

RESEARCH INTEGRITY

Fostering reproducibility in industry-academia research

Sharing can pose challenges for collaborations

By **B. R. Jasny**,¹ **N. Wigginton**,² **M. McNutt**,^{3*} **T. Bubela**,⁴ **S. Buck**,⁵ **R. Cook-Deegan**,⁶ **T. Gardner**,⁷ **B. Hanson**,⁸ **C. Hustad**,⁹ **V. Kiermer**,¹⁰ **D. Lazer**,¹¹ **A. Lupia**,² **A. Manrai**,¹² **L. McConnell**,¹³ **K. Noonan**,¹⁴ **E. Phimister**,¹⁵ **B. Simon**,¹⁶ **K. Strandburg**,¹⁷ **Z. Summers**,¹⁸ **D. Watts**¹⁹

Many companies have proprietary resources and/or data that are indispensable for research, and academics provide the creative fuel for much early-stage research that leads to industrial innovation. It is essential to the health of the research enterprise that collaborations between industrial and university researchers flourish. This system of collaboration is under strain. Financial motivations driving product development have led to concerns that industry-sponsored research comes at the expense of transparency (1). Yet many industry researchers distrust quality control in academia (2) and question whether academics value reproducibility as much as rapid publication. Cultural differences between industry and academia can create or increase difficulties in reproducing research findings. We discuss key aspects of this problem that industry-academia collaborations must address and for which other stakeholders, from funding agencies to journals, can provide leadership and support.

Here we are not talking about irreproducibility caused by fundamental gaps in knowledge, which are intrinsic to the nature of scientific research, but situations in which incomplete communication and sharing of techniques, data, or materials interferes with independent validation or future investigations. Irreproducibility has serious economic consequences. Representatives of venture firms and industries such as biopharma argue that they must replicate findings from academic research before investing. For preclinical research, this can involve, on average, two to six researchers, 1 to 2 years, and \$500,000 to

\$2,000,000 per project (2). For academic scientists, an inability to trust research findings means an erosion of confidence from the scientific community, decision-makers, and the general public, as well as the waste of scarce resources.

BARRIERS TO SHARING

Efforts to promote reproducible research have varied. One widely supported strategy is to increase the availability of data produced in studies, along with computer code written to clean and analyze data. Publishers and funders have instituted policies mandating data deposition or data management plans; however, success has not been uniform.

There are disincentives to open sharing of information. For academic research, rewards come from public presentations and publications that lead to recognition within the community, grants, and tenure. The emphasis on publications to reap academic rewards means that academic researchers can be reluctant to release information or even to fully describe their work. In industry, publishing is typically not a high priority; the goals are the provision of a product (whether goods or services) that will outstrip competitors and provide monetary rewards to investors. The need to obtain patents or maintain trade secrets to protect intellectual property (IP) can provide a strong financial incentive to not disclose or share information. Corporations see relatively little advantage to releasing data for research purposes, so any nonzero risk of consequences (even if only hypothetical) can be sufficient to shut down such efforts.

Certain barriers to data access and

sharing confront both industry and academia, e.g., protecting privacy (3). There are legitimate concerns about privacy of research participants, or those whose data is collected and used for research. This includes not only health-related data used in preclinical and clinical investigations, but also data used in studies of complex social systems where, for example, the pooling of massive data sets of cell phone use with credit card information might conceivably reveal confidential information. Use of proprietary data obtained via researcher-company agreements that restrict data access by others is being debated in the social science community (4).

Sharing also has transactional costs, such as time and resources required to prepare data and materials for sharing, determining how and where to share, storing and curating the data, and following up to ensure that shared information is being correctly understood.

BUILDING BETTER PARTNERSHIPS

Despite resources available to help guide formation of industry-academia partnerships (5, 6), most such agreements lack uniform approaches. We identify key issues to be addressed by all parties in these partnerships (see the table). Some of the issues apply to any collaboration whether with industry, government, nongovernmental organizations, or academics, but the points bear repeating as problems continue to emerge. All partners should identify any IP concerns before engaging, and outline the boundaries of the work (what should be done in-house and what can be shared) before entering into any external agreement. Industry representatives must engage university representatives (i.e., dedicated staff in corporate relations, sponsored projects, and/or technology transfer offices) directly. Drafting contracts or even memoranda of understanding with individual researchers will likely be problematic because individuals may or may not have legal authority to grant licenses or make promises about IP rights arising from work performed by university personnel.

Early-career researchers will typically have incentives to publicize their findings in talks and publications, although some might want to have exposure to industry research projects for which publication is not an option. Industry-academia partnerships involving students and postdocs will there-

¹Science, Washington, DC 20005, USA. ²University of Michigan, Ann Arbor, MI 48109, USA. ³National Academy of Sciences, Washington, DC 20001, USA. ⁴Faculty of Health Sciences, Simon Fraser University, British Columbia, BC V5A 1S6 Canada. ⁵Laura and John Arnold Foundation, Houston, TX 77056, USA. ⁶Consortium for Science Policy and Outcomes at Arizona State University, Washington, DC 20009, USA. ⁷Riffyn, Inc., Oakland, CA 94612, USA. ⁸American Geophysical Union, Washington, DC 20009, USA. ⁹Merck & Co., Inc., Kenilworth, NJ 07033, USA. ¹⁰Public Library of Science (PLOS), San Francisco, CA 94111, USA. ¹¹Northeastern University, Boston, MA 02115, USA. ¹²Harvard Medical School, Boston, MA 02115, USA. ¹³Bayer U.S., Research Triangle Park, NC 27709, USA. ¹⁴McDonnell Boehnen Hulbert & Berghoff LLP, Chicago, IL 60606, USA. ¹⁵The New England Journal of Medicine, Boston, MA 02115, USA. ¹⁶Thomas Jefferson School of Law, San Diego, CA 92101, USA. ¹⁷New York University School of Law, New York, NY 10012, USA. ¹⁸ExxonMobil Research and Engineering Company, Annandale NJ 08801, USA. ¹⁹Microsoft Research, New York, NY 10003, USA. *Email: naspresident@nas.edu

fore likely be more successful with exploratory research projects instead of targeted or trade-secret research.

If publication is a desired end result, it will be important for academic participants to understand whether or not their industry partner has the right to review (and potentially delay or veto) manuscripts before publication. A recent study of archived research protocols and journal publications of randomized clinical trials revealed that for more than 80% of trials, the industry partner had such rights (7). In the interest of integrity, transparency, and reproducibility, researchers and academic medical centers should be careful to retain academic freedom and fully declare any potential conflicts in papers or presentations.

Once the collaborative research agreement is executed, the terms need to be explicitly communicated among the re-

Another impediment that would require open communication between academic and industry researchers is concern that the joint research might expose some unintentional failing in industry processes or products that will harm the organization's reputation. One possible approach to this would be to suggest that industry be given time to devise and implement corrections to the problem that are published along with the exposition of the problem, in the same way that software patches are rolled out with notifications of the vulnerabilities.

ROLES FOR OTHER STAKEHOLDERS

Universities. Students and faculty in academia, and practicing scientists in industry, have not been adequately trained in long-established best practices for experimental design, data science, and statistical analysis. Opportunities to incorporate

of research and a variety of contributions for hiring and promotion in each school or department, rather than relying solely on quantity of publications or citation indicators. Robust partnerships with industry should also be considered. Where appropriate, universities should recognize contributions of a range of researchers, including data providers, method developers, technology innovators, and entrepreneurs.

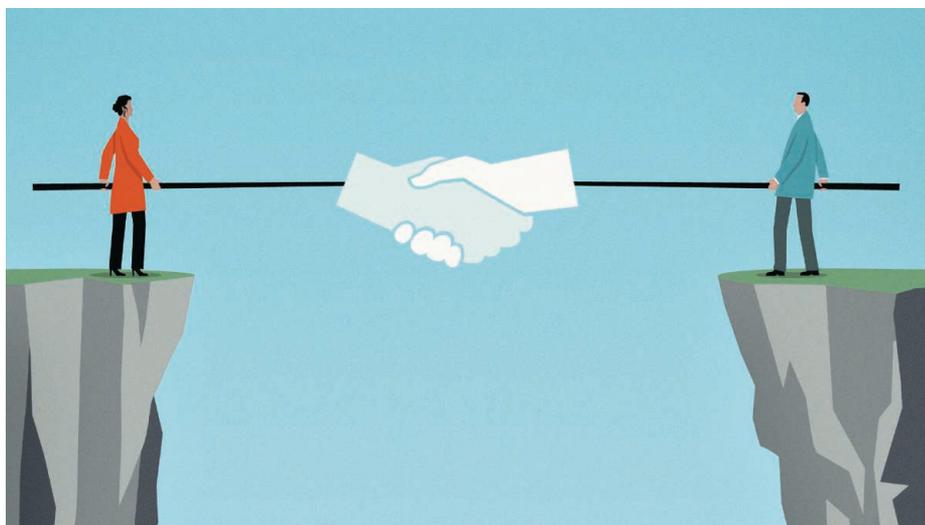
Funding agencies. Major funding agencies (such as the U.S. National Institutes of Health and National Science Foundation) have begun to require that prospective grantees provide information regarding reproducibility. If all funding agencies were to require adherence to uniform standards, including a section in every proposal for how the research team plans to test the reproducibility of its findings, universities would have a strong incentive to comply quickly. Follow-up monitoring after the receipt of grants and during the course of the funded research to ensure adherence to standards for data-sharing and reproducibility is recommended. Funding agencies could recognize and reward principal investigators or universities to incentivize data-sharing.

Funding agencies can do more in support of data and material repositories (10); these are key resources in enabling tests of reproducibility. To this end, as well, funders can and should provide ongoing support for software, open notebooks, data management systems, and other infrastructure.

Finally, although standard setting is not a trivial endeavor, funding agencies should support community-endorsed, discipline-specific standards to define what constitutes useful and necessary metadata as part of data deposition.

Journals. As more journals strive for transparency and mandate that data upon which conclusions are based be shared, it is important that all collaborators on a project, both academic and industry-based, understand the ramifications of these requirements in their dissemination plans. The consequences of ignoring that can be severe. Journals have rejected submissions in late stages of revision when it became apparent that binding agreements between academic authors and industry partners prevented adequate sharing of data and methodologies, or the disclosure of funding sources.

Although some journals remain more noncommittal about data sharing (12), a growing number of journals and research organizations have successfully adopted policies that require data sharing upon publication (see, for example, information for contributors/authors at *Science*, *Nature*,



searchers, and regular meetings must occur with all teams involved to ensure that expectations are being met and that there are no misunderstandings. For larger-scale collaborations, such as consortia of universities, government entities, and/or companies, this becomes even more important. The University of Michigan, for example, hosts a consortium of over 65 companies working together in a precompetitive environment around automated and Internet-connected vehicles (mcity.umich.edu). Regular meetings with groups representing many industries, combined with identical terms and conditions for disclosure and data-sharing, have helped ease concerns of releasing proprietary data and/or trade secrets. Likewise, the Structural Genomics Consortium has worked with firms and academics to determine structures of macromolecules and in precompetitive demonstration of promising clinical applications (8).

the incentive structures, lessons learned, and quality assurance practices of industry into academic education need to be more fully explored.

University administrators could take a more active role by randomly auditing faculty research (performing quality assurance checks) to assess best practices for reproducibility. Faculty and administrators would have to view such programs as constructive, not punitive, with an eye toward minimizing faculty burden. Alternatively, universities receiving corporate sponsorship for research could be required to return funding if the results prove to be irreproducible—or research agreements could incorporate bonuses to universities for demonstrably reproducible work (9).

The way the tenure system is operationalized in many universities is antagonistic to reproducible research. Universities need to adopt better metrics for assessing quality

and *PLOS*). In political science, prominent journals such as the *American Journal of Political Science* are requiring that once a paper is accepted, the authors submit a replication data set or archive that will be reviewed and made available at the site upon publication (13).

When it is not possible to de-identify protected participant-level data, journals have insisted on other approaches, such as public sharing of aggregated data or a statement of how qualified researchers can gain access to the full data set. For example, *Science* magazine has, after consultation with its board and reviewers, allowed authors, on a case-by-case basis to provide aggregated data or provide a randomly sampled subset of data for replication to protect privacy of research subjects (14, 15).

A step toward consensus among journals was the development of Transparency and Openness Promotion (TOP) guidelines for journals (11). These guidelines identified different levels of transparency and disclosure that could be mandated by journals, with the highest level requiring that data, code, and materials be submitted to a repository and also that the analysis be verified by independent replication before publication. Implementation of the TOP guidelines, however, is still a work in progress and cannot by itself be a complete solution. Industry and academic researchers are wary of more administrative burdens, and journal editors have been reluctant to impose norms in the absence of community consensus.

Progress toward achieving consensus within particular communities has occurred as well. The International Council of Medical Journal Editors recently announced that, as of July 2018, submitted reports of clinical trials must at a minimum include a data-sharing plan, and reports of clinical trials beginning after January 2019 will not be considered unless a data-sharing plan is provided in a public registry prior to enrollment of the first patient (16). The Coalition on Publishing Data in the Earth and Space Sciences is developing a site for best practices as they evolve (17). Continued dialogue within the publishing community will help provide leadership in promoting best practices.

Journals should continue to clarify guidelines about their data availability requirements and legitimate exceptions to those requirements. The goal is for guidelines to reflect modern standards and expectations on reproducibility and to strive toward higher levels of compliance in the TOP guidelines. Journals should explicitly state that funding sources, including in-kind contributions, and industry relations must be disclosed and that information regarding

Partnership issues

Both sides of industry-academia partnerships need to explicitly address key issues in writing before the partnership begins, including:

- Has a research plan, including agreed-on deliverables and milestones, been included, and do the academic researchers agree with all its provisions?
- What data and/or materials are needed to accomplish the goals of the project? What material transfer agreements (MTAs) will be needed?
- Who will be generating data?
- What will be proprietary? Industry partners must make clear what data are proprietary before they are even collected and analyzed, and if de-identification (e.g., of research participants or origins of the data) will obviate concerns.
- Even for proprietary data, to what extent can industry agree to make data available for replication purposes in restricted-access data enclaves?
- Can results of the research be presented at public meetings? What approvals are needed? What is protected by nondisclosure agreements and what can be public?
- Will data be released, and what form can the data take if made publicly available (e.g., raw data, de-identified data, aggregated data)? Data release needs to be timely, intentional, and done in compliance with agreements.
- Can released data be used to extend research findings or only to reproduce them?
- Can results be published? What approvals are needed for publication? Are there any restrictions on publication?
- What happens to follow-up research by the academic partner that is based on the results of the collaboration and/or has been done with industry funding?
- Are patents expected as a result of the research?
- What other institutional partners are performing similar work that is related to this agreement and could conflict with a nascent partnership?
- What happens in the case of a dispute around the terms of the agreement?

material transfer agreements (MTAs) and other restrictions on presentation and sharing of research results be given to journal editors at the time of submission. Editors must enforce guidelines and policies by rejecting manuscripts that do not comply. Like funding agencies, journals and professional societies should acknowledge and reward good behavior (18).

The Creative Commons license CC0 (which waives all rights of data authors) is attractive as it does not require data sets to have a “provenance” trail and can thus ease automated mining of data. However, lack of provenance tracking in CC0 creates challenges for data evaluation, interpretation of analyses, and accreditation of data generators, thus making CC-BY (in which author attribution is required) attractive. Both continue to be discussed as aspirational goals.

Irreproducible research wastes time, money, and resources. Academic researchers, universities, and other institutions, industry, funding agencies, and editors all have a role to play in raising research standards and creating an environment of trust between communities. ■

REFERENCES AND NOTES

1. J. C. Begley *et al.*, *PLOS ONE* **12**, e0175643 (2017).
2. L. P. Freedman, I. M. Cockburn, T. S. Simcoe, *PLOS Biol.* **13**, e1002165 (2015).
3. L. L. Rodriguez, L. D. Brooks, J. H. Greenberg, E. D. Green, *Science* **339**, 6117 (2013).
4. J. Markoff, “Trove of personal data, forbidden to researchers”; <https://nyti.ms/2pwqmf6> (2012).
5. University Industry Demonstration Partnership (UIDP), Researcher guidebook; <https://www.uidp.org/publication/researcher-guidebook-and-quick-guide/> (2012).
6. National Council of University Research Administrators (NCURA), Guiding principles for university-industry endeavors; http://sites.nationalacademies.org/cs/groups/pgasite/documents/webpage/pga_044335.pdf (2006).
7. B. Kasenda *et al.*, *PLOS Med.* **13**, e1002046 (2016).
8. E. Low, C. Bountra, W. H. Lee, *ecancermedicalscience* **10**, ed57 (2016).
9. M. Rosenblatt, *Sci. Transl. Med.* **8**, 336 (2016).
10. www.icpsr.umich.edu/files/ICPSR/pdf/DomainRepositoriesCTA16Sep2013.pdf
11. B. A. Nosek *et al.*, *Science* **348**, 1422 (2015).
12. www.elsevier.com/about/our-business/policies/research-data/research-data-faqs; accessed 8 July 2017.
13. www.insidehighered.com/blogs/rethinking-research/should-journals-be-responsible-reproducibility
14. E. Bakshy *et al.*, *Science* **348**, 1130 (2015).
15. Y. A. de Montjoye *et al.*, *Science* **347**, 536 (2015).
16. www.nejm.org/doi/full/10.1056/NEJMe1705439#t=article
17. www.copdess.org/copdess-suggested-author-instructions-and-best-practices-for-journals/
18. M. R. Munafò *et al.*, *Nat. Hum. Behav.* **1**, 0021 (2017).

ACKNOWLEDGMENTS

We thank participants of a workshop on this topic held in Washington, DC, on 24 to 25 October 2016 and supported by a grant from the Laura and John Arnold Foundation to the American Association for the Advancement of Science. Views expressed herein represent those of the authors and do not necessarily reflect views of their employers or sponsor. The authors had no input into the handling of the paper after submission or supervisory authority over anyone at the journal.

10.1126/science.aan4906

Fostering reproducibility in industry-academia research

B. R. Jasny, N. Wigginton, M. McNutt, T. Bubela, S. Buck, R. Cook-Deegan, T. Gardner, B. Hanson, C. Hustad, V. Kiermer, D. Lazer, A. Lupia, A. Manrai, L. McConnell, K. Noonan, E. Phimister, B. Simon, K. Strandburg, Z. Summers and D. Watts

Science **357** (6353), 759-761.
DOI: 10.1126/science.aan4906

ARTICLE TOOLS

<http://science.sciencemag.org/content/357/6353/759>

REFERENCES

This article cites 10 articles, 3 of which you can access for free
<http://science.sciencemag.org/content/357/6353/759#BIBL>

PERMISSIONS

<http://www.sciencemag.org/help/reprints-and-permissions>

Use of this article is subject to the [Terms of Service](#)

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. 2017 © The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works. The title *Science* is a registered trademark of AAAS.