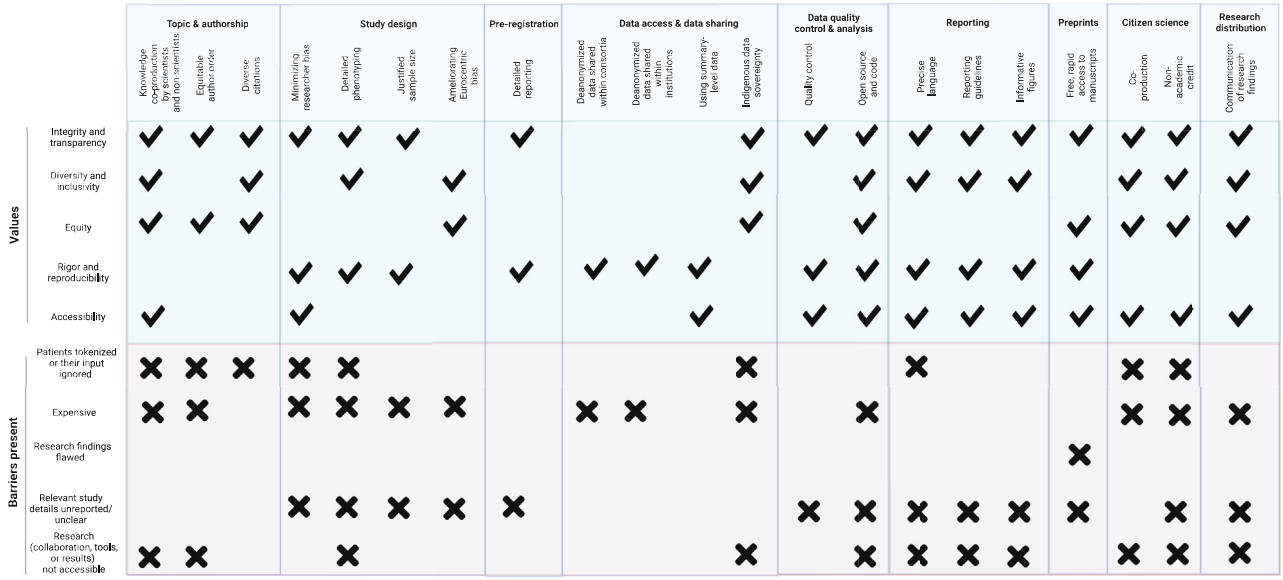

Figure 1. Open science practices at every research step, values which these actions support, and barriers to open science implementation. Figure created with Biorender.com.

Equity in Authorship and Author Order

Transparency and equity should be considered in authorship decisions. Authorship disclosures may be standardized forms of contributions, e.g., Contributor Roles Taxonomy (CRediT) (27). However, authorship is academic currency, but author order is still unequally attributed to men over women (28). Liboiron *et al.* (29) argued for equity as the guiding principle in establishing author order, considering care work and unequal division of labor. Care work (mentoring and teaching) is poorly acknowledged and disproportionately delegated to women and early-career or marginalized scholars (30). However, care work is vital to early-career scholars' skillsets and well-being in research groups (31). Thus, we argue that care work must be credited, e.g., using CRediT (27).

To improve transparency and equity, we suggest that, prior to any project, researchers declare detailed criteria for determining author order (32–34), such as extent of contribution, author seniority, and the degree/rarity or number of biological samples contributed by each collaborator. Policies could define a significant contribution, e.g., whether author order is based on the number of completed tasks and how these tasks are quantified. This is crucial for studies that require different types of expertise.

Another solution is designating multiple corresponding authors for specific analyses, thereby distributing extra credit among authors who are responsible for technical contributions (e.g., programmers, statisticians). These authors are often placed in less desirable middle positions in author lists (35) or, historically in the case of female programmers, given acknowledgments but not authorship (36). Unfortunately, joint authorships are not consistently acknowledged (37,38). A novel approach proposes crediting authors throughout the manuscript text, thus making the contributions of each author clearer to readers, including hiring committees (39). While implementing this approach may prove complicated, we agree that changes in hiring practices are necessary. Diverse contributions should not be underappreciated if they do not result in first or senior authorship.

Equality in Citations

Genetic studies fail to represent global genetic diversity and come predominantly from Western sites (40), which have effectively monopolized the literature. Currently limited ways of redressing this imbalance are transparency about how a lack of global perspective impacts results and limits generalizability of study findings and referencing both large-scale, European ancestry-skewed studies and smaller studies of more diverse populations. Where studies are unavailable, researchers should cite theoretical work, protocols, and software by underrepresented scholars.

STUDY DESIGN

Next, we briefly outline key study design considerations related to research clarity, reproducibility, replicability, and diversity.

Preventing Researcher Bias in Secondary Data Analysis

Secondary data analysis is common in psychiatric genetics. However, knowledge of data before analysis may introduce cognitive bias, thus distorting conclusions. Several solutions may limit bias (41): 1) a preregistration (Preregistration and Registered Reports); 2) sampling (e.g., hold out samples for exploratory research); 3) blinded analyses (42) or scrambling data (41); 4) multiverse analyses (testing whether findings remain consistent across multiple paradigms, e.g., frequentist

and Bayesian analyses) (43); or 5) coordinated analyses (testing models across separate large-scale cohorts) (44).

Clear Phenotype Specification

Heterogeneous psychiatric phenotypes are frequently defined under a single diagnosis (45). Granular phenotypes (symptoms, brain structures, endophenotypes), which are difficult to obtain retrospectively, may be subject to stricter data sharing restrictions or may be difficult to harmonize across studies. Detailed publication of all phenotypic information, where appropriate, may partially address concerns about heterogeneity.

Another issue is phenotyping quality in genomic datasets that are linked to electronic health records. Electronic health records may reflect diagnostic difficulties (e.g., rare phenotypes), stigma, or race- or gender-specific bias in admission, diagnosis, and treatment. These biases may be ameliorated through 1) replication of phenotypic and genetic associations; 2) accounting for order, length, and time between episodes and treatment to avoid conflating unrelated episodes into a diagnosis; and 3) using known disorder biology to verify diagnoses, e.g., using polygenic scores or genetic correlations to disentangle whether diagnoses match their predicted underlying biology (46).

Studies should report what details were requested, whether questionnaires were applied across the cohort (e.g., whether trauma histories were obtained from posttraumatic stress disorder controls or only cases), who determined the phenotypes, and potential inaccuracies. It is also crucial to state whether results were compared across samples with different phenotype operationalizations. For instance, a GWAS of age of onset in psychiatric and nonpsychiatric disorders yielded different significant hits according to the age of onset definitions that were applied (47). Finally, data collection protocols should emphasize the maximum level of detail so that analysts can eventually understand sources of variability in the data (48).

Sample Size Justifications

Sample size justifications should include power calculations or clarification of why power calculations are absent (a whole population study; resource constraints; heuristics based on literature or norms in the field; or no reason to specify power, e.g., no clear inference goal) (49).

Ameliorating Eurocentric Bias

In 2017, 88% of GWAS findings came from European ancestry samples. Indigenous samples represented only 0.02% of GWAS findings (40). Genetic association statistics are not generalizable across ancestries (50). Eurocentric bias is also present in pharmacogenetics. While 30% of studies use European ancestry samples, ancestry reporting is frequently incomplete or vague. This results in biased assessments of drug targets, which are potentially not relevant across populations (51).

To ameliorate Eurocentric bias, researchers should analyze data from minoritized populations. Data from different ancestries should be analyzed separately. Researchers should acknowledge underpowered samples (52) or use well-powered international biobank data or summary statistics from biobank networks (8). Researchers should also implement pipelines for multiancestral data analysis (53–56).

PREREGISTRATION AND REGISTERED REPORTS

Preregistration specifies hypotheses (or research questions for hypothesis-free designs), study design, and analysis plans before data collection or secondary data analyses begin (57). Registered reports (RRs) are peer-reviewed, preregistered methods and data collection protocols (58) that result in publication regardless of study outcome. We encourage journals to publish RRs, null results, and nonreplications. Guidance on preregistrations/RRs, including in consortia, is detailed elsewhere (57,59–63).

Despite increasing guidance, not all aspects of preregistration have been resolved. Preregistrations require substantial effort, but they can be completed using research plans required for proposals to ethics boards (64) or consortia for data access (59). Another issue is the cost of open RR publications, which is paid early in the process. This may prevent researchers from choosing RRs. We call on journals to provide clear RR pricing guidance and on funders to provide incentives (e.g., https://www.cancerresearchuk.org/funding-for-researchers/how-we-deliver-research/positive-research-culture/registered-reports).

DATA ACCESS AND DATA SHARING

Open annotated data and metadata are the gold standards for replicability. However, reasonable privacy concerns prevent making much genetic data fully open. The European Commission Directorate-General for Research and Innovation has called for data "as open as possible and as closed as necessary": "open" where possible for reuse, but "closed" to protect privacy (65). Studies of public perception of genomic data sharing in 22 countries demonstrated that attitudes vary on increasing public trust in the process (66) and willingness to share data with for-profit researchers and medical doctors (67) in return for results (68). Thus, we argue that participant safety and wishes must be prioritized and should not be assumed to be universal.

Nonetheless, respect for study participants is at odds with full data access. The onus falls on researchers to optimize approaches for making data accessible without risking identification of participants or to develop methods whereby summary statistics (which aggregate cohort-level genetic information) are sufficient to perform analyses. For multiancestral projects, ancestry-specific summary statistics are also necessary (69).

Within Consortia

The "secondary analysis proposal" model gives access to deidentified individual-level genotype and phenotype data to researchers with an approved analysis plan [implemented, for example, by the PGC (70)]. Where external users cannot easily access data, sites may employ a confederate model whereby analytical procedures and scripts are shared or site analysts run analyses and share summary data.

Within Institutions

Data sharing between teams may accelerate collaboration, reduce time spent on quality control (Data Quality Control and Analysis), and enable researchers to detect errors prior to publication. Institutions should incentivize such data sharing. A blueprint is the Quality-Ethics-Open Science-Translation Center (Berlin Institute of Health, Germany), which funds institution-wide data sharing (71).

A novel approach to data sharing is adversarial collaborations. Adversarial collaborations have research teams jointly access the data but test opposing hypotheses. Teams also reproduce analyses of their opponents and edit each other's descriptions of theory to prevent misrepresentation of opposing views (72).

Summary-Level Data

Alternatively, researchers may prioritize analyzing open summary statistics, including associated statistics such as linkage disequilibrium matrices. Gold standard approaches include, e.g., GCTA-COJO (genome-wide complex trait analysis-conditional and joint analysis) (73) for conditional analysis of GWAS summary statistics; FINEMAP (74) and PAINTOR (75) for fine-mapping loci; COLOC (76–78) and summary data-based Mendelian randomization (79) for expression quantitative trait loci annotation of GWAS summary statistics; and S-PrediXcan (80) for transcriptomic imputation. However, we note that these approaches perform best with linkage disequilibrium matrices of the original data.

Another issue may be limited access to summary-level data from commercial entities (17). We believe that journals could enforce data sharing, e.g., by accepting a publication only after its authors have provided complete summary statistics.

Indigenous Data Sovereignty

Indigenous data sovereignty (IDS) describes the right of Indigenous peoples to control access to data and samples that are collected on their lands and from their communities (81,82). Although IDS precludes broad data access, the benefits to open science are nevertheless myriad: inclusive research development, innovation, and improved citizen engagement (because Indigenous communities fear that fully open data could result in misinterpretation of eventual results and increased stigma) (81,82). Increased IDS will enhance accountability through transparency of data and sample handling and improve equity when Indigenous communities benefit through publication, research, or commercialization. Increased Indigenous participation would increase diversity and inclusion in genetics (81). These values, together with degrees of improved transparency and equity, are also open science values.

IDS guidance is summarized in the CARE Principles for Indigenous Data Governance (collective benefit, authority to control, responsibility and ethics) (83). CARE principles reflect tribal expectations of research conduct: tribal consent and input are followed throughout, e.g., when and how samples are returned to honor participants' spiritual practices; research benefits reflect tribal priorities; Indigenous individuals are employed in research projects and compensated; research materials and data belong to tribes who govern access, review research progress, provide institutional review boards, and grant secondary uses of data and research materials; findings are returned to communities; researchers respect Indigenous cultural and spiritual values; researchers and Indigenous communities are equal collaborators (84).

DATA QUALITY CONTROL AND ANALYSIS

Quality Control

Genetic data quality remains inconsistent (17). We suggest consistently following and describing completed quality control and imputation protocols and explaining any deviations from protocols in manuscripts and open datasets (which frequently lack information on the quality control protocol that was applied).

Open Source

Open source tools are commonly used and potentially improve replicability and reproducibility [e.g., PLINK (22,85)]. However, while software used for analysis is often shared, software and scripts for preprocessing data are not, which limits reproducibility (86).

Detailed Record of Analytical Steps and Open Code

While preregistrations (Preregistration and RRs) ideally provide specific research plans, these plans can expand (e.g., following reviewer suggestions). Researchers should keep detailed, transparent analysis records, e.g., in open online laboratory notebooks (87). Code should be openly deposited with version control and a codebook in which all variables are summarized (88). Workflows should be automated (files are not edited manually but result from how the script executes), with labeled input and output files. Relationships between data, data preprocessing, software, and analysis outcomes should be explained in a compendium article (89), e.g., an interactive website Jupyter notebook, or R Markdown files, which are convertible into books, articles, or websites (90).

Bioinformatics Tools

Development and maintenance of bioinformatics tools provides research tools and hands-on software development training, which most biologists (or psychiatry researchers) are not trained in. However, tool development is unscalable without continuous funding for maintenance, reproducibility, and training. Tool development is also not credited with coveted senior authorship, and software is not consistently cited (91). We echo the call from the bioinformatics community for consistent funding and counting software development toward career progression. To that end, researchers could make their bioinformatic work citable with persistent handles (e.g., generated through Zenodo, https://zenodo.org/) and consistently attribute authorship of software, data visualization tools, etc.

REPORTING

Language

Detailed reporting supports transparency, replicability, and reproducibility but may also be in tension with open science

values. Using language that participants prefer (to describe their experience, communities, etc.) improves inclusivity and reproduces information shared by participants. However, using unstandardized descriptors contradicts facilitating reproducibility through shared terminology. Consequently, we suggest that researchers provide descriptions from participants, operationalizations (e.g., diagnostic criteria), or, if these are unavailable, inclusion/exclusion criteria (e.g., self-diagnosis, disorder for which formal diagnostic criteria are lacking, symptom types, frequency, intensity, etc.). Below, we also briefly outline discussions about language choices for commonly reported data as a starting point for researchers.

Ancestry, Race, and Ethnicity. The use of race as interchangeable with ancestry and ethnicity is decreasing (92) but has not stopped. The terms have different meanings. Race is a sociopolitical concept whereby group identification is based on often stereotypical or politically influenced defining physical characteristics, e.g., skin pigmentation (54,93). Ethnicity, on the other hand, is a category that is based on group belonging through shared language and traditions. Ethnic groups may share genetic factors because of similar ancestral origins or be self-identified based on shared culture (54,94). Genetic ancestry refers to populations from which an individual has descended, which is reflected in DNA inherited from recent biological ancestors (54). Biogeographical ancestry labeling (African, Asian, and European) has increased (92). However, it has been criticized for being reminiscent of historical racial taxonomies (95), not capturing heterogeneity within local subpopulations (96), and being inconsistent with global genetic variation, which is continuous, not discrete continental (93). In addition, individuals who are not monoracial are frequently grouped into monolithic, imprecise, or othering categories, e.g., multicultural, admixed, bi/multiracial, and other (97). These descriptors are unacceptable.

At a minimum, researchers should follow antiracist genetics publication recommendations (94), which require clearly stated definitions and reasons for addressing race and ancestry. Race impacts health care delivery, disorder etiologies, and health outcomes. Accounting for race/ethnicity in modeling is necessary, but explicitly as a marker of inequality, never as a proxy for genetic ancestry (54). Ideally, researchers should work in interdisciplinary teams with affected communities (93) and consider outlining their own identities and motivations in manuscripts to build trust with affected communities (97).

Sex and Gender. Sex and gender are not synonymous, essentialist constructs. Gender encompasses sociopolitically constructed roles, behaviors, and identities (98). Sex is a set of physical and physiological variables (genitalia, gametes, karyotypes, gene expression, hormones) (98) that do not necessarily or fully determine sex (99). Common issues with reporting gender/sex differences include inconsistent or conflated use of sex and gender; sex or sample sizes per sex not reported; statistical evidence for claimed gender differences not provided; and no stated criteria for determining or ascertaining sex/gender (100,101).

Diagnosis and Patient Identity. Two general ways of describing patients are person-first language (e.g., a person with autism) and identity-first language (e.g., an autistic person). Person-first language prioritizes an individual, while communities, e.g., the English-speaking autism community, argue that identity-first descriptions emphasize that diagnoses can be proudly integral to one's identity (102). Participant preference should be respected; where unknown, researchers should be sensitive to historical agendas, priorities, and different experiences within the community (103). We also suggest that researchers transparently report how descriptions were determined.

Reporting Guidelines

Health care studies insufficiently report statistical methods, missing values, or reporting guidelines followed (104). Multiple reporting standards exist for common genetic designs (105–108). Furthermore, researchers should ensure that their results are not overinterpreted. Researchers should state why the evidence may be causal and clearly communicate any uncertainty (109–111), including in press releases.

Figures

Good figures clearly communicate approach, results, and key messages, thus improving transparency and reproducibility. To make figures accessible, researchers should address font size, readability, and color selection for individuals with visual impairments or color blindness. Accessibility and inclusivity also extend to imagery, mindful of lived experience. For example, eating disorder studies should avoid potentially triggering visuals, e.g., scales or distorted body images in mirrors. Graphics for case-control groups should not depict a singular gender presentation/race/body type as a sample case. Differences between cases and controls could be labeled with neutral colors (not green and red, implying good and bad for controls and cases; these colors are also unreadable to individuals with color-vision deficiencies). We provide free resources for creating accessible figures in Table S1.

Crucially, researchers should be aware of how data visualizations may be misinterpreted through a racist lens and coopted by far-right extremists to further white supremacy, especially given that much of the genetic research is Eurocentric (Study Design). Researchers have to anticipate how their visualizations, out of context and without labels, could be misconstrued. A preventive solution would be to standardize plot presentation in the field (112).

PREPRINTS

Preprints, nonpeer reviewed papers, are now integral to fast and free research sharing (113). Preprints also streamline submission to journals. In bioRxiv, 139 journals relevant to psychiatric genetics, psychiatry, and bioinformatics endorsed direct journal transfer. However, the preprint explosion during the COVID-19 pandemic highlighted limitations of rapid sharing, which we suggest that researchers be mindful of. Rapidly generated preprints, some of which may be methodologically flawed or fraudulent, may end up being integrated into unreliable meta-analyses or flawed health policy. Furthermore, flawed studies may lead to wasteful or

harmful follow-up research even after retraction of the original preprints (114).

CITIZEN SCIENCE AND PARTICIPANT INCLUSION

Knowledge coproduction (between patient experts and researchers) may be difficult due to pressures of timescales and funding (115). More extremely, coproduction could leave patients feeling tokenized or invalidated when their experience is deemed wrong against the views of psychiatrists or Eurocentric models of knowledge (116). Specific approaches may foster inclusive collaborations (115,116) including the following: 1) resources dedicated to supporting relationships with patient experts during and between projects; 2) open events where patients and researchers explore a specific key issue to identify creative solutions; 3) published evidence of the impact of coproduced research, supporting continuous collaborations; and 4) careful consultations between researchers and patient experts. Researchers should prepare to address boundaries to participation, conflicting views, and difficult journeys through health services (e.g., due to racism or coercion) (116,117). Patient experts should not be expected to share sensitive community knowledge. It should be clear how researchers will use knowledge that is obtained from patient experts (118).

In citizen science, nonacademic members of the public are active research collaborators, not passive data donors. Collaborations should reflect citizen science values: autonomy, fun, respect, altruism, inclusivity, openness, reciprocity, and solidarity (119,120). Researchers should plan how to navigate divergent collaborator priorities, e.g., when and how to release results to comply with institutional or regulatory requirements (121). In terms of compensation, citizen scientists wish for accessible data, clearly communicated findings, and acknowledgments in publications (122). Credit or intellectual property compensation, to be legally binding, needs to be detailed in contracts or institutional policy (121).

RESEARCH DISTRIBUTION

Researchers should consider how to communicate findings in such a way as to uphold open science values. While psychiatric geneticists support returning results to patients, most do not deem their knowledge of the process to be adequate (123) [although see recent international guidelines (124–126)]. In addition, scientists incorrectly assume that publication in peer-reviewed journals makes results accessible (121). Scientific publications may be publicly unavailable or written in a language that is difficult for nonspecialists to understand. To ensure access, American genomic citizen science projects provide links to open access publications on project websites (127).

Writing manuscripts in a nontechnical manner would be a good general practice. However, we acknowledge that it can be challenging. Extra (technical) details of analyses are required for replicability/reproducibility, but they decrease the overall accessibility of manuscripts. Several solutions could be put in place; for example, manuscripts could be accompanied by nontechnical press releases sent to participants in automated e-mails or letters. Researchers already explain their work or summarize talks (128) using blogs or social media posts (129). Accessible materials explaining research to nonacademic readers could be provided, e.g., on manuscript companion websites and websites of research societies (e.g., https://ispg.net/resources/educational-presentations, https://pgc.unc.edu/for-the-public/basic-genetics/), laboratories, or patient advocacy groups.

An alternative to written communication is live events. Accessibility practices include live captions/transcriptions, inclusive language, and translations, including sign language; affordable online conferences and ensuring that scholars who require visas can attend meetings; even division of time among speakers; and implementing a code of conduct (130). Alternatively, nonreal-time web conferences allow attendees to watch presentations, while remaining friendly toward time zones and caretaking commitments. Nonlive conferences also do not require a stable Internet connection (131).

Finally, because psychiatric genetics requires multidisciplinary technical training, open teaching materials would support learning about the newest techniques. Platforms such as Open Science Framework (https://osf.io/) or Zenodo permit free content uploads (slides, posters, and recordings of talks, which should include captions, transcripts, and translations). These teaching materials can have citable digital object identifiers generated. Textbooks can also be easily uploaded online from R Markdown files (Data Quality Control and Analysis).

CONCLUSIONS

Methodological rigor and careful collection of open data and their transparent, easy-to-follow presentation are key to improving the reproducibility, replicability, and accessibility of psychiatric genetics studies. However, psychiatric genetics remaining at the forefront of open science will depend on the ongoing support of and advocacy for collaborations and credit for junior, minority, technical, and nonprofessional researchers. Continuous funding is needed for research with non-European ancestry samples in a manner which respects the rights and privacy of underrepresented groups, following the increasingly embraced call to make data “as open as possible, as closed as necessary.” Funding is also required for the continuous software development that is necessary to analyze these complex data. Finally, consistent funding and institutional change of research norms would support researchers by providing them with extra time needed to run reproducible and replicable studies, providing clear research explanations for the public, creating open teaching materials, and maintaining ongoing collaborations with patient experts.

ACKNOWLEDGMENTS AND DISCLOSURES

APK was funded in part by the National Institute for Health and Care Research Maudsley Biomedical Research Centre at South London and Maudsley National Health Service Foundation Trust and King’s College London, and the Klarman Family Foundation. JSJ was funded by National Institute of Mental Health (Grant No. R01MH124839) and the Klarman Family Foundation. LMH was funded by National Institute of Mental Health (Grant Nos. R01MH118278 and R01MH124839), National Institute of Environmental Health Sciences (Grant No. R01ES033630), and the Klarman Family Foundation. The views expressed are those of the authors and not necessarily those of the funders.

We thank Carina Seah, Rebecca Signer, and Kayla Townsley for their helpful feedback.

APK and LMH conceptualized the study; APK, JSJ, and LMH wrote the original draft; APK, JSJ, and LMH reviewed and edited; APK created visualizations; APK administered the project; JSJ provided resources; LMH supervised the project.

The authors report no biomedical financial interests or potential conflicts of interest.

ARTICLE INFORMATION

From the Pamela Sklar Division of Psychiatric Genomics, Icahn School of Medicine at Mount Sinai, New York, New York (APK, JSJ, LMH); Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, New York (APK, LMH); Seaver Autism Center for Research and Treatment, Icahn School of Medicine at Mount Sinai, New York, New York (APK, LMH); Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, New York (APK, LMH); Department of Psychosis Studies, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, United Kingdom (APK); Psychiatry Department, The University of North Carolina at Chapel Hill School of Medicine, Chapel Hill, North Carolina (JSJ); and Department of Psychiatry, Yale University, New Haven, Connecticut (LMH).

Address correspondence to Laura M. Huckins, Ph.D., at laura.huckins@yale.edu.

Received May 25, 2023; revised Aug 4, 2023; accepted Aug 11, 2023.

Supplementary material cited in this article is available online at https://doi.org/10.1016/j.bpsgos.2023.08.007.

REFERENCES

1. Munafò MR, Nosek BA, Bishop DVM, Button KS, Chambers CD, Percie du Sert N, *et al.* (2017): A manifesto for reproducible science. Nat Hum Behav 1:0021.
2. Leonelli S, Spichtinger D, Prainsack B (2015): Sticks *and* carrots: Encouraging open science at its source. Geo 2:12–16.
3. Robinson D (2018): Open Source Alliance, What is open? | OSAOS [The Open Source Alliance for Open Scholarship] handbook. Available at: https://osaos.codeforscience.org/what-is-open/. Accessed September 22, 2022.
4. Murphy MC, Mejia AF, Mejia J, Yan X, Cheryan S, Dasgupta N, *et al.* (2020): Open science, communal culture, and women's participation in the movement to improve science. Proc Natl Acad Sci USA 117:24154–24164.
5. Sullivan PF, Agrawal A, Bulik CM, Andreassen OA, Børglum AD, Breen G, *et al.* (2018): Psychiatric genomics: An update and an agenda. Am J Psychiatry 175:15–27.
6. Uffelmann E, Huang QQ, Munung NS, de Vries J, Okada Y, Martin AR, *et al.* (2021): Genome-wide association studies. Nat Rev Methods Primers 1:59.
7. GTEx Consortium (2020): The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science 369:1318–1330.
8. Zhou W, Kanai M, Wu K-HH, Rasheed H, Tsuo K, Hirbo JB, *et al.* (2022): Global Biobank Meta-analysis Initiative: Powering genetic discovery across human disease. Cell Genomics 2:100192.
9. Abdill RJ, Blekhman R (2019): Tracking the popularity and outcomes of all bioRxiv preprints. eLife 8:e45133.
10. Johnson EC, Border R, Melroy-Greif WE, de Leeuw CA, Ehringer MA, Keller MC (2017): No evidence that schizophrenia candidate genes are more associated with schizophrenia than noncandidate genes. Biol Psychiatry 82:702–708.
11. Border R, Johnson EC, Evans LM, Smolen A, Berley N, Sullivan PF, Keller MC (2019): No support for historical candidate gene or candidate gene-by-interaction hypotheses for major depression across multiple large samples. Am J Psychiatry 176:376–387.
12. Sullivan PF (2007): Spurious genetic associations. Biol Psychiatry 61:1121–1126.
13. Duncan LE, Keller MC (2011): A critical review of the first 10 years of candidate gene-by-environment interaction research in psychiatry. Am J Psychiatry 168:1041–1049.
14. Claussnitzer M, Cho JH, Collins R, Cox NJ, Dermitzakis ET, Hurles ME, *et al.* (2020): A brief history of human disease genetics. Nature 577:179–189.
15. Colhoun HM, McKeigue PM, Davey Smith GD (2003): Problems of reporting genetic associations with complex outcomes. Lancet 361:865–872.
16. Grabe HJ, Van der Auwera S (2020): Gene-environment interaction in psychiatry. In: Bernhard TB, editor. Personalized Psychiatry. London: Academic Press, 363–373.
17. Burt C, Munafò M (2021): Has GWAS lost its status as a paragon of open science? PLoS Biol 19:e3001242.
18. Bell V (2017): Open science in mental health research. Lancet Psychiatry 4:525–526.
19. Burke NL, Frank GKW, Hilbert A, Hildebrandt T, Klump KL, Thomas JJ, *et al.* (2021): Open science practices for eating disorders research. Int J Eat Disord 54:1719–1729.
20. Scheibein F, Donnelly W, Wells JS (2022): Assessing open science and citizen science in addictions and substance use research: A scoping review. Int J Drug Policy 100:103505.
21. Marigorta UM, Rodríguez JA, Gibson G, Navarro A (2018): Replicability and prediction: Lessons and challenges from GWAS. Trends Genet 34:504–517.
22. Lin X (2020): Learning lessons on reproducibility and replicability in large scale genome-wide association studies. Harv Data Sci Rev 2.
23. Carter CS, Bearden CE, Bullmore ET, Geschwind DH, Glahn DC, Gur RE, *et al.* (2017): Enhancing the informativeness and replicability of imaging genomics studies. Biol Psychiatry 82:157–164.
24. Hübel C, Leppä V, Breen G, Bulik CM (2018): Rigor and reproducibility in genetic research on eating disorders. Int J Eat Disord 51:593–607.
25. Pratte MM, Audette-Chapdelaine S, Auger AM, Wilhelmy C, Brodeur M (2023): Researchers' experiences with patient engagement in health research: A scoping review and thematic synthesis. Res Involv Engagem 9:22.
26. Garrison NA, Hudson M, Ballantyne LL, Garba I, Martinez A, Taualii M, *et al.* (2019): Genomic research through an Indigenous lens: Understanding the expectations. Annu Rev Genomics Hum Genet 20:495–517.
27. Brand A, Allen L, Altman M, Hlava M, Scott J (2015): Beyond authorship: Attribution, contribution, collaboration, and credit. Learn Pub 28:151–155.
28. Ross MB, Glennon BM, Murciano-Goroff R, Berkes EG, Weinberg BA, Lane JI (2022): Women are credited less in science than men. Nature 608:135–145.
29. Liboiron M, Ammendolia J, Winsor K, Zahara A, Bradshaw H, Melvin J, *et al.* (2017): Equity in author order: A feminist laboratory's approach. Catalyst 3:1–17.
30. Heijstra TM, Steinthorsdóttir FS, Einarsdóttir T (2017): Academic career making and the double-edged role of academic housework. Gend Educ 29:764–780.
31. Limas JC, Corcoran LC, Baker AN, Cartaya AE, Ayres ZJ (2022): The impact of research culture on mental health & diversity in STEM. Chemistry 28:e202102957.
32. Veldink J, Al-Chalabi A (2017): Project MinE Publication Policy. Available at: https://www.projectmine.com/wp-content/uploads/2016/06/ProjectMinE_publicationpolicy_v2.pdf. Accessed May 4, 2022.
33. Corvin A (2015): Schizophrenia Working Group of the PGC [Psychiatric Genomics Consortium] revision of authorship. Available at: https://www.med.unc.edu/pgc/wp-content/uploads/sites/959/2019/01/SCZ_authorship.docx. Accessed May 5, 2022.
34. PTSD Working Group of the PGC (2015): PTSD [post-traumatic stress disorder] Working Group of the PGC [Psychiatric Genomics Consortium] draft authorship. Available at: https://pgc-ptsd.com/wp-content/uploads/2017/06/Authorship-Guidelines-PGC-PTSD.pdf. Accessed May 5, 2022.
35. Larivière V, Desrochers N, Macaluso B, Mongeon P, Paul-Hus A, Sugimoto CR (2016): Contributorship and division of labor in knowledge production. Soc Stud Sci 46:417–435.

36. Dung SK, López A, Barragan EL, Reyes RJ, Thu R, Castellanos E, *et al.* (2019): Illuminating women's hidden contribution to historical theoretical population genetics. Genetics 211:363–366.
37. Lount RB, Pettit NC (2023): Shared first authorship should be declared on academic CVs. Nat Hum Behav 7:659-659.
38. Lapidow A, Scudder P (2019): Shared first authorship. J Med Libr Assoc 107:618–620.
39. Rechavi O, Tomancak P (2023): Who did what: Changing how science papers are written to detail author contributions. Nat Rev Mol Cell Biol 24:519–520.
40. Mills MC, Rahal C (2019): A scientometric review of genome-wide association studies. Commun Biol 2:9.
41. Baldwin JR, Pingault JB, Schoeler T, Sallis HM, Munafò MR (2022): Protecting against researcher bias in secondary data analysis: Challenges and potential solutions. Eur J Epidemiol 37:1–10.
42. MacCoun RJ, Perlmutter S (2017): Blind analysis as a correction for confirmatory bias in physics and in psychology. In: Lilienfeld SO, Waldman ID, editors. Psychological Science Under Scrutiny. Hoboken, NJ: John Wiley & Sons, Inc., 295–322
43. Steegen S, Tuerlinckx F, Gelman A, Vanpaemel W (2016): Increasing transparency through a multiverse analysis. Perspect Psychol Sci 11:702–712.
44. Hofer SM, Piccinin AM (2009): Integrative data analysis through coordination of measurement and analysis protocol across independent longitudinal studies. Psychol Methods 14:150–164.
45. Cai N, Choi KW, Fried EI (2020): Reviewing the genetics of heterogeneity in depression: Operationalizations, manifestations and etiologies. Hum Mol Genet 29:R10–R18.
46. Dueñas HR, Seah C, Johnson JS, Huckins LM (2020): Implicit bias of encoded variables: Frameworks for addressing structured bias in EHR–GWAS data. Hum Mol Genet 29:R33–41.
47. Feng Y-CA, Ge T, Cordioli M, FinnGen GannaA, Smoller JW, Neale BM (2020): Findings and insights from the genetic investigation of age of first reported occurrence for complex disorders in the UK Biobank and FinnGen. medRxiv. https://doi.org/10.1101/2020.11.20.20234302.
48. Lambert CG, Black LJ (2012): Learning from our GWAS mistakes: From experimental design to scientific method. Biostatistics 13:195–203.
49. Lakens D (2022): Sample size justification. Collabra Psychol 8:33267.
50. Ding Y, Hou K, Xu Z, Pimplaskar A, Petter E, Boulier K, *et al.* (2023): Polygenic scoring accuracy varies across the genetic ancestry continuum. Nature 618:774–781.
51. Popejoy AB (2019): Diversity in precision medicine and pharmacogenetics: Methodological and conceptual considerations for broadening participation. Pharmgenomics Pers Med 12:257–271.
52. Ben-Eghan C, Sun R, Hleap JS, Diaz-Papkovich A, Munter HM, Grant AV, *et al.* (2020): Don't ignore genetic data from minority populations. Nature 585:184–186.
53. Liu D, Meyer D, Fennessy B, Feng C, Cheng E, Johnson JS, *et al.* (2022): Rare schizophrenia risk variant burden is conserved in diverse human populations. medRxiv. https://doi.org/10.1101/2022.01.03.22268662.
54. Peterson RE, Kuchenbaecker K, Walters RK, Chen CY, Popejoy AB, Periyasamy S, *et al.* (2019): Genome-wide association studies in ancestrally diverse populations: Opportunities, methods, pitfalls, and recommendations. Cell 179:589–603.
55. Highland HM, Wojcik GL, Graff M, Nishimura KK, Hodonsky CJ, Baldassari AR, *et al.* (2022): Predicted gene expression in ancestrally diverse populations leads to discovery of susceptibility loci for lifestyle and cardiometabolic traits. Am J Hum Genet 109:669–679.
56. Bhattacharya A, Hirbo JB, Zhou D, Zhou W, Zheng J, Kanai M, *et al.* (2022): Best practices for multi-ancestry, meta-analytic transcriptome-wide association studies: Lessons from the Global Biobank Meta-analysis Initiative. Cell Genomics 2:100180.
57. Nosek BA, Ebersole CR, DeHaven AC, Mellor DT (2018): The preregistration revolution. Proc Natl Acad Sci USA 115:2600–2606.
58. Chambers CD, Tzavella L (2022): The past, present and future of Registered Reports. Nat Hum Behav 6:29–42.
59. Zugman A, Harrewijn A, Cardinale EM, Zwiebel H, Freitag GF, Werwath KE, *et al.* (2022): Mega-analysis methods in ENIGMA: The experience of the generalized anxiety disorder working group. Hum Brain Mapp 43:255–277.
60. Haroz S (2022): Comparison of preregistration platforms. MetaArXiv. https://doi.org/10.31222/osf.io/zry2u.
61. Kiyonaga A, Scimeca JM (2019): Practical considerations for navigating Registered Reports. Trends Neurosci 42:568–572.
62. Bakker M, Veldkamp CLS, van Assen MALM, Crompvoets EAV, Ong HH, Nosek BA, *et al.* (2020): Ensuring the quality and specificity of preregistrations. PLS Biol 18:e3000937.
63. Van Den Akker OR, Weston S, Campbell L, Chopik B, Damian R, Davis-Kean P, *et al.* (2021): Preregistration of secondary data analysis: A template and tutorial. Meta-Psychol 5.
64. Evans TR, Branney P, Clements A, Hatton E (2023): Improving evidence-based practice through preregistration of applied research: Barriers and recommendations. Account Res 30:88–108.
65. Landi A, Thompson M, Giannuzzi V, Bonifazi F, Labastida I, da Silva Santos LOB, Roos M (2020): The "A" of FAIR – As open as possible, as closed as necessary. Data Intellegence 2:47–55.
66. Milne R, Morley KI, Almarri MA, Anwer S, Atutornu J, Baranova EE, *et al.* (2021): Demonstrating trustworthiness when collecting and sharing genomic data: Public views across 22 countries. Genome Med 13:92.
67. Middleton A, Milne R, Almarri MA, Anwer S, Atutornu J, Baranova EE, *et al.* (2020): Global public perceptions of genomic data sharing: What shapes the willingness to donate DNA and health data? Am J Hum Genet 107:743–752.
68. Milne R, Morley KI, Almarri MA, Atutornu J, Baranova EE, Bevan P, *et al.* (2022): Return of genomic results does not motivate intent to participate in research for all: Perspectives across 22 countries. Genet Med 24:1120–1129.
69. Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ (2019): Clinical use of current polygenic risk scores may exacerbate health disparities. Nat Genet 51:584–591.
70. Davis L (2019): The PGC data access portal and genomic privacy: Data sharing procedures to satisfy all communities. Eur Neuropsychopharmacol 29:S714.
71. Strech D, Weissgerber T, Dirnagl U, QUEST Group (2020): Improving the trustworthiness, usefulness, and ethics of biomedical research through an innovative and comprehensive institutional initiative. PLoS Biol 18:e3000576.
72. Martschenko DO, Trejo S (2022): Ethical, anticipatory genomics research on human behavior means celebrating disagreement. HGG Adv 3:100080.
73. Yang J, Ferreira T, Morris AP, Medland SE, Madden PAF, Heath AC, *et al.* (2012): Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. Nat Genet 44:369–375.
74. Benner C, Spencer CC, Havulinna AS, Salomaa V, Ripatti S, Pirinen M (2016): FINEMAP: Efficient variable selection using summary data from genome-wide association studies. Bioinformatics 32:1493–1501.
75. Kichaev G, Yang WY, Lindstrom S, Hormozdiari F, Eskin E, Price AL, *et al.* (2014): Integrating functional data to prioritize causal variants in statistical fine-mapping studies. PLoS Genet 10:e1004722.
76. Wallace C (2020): Eliciting priors and relaxing the single causal variant assumption in colocalisation analyses. PLoS Genet 16:e1008720.
77. Wallace C (2021): A more accurate method for colocalisation analysis allowing for multiple causal variants. PLoS Genet 17:e1009440.
78. Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, Plagnol V (2014): Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS Genet 10:e1004383.
79. Zhu Z, Zhang F, Hu H, Bakshi A, Robinson MR, Powell JE, *et al.* (2016): Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. Nat Genet 48:481–487.

80. Barbeira AN, Dickinson SP, Bonazzola R, Zheng J, Wheeler HE, Torres JM, *et al.* (2018): Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. Nat Commun 9:1825.
81. Hudson M, Garrison NA, Sterling R, Caron NR, Fox K, Yracheta J, *et al.* (2020): Rights, interests and expectations: Indigenous perspectives on unrestricted access to genomic data. Nat Rev Genet 21:377–384.
82. Rainie SC, Kukutai T, Walter M, Figueroa-Rodríguez OL, Walker J, Axelsson P (2019): Indigenous data sovereignty. In: Davies T, Walker S, Rubinstein M, Perini F, editors. The State of Open Data: Histories and Horizons. Cape Town and Ottawa: African Minds and International Development Research Centre, 300–319.
83. Carroll SR, Garba I, Figueroa-Rodríguez OL, Holbrook J, Lovett R, Materechera S, *et al.* (2020): The CARE principles for Indigenous data governance. Data Sci J 19:43.
84. Carroll SR, Garba I, Plevel R, Small-Rodriguez D, Hiratsuka VY, Hudson M, Garrison NA (2022): Using Indigenous standards to implement the CARE principles: Setting expectations through tribal research codes. Front Genet 13:823309.
85. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ (2015): Second-generation PLINK: Rising to the challenge of larger and richer datasets. GigaScience 4:7.
86. Peng RD, Dominici F, Zeger SL (2006): Reproducible epidemiologic research. Am J Epidemiol 163:783–789.
87. Schapira M, Open Lab Notebook Consortium, Harding RJ (2019): Open laboratory notebooks: Good for science, good for society, good for scientists. F1000Res 8:87.
88. Arslan RC (2019): How to automatically document data with the *codebook* package to facilitate data reuse. Adv Methods Pract Psychol Sci 2:169–187.
89. Peng RD (2011): Reproducible research in computational science. Science 334:1226–1227.
90. Xie Y, Allaire JJ, Grolemund G (2019): R Markdown: The Definitive Guide. Boca Raton, FL: Chapman & Hall/CRC.
91. Prins P, de Ligt J, Tarasov A, Jansen RC, Cuppen E, Bourne PE (2015): Toward effective software solutions for big biology. Nat Biotechnol 33:686–687.
92. Byeon YJJ, Islamaj R, Yeganova L, Wilbur WJ, Lu Z, Brody LC, Bonham VL (2021): Evolving use of ancestry, ethnicity, and race in genetics research-A survey spanning seven decades. Am J Hum Genet 108:2215–2223.
93. Lewis ACF, Molina SJ, Appelbaum PS, Dauda B, Di Rienzo A, Fuentes A, *et al.* (2022): Getting genetic ancestry right for science and society. Science 376:250–252.
94. Brothers KB, Bennett RL, Cho MK (2021): Taking an antiracist posture in scientific publications in human genetics and genomics. Genet Med 23:1004–1007.
95. Rajagopalan R, Fujimura JH (2012): Making history via DNA, making DNA from history: Deconstructing the race-disease connection in admixture mapping. In: Wailoo K, Nelson A, Lee C, editors. Genetics and the Unsettled Past. New Brunswick, NJ: Rutgers University Press, 143–163.
96. Weiss KM, Long JC (2009): Non-Darwinian estimation: My ancestors, my genes' ancestors. Genome Res 19:703–710.
97. Martschenko DO, Wand H, Young JL, Wojcik GL (2023): Including multiracial individuals is crucial for race, ethnicity and ancestry frameworks in genetics and genomics. Nat Genet 55:895–900.
98. Heidari S, Babor TF, De Castro P, Tort S, Curno M (2016): Sex and gender equity in research: Rationale for the SAGER guidelines and recommended use. Res Integr Peer Rev 1:2.
99. Miyagi M, Guthman EM, Sun SED (2021): Transgender rights rely on inclusive language. Science 374:1568–1569.
100. Garcia-Sifuentes Y, Maney DL (2021): Reporting and misreporting of sex differences in the biological sciences. eLife 10:e70817.
101. Rechlin RK, Splinter TFL, Hodges TE, Albert AY, Galea LAM (2022): An analysis of neuroscience and psychiatry papers published from 2009 and 2019 outlines opportunities for increasing discovery of sex differences. Nat Commun 13:2137.
102. Monk R, Whitehouse AJO, Waddington H (2022): The use of language in autism research. Trends Neurosci 45:791–793.
103. Vivanti G (2020): Ask the Editor: What is the most appropriate way to talk about individuals with a diagnosis of autism? J Autism Dev Disord 50:691–693.
104. Held U, Steigmiller K, Hediger M, Cammann VL, Garaiman A, Halvachizadeh S, *et al.* (2022): The incremental value of the contribution of a biostatistician to the reporting quality in health research-A retrospective, single center, observational cohort study. PLoS One 17:e0264819.
105. Wand H, Lambert SA, Tamburro C, Iacocca MA, O'Sullivan JW, Sillari C, *et al.* (2021): Improving reporting standards for polygenic scores in risk prediction studies. Nature 591:211–219.
106. Skrivankova VW, Richmond RC, Woolf BAR, Yarmolinsky J, Davies NM, Swanson SA, *et al.* (2021): Strengthening the reporting of observational studies in epidemiology using Mendelian randomization: The STROBE-MR statement. JAMA 326:1614–1621.
107. Little J, Higgins JPT, Ioannidis JPA, Moher D, Gagnon F, von Elm E, *et al.* (2009): STrengthening the REporting of Genetic Association Studies (STREGA): An extension of the STROBE statement. PLoS Med 6:e22.
108. Little J, Higgins J, editors. (2006). The HuGENet™ HuGE Review Handbook, Version 1.0. Available at: http://www.medicine.uottawa.ca/public-health-genomics/web/assets/documents/HuGE_Review_Handbook_V1_0.pdf. Accessed July 23, 2023.
109. Haber NA, Wieten SE, Rohrer JM, Arah OA, Tennant PWG, Stuart EA, *et al.* (2022): Causal and associational language in observational health research: A systematic evaluation. Am J Epidemiol 191:2084–2097.
110. Ioannidis JP, Boffetta P, Little J, O'Brien TR, Uitterlinden AG, Vineis P, *et al.* (2008): Assessment of cumulative evidence on genetic associations: Interim guidelines. Int J Epidemiol 37:120–132.
111. van der Bles AM, van der Linden S, Freeman ALJ, Mitchell J, Galvao AB, Zaval L, Spiegelhalter DJ (2019): Communicating uncertainty about facts, numbers and science. R Soc Open Sci 6:181870.
112. Carlson J, Henn BM, Al-Hindi DR, Ramachandran S (2022): Counter the weaponization of genetics research by extremists. Nature 610:444–447.
113. Kirkham JJ, Penfold NC, Murphy F, Boutron I, Ioannidis JP, Polka J, Moher D (2020): Systematic examination of preprint platforms for use in the medical and biomedical sciences setting. BMJ, (Open) 10:e041849.
114. Watson C (2022): Rise of the preprint: How rapid data sharing during COVID-19 has changed science forever. Nat Med 28:2–5.
115. Staniszewska S, Hickey G, Coutts P, Thurman B, Coldham T (2022): Co-production: A kind revolution. Res Involv Engagem 8:4.
116. Rose D, Kalathil J (2019): Power, privilege and knowledge: The untenable promise of co-production in mental "health." Front Sociol 4:57.
117. Blakey H (2005): Participation . . . why bother?: The views of Black and minority ethnic mental health service users on participation in the NHS in Bradford. University of Bradford, Department of Peace Studies, International Centre for Participation Studies. ICPS Working Paper 2. Available at: http://hdl.handle.net/10454/3798. Accessed February 15, 2022.
118. Montgomery L, Hartley J, Neylon C, Gillies M, Gray E, Herrmann-Pillath C, *et al.* (2021): Diversity. In: Open Knowledge Institutions: Reinventing Universities. Cambridge, MA: The MIT Press, 47–65.
119. Eitzel MV, Cappadonna JL, Santos-Lang C, Duerr RE, Virapongse A, West SE, *et al.* (2017): Citizen science terminology matters: Exploring key terms. CSTP 2:1.
120. Guerrini CJ, Trejo M, Canfield I, McGuire AL (2022): Core values of genomic citizen science: Results from a qualitative interview study. BioSocieties 17:203–228.
121. Guerrini CJ, Contreras JL (2020): Credit for and control of research outputs in genomic citizen science. Annu Rev Genomics Hum Genet 21:465–489.

122. de Vries M, Land-Zandstra A, Smeets I (2019): Citizen scientists' preferences for communication of scientific output: A literature review. CSTM 4:2.
123. Lázaro-Muñoz G, Torgerson L, Pereira S (2021): Return of results in a global survey of psychiatric genetics researchers: Practices, attitudes, and knowledge. Genet Med 23:298–305.
124. Matimba A, Ali S, Littler K, Madden E, Marshall P, McCurdy S, *et al.* (2022): Guideline for feedback of individual genetic research findings for genomics research in Africa. BMJ Glob Health 7:e007184.
125. Aizawa Y, Nagami F, Ohashi N, Kato K (2020): A proposal on the first Japanese practical guidance for the return of individual genomic results in research settings. J Hum Genet 65:251–261.
126. Lewis ACF, Knoppers BM, Green RC (2021): An international policy on returning genomic research results. Genome Med 13:115.
127. Guerrini CJ, Lewellyn M, Majumder MA, Trejo M, Canfield I, McGuire AL (2019): Donors, authors, and owners: How is genomic citizen science addressing interests in research outputs? BMC Med Ethics 20:84.
128. Ekins S, Perlstein EO (2014): Ten simple rules of live tweeting at scientific conferences. PLoS Comput Biol 10:e1003789.
129. Cheplygina V, Hermans F, Albers C, Bielczyk N, Smeets I (2020): Ten simple rules for getting started on Twitter as a scientist. PLoS Comput Biol 16:e1007513.
130. Whitaker KJ, Guest O (2020): #bropenscience is broken science. Psychologist 33:34–37.
131. Arnal A, Epifanio I, Gregori P, Martínez V (2020): Ten simple rules for organizing a non-real-time web conference. PLoS Comput Biol 16: e1007667.