OPINION ARTICLE

The imperative to find the courage to redesign the biomedical research enterprise [version 1; peer review: 1 approved with reservations, 1 not approved]

Bibiana Bielekova ¹, Shannon Brownlee^{2,3}

¹Neuroimmunological Diseases Section, National Institute of Allergy and Infectious Diseases, USA, Bethesda, MD, 20892, USA

²Lown Institute, Brookline, MA, USA

³School of Public Health, George Washington University, Washington, DC, USA

V1 First published: 23 Jul 2021, 10:641
<https://doi.org/10.12688/f1000research.53654.1>
 Latest published: 23 Jul 2021, 10:641
<https://doi.org/10.12688/f1000research.53654.1>

Abstract

Medical research aims to improve health for everyone. While its advances are undeniable, the pace and cost of the progress are not optimal. For example, independent analyses concluded that at least half of the published biomedical research findings are irreproducible, and most scientific papers are never read or cited. This paper examines biomedical research holistically, as a system of incentives that shape the behavior of scientists, administrators, publishers, and funders, and are detrimental to medical progress. We identify opportunities to change and improve those incentives by altering the way research output is disseminated and evaluated, and recommend transparent, data-driven measures of methodological rigor, reproducibility, and societal value of scientific discoveries. Embracing these opportunities would maximize our investments in biomedical research and optimize its value to human health, while simultaneously increasing freedom, creativity, and satisfaction of the scientific workforce.

Keywords

research reproducibility, science administration, scientific publication system, incentive structure in science

This article is included in the **Research on Research, Policy & Culture** gateway.

Open Peer Review

Approval Status ? X

	1	2
version 1		
23 Jul 2021	view	view

1. **Paul P. Glasziou** , Bond University, Gold Coast, Australia

2. **Andrew Lee Hufton** , Leibniz-Institut DSMZ, Braunschweig, Germany

Any reports and responses or comments on the article can be found at the end of the article.

Corresponding author: Bibiana Bielekova (Bibi.Bielekova@nih.gov)

Author roles: **Bielekova B:** Conceptualization, Data Curation, Formal Analysis, Methodology, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Brownlee S:** Writing – Review & Editing

Competing interests: Dr. Bielekova is a NIH employee. No other competing interests were disclosed.

Grant information: This study was supported (in part) by the Intramural Research Program of the National Institute Allergy and Infectious Diseases (NIAID) of the National Institutes of Health (NIH) and the Lown Institute.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2021 Bielekova B and Brownlee S. This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The author(s) is/are employees of the US Government and therefore domestic copyright protection in USA does not apply to this work. The work may be protected under the copyright laws of other jurisdictions when used in those jurisdictions.

How to cite this article: Bielekova B and Brownlee S. **The imperative to find the courage to redesign the biomedical research enterprise [version 1; peer review: 1 approved with reservations, 1 not approved]** F1000Research 2021, 10:641 <https://doi.org/10.12688/f1000research.53654.1>

First published: 23 Jul 2021, 10:641 <https://doi.org/10.12688/f1000research.53654.1>

Introduction

At least half of the human and financial resources currently devoted to the biomedical academic research generates findings that may not enhance collective knowledge or are unlikely to improve health. Every independent attempt to reproduce published biomedical scientific studies found stunning irreproducibility rates ranging from 51 to 89%.^{1–7} In a survey of more than 1500 researchers, 90% of respondents believe there is a reproducibility crisis in biomedical sciences.⁸

Current solutions to this crisis focus on encouraging better training of scientists in the methods of reproducible research and transparent reporting of methodological details.^{9,10} While these efforts target important aspects of the problem, they fail to address its roots.

This paper treats the research enterprise as a system and examines how that system shapes the behavior of scientists and administrators. It asks whether the current design of publicly funded biomedical research maximizes its societal value and considers modifications to the organizational structure to enhance research reproducibility and its relevance for human health.

What is the desired outcome or mission of biomedical research?

The mission of the US National Institutes of Health (NIH), the largest publicly funded biomedical research enterprise in the world, is to “Uncover new knowledge that will lead to better health for everyone.” One could assume that “uncovering new knowledge” should be optimized in terms of speed and value to improving health. To “lead to better health for everyone,” the knowledge must be both reproducible, which ensures that we can trust the results and build upon them, and relevant to human health.¹¹ To achieve optimal speed and value in advancing the NIH’s mission, a system must be organized around the pursuit of these implicit goals, which presently it is not. One way to see this is to explore the differences between the current biomedical research enterprise, and uniquely successful past public projects such as the Manhattan Project,¹² and the Apollo Mission.¹³

The Manhattan and Apollo projects were entirely mission-driven with defined, concrete end-products. Accomplishment of their missions, in the expected timeframe, was the definition of success for all actors from funders to administrators and scientists (Figure 1A). Every expense, every human resource decision, every regulation imposed was judged based on its relevance to the mission: will it bring us closer or further away from the goal? Thus, the administrators and funders of successful organizations facilitate innovation by dismantling barriers and producing environments conducive to creativity and productivity.

Figure 1. Contrasting relationship between elements and protocols in different human systems.

- The successful institutions or temporary endeavors such as the Manhattan or NASA’s Apollo projects are strictly mission-driven: all players (i.e., funders, administrators and scientists) are awarded based on achieving defined organizational goals. This leads to synergy between players, creating an environment that seeks and values creativity, diversity, cooperation, and shared responsibility.
- In most current, publicly funded biomedical research institutions, the mission exerts only a minor influence on daily decisions. Goals and award structures are largely separate for different actors. This creates inefficiencies, administrative burdens, and wastes resources.

For example, at the inception of the Apollo project, the paucity of skilled aeronautic engineers was solved by the administrators flying to Canada, recruiting Canadian engineers and facilitating their integration to NASA teams.¹³ There was a concerted effort to foster collaboration within the endeavor, challenging the current belief that internal competition generates efficient organizations.^{14–18} Analogously, funders of productive organizations either directly lead the effort or facilitate and monitor accomplishment of the mission, consistent with their (often financial) interest in the outcome.

In stark contrast, while the current publicly funded biomedical research institutions defined the broad aims (to improve health), there are few concrete steps or timelines, and these exert only minor influence on daily decisions by researchers, administrators, and funders. Although researchers need freedom to pursue novel hypotheses, as these often lead to important discoveries, the award structures must unite the various actors within the system behind the organizational mission. Disturbingly, this is currently not true, as the goals and awards are often completely disconnected from the mission (Figure 1B).

For example, the NIH intramural research program has close to 100 different electronic systems. Most fulfill necessary administrative or regulatory needs, but generally disregard the burdens their mandated use imposes on researchers. Because these systems largely do not communicate with each other, researchers must repeatedly input identical data in different formats and at different time-intervals mandated by each electronic system. Thus, medical providers and researchers with advanced degrees perform repetitive tasks that can be completed much faster and without errors by the computers. Even worse, the lack of harmonization creates an almost comical level of complexity, where the researchers must “remember” that the same event, for example a side effect of a medication, is considered “serious” by one regulatory system and must be imputed immediately, “expected” by another system and should be imputed as aggregate yearly data, and “serious but expected” by the third system. These arbitrary complexities cause mistakes, which must be found and reported, creating a vicious cycle of useless paperwork.

One reason for such inefficiencies is the fact that administrators are judged not on the output of the scientists in their division, much less the reproducibility of the work coming out of their labs, but rather on the administrator’s skill in accomplishing administrative tasks and, often, their ability to raise money for the institution. No administrator is punished for being unwilling to remove an impediment to efficient scientific production, such as the NIH’s byzantine electronic systems. On the contrary, the disincentives of potential failure and the effort needed for transformative leadership eventually stifles even the most mission-driven administrator.

Examining incentive structure for scientists

Similarly, the research teams, the “engines” discovering new knowledge, also operate within a reward structure which is not conducive to producing the most relevant and reproducible research possible.

Research teams typically consist of three to 30 people led by a single principal investigator (PI). Up to 80% of the team’s workforce are trainees. Their inevitable turnover impedes the continuity of long-term projects, such as clinical trials. The peak of a career trajectory for an academic researcher is to attain a (scarce) tenured PI position with an independent laboratory and budget. The currency of their realm is publications: how many articles are published; the prestige of the journals in which the articles are published; and the fame these articles generate.

Astonishingly, research reproducibility and its value to health are almost never considered among hiring, promotion and tenure criteria.¹⁹ There are few penalties for scientists who routinely publish irreproducible results except in cases of outright fraud or research misconduct. This gross, de-facto misalignment of the researcher’s incentives causes the irreproducibility crisis.

Incremental changes versus system re-design?

Numerous authors and organizations have suggested or implemented incremental enhancements of research enterprise. Partially successful fixes, such as pre-print services, and journals that embrace open reviews and/or implement publication checklists that provide transparency in disclosing methodological details, are steps in the right direction, but thus far have failed to make a significant dent in irreproducibility. These incremental solutions often cannot become permanent in the face of existing incentives, as was the case of the PubMed Commons post-publication open peer review,²⁰ which was discontinued because its use offered no advantage to scientists’ careers.

Change is difficult. Even people deeply dissatisfied with the status quo fear its disruptive consequences. Furthermore, redesigning the publicly funded biomedical research enterprise is a daunting task, as no single entity oversees research in its entirety. Nevertheless, NIH is the leading force in the US, and as demonstrated in the coronavirus disease 2019 (COVID-19) pandemic, is best positioned to establish fruitful collaboration among funding agencies, academic institutions, and private entities to lead transformative initiatives.

Reforming the scientific enterprise into a self-regulating system

A mission-driven approach to scientific reform includes the following desired outcomes:

1. Incentivize reproducible, impactful research that is efficiently performed and rapidly disseminated.
2. Turn scientists into collaborators by rewarding sharing of resources, dynamic data re-analyses, and transparent criticism and correction of the methods, results, and interpretations.
3. Re-envision scientific teams by utilizing individuals' strengths, by fairly and transparently distributing benefits from scientific accomplishments to team members, and by facilitating dynamic collaborations.

To attain these outcomes, the greatest opportunity is leveraging technological advances that can integrate the existing system of cataloguing scientific publications (e.g. PubMed) with objective, data-driven assessment of scientific papers' technical quality, generalizability, and resulting societal value. We call such a dynamic data integration the "biomedical research network" (BRN, [Figure 2](#)). This free, transparently curated resource would transform the way in which scientists (and their institutions) are judged and rewarded, by their funders and the public, thereby affecting all outcomes listed above.

BRN and its metrics

We envision three ways to measure contributions to biomedical research objectively: a methodological rigor (MR) score; a reproducibility (R) score; and a societal impact (SI) score. Every publication would be assessed by each of these metrics.

Figure 2. Simplified Interleukin 17 (IL17) branch of the BRN with examples of calculation of the Biomedical Research Network (BRN) metrics for three related studies. A. The BRN represents a massive multi-dimensional network of all biomedical science studies. Electronic navigation allows to zoom in a specific branch of scientific investigation, exemplified by IL17 subnetwork (B). The IL17 branch is initiated by the discovery/sequencing of the IL17 gene in 1993. There are many different new branches of investigation related to IL17, but the branch that links IL17 to autoimmunity is the most successful. Although IL17 has been associated with many different animal models of autoimmunity and with more human diseases, only three subdivisions are highlighted due to space constraints: IL17 and ankylosing spondylitis (AS), IL17 and psoriasis, and IL17 and multiple sclerosis (MS). Of these, the IL17 and AS, and IL17 and psoriasis branches are highly successful, with proven therapeutic efficacy of many direct and indirect IL17 inhibitors. Due to this commercial and public health impact, the publications forming branches highlighted in green color accumulate positive Societal Impact (SI) scores, whereas the IL17 and MS branch thus far had no commercial success; therefore, its publications do not partake in positive SI scores allocation. The size of the circles of the defining papers (network "hubs" exemplified by publication of the discovery of IL17, 1st publication that associated IL17 with autoimmunity, 1st publication that associated IL17 with human psoriasis etc.) is proportional to the assigned SI scores: discovery of IL17 has the highest Scientific Impact subscore; whereas positive clinical trials of IL17-blocking agents in AS and psoriasis have a higher Commercial impact subscore of SI scores. The Studies #1-3 are described in detail in the text of the article, where the studies #2-3 attempted to reproduce findings of the high-impact Study#1 but failed. The font color in the Integrated knowledge diagram is proportional to the MR scores of the three publications.

Methodological rigor (MR) score

The first metric, the MR score, is the most critical and the easiest to implement. It would be based on established attributes of reproducible research, such as blinding, power calculation, randomization, use of controls, validation of results in an independent cohort, availability of raw data and analysis codes.^{4,21,22} Though the quickest way forward is to assign an MR score to a paper based on a self-reported checklist by the author(s), this may generate bias and cannot be applied to already-published papers. Thus, at the BRN's inception, reviewers would need to validate the self-reported MR score against the Methodology section of the paper. The latter can be accomplished through an open review system, associated e.g., with free/public pre-print service such as [medRxiv](#). MedRxiv is run by a consortium of two academic institutions and the *BMJ*, and funded by a private foundation, the Chan Zuckerberg Initiative. MedRxiv has proven to be an invaluable aggregator of science associated with the COVID-19 pandemic. Reviewers' curated MR scores would serve as the dataset for the development, optimization, and validation of an automated, word-recognition based algorithm, which assigns MR score by scanning the Methods section of scientific papers.

Scoring of all publications can produce an MR score for individual scientists, averaged across their publications. This would incentivize all co-authors to support methodological rigor for every experiment, effectively self-regulating high-quality experimental design. The MR score would also help non-experts to judge the technical rigor of the presented work.

Scientists may have two objections: one, the retrospective grading may not accurately reflect the quality of the original experimental design, because in the past, journals restricted word count for the Methods section. This is remediated by standardizing an individual paper's MR score to the yearly averages, a solution used in standardized testing. Two, the authors may disagree with an assigned score. True mistakes, if validated in an open review, would improve the automated algorithm, possibly through public competition(s) similar to the [Netflix prize](#).

The evolution of MR scores on a population level would objectively measure the success of the BRN strategy to combat irreproducibility. Improvements in a scientist's standardized MR scores could represent professional growth. Additional training initiatives intended to improve methodological deficiencies and currently imposed on all scientists, could be more effectively targeted to the scientists with consistently low MR scores.

Reproducibility score (R) score

Unless we demand pre-registration of all experimental designs akin to pre-registration of clinical trials in [www.clinicaltrials.gov](#), the MR score will depend on what scientists choose to disclose. Some fear that pre-registration of all experiments will increase administrative burden²³ and stifle innovation. Others believe that a lack of preregistration substitutes scientific predictions by less credible "postdictions",²⁴ exacerbating irreproducibility. A solution that integrates both views is to objectively assess reproducibility of results, or in broad understanding, the "scientific truth"²⁵ of each study, post-publication.

Using language recognition, the BRN would assemble all related publications to a sub-network ([Figure 2B](#)), where the publication describing the crucial discovery that spurred a subsequent line of investigation would represent the main hub located at the center of the subnetwork. For example, the discovery of cytokine IL17 would sit at the hub of a subnetwork, and subsequent papers describing its role in autoimmune diseases would be one of the most successful "branches" of this IL17 subnetwork. Each new, related publication would be placed in the periphery of such a subnetwork, while minimizing the network "distance" between closely related research papers.

To illustrate how the BRN associates related articles and uses their MR score to assign a Reproducibility, or R score, consider three related articles that investigate the role of IL17F in predicting therapeutic response to the drug interferon beta (IFNb) for treating multiple sclerosis (MS). Study #1²⁶ was published in the high impact journal, *Nature Medicine*. By using only 26 MS patients and not employing methods to prevent bias, this study expands observations in mouse models to humans, concluding that high blood levels of IL17F predict poor response to IFNb. Two validation studies (Studies #2²⁷ and #3²⁸) that collectively (and blindly) analyzed samples from 357 randomly selected MS patients demonstrated unequivocally that serum levels of IL17F do not predict IFNb response in MS. In contrast to PubMed, which does not automatically associate related studies, the BRN algorithm places the validation studies in proximity of the Study #1 (i.e., on the same branch). Additionally, the BRN automatically integrates knowledge ([Figure 2B](#)) from related studies: because Studies #2 and #3 have high MR scores and validate each other, both receive high R scores. In contrast, Study #1 receives a low R score because it has a low MR score and was, in fact, not reproducible. To compute the R score, the algorithm compares the results of the two congruent studies "weighted" by their MR scores against the MR score of the conflicting study, to assign the probability that the specific result is true: the probability that serum IL17F

levels predict therapeutic response to IFN β based on this integrated knowledge is very low – let's say 0.1%; therefore, the Study#1 receives an R score of 0.001, whereas Studies#2 and #3 receive R scores of 0.999.

The goal is to employ technology to curate knowledge, while rewarding scientists for greater rigor. Because experimental rigor positively correlates with reproducibility,²¹ receiving consistently low R scores for consistently high MR scores will question accuracy of self-reported experimental designs. Such discrepancies would become exceedingly rare, as the mere understanding that objective curation will identify contradictions would naturally limit any temptation to cheat.

The most obvious objection against implementing the R score is its potential inaccuracy. Like every machine-learning algorithm, generated R scores will not be 100% accurate at its first iteration, but access to large, curated datasets (provided by author's contestation and human reviewers' adjudication of R scores), will allow data-driven optimizations. Ultimately, this will generate algorithm(s) that consistently outperform humans.²⁹

Societal impact (SI) score and its scientific, commercial, and public health contributions

Innovation and relevance of scientific discoveries to health are essential attributes of the NIH mission. Indeed, science can be well-designed, reproducible, but still irrelevant.¹¹ To assure that biomedical science is innovative and generates societal value, we must learn to quantify its worth objectively.

The impact of a discovery on science itself can be measured within the BRN as a growth of a new branch of scientific explorations. For example, basic scientific discoveries, such as the reprogramming of somatic cells to induced pluripotent stem cells (iPS;³⁰) or CRISPR-CAS9 genetic editing both established new branches of science.³¹ Thus, their scientific impact is enormous. To assess the public health and commercial impacts of scientific discoveries, the BRN must link to databases of patents, newly approved drugs, medical products, and new technologies. This will trace the origin of societal and public health advances back to individual scientists in order to develop the integrated SI scores. The goal of the SI score is to objectively measure the impact of individual discoveries and cumulative impact of individual scientists or even scientific institutions, which can be used to reward institutional leaders. Current attempts to assign value to a publication or a researcher, such as NIH's Relative Citation Ratio are insufficient, because they rely on the number of accrued citations. In our concrete example, the irreproducible publication in the prestigious journal is easily found in internet searches and continues to accrue citations years after its findings were unequivocally refuted: Study #1²⁶ accrued (as of May 2021) 590 citations, while the two congruent validation studies, published in less prestigious journals, accrued only 114 citations in total. This example is hardly an exception: a major effort to reproduce important scientific discoveries found non-reproduced articles published in higher impact journals and were accruing more citations than reproducible articles.³

Undoubtedly, the SI score is the most difficult metric to derive. The challenge is to assign the accurate proportion of any given publication's SI score to individual discoveries, as no discovery is made in isolation (exemplified in CRISPR technology, see the [Broad Institute's CRISPR Timeline](#)). In other words, as a pie grows (i.e., measured by the growth of subsequent publications and their societal impact), the number of scientists credited for the share of a pie also expands proportionally to the calculated impact of their discoveries. The SI score will likely receive the greatest push-back, as scientists have limited influence over commercialization of their discoveries and comparing societal impact of discoveries in different scientific fields may seem unsurmountable. But should we shrink from the enormity of this task and accept that the opinion of a few "experts" who nominate and select scientists for prestigious awards, is the best we can ever do?

Linking the BRN with the reform of the scientific publication system

The overarching goal of the BRN is to align research incentives to promote scientific breakthroughs with meaningful societal impacts. Current emphasis on publications, detached from their scientific rigor, reproducibility, and societal impact has been detrimental: thousands of scientists publish a paper every five days,³² most of which are unread and never cited (see³³ and the NIH's [iCite database](#)). No human being can keep up with this exponential growth of publications, even in his/her research field. Without transparent algorithms that curate this publication excess, work may be unnecessarily duplicative and findings with potential societal value may be difficult to find in search engines, slowing medical progress.

Additionally, publications must become the means, while transformative advances in basic sciences and medical cures are the ends. Therefore, although the BRN can be developed in parallel with the existing publication and dissemination system, reforming that system offers major advantages. First, the current unacceptably long and pointlessly expensive publication process stifles innovation. The public may be unaware that most scientific articles take years to publish, with research teams reformatting rejected papers to conform to requirements of different journals and paying thousands of

USD per publication, even if these are no longer printed and only published electronically. This causes frustration, and wastes resources that should be dedicated to new discoveries. Making the BRN a universal, immediate and free dissemination system for biomedical sciences (e.g., by merging PubMed with current pre-print services such as bioRxiv and medRxiv) would avoid wasted time, labor and money.

Open peer review

Second, a single-blinded review system is morally indefensible. It empowers reviewers to determine the fate of the publication without accepting public responsibility (or reward) for the review. This system institutionalizes bias by limiting scientific contributions from women and minorities, and has shown itself to be incapable of weeding out irreproducible science.³

Open peer review is often disparaged on the assumption that few scientists would choose to participate, fearing retaliation for a critical review. If this is true, the scientific culture is broken. Engaging in discourse on validity and interpretation of scientific discoveries is a precondition to scientific progress. Providing constructive criticism of presented work and identifying ways for its improvement, while respecting the dignity of the scientists who produced it, underlies scientific maturity. This is very different from what can occur now, the writing of anonymous reviews containing publicly indefensible arguments and refusing to take personal responsibility for their accuracy.

The validity of science is the responsibility of all scientists. The irreproducibility crisis reflects on all of us, whether we adhere to sound or poor experimental design. We are the ones who form and embrace scientific culture and the only ones who can reform it.

Reviewer impact (RI) score

If we want to achieve sustainable open peer review,³⁴ we must reward scientists for their willingness to provide critical, creative thinking. Let's return to the example described in the legend of Figure 2 and consider three hypothetical scenarios: in scenario #1, Study #1 is submitted to the BRN and a reviewer alerts the authors to the methodological flaw through open review. The authors acknowledge their mistake and amend or retract the problematic experiment in their paper. Because there were no negative societal consequences, the authors' MR score would go down slightly for submitting a flawed experiment, and the reviewer would accrue positive RI score for identifying the paper's shortcoming and triggering a correction.

In scenario #2, the authors submit a flawed experiment but choose to ignore the raised criticism. With the critical review in the public domain, an economical, well-designed (i.e., of high MR score) independent validation study is performed and fails to reproduce original finding. Now, the authors accrue not only a low MR score but also a low R score and negative SI score, equivalent to the value of wasted resources.

Finally, in scenario #3, which in many ways reflects the status quo, all reviewers fear the consequences of posting a negative review of the experiment coming from an influential scientist. The reported findings appear to be so important that a pharma company decides to perform a new clinical trial based on the published data. After considerable expenditures, the trial unequivocally demonstrates the irreproducibility of the original publication. The authors would accumulate not only low MR and R scores, but also highly negative SI score due to accumulated costs of wasted resources. Everybody loses. Through transparent, responsible review we can correct mistakes while they are still small, help each other grow as scientists and maximize the value of our discoveries.

Scientists should freely publish their experiments, and any reader should be able to transparently (and succinctly) critique them, accepting full responsibility for the review. Reviewers with habitually frivolous criticism not supported by data or scientific understanding would themselves accrue low cumulative RI scores, prompting them to change this behavior. By contrast, critiques from scientists with high cumulative RI scores would impel the authors to carefully consider the reviewers' arguments and perform additional experiments to ensure the validity of findings.

The BRN would automatically limit irreproducible science if career rewards depended on high MR, R, and SI metrics. If performing bad science were detrimental to one's career, why would anybody conduct much less try to publish flawed experiments? And, if scientists read papers anyway (which they must do) and rating them were a simple task that promoted their careers, why would they not review?

Finally, an open review process would level the playing field of scientific education by allowing all trainees and even journalists to learn critical thinking from the greatest minds. Everybody wins.

Dynamic publications

We also wholeheartedly embrace the concept of dynamic publications.³⁵ Publications should be transparently modified (with the modification history recorded in the BRN) based on new experiments or evolved knowledge. Such modifications may enhance clarity, adjust conclusions, and add, improve, or retract experiments in response to open review. Furthermore, scientists who were not original authors should be able to attach results of subsequent experiments to an existing publication, rather than publishing independent papers. For example, an investigator publishes an intriguing finding that lacks an independent validation cohort. Another scientist has such a cohort and validates the published findings. Rather than submitting this as a new publication (and thus needlessly duplicating information already contained in the introduction and discussion of the original publication), this independent lab could simply submit the results of the validation experiment and its related methods to the BRN, and link it to the previous paper. If the results subsequently translate into societal value, all contributing scientists will share the benefit proportional to their contribution. This way, many scientists can efficiently strengthen or refute existing publications, solving the current problem of (not) publishing negative data³⁶ and replication studies, which are essential for assessing the generalizability of discoveries.^{37,38}

Sharing raw data and dynamic re-analyses and re-use of published cohorts

Related to the publication process is the issue of depositing full datasets into the public domain, not only to verify published results but also for the datasets to be reused to test new hypotheses. The societal benefit of sharing raw data is widely recognized^{39,40} but practically left at the discretion of the scientists. For example, the NIH encourages data sharing, and many journals request that the authors include a statement on data availability, such as: “Anonymized data not published within this article will be made available by request from any qualified investigator.” Such requests are not always granted and the declarations in the publication cannot be enforced.⁴¹

The reason for the lack of data sharing, especially from well-designed clinical studies, is that these datasets remain valuable; they allow scientists to ask new questions and publish new papers—the same value society loses if the data are not shared. Because different researchers ask different questions, some essential analyses may never be performed. Acknowledging truth on both sides, the solution is to incentivize data-sharing by assigning SI score to the investigators who generated valuable datasets whenever their reuse led to new societal benefits. If one team published raw data after spending 10 years generating an extremely well-characterized longitudinal cohort, and a second team re-analyzes these data and gains valuable insights, then both teams should share the resulting SI score for such invention, proportional to their efforts and intellectual contributions. Everybody wins, especially patients.

Re-envisioning the role of publication industry experts in the research enterprise

Would the BRN destroy the current publication industry? Not necessarily. First, the BRN may be based on public-private partnership and employ current publication industry experts. Second, the BRN would provide new job opportunities, such as science writers analyzing aggregate knowledge to identify new trends and knowledge gaps, or to promote commercialization of discoveries by guiding investment funds. These would allow the current scientific publication workforce to directly advance science and health.

Re-envisioning scientific teams by mapping individuals' strengths, by fairly and transparently distributing benefits from scientific accomplishments and by facilitating dynamic collaborations

Successful research requires a diverse workforce with complementary skills, as again highlighted by the Manhattan and Apollo projects. When people do their job expertly, they become invaluable.

Current co-authorship criteria on scientific publications favor intellectual over manual or technical contributions. But what prevents us from fairly crediting the work of all team-members? The current acknowledgement style is needlessly vague: e.g., writing that persons #1-3 contributed to data analysis does not give justice to the reality if person #1 performed 90%, person #2, 9% and person #3, 1% of the work. The BRN could measure scientific contributions comprehensively, providing person-specific aggregate data, essential for junior scientists and support staff. This level of transparency and granularity, if linked to reward structures (recruitment, promotion, salary) would inspire the scientific workforce to assume more responsibilities, enhance their skillset and increase productivity. This productivity and cumulative societal value of institutional discoveries measured by the BRN metrics could re-align incentive structures of administrators and empower funders to make more impactful funding decisions.

Conclusion

Biomedical data has become unmanageably vast, and much of it is not contributing to new knowledge or better health, which are the stated goals of the NIH and undoubtedly of many if not most biomedical researchers. We need advanced computer algorithms to help us make sense of all that data, as well as a new set of incentives to encourage all researchers to produce replicable results. Everything described in this article is technologically feasible. The COVID-19 pandemic

demonstrated that medical innovation can be both fast and transformative, as long as leaders are willing to dismantle administrative barriers and foster an environment conducive to creativity and collaboration. Courageous leadership in science administration can propel us to a new era of accelerated biomedical advances, that will achieve the mission of the NIH: to improve health for everyone.

Data availability

No data are associated with this article.

Acknowledgments

The authors express gratitude to Drs. Gary Birnbaum and Alison Wichman for their thoughtful suggestions and Brian Brown, NIH Library Editing Service, for reviewing the manuscript. The opinions expressed in this paper are those of the authors only and do not represent the official stance of their institutions.

References

- Ioannidis JP: **An epidemic of false claims. Competition and conflicts of interest distort too many medical findings.** *Sci Am.* 2011; **304**(6): 16.
[PubMed Abstract](#)
- Ioannidis JP: **Why most published research findings are false.** *PLoS Med.* 2005; **2**(8): e124.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Begley CG, Ellis LM: **Drug development: Raise standards for preclinical cancer research.** *Nature.* 2012; **483**(7391): 531–3.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Begley CG, Ioannidis JP: **Reproducibility in science: improving the standard for basic and preclinical research.** *Circ Res.* 2015; **116**(1): 116–26.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Prinz F, Schlange T, Asadullah K: **Believe it or not: how much can we rely on published data on potential drug targets?** *Nat Rev Drug Discov.* 2011; **10**(9): 712.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Freedman LP, Cockburn IM, Simcoe TS: **The Economics of Reproducibility in Preclinical Research.** *PLoS Biol.* 2015; **13**(6): e1002165.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Gall T, Ioannidis JPA, Maniatis Z: **The credibility crisis in research: Can economics tools help?** *PLoS Biol.* 2017; **15**(4): e2001846.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Baker M: **1,500 scientists lift the lid on reproducibility.** *Nature.* 2016; **533**(7604): 452–4.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Landis SC, et al.: **A call for transparent reporting to optimize the predictive value of preclinical research.** *Nature.* 2012; **490**(7419): 187–91.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Collins FS, Tabak LA: **Policy: NIH plans to enhance reproducibility.** *Nature.* 2014; **505**(7485): 612–3.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Ioannidis JP: **Why Most Clinical Research Is Not Useful.** *PLoS Med.* 2016; **13**(6): e1002049.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Feynman RP: *Surely you're joking, Mr. Feynman!* New York, NY, USA: W.W. Norton & Company; 1977.
- Murray C, Bly Cox C: *Apollo.* Burkittsville, MD, USA: Cox and Murray, Inc.; 2004.
- Logan D, King J, Fischer-Wright H: *Tribal leadership: Leveraging natural groups to build a thriving organization.* Harper Business; 2011.
- Dalio R: *Principles.* New York, NY, USA: Simon & Schuster; 2017.
- Hsieh T: *Delivering Happiness: A path to profits, passion and purpose.* New York, NY, USA: Hachette Book Group; 2010.
- Partridge D: *People over profit.* Nashville, TN, USA: Nelson Books; 2015.
- Catmull E: *Creativity, Inc.* New York, NY, USA: Random House; 2014.
- Moher D, et al.: **Assessing scientists for hiring, promotion, and tenure.** *PLoS Biol.* 2018; **16**(3): e2004089.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Meadows DH: *Thinking in Systems: a primer.* Writng D. White River Junction, VT: Sustainability Institute; 2008; 215.
- Hackam DG, Redelmeier DA: **Translation of research evidence from animals to humans.** *JAMA.* 2006; **296**(14): 1731–2.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Ioannidis JP: **How to make more published research true.** *PLoS Med.* 2014; **11**(10): e1001747.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Hunter P: **The reproducibility “crisis”: Reaction to replication crisis should not stifle innovation.** *EMBO Rep.* 2017; **18**(9): 1493–6.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Nosek BA, et al.: **The preregistration revolution.** *Proc Natl Acad Sci U S A.* 2018; **115**(11): 2600–6.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Goodman SN, Fanelli D, Ioannidis JP: **What does research reproducibility mean?** *Sci Transl Med.* 2016; **8**(341): 341ps12.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Axtell RC, et al.: **T helper type 1 and 17 cells determine efficacy of interferon-beta in multiple sclerosis and experimental encephalomyelitis.** *Nat Med.* 2010; **16**(4): 406–12.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Hartung HP, et al.: **Interleukin 17F level and interferon beta response in patients with multiple sclerosis.** *JAMA Neurol.* 2013; **70**(8): 1017–21.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Bushnell SE, et al.: **Serum IL-17F does not predict poor response to IM IFNbeta-1a in relapsing-remitting MS.** *Neurology.* 2012; **79**(6): 531–7.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Silver D, et al.: **A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play.** *Science.* 2018; **362**(6419): 1140–4.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Takahashi K, et al.: **Induction of pluripotent stem cells from adult human fibroblasts by defined factors.** *Cell.* 2007; **131**(5): 861–72.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Scudellari M: **How iPS cells changed the world.** *Nature.* 2016; **534**(7607): 310–2.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Ioannidis JPA, Klavans R, Boyack KW: **Thousands of scientists publish a paper every five days.** *Nature.* 2018; **561**(7722): 167–9.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Franca TFA, Monserrat JM: **To Read More Papers, or to Read Papers Better? A Crucial Point for the Reproducibility Crisis.** *Bioessays.* 2019; **41**(1): e1800206.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Flier JS: **Irreproducibility of published bioscience research: Diagnosis, pathogenesis and therapy.** *Mol Metab.* 2017; **6**(1): 2–9.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Ioannidis JP, Tatsioni A, Karassa FB: **Published articles should not be dead and buried: introducing research updates.** *Eur J Clin Invest.* 2010; **40**(9): 767–9.
[PubMed Abstract](#) | [Publisher Full Text](#)

36. Ioannidis JP: **Journals should publish all “null” results and should sparingly publish “positive” results.** *Cancer Epidemiol Biomarkers Prev.* 2006; **15**(1): 186.
[PubMed Abstract](#) | [Publisher Full Text](#)
37. Ioannidis JPA: **Why replication has more scientific value than original discovery.** *Behav Brain Sci.* 2018; **41**: e137.
[PubMed Abstract](#) | [Publisher Full Text](#)
38. Ioannidis JP: **Translational research may be most successful when it fails.** *Hastings Cent Rep.* 2015; **45**(2): 39–40.
[PubMed Abstract](#) | [Publisher Full Text](#)
39. Taichman DB, *et al.*: **Sharing Clinical Trial Data—A Proposal from the International Committee of Medical Journal Editors.** *N Engl J Med.* 2016; **374**(4): 384–6.
[PubMed Abstract](#) | [Publisher Full Text](#)
40. McCarthy M: **Sharing clinical trial data should become the “expected norm,” US panel says.** *BMJ.* 2015; **350**: h240.
[PubMed Abstract](#) | [Publisher Full Text](#)
41. Miyakawa T: **No raw data, no science: another possible source of the reproducibility crisis.** *Mol Brain.* 2020; **13**(1): 24.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Open Peer Review

Current Peer Review Status: ? ✖

Version 1

Reviewer Report 15 December 2021

<https://doi.org/10.5256/f1000research.57067.r101577>

© 2021 Hufton A. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Andrew Lee Hufton

Leibniz-Institut DSMZ, Braunschweig, Germany

Thank you for inviting me to review the Opinion Article by Drs. Bielekova and Brownlee, entitled "The imperative to find the courage to redesign the biomedical research enterprise".

In this piece, the authors argue that the incentive structure of modern biomedical research does not align with its supposed mission, and propose a series of changes that could reform this incentive structure in a way that they believe will "maximize our investments in biomedical research and optimize its value to human health".

The authors build their argument on the now quite mature literature documenting the so-called "reproducibility crisis". I tend to agree with the authors that this is a serious and urgent issue, and indeed many of the reforms proposed by the authors strike me as sensible. Despite this, however, I find a number of elements of the authors' argument hard to follow or poorly supported.

The authors begin by using the NIH as their dominant institutional example, and compare it to the Apollo Mission and the Manhattan Project, which they present as two enterprises with clearer, mission-driven management. I find this comparison, at minimum, unhelpful. First, let's be clear that the mission of one of these projects was the production of an instrument of mass death, which makes for a poor comparison to the NIH. That aside, both were much more defined projects with clear end-points, making any comparison to the NIH hard for me to really engage with. I think, in any analysis, the NIH and the research it has funded would be found to have had a far more positive impact on humanity than either of these projects.

I would prefer to see a more informed comparison between the NIH and other organizations of similar scale or purpose. Can the authors support their argument that the NIH's system of organization is non-optimal by providing more relevant examples of organizations that have been able to more efficiently advance human health research?

The authors then propose a framework of radical changes to the way that scientific publications are published, assessed, and to how they are used to evaluate research programs and

researchers. Before I address the authors' specific proposal, I would like to note that at this point I have lost the connection to the authors' original argument that the NIH's organizational system should be more mission-driven. The changes being proposed here reach far beyond the NIH's scope and its leverage to implement.

Central to the authors' proposal is a network based, multi-criteria method for assessing the value of scientific publications, which they call the BRN for "biomedical research network". I like some of the ideas embedded in this concept, but as a whole I remain far from convinced that this is technically feasible or that the ratings of a such system would better align with the NIH's mission. I appreciate that this is an opinion paper, but the BRN proposal being put forward is relatively specific, and I think should be supported with some evidence of feasibility for a scholarly work of this kind.

I question whether reviewers could reliably quantify "methodological rigor", in a manner that would be meaningful across topics and disciplines. I am also not convinced that an automated system could identify related publications with sufficient semantic rigor to allow any calculations of reproducibility, in the manner proposed by the authors. Similarly, how would one rate social or commercial impact objectively in the short term, when these kinds of impacts often take years or even decades to mature?

These concerns are not trivial in my mind, since we have already seen the harm that poorly designed metrics have done to the scientific enterprise. I share the authors' concern about the reproducibility crisis, but I think there are very legitimate concerns in the scientific community that a rush to create new reproducibility metrics, could end up causing further harm, for example, by de-incentivizing cutting edge or interdisciplinary research where standards are still evolving and sparser research networks means that reproducibility is harder to assess.

In the rest of the paper, the authors make further recommendations, endorsing rating reviewers, open peer-review, dynamic publications, and data sharing, among other ideas. These are sensible recommendations, although I feel the authors do not manage to relate all of them directly back to their idea that biomedical science organizations should be more mission-driven.

Overall, I am sympathetic to many of the authors' ideas and the overall sense of urgency that pervades the piece, but I am not convinced at this stage that the authors have made a coherent argument for how biomedical research could be reformed.

My advice, were the authors to consider developing this further, would be to think more critically about how the NIH or a similar organization could specifically enact change, using the tools and leverage available to it, and expounding more on how such an organization could concretely use mission-driven assessments to improve its outputs. As a reader, I personally would like to see this supported with more relevant organizational management comparisons, and some consideration of the costs of the proposed reforms relative to current organizational systems.

The BRN itself, while interesting, feels too underdeveloped at this stage. For such a proposal, I would like to see at least some evidence of feasibility, including numerical rankings and networking-building with actual papers.

Finally, I would like to note that this paper at present is generally well written, after setting aside

my more conceptual criticisms above.

Is the topic of the opinion article discussed accurately in the context of the current literature?

Partly

Are all factual statements correct and adequately supported by citations?

Partly

Are arguments sufficiently supported by evidence from the published literature?

No

Are the conclusions drawn balanced and justified on the basis of the presented arguments?

No

Competing Interests: I worked for Springer Nature, a major scientific publisher, until March 2021.

Reviewer Expertise: Genetics & computational biology, scientific publishing, data sharing, open science

I confirm that I have read this submission and believe that I have an appropriate level of expertise to state that I do not consider it to be of an acceptable scientific standard, for reasons outlined above.

Reviewer Report 27 August 2021

<https://doi.org/10.5256/f1000research.57067.r90258>

© 2021 Glasziou P. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Paul P. Glasziou

Centre for Research in Evidence-Based Practice, Bond University, Gold Coast, Qld, Australia

This is an interesting and valuable article which looks at the current problems in the health and medical research system such as (lack of) reproducibility, accessibility, and relevance. Importantly, the paper considers the research enterprise as a system - of researchers, funders, publishers etc., with varying incentives and motivations - and examines potential solutions at this systems level. This is neatly illustrated by Figure 1 showing the poor alignment of the various elements and actors within the research system. There are the seeds of many good ideas within this paper. However considerable meta-research would be necessary before implementing the metrics and solutions proposed - for the detailed development and testing.

I had a few suggestions to consider for changes:

1. Alongside the Manhattan project an Apollo mission examples, given the current pandemic it

would be useful to mention the CEPI vaccine collaboration - Coalition for Epidemic Preparedness Innovations (CEPI) which was launched at Davos 2017 and without which we would not have the vaccines we now have - [CEPI](#)

2. The proposed solutions rightly focus on aligning incentives with the need for impactful research requiring collaborations, and well-functioning teams. In "COM-B" terms (Capability + Opportunity + Motivation = Behaviour) the authors' focus is particularly on Motivation. However, it would be helpful to consider some of the Capability and Opportunity elements, such as skills, training, infrastructure, tools, access, etc.
3. The authors also proposes three new measures: methodological rigour score, a reproducibility score, and a societal impact score - again aimed at incentives/motivation. This is worthwhile, but a complex task and some suggestions on who and how the needed meta-research and experimental studies could be done would be helpful.
4. There are a number of important initiatives already underway which are aligned with the ideas presented here and it would be useful to set those out in a table, with examples such as:
  - [DORA](#): The Declaration on Research Assessment (DORA) which set out principles addressing the need to improve the ways in which researchers and the outputs of scholarly research are evaluated.
  - [The Hong Kong principles](#): which provide a more detailed set of criteria for assessing researchers for promotion and tenure.
  - [The EQUATOR Network centres](#) - for Enhancing the QUALity and Transparency Of health Research which develop and disseminate the reporting guidelines (CONSORT, PRISMA, etc).
  - [EVIR forum](#) - a forum of over 40 research funders founded in 2017 to help health-related research funders increase the value of their research.
  - [The evidence based research network](#) - which aims to reduce waste in research by promoting: (a) No new studies without prior systematic review of existing evidence, (b) Efficient production, updating and dissemination of systematic reviews.

Is the topic of the opinion article discussed accurately in the context of the current literature?

Yes

Are all factual statements correct and adequately supported by citations?

Yes

Are arguments sufficiently supported by evidence from the published literature?

Yes

Are the conclusions drawn balanced and justified on the basis of the presented arguments?

Yes

Competing Interests: I am a co-director of one of 5 EQUATOR Centres working on reporting guidelines.

Reviewer Expertise: Clinical Epidemiology

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Comments on this article

Version 1

Reader Comment 10 Feb 2022

Anita Bandrowski, UCSD, USA

I have read your paper with interest. However, you did not seem to go past the need for a tool into looking at what tools are already available and how they are being used.

This statement seems well founded: "The first metric, the MR score, is the most critical and the easiest to implement. It would be based on established attributes of reproducible research, such as blinding, power calculation, randomization, use of controls, validation of results in an independent cohort, availability of raw data and analysis codes.^{4,21,22}"

However, while your paper discusses the need for an AI check, it seems that you are missing several things. First, is the fact that the AI community already has tools implemented for "blinding, power calculation, randomization, use of controls, availability of raw data and analysis codes" and have assessed the totality of the open access literature based on all of these metrics, see [Menke et al., 2020](#), [Serghiou et al., 2021](#). These and other tools that assess the quality of different aspects of reproducibility are currently working together to assess COVID preprints on BioRxiv and MedRxiv, see [Weissgerber et al., 2021](#).

Second, the nuance that is needed to implement this kind of a check is completely missed by this paper. Different fields, outside of medicine, have different standards for rigor and all of those are missed here. Even within medicine, a rigorous case study will need to be evaluated a little differently from a clinical trial report. The AI community has come up with tools that can detect features of manuscripts that may or may not be associated with reproducibility, however coming up with any tool or score that can detect reproducibility of the paper is probably out of reach of current AI. Perhaps [DARPA's SCORE project](#) will change my mind, but as a reproducibility researcher I currently have my doubts.

This article would be a lot more compelling if it could be grounded in what is currently available and possible with AI.

Competing Interests: No competing interests were disclosed.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research