laura and john arnold foundation

Comments in Response to *Request for Information re: Strategy for American Innovation*

by

Stuart Buck
Vice President of Research Integrity, Laura and John Arnold Foundation
September 23, 2014

I. Introduction

The Laura and John Arnold Foundation (LJAF) submits these comments in response to the Request for Information by the White House Office of Science and Technology Policy and the National Economic Council. In particular, we intend to address the following question: "Given recent evidence of the irreproducibility of a surprising number of published scientific findings, how can the Federal Government leverage its role as a significant funder of scientific research to most effectively address the problem?"

Our foundation is a leading advocate for open science and improved research standards across various disciplines. Among other things, LJAF is the primary funder of the Meta-Research Innovation Center at Stanford University, an organization focused on improving the quality of medical research; and the Center for Open Science, an organization that works to promote transparency in scientific research and that has sponsored replication projects in both psychology and cancer cell biology.

By now, everyone has heard that many highly cited lab experiments on drug targets cannot be reproduced, even with the cooperation of the original investigators.[1] But this is just the tip of the iceberg. Publication bias and questionable research practices have been identified in virtually every field imaginable, such as clinical trials in medicine,[2] antibiotic resistance,[3] high-throughput bioinformatics,[4] neuroimaging,[5] ecology and evolution,[6] cognitive

[1] F. Prinz, T. Schlange, & K. Asadullah, "Believe it or not: how much can we rely on published data on potential drug targets?," *Nature Reviews Drug Discovery* 10 (Sept. 2011): 712.

[2] S. Hopewell et al., "Publication bias in clinical trials due to statistical significance or direction of trial results," Cochrane Database of Systematic Reviews 1 (2009).

[3] J. Caryl, "My published negative result," *The Gene Gym* (11 Oct. 2012). Available at http://www.scilogs.com/the_gene_gym/my-published-negative-result/.

[4] K. A. Baggerly & K. R. Coombes, "Deriving Chemosensitivity from Cell Lines: Forensic Bioinformatics and Reproducible Research in High-Throughput Biology," *Annals of Applied Statistics* 3 (2009): 1309-1334.

science,[7] public health and epidemiological research,[8] genetic research on fruit flies,[9] animal research on potential stroke treatments,[10] parenting programs,[11] economics,[12] political science,[13] psychology,[14] psychiatry and psychotherapy,[15]

---

[5] "Repeat after me: Replication in clinical neuroimaging is critical," available at http://www.sciencedirect.com/science/article/pii/S2213158213000090; J. Carp, "Better living through transparency: Improving the reproducibility of fMRI results through comprehensive methods reporting," *Cognitive, Affective & Behavioral Neuroscience* 13 (Sept. 2013): 660-66; R. G. Jennings & J. D. Van Horn, "Publication bias in neuroimaging research: implications for metaanalyses," *Neuroinformatics* 10 (2012): 67–80.

[6] M.D. Jennions & A. P. Møller, "Publication bias in ecology and evolution: an empirical assessment using the 'trim and fill' method," *Biological Reviews of the Cambridge Philosophical Society* 77 (2002): 211–222.

[7] J. Ioannidis et al., "Publication and other reporting biases in cognitive sciences: detection, prevalence, and prevention," *Trends in Cognitive Sciences* 18 (2014): 235–241.

[8] S. Young & A. Karr, "Deming, data and observational studies: A process out of control and needing fixing," *Significance* (2011): 116-120; T. Churches, "The benefits of reproducible research: a public health example," available at https://github.com/timchurches/meta-analyses/tree/master/benefits-of-reproducible-research.

[9] D. L. Stern, "Reported *Drosophila* courtship song rhythms are artifacts of data analysis," *BMC Biology* 12 (2014): 38.

[10] J. A. Hirst et al., "The Need for Randomization in Animal Trials: An Overview of Systematic Reviews," *PLoS ONE* (June 6, 2014), DOI: 10.1371/journal.pone.0098856; E.S. Sena et al., "Factors affecting the apparent efficacy and safety of tissue plasminogen activator in thrombotic occlusion models of stroke: systematic review and meta-analysis," Journal of Cerebral Blood Flow and Metabolism 30 (2010): 1905–1913.

[11] P. Wilson et al., "How evidence-based is an 'evidence-based parenting program'? A PRISMA systematic review and meta-analysis of Triple P," *BMC Medicine* 10 (2012): 130.

[12] D. S. Hamermesh, "Replication in Economics," *Canadian Journal of Economics* 40 (2006): 715-33; J. Ioannidis & C, Doucouliagos, "What's to know about the credibility of empirical economics?," *Journal of Economic Surveys* 27 (2013): 997–1004; S. Necker, "Scientific misbehavior in economics," *Research Policy* (18 June 2014), doi: 10.1016/j.respol.2014.05.002.

[13] J. Esarey & A. Wu, "The Fault in our Stars: Measuring and Mitigating 'Significance Bias' in Published Work," working paper (11 Nov. 2013), available at http://jee3.web.rice.edu/significance-bias.pdf/.

[14] C. J. Ferguson & M. Heene, "A vast graveyard of undead theories publication bias and psychological science's aversion to the null," *Perspectives on Psychological Science* 7 (2012): 555–561; J. P. Simmons, L. D. Nelson, & U. Simonsohn, "False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant," *Psychological Science* 22 (2011): 1359–1366.

[15] E. Turner, "Publication Bias, with a Focus on Psychiatry: Causes and Solutions," *CNS Drugs* 27 (2013): 457-68; J. Coyne, "Salvaging psychotherapy research: a manifesto," (10 June 2014),

education,[16] sociology,[17] computer science,[18] nanochemistry,[19] computational astronomy,[20] and physics.[21]

We therefore applaud the White House's interest in scientific reproducibility. **This issue should be a major agenda and budgetary item for every federal agency that funds research, without exception.**

II. Possible Solutions

While space does not permit a thorough discussion of every possible remedy for every scientific field,[22] a few broad ideas would improve almost all of science.

A. Mandate Greater Use of Pre-Registration

At the outset of any experiment or observational study that aims to confirm any particular hypothesis, scholars should specify and publicly pre-register as much of their research design as is possible.

Pre-registration is the standard practice in clinical trials, particularly since the Food and Drug Administration (FDA) Amendments Act of 2007, and is

---

available at http://blogs.plos.org/mindthebrain/2014/06/10/salvaging-psychotherapy-research-manifesto/.

[16] T. D. Pigott et al., "Outcome-Reporting Bias in Education Research," *Educational Researcher* 42 (3 Oct. 2013): 424-32.

[17] A. S. Gerber & N. Malhotra, "Publication bias in empirical sociological research: Do arbitrary significance levels distort published results?," *Sociological Methods & Research* 37 (2008): 3–30.

[18] "Examining 'Reproducibility in Computer Science,'" http://cs.brown.edu/~sk/Memos/Examining-Reproducibility/.

[19] G. Ozin, "Nanochemistry Reproducibility" (19 Sept. 2013), http://www.materialsviews.com/nanochemistry-reproducibility/.

[20] L. Shamir et al., "Practices in source code sharing in astrophysics," *Astronomy and Computing* 1 (Feb. 2013): 54-58.

[21] E. Gross & O. Vitells, "Trial factors for the look elsewhere effect in high energy physics,"*The European Physical Journal C - Particles and Fields* 70 (2010): 525–530; J. R. Klein & A. Roodman, "Blind analysis in nuclear and particle physics," *Annual Reviews of Nuclear and Particle Science* 55 (2005): 141–163.

[22] Many other ideas should be considered, such as uniquely identifying and validating reagents used in biomedical experiments. See N. A. Vasilevsky, "On the reproducibility of science: unique identification of research resources in the biomedical literature," *PeerJ* 1 (2013): e148, http://dx.doi.org/10.7717/peerj.148.

becoming more common in social science as well.[23] The rationale both in clinical trials and elsewhere is that studies are much more informative and reliable if the following are specified ahead of time: the main and supporting hypotheses, primary outcomes, statistical methods, sample size, subgroup analyses (if any), exclusion criteria, statistical power analysis, and more.[24]

Without such pre-specification, researchers have free rein to engage in ad hoc data mining, which is rarely reliable or reproducible because its results are more likely to be spurious positives. Federally mandated pre-registration could increase the reliability of many research fields.

B. Mandate improved experimental design and reporting standards.

Many problems of reproducibility arise from poor experimental design and poor reporting standards. Each federal agency that sponsors research should make improvements along the following lines.

First, the actual design of an experiment can be the difference between science that is reproducible or not. For example, one set of scholars argues that most pre-clinical animal research is so bad that it should "not be allowed to constitute part of the rationale for human trials," as it lacks such crucial elements as "randomization, allocation concealment, and blind outcome assessment."[25] In that field, then, federal agencies should mandate that researchers randomize animals to the different treatments or conditions that they are studying, blind the analysts to which animals were in each treatment group, and calculate sample size in advance so as to increase statistical power to a minimum of 80 percent. Beyond animal research, federal agencies should convene expert statisticians and subject matter experts to set minimum standards on how experiments must be designed whenever federal funding is used.

---

[23] Examples can be seen on https://www.socialscienceregistry.org/ or https://osf.io/hxeza/.

[24] E. Miguel et al., "Promoting Transparency in Social Science Research," *Science* 343 (3 Jan. 2014): 30-31. Available at http://www.sciencemag.org/content/343/6166/30.full.pdf?keytype=ref&siteid=sci&ijkey=TMhxM94eiQUc2; M. Humphreys, R. S. de la Sierra, & P. Windt, "Fishing," working paper (8 May 2012), available at http://www.columbia.edu/~mh2245/papers1/PA_2012b.pdf.

[25] J. A. Hirst et al., "The Need for Randomization in Animal Trials: An Overview of Systematic Reviews," *PLoS ONE* (6 June 2014), doi: 10.1371/journal.pone.0098856. See also Malcolm R. Macleod, "Preclinical research: Design animal studies better," *Nature* 510 (5 June 2014): 35.

Second, reporting standards can be improved dramatically. The CONSORT guidelines for clinical trial reporting[26] and the MIAME guidelines for microarray experiments[27] are widely applauded for standardizing the information that is collected and reported in those research areas, making it much easier for other scholars to understand the experimental procedure and to do meta-analyses on the results. Federal agencies could adopt similar standards for other fields, e.g., the ARRIVE guidelines for pre-clinical animal research.[28]

Finally, as Dartmouth's Brendan Nyhan has recently argued,[29] federal agencies should reward (through proposal scoring or supplemental grant awards) scholars who publish via the new "Registered Report" format. In this publication format, a scholar submits the experimental design to a journal prior to running the experiment and collecting results. Peer review occurs on the basis of how rigorous the experimental design is, rather than on whether the experiment shows "positive" results. As Nyhan says, "This procedure encourages authors and reviewers to develop the strongest possible designs — including those that replicate previously published studies — and eliminates perverse incentives to find or emphasize significant results after the fact." This procedure is already in use at several psychology and neuroscience journals, and should be extended much more broadly via federal incentives.[30]

C. *Mandate the Sharing of Data and Code*

A central principle of scientific advancement is the sharing of data so that other scientists can more easily build upon previous work. As *Science* has said, "Making data widely available is an essential element of scientific research."[31] Sharing data has led to many scientific advances, particularly in genetics. Since 2007 alone, the National Institutes of Health (NIH) database of Genotypes and Phenotypes has allowed "2,221 investigators access to 304 studies, resulting in 924

---

[26] http://www.consort-statement.org/.

[27] http://www.ncbi.nlm.nih.gov/geo/info/MIAME.html.

[28] https://www.nc3rs.org.uk/arrive-guidelines.

[29] B. Nyhan, "To Get more Out of Science, Show the Rejected Research," *New York Times' The Upshot* (18 Sept. 2014), available at http://www.nytimes.com/2014/09/19/upshot/to-get-more-out-of-science-show-the-rejected-research.html?src=twr&abt=0002&abg=1.

[30] See https://osf.io/8mpji/wiki/faqs for a much more detailed explanation of how the new publication format works.

[31] B. Hanson, A. Sugden, & B. Alberts, "Making Data Maximally Available," *Science* 331 (11 Feb. 2011): 649.

publications and significant scientific advances."[32] For example, a recent reanalysis of data made significant advances in our understanding of which genetic loci are associated with esophageal cancer.[33]

By contrast, failure to share data can halt scientific progress. Recently, for example, researchers tried to do a meta-analysis of techniques for treating newborn infants who have trouble regulating their breathing, reflexes, etc., in the hopes of developing a prognostic tool for doctors to use. They found the meta-analysis impossible to carry out: over 60 percent of the data was unavailable because researchers either ignored the request or outright refused to share.[34]

Unfortunately, current data-sharing policy for federally funded research is far too narrow and easy to circumvent. The NIH, for example, currently requires that grants funded at $500,000 or more in direct costs per year must have a data management plan. But many grants whose data would be extremely valuable to the progress of science are funded at less than $500,000 per year. Worse, because data-sharing is not mandatory by default, investigators often get away with a data management plan that declines to share data or that shares only under such narrow circumstances that it is virtually worthless. To cap things off, even where the investigator theoretically agrees to share data, that is no guarantee that sharing will actually occur. People in multiple disciplines say that they have requested data from an NIH-funded grant and were unsuccessful even though the investigator had purported to be willing to share data on request.

Such a narrow and toothless rule should be tightened up considerably. **That is, the default rule for all research should be that the raw data created pursuant to any federal grant must be shared, along with the computer code used to transform and analyze it.**[35] The only exception should be for rare situations where the raw data is too voluminous (e.g., the Large Hadron Collider or certain astronomy projects); in these cases, the researchers should be mandated to share the processed or sampled data that they themselves use for any analytic purposes.

---

[32] D. N. Paltoo et al., "Data use under the NIH GWAS Data Sharing Policy and future directions," *Nature Genetics* 46 (27 Aug. 2014): 934-38, doi:10.1038/ng.3062.

[33] C. Wu et al., "Joint analysis of three genome-wide association studies of esophageal squamous cell carcinoma in Chinese populations," *Nature Genetics* 46 (2014): 1001-06. Available at http://www.nature.com/ng/journal/v46/n9/full/ng.3064.html.

[34] G. J. Jaspers & P. LJ Degraeuwe, "A failed attempt to conduct an individual patient data meta-analysis," *Systematic Review* 3 (4 Sept. 2014): 97.

[35] When journals make data sharing mandatory as opposed to optional, the result is a 1,000-fold difference in the availability of data. See T. H. Vines et al., "Mandated data archiving greatly improves access to research data," *FASEB* Journal 27 (April 2013): 1304-08. Available at http://www.fasebj.org/content/27/4/1304.

Likewise, computer code used to process and analyze data should routinely be made public in case anyone else is interested in reviewing it for correctness.[36] A good example can be found in the reproducibility policy for the journal *Biostatistics*, which requires authors to "submit all the necessary materials" so that an editor "can execute the code on analytic data sets and produce output similar to that obtained by the author."[37] Even prestigious scholars have been tripped up by coding errors,[38] and making code available allows independent researchers the chance to exercise oversight over poorly written code. The very possibility that other researchers might examine an investigator's code will incentivize that investigator from the outset to be meticulous about testing and writing careful comments.

Moreover, to the greatest extent possible, the entire scientific workflow should be preserved in an open manner. For example, in all fields where lab notebooks are used, government-funded research projects should require the use of open electronic notebooks, so that other researchers will be much better informed about how to replicate and extend the findings.[39] In many disciplines, an open software tool like the Open Science Framework would make it possible to preserve every document, dataset, and code script associated with a research project, along with every change made to each file during the project's lifespan.

These same rules on sharing data and code should apply to clinical trials in medicine. Clinical trials are often thought to be more reproducible than other research fields because they use the rigorous method of randomized trials and are usually conducted under the watchful oversight of the FDA. Even so, misinformation about pharmaceutical effectiveness abounds, because many trials, including trials funded by federal tax dollars, are never published in the medical literature.[40] Worse, even the trials that are published can be misleading, because a

---

[36] Z. Merali, ". . . Error . . . why scientific programming does not compute," *Nature* 467 (14 Oct. 2010): 775-77.

[37] See http://biostatistics.oxfordjournals.org/content/10/3/405.full.

[38] G. Miller et al, "A Scientist's Nightmare: Software Problem Leads to Five Retractions," *Science* 314 (22 Dec. 2006): 1856-57.

[39] A. Mascarelli, "Jump off the page," *Nature* 507 (2014): 523-25.

[40] Ross et al. (2009) found that only 40% of industry-funded trials and 47% of NIH-funded trials are published. J. Ross et al., "Trial Publication after Registration in ClinicalTrials.Gov: A Cross-Sectional Analysis," *PLoS Medicine* (8 Sept. 2009), at http://www.plosmedicine.org/article/info%3Adoi%2F10.1371%2Fjournal.pmed.1000144. A subsequent study found that a third of NIH-funded trials remained unpublished even after a median 51 months since the trials were completed. Ross et al., "Publication of NIH funded trials registered in ClinicalTrials.gov: cross sectional analysis," *BMJ* 344 (3 Jan. 2012), at http://www.bmj.com/content/344/bmj.d7292.

published article will be selectively written so as to highlight a desired finding.[41] Even government agencies, including the Centers for Disease Control, have published incorrect information when they had nothing to go on but the cherry-picked clinical trial results that make it into the published literature.[42] Due to inaccurate and incomplete information caused by a lack of transparency, patients and doctors are misled into using the wrong treatments.

Data from NIH-funded clinical trials should be made available to all reasonable requests from qualified analysts who agree to the appropriate confidentiality protections.[43] This would be the "gold standard" for turning clinical trials into reproducible research,[44] and an Institute of Medicine committee is currently considering guidelines on how best to release such data.[45] This is not a novel idea, however: since 2012, the *BMJ* (formerly *British Medical Journal*) has had this requirement for any clinical trials that it publishes.[46] Even with "failed" clinical trials, datasets can be examined to look for genetic and other characteristics of so-called super-responders, i.e., people in the treatment group who had surprising recoveries from a terminal disease.[47] As a distinctly second-best alternative, NIH-funded trials should at least make clinical study reports available; such reports

---

[41] Dwan et al., "Evidence for the Selective Reporting of Analyses and Discrepancies in Clinical Trials: A Systematic Review of Cohort Studies of Clinical Trials," *PLoS Medicine* (24 June 2014), at http://www.plosmedicine.org/article/info%3Adoi%2F10.1371%2Fjournal.pmed.1001666.

[42] To this day, the CDC's website claims that Tamiflu can prevent flu and decrease the risk of death (see http://www.cdc.gov/flu/professionals/antivirals/summary-clinicians.htm), even though a full review of the evidence indicates that those claims are overstated. Jefferson et al., "Oseltamivir for influenza in adults and children: systematic review of clinical study reports and summary of regulatory comments," *BMJ* 348 (9 April 2014), at http://www.bmj.com/content/348/bmj.g2545.

[43] A. J. Vickers, "Whose data set is it anyway? Sharing raw data from randomized trials," *Trials* 7 (2006): 15; M. A. Rodwin & J. D. Abramson, "Clinical Trial Data as a Public Good," *JAMA* 308 (5 Sept. 2012): 871-72.

[44] P. C Gøtzsche, "Why we need easy access to all data from all clinical trials and how to accomplish it," *Trials* 12 no. 249 (2011), at http://www.trialsjournal.com/content/12/1/249.

[45] See http://www.iom.edu/Activities/Research/SharingClinicalTrialData.aspx.

[46] K. Thomas, "Medical Journal to Require More Details on Drug Trials," *New York Times* (31 Oct. 2012), available at http://www.nytimes.com/2012/11/01/business/british-medical-journal-to-require-detailed-clinical-trial-data.html?_r=1&.

[47] H. Ledford, "Cancer researchers revisit 'failed' clinical trials," *Nature News & Comment* (18 April 2013).

provide information on the outcomes and side effects in clinical trials that are left out of the main published literature.[48]

One objection is that patient-level data are subject to privacy protections under the Health Insurance Portability and Accountability Act (HIPAA).[49] Nonetheless, it is possible to de-identify and further statistically anonymize datasets and then to make them available to other researchers subject to a confidentiality agreement. Indeed, several drug companies, such as GlaxoSmithKline, are voluntarily making some of their clinical trial datasets available to outside researchers. NIH-funded clinical trials should do no less.

D. Fund More Replications

Government agencies should fund many more replication experiments, particularly in areas like cancer cell biology or pre-clinical animal research. Even if initial studies have a high risk of being false positives, well-designed replication experiments can sharply reduce the overall false positive rate.[50]

LJAF has funded large replication projects in psychology and cancer cell biology that are the first of their kind in each field. But those projects are a drop in the bucket compared to what the NIH and the National Science Foundation (NSF) could fund *every year* if they devoted even *one percent* of their funding to replication experiments. Agencies should run systematic audits to decide on important experimental results that deserve replication, and fund such replication experiments either via contracts or as a condition for grant renewal.

Agencies should also fund re-analyses of existing datasets, particularly for expensive clinical trials where it is important that the findings' robustness be checked by independent investigators. This is rarely done at present; a recent team of researchers could find only 37 such re-analyses published in the past 35 years, 13 of which led to conclusions that were different from the original publication.[51]

E. Strengthen Monitoring and Enforcement Power

---

[48] B. Wieseler et al., "Completeness of Reporting of Patient-Relevant Clinical Trial Outcomes: Comparison of Unpublished Clinical Study Reports with Publicly Available Data," *PLoS Medicine* (8 Oct. 2013), doi:10.1371/journal.pmed.1001526.

[49] For a good discussion of the risks, see M. Mello et al., "Preparing for Responsible Sharing of Clinical Trial Data," *New England Journal of Medicine* 369 (2013): 1651-1658, at http://www.nejm.org/doi/full/10.1056/NEJMhle1309073.

[50] R. Moonesinghe et al., "Most Published Research Findings Are False – But a Little Replication Goes a Long Way," *PLoS Medicine* (27 Feb. 2007), doi: 10.1371/journal.pmed.0040028.

[51] Ebrahim et al., "Reanalyses of Randomized Clinical Trial Data," *JAMA* 312 (2014): 213, doi:10.1001/jama.2014.9646.

laura and john arnold foundation

For any solution that is adopted, federal enforcement power should be strengthened. Government funders should be empowered to undertake audits of researchers (R01 grants in particular) to make sure that all policies related to reproducibility are being followed. There should be serious statutory penalties (such as funding freezes or even restitution) if researchers are wilfully falling short on their commitments. This stick should be joined by a carrot as well: agency reviewers who score grant proposals could give *substantial* credit to researchers who have pre-registered previous experiments and shared data and code, as well as to researchers whose main contribution was writing software or creating a dataset in and of itself.

III. Conclusion

LJAF strongly encourages the White House to make scientific reproducibility a key part of its innovation agenda. We have outlined several possible solutions and look forward to being part of the discussion that emerges over the next few years.

2800 POST OAK BLVD., SUITE 225 / HOUSTON, TEXAS 77056 / P: 713.554.1349 / F: 832.615.8669