**Journal of Clinical and Translational Research** 2017; 3(S2): 1-4

**Journal of Clinical and Translational Research**

Journal homepage: http://www.jctres.com/en/home

**EDITORIAL**

# The need for reporting negative results – a 90 year update

Brian D. Earp

Ethics advisory editor

*Departments of Psychology and Philosophy, Yale University, New Haven, Connecticut, United States*
*brian.earp@gmail.com*

In January of 1927, Dr. Richard D. Mudd of Detroit published a letter in the Journal of the American Medical Association, seeking to vindicate his grandfather, Dr. Samuel A. Mudd, against charges of conspiring in a murder [1]. The victim was U.S. President Abraham Lincoln; the murderer, actor John Wilkes Booth (see Appendix). In this editorial, I, an erstwhile actor, would like to vindicate my own grandfather, Dr. John Rosslyn Earp, for a letter he published on the same day, just one column over, in the very same issue of the journal [2]. But I mean "vindicate" in its other sense—to prove correct—as we shall see.

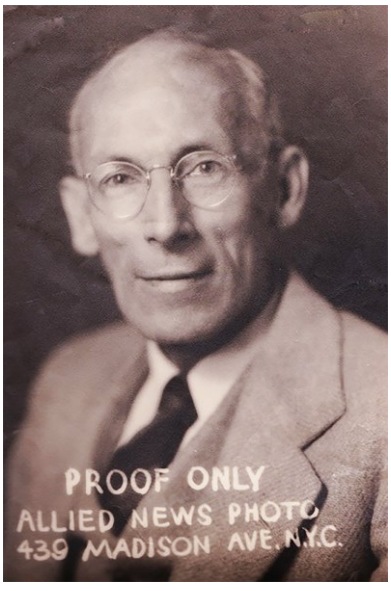



**Figure 1.** Photograph of John Rosslyn Earp, taken circa 1930

I never knew my grandfather. He died in 1941 at the age of 49, more than four decades before I was born. My father, his son, hardly knew him either: he was only 7 when "Ros" passed away from longstanding health problems, leaving him and his siblings to the care of their mother. I had been told that Grandpa Earp—no relation to Wyatt—was at one point the Director of Public Health for the State of New Mexico [3]. I knew that he'd emigrated from somewhere in England around the turn of the last century. That, and an impression I had from an old photographic proof balanced atop a bookcase in my childhood home, was about it (Figure 1).

In 2013, I took a break from my acting career to study the history and philosophy of science at the University of Cambridge.[1] My preoccupation at the time, which has not abated, was the public and professional "crisis of confidence" affecting among other fields medicine and social psychology [4-6]. The term "crisis of confidence" refers to the "unprecedented level of doubt" experienced by many contemporary scientists about the reliability of reported findings in the literature [7].

Why all the doubt? There are several reasons. Anonymous surveys of practicing scientists have shown widespread use of "questionable research practices," including "p-hacking," selective reporting of measures or outcomes, and HARKing —hypothesizing after the results are known—all of which increase the likelihood of generating Type 1 errors [8-11]. Moreover, critiques have been raised about the reward structure of science which favors non-stop "productivity" and headline-grabbing conclusions over painstaking methodology [12-15]. And a series of high-profile apparent failures to replicate major findings from prior studies has sent shockwaves through the scientific community [16,17].

All of this has combined to create a sense of genuine worry: how much of what we think we know do we actually know? Controversially, at least one prominent meta-scientist, John Ioannidis, has estimated that "most published research findings are false" [18].

The hardest-hit field seems to be psychology (which to its credit has also taken up the vanguard for reform) [19,20], with

1 After I arrived, I got a phone call from my father. "You know, Brian, now that I think about it, I seem to remember that your grandpa used to be a student at Cambridge, too, before he came to America." Sure enough, an email sent to a university archivist resulted in a record for John Rosslyn Earp: he had been at St. Johns—the college right next door to where I was studying at Trinity—almost exactly a century before.


DOI: http://dx.doi.org/10.18053/jctres.03.2017S2.001

biomedicine and related disciplines trailing not so far behind [21-23]. Since I had studied the former subject as an undergraduate student, I was familiar with an eerily similar crisis in that field from the 1970s, as a result of which leading practitioners sought to root out problems in the way they conducted, evaluated, and published their empirical research [24]. One of the biggest problems to get spotlight treatment was the failure of most journals to publish "negative" results.

In a now-famous article published in 1975, Professor Anthony Greenwald, then of Ohio State University, discussed what he called the "Consequences of prejudice against the null hypothesis" [25]. As he wrote, the lack of a dependable "home" for negative findings creates "a dysfunctional research-publication system." Not only are there "relatively few publications on problems for which the null hypothesis is (at least to a reasonable approximation) true," but, even among those, "a high proportion will erroneously reject the null hypothesis."

In short, Greenwald identified what is now termed "publication bias" in favor of "statistically significant" findings—a bias that has featured prominently in contemporary discussions about the potential causes of the so-called "replication crisis" [26–28].

The idea is simple. If 20 labs, say, run essentially the same experiment, and only one of them gets it to "work," chances are good that the apparent finding from this one "lucky" lab is actually a statistical fluke. But since journals—and especially high-impact journals—have had a historical tendency to publish only positive findings, it is this probably-a-fluke result that will end up enshrined in the scientific record [29].

The "negative" results, by contrast, from the 19 other labs in our dummy example—or perhaps the 19 previous versions of the same study from the original lab, recast as "pilot" experiments when they didn't pan out—won't typically be written up and submitted, much less published in a prominent journal. Instead, they get "filed away" in the researcher's bottom drawer (the so-called "file drawer" problem), never to be seen again [30,31].

The literature, then, gets skewed in the direction of impressive-looking errors, which, for obvious reasons, can't be replicated later on. In a clinical context, this "skew" may have serious ethical implications for the protection of patient health and well-being. As the editor-in-chief of this journal notes, "selective publication [of] trials can skew the apparent risk-benefit ratio of the drug towards the latter and generate an unrealistic bias, thereby potentially slanting the accuracy of evidence-based medicine" [32].

Needless to say, medical treatments need to be based on accurate research. Basing them on something else is not only unethical (because of the unjustified risk it poses to patients and study participants); it is also an extraordinary waste of resources [33]. Selectively publishing "positive" findings makes these problems worse.

So what can be done? In the course of researching this issue, I stumbled across a paper with a pertinent title that I thought might offer a solution: "The Need for Reporting Negative Results." The source? Journal of the American Medical Association—volume 88, number 2. The year? 1927. The author? J. R. Earp, my grandfather [2].

I had no idea he had ever written on the subject (to speak of chills and spines is to get it right). What follows then is his prophetic letter in full, with a few minor edits for ease of reading:

> *To the Editor:—One of the things we practitioners sometimes neglect is the reporting of failures. In THE JOURNAL, Oct. 2, 1926, Dr. Richard L. Sutton, with proper scientific reserve, reported the treatment of six consecutive cases of warts with intramuscular injections of sulpharsphenamine. As a result of this communication, I venture to guess that not less than a hundred physicians, perhaps several hundred, injected sulpharsphenamine into patients with warts. Supposing that 99 per cent get negative results, what happens? Each of them gives up the method as a failure and does not say anything more about it, and the treatment remains on record as an undisputed success. Possibly 1 per cent who meet with success will communicate with Dr. Sutton, so that by and by he will have quite an impressive series of cases, comparable with the mercurochrome successes published in a recent number of THE JOURNAL. ...*
>
> *To practice what I am preaching, let me now report that on November 30, I injected 0.4 g of sulpharsphenamine [into] the left buttock of E. M. B., a girl, aged 18, who was at that date complaining of the presence of twenty-four warts distributed mostly over the hands and arms. At the present date, there are twenty-eight warts, and evidence of regressive changes in the original twenty-four has not been seen.*

The problem is plain to see; the "need for reporting negative results" is equally apparent [34]. But one-off letters to the editor by conscientious doctors like my grandfather will not suffice to address the root of the problem. What is needed is top-down leadership from journals themselves: not only passively allowing for the submission of negative findings, but actively welcoming them and even seeking them out. In fact, it should be no harder to publish a high-quality study with "null" results—including unsuccessful attempts at replication—than a high-quality study that purports to show an effect.

There are some signs of progress. Articles with "replication" in the title are now being published on a regular basis [35–42]; there is even a dedicated Journal of Articles in Support of the Null Hypothesis (although it is not especially well-known). But there is still a lot of room for improvement. In a recent review of 1151 journals, researchers found that only 3% explicitly stated that they accepted replications; 63% did not state as much but also did not discourage them; 33% discouraged them implicitly by stressing novelty in solicited submissions; and 1% actively frowned on replications by stating that they did not publish them [43].

 DOI: http://dx.doi.org/10.18053/jctres.03.2017S2.001

Against this backdrop, where does the Journal of Clinical and Translational Research (JCTR) stand? In the founding editorial for this journal, the editor states that JCTR encourages the publication of negative results for two main reasons in addition to counteracting the "skewing" problem already mentioned [32]:

> *(1) publication of negative data, especially when obtained in a technically sound study … provides cues as to why a certain procedure or process did not work and steers research efforts away from failure. In that sense, something not working can be considered 'part' of the mechanism.*
>
> *(2) negative results prevent colleagues from conducting redundant work, saving animals and valuable resources in the process. An expedient trajectory to the clinical setting, during which redundancy is minimized, is ultimately beneficial for everyone involved in translational and clinical research as well as the target group (i.e., patients).*

It is with these points in mind that I am happy to introduce, on behalf of my co-editors Emma Bruns and Michal Heger—as well as the entire journal staff—this special issue dedicated entirely to the publication of negative results. Though I never had a chance to meet him, something tells me Grandpa would be proud.

## References


[1] Mudd RD. Dr. Mudd and the death of Lincoln. JAMA. 1927;88:119.

[2] Earp JR. The need for reporting negative results. JAMA. 1927;88:119.

[3] Editor. News from the field. Am J Public Health. 1937;27:755–758.

[4] Baker M. Is there a reproducibility crisis? Nature. 2016;533:452–454.

[5] Earp BD, Trafimow D. Replication, falsification, and the crisis of confidence in social psychology. Front Psychol. 2015;6:1–11.

[6] Nosek BA, Errington TM. Making sense of replications. eLife. 2017;6:e23383.

[7] Pashler H, Wagenmakers E. Editors' introduction to the special section on replicability in psychological science: a crisis of confidence? Perspect Psychol Sci. 2012;7:528–530.

[8] John LK, Loewenstein G, Prelec D. Measuring the prevalence of questionable research practices with incentives for truth telling. Psychol Sci. 2012;23:524–532.

[9] Kerr NL. HARKing: hypothesizing after the results are known. Personal Soc Psychol Rev. 1998;2:196–217.

[10] Head ML, Holman L, Lanfear R, Kahn AT, Jennions MD. The extent and consequences of p-hacking in science. PLOS Biol. 2015;13:e1002106.

[11] Trafimow D, Earp BD. Null hypothesis significance testing and Type I error: the domain problem. New Ideas in Psychology. 2017;45:19-27.

[12] Nosek BA, Spies JR, Motyl M. Scientific utopia II: restructuring incentives and practices to promote truth over publishability. Perspect Psychol Sci. 2012;7:615–631.

[13] Munafò MR, Nosek BA, Bishop DVM, Button KS, Chambers CD, Sert NP du, Simonsohn U, Wagenmakers E, Ware JJ, Ioannidis JPA. A manifesto for reproducible science. Nat Hum Behav. 2017;1:1–9.

[14] Everett JAC, Earp BD. A tragedy of the (academic) commons: interpreting the replication crisis in psychology as a social dilemma for early-career researchers. Front Psychol. 2015;6:1–4.

[15] Earp BD. The unbearable asymmetry of bullshit. Health Watch. 2016;Spring(101):4–5.

[16] Yong E. Replication studies: bad copy. Nat News. 2012;485:298–300.

[17] Earp BD. What did the OSC replication initiative reveal about the crisis in psychology? BMC Psychol. 2016;4:1–19.

[18] Ioannidis JPA. Why most published research findings are false. PLoS Med. 2005;2:e124.

[19] Chambers C. The changing face of psychology. The Guardian. 2014 Jan 24 https://www.theguardian.com/science/head- quarters/2014/jan/24/the-changing-face-of-psychology

[20] LeBel EP, Vanpaemel W, McCarthy RJ, Earp BD, Elson M. A unified framework to quantify the trustworthiness of empirical research. PsyArXiv. 2017; https://osf.io/preprints/ psyarxiv/uwmr8

[21] Engber D. Cancer research is broken. Slate. 2016 Apr 19. http://www.slate.com/articles/health_and_science/future_tense/2016/04/biomedicine_facing_a_worse_replication_crisis_than_the_one_plaguing_psychology.html

[22] Collins FS, Tabak LA. NIH plans to enhance reproducibility. Nature. 2014;505:612–613.

[23] Lose G and Klarskov N. Why published research is untrustworthy. Int Urogynecol J. 2017; in press.

[24] Elms AC. The crisis of confidence in social psychology. Am Psychol. 1975;30:967–976.

[25] Greenwald AG. Consequences of prejudice against the null hypothesis. Psychol Bull. 1975;82:1–20.

[26] Easterbrook PJ, Gopalan R, Berlin JA, Matthews DR. Publication bias in clinical research. The Lancet. 1991;337:867–872.

[27] Francis G. Replication, statistical consistency, and publication bias. J Math Psychol. 2013;57:153–69.

[28] Bakker M, van Dijk A, Wicherts JM. The rules of the game called psychological science. Perspect Psychol Sci. 2012;7:543–554.

[29] Earp BD, Wilkinson D. The publication symmetry test: a simple editorial heuristic to combat publication bias. J Clin Transl Res. 2017; 3: in press.

[30] Rosenthal R. The file drawer problem and tolerance for null results. Psychol Bull. 1979;86:638–41.

[31] Pautasso M. Worsening file-drawer problem in the abstracts of natural, medical and social science databases. Scientometrics. 2010;85:193–202.

[32] Heger M. Editor's inaugural issue foreword: perspectives on translational and clinical research. J Clin Transl Res. 2015;1: 1– 5.

[33] Glasziou P, Altman DG, Bossuyt P, Boutron I, Clarke M, Julious S, Michie S, Moher D, Wager E. Reducing waste from incomplete or unusable reports of biomedical research. The Lancet. 2014;383:267–276.

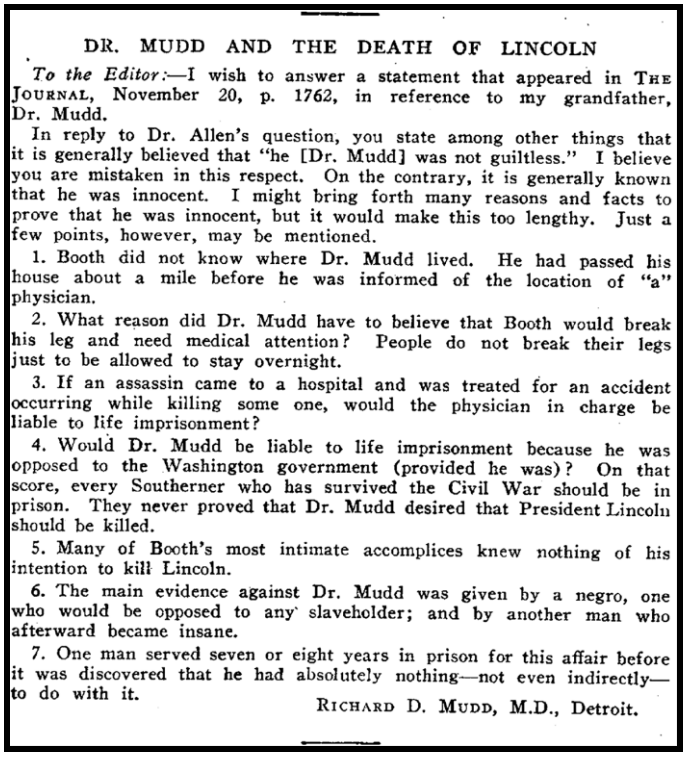
[34] Earp BD, Everett JAC. How to fix psychology's replication crisis. The Chronicle of Higher Education. 2015 Oct 25. http://www.chronicle.com/article/How-to-Fix- psychologys/233857


 DOI: http://dx.doi.org/10.18053/jctres.03.2017S2.001

[35] Boekel W, Wagenmakers EJ, Belay L, Verhagen J, Brown S, Forstmann BU. A purely confirmatory replication study of structural brain-behavior correlations. Cortex. 2015;66:115–133.
[36] Bostyn DH, Roets A. Trust, trolleys and social dilemmas: a replication study. J Exp Psychol Gen. 2017;146:e1–7.
[37] Castro VM, Kong SW, Clements CC, Brady R, Kaimal AJ, Doyle AE, Robinson EB, Churchill SE, Kohane IS, Perlis RH. Absence of evidence for increase in risk for autism or attention-deficit hyperactivity disorder following antidepressant exposure during pregnancy: a replication study. Transl Psychiatry. 2016;6:e708.
[38] Earp BD, Everett JAC, Madva EN, Hamlin JK. Out, damned spot: Can the "Macbeth Effect" be replicated? Basic Appl Soc Psychol. 2014;36:91–98.
[39] Radke S, de Bruijn ERA. Does oxytocin affect mind-reading? A replication study. Psychoneuroendocrinology. 2015;60:75–81.
[40] Renes RA, van der Weiden A, Prikken M, Kahn RS, Aarts H, van Haren NEM. Abnormalities in the experience of self-agency in schizophrenia: a replication study. Schizophr Res. 2015;164:210–213.
[41] Simeoni S, Hannah R, Daisuke S, Kawakami M, Gigli GL, Rothwell JC. Effects of quadripulse stimulation on human motor cortex excitability: a replication study. Brain Stimul. 2016;9:148–150.
[42] Gil-Gómez de Liaño B, Stablum F, Umiltà C. Can concurrent memory load reduce distraction? A replication study and beyond. J Exp Psychol Gen. 2016;145:e1.
[43] Martin GN, Clarke RM. Are psychology journals anti-replication? A snapshot of editorial practices. Front Psychol. 2017;8:1–6.

## APPENDIX

Letter from Dr. Mudd. JAMA. 1927;88:119.



DR. MUDD AND THE DEATH OF LINCOLN

*To the Editor:*—I wish to answer a statement that appeared in THE JOURNAL, November 20, p. 1762, in reference to my grandfather, Dr. Mudd.

In reply to Dr. Allen's question, you state among other things that it is generally believed that "he [Dr. Mudd] was not guiltless." I believe you are mistaken in this respect. On the contrary, it is generally known that he was innocent. I might bring forth many reasons and facts to prove that he was innocent, but it would make this too lengthy. Just a few points, however, may be mentioned.

1. Booth did not know where Dr. Mudd lived. He had passed his house about a mile before he was informed of the location of "a" physician.

2. What reason did Dr. Mudd have to believe that Booth would break his leg and need medical attention? People do not break their legs just to be allowed to stay overnight.

3. If an assassin came to a hospital and was treated for an accident occurring while killing some one, would the physician in charge be liable to life imprisonment?

4. Would Dr. Mudd be liable to life imprisonment because he was opposed to the Washington government (provided he was)? On that score, every Southerner who has survived the Civil War should be in prison. They never proved that Dr. Mudd desired that President Lincoln should be killed.

5. Many of Booth's most intimate accomplices knew nothing of his intention to kill Lincoln.

6. The main evidence against Dr. Mudd was given by a negro, one who would be opposed to any slaveholder; and by another man who afterward became insane.

7. One man served seven or eight years in prison for this affair before it was discovered that he had absolutely nothing—not even indirectly—to do with it.

RICHARD D. MUDD, M.D., Detroit.

 DOI: http://dx.doi.org/10.18053/jctres.03.2017S2.001