

Virtual Mentor

American Medical Association Journal of Ethics
January 2013, Volume 15, Number 1: 29-33.

JOURNAL DISCUSSION

When Research Evidence is Misleading

Chetan Huded, MD, Jill Rosno, MD, and Vinay Prasad, MD

Ioannidis JP. Why most published research findings are false. *PLoS Med.* 2005;2(8):e124.

Each year, millions of research hypotheses are tested. Datasets are analyzed in ad hoc and exploratory ways. Quasi-experimental, single-center, before and after studies are enthusiastically performed. Patient databases are vigorously searched for statistically significant associations. For certain “hot” topics, twenty independent teams may explore a set of related questions. From all of these efforts, a glut of posters, presentations, and papers emerges. This scholarship provides the foundation for countless medical practices—the basis for the widespread acceptance of novel interventions. But are all of the proffered conclusions correct? Are even most of them?

John P. A. Ioannidis’s now famous work, “Why Most Published Research Findings Are False,” makes the case that the answer is no [1]. Ioannidis uses mathematical modeling to support this claim. The core of his argument—that a scientific publication can be compared to a diagnostic test, and that we can ask of it, “how does a positive finding change the probability of the claim we are evaluating?”—has a simple elegance.

Ioannidis asks us to think broadly about the truth of a claim, contemplating not only the study in question but the totality of the evidence. He uses the concept of positive predictive value (PPV)—that is, the probability that a positive study result is a true positive—as the foundation of his analysis. He demonstrates that, under common statistical assumptions, the PPV of a study is actually less than 50 percent if only 5 of 100 hypotheses relating to a field or area of study are true. This means that, under these circumstances—which seem reasonable in many medical fields—a simple coin flip may be as useful as a positive research finding in determining the validity of a claim.

Bias, however, is a ubiquitous threat to the validity of study results, and testing by multiple independent teams can further degrade research findings. As bias, such as that resulting from the use of poor controls or study endpoints, increases, the PPV of a study decreases. Similarly, the global pursuit of positive research findings by multiple investigators on a single subject also decreases the likelihood of true research findings due to a phenomenon called multiple hypothesis testing.

From this framework, Ioannidis draws 6 conclusions:

1. The smaller the sample size of a study, the lower its PPV.
2. Fields with lower effect sizes (e.g., the degree of benefit/harm afforded by a treatment or diagnosis) suffer from lower PPV.
3. The prestudy probability of a true finding influences the PPV of the study—thus, fields with many already tested hypotheses are less likely to yield true findings.
4. Increased flexibility in study design, endpoints, and analysis affords more opportunities for bias and false results.
5. The greater the financial and other external influences, the greater the bias and lower the PPV.
6. There is a paradoxical relationship between the number of independent investigators on a topic and the PPV for a given study in that field—the hotter the field, the more spurious the science.

The consequences of all this bad science are not felt by researchers—whose careers may be propelled by these erroneous results—but by the patients who are subject to medical practices the validity of which is uncertain. Certain medical practices have risen to prominence based on false preliminary results only to be contradicted years later by robust, randomized controlled trials. Examples from recent history include the use of hormone replacement therapy for postmenopausal women [2], stenting for stable coronary artery disease [3], and fenofibrate to treat hyperlipidemia [4].

We have previously called this phenomenon “medical reversal” [5]. Reviewing all original articles published in the *New England Journal of Medicine* during 2009, we found that 46 percent (16 of 35) of articles challenged standard of care and constituted reversals. These reversals spanned the range of medical practice, including screening practices (prostate-specific antigen testing), medications (statins in hemodialysis patients), and even procedural interventions (such as percutaneous coronary intervention in stable angina).

Ioannidis has also provided an empirical estimate of contradiction in medical research [6]. He reviewed the conclusions of highly cited medical studies from prominent journals over 13 years and tracked research on those same topics over time. He found that 16 percent (7 of 45) of the highly cited papers were later contradicted and another 16 percent found stronger effects than subsequent studies. Observational studies were most likely to be later contradicted (5 of 6 observational studies versus 9 of 39 randomized trials). Among randomized trials, the only predictor of contradiction was sample size (the median sample size was 624 among contradicted findings as opposed to 2,165 among validated findings), a finding that supports Ioannidis’s conclusions 1 and 2. Sufficient sample size and statistical power are invaluable for reproducibility.

How can we fix the problem of all of this incorrect science in medicine’s “evidence base”? With Ioannidis, we have proposed [7] that trial funding for phase-3 studies—

which test the efficacy of medical products—be pooled. Scientific bodies without conflicts of interest should prioritize research questions and design and conduct clinical trials. This recommendation would address several problems in medicine [8, 9] including forcing trials to address basic unanswered questions for common diseases rather than simply advancing the market share of specific products.

Additionally, our proposal would dramatically reduce medical reversal by favoring large-scale randomized trials over countless, scattered lesser ones [10]. Not all RCTs are the same. When RCTs have small sample sizes and financial conflicts of interest and there are multiple studies investigating a single intervention, they are more likely to be in error. (For an illustration of this, look at the number of RCTs that have been conducted for the drug bevacizumab, for which the FDA withdrew its indication for metastatic breast cancer last year [11].)

For observational studies, careful selection of study hypotheses is crucial to increasing the truth and durability of research findings. Hypotheses must be clearly defined prior to data collection and analyses to minimize bias. Krumholz has proposed that such studies document not only the final methods but the history of the methods, accounting for how they were adjusted or changed during the study detailing any and all exploratory analyses [12]. Registration of protocols for observational studies currently remains optional, unlike clinical trial registration, which is required prior to patient enrollment. We agree with suggestions to formally establish a registry of observational analyses [13]. Bias in observational analyses currently presents challenges in reliably basing medical practices on this type of work alone. Only time will tell if this is surmountable.

Finally, Ioannidis's conclusions pertain to concrete ethical choices that budding physicians and researchers make, although they may not see them as such. Our system of medical education and postgraduate medical training rewards the accumulation of publications (abstracts, posters, presentations, and papers) rather than the pursuit of truths. Even students who ultimately pursue private practice careers often engage in research to build their curriculum vitae. Many educators feel that this process is acceptable—better for physicians to gain exposure to research, even if they don't ultimately pursue it, and, anyway, what's the harm? Here, we show the harm. The current system increases the number of publications in a given field while muddying true associations. Ioannidis has criticized medical conferences for promulgating poor science by selecting—on the basis of several-hundred-word abstracts—research that is often not published after more extensive peer review [14]. Students and trainees should consider the inevitable ethical question: if most research findings are false—how vigorously should I advertise my own?

The prevalence of false research findings and their adoption into mainstream practice carries heavy consequences. In our era of soaring health care costs, we cannot afford to implement unnecessary and costly interventions in the absence of sound evidence. More importantly, the use of unfounded interventions puts millions of patients at risk for harm without benefit. Ioannidis's piece reminds us that we have an ethical

responsibility to adhere to high-quality clinical investigation in order to eliminate waste and promote the health and safety of our patients.

References

1. Ioannidis JP. Why most published research findings are false. *PLoS Med.* 2005;2(8):e124.
2. Rossouw JE, Anderson GL, Prentice RL, et al. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results from the Women's Health Initiative randomized controlled trial. *JAMA.* 2002;288(3):321-333.
3. Boden WE, O'Rourke RA, Teo KK, et al. Optimal medical therapy with or without PCI for stable coronary disease. *N Engl J Med.* 2007;356(15):1503-1516.
4. The ACCORD Study Group, Ginsberg HN, Elam MB, et al. Effects of combination lipid therapy in type 2 diabetes mellitus. *N Engl J Med.* 2010;362(17):1563-1574.
5. Prasad V, Gall V, Cifu A. The frequency of medical reversal. *Arch Intern Med.* 2011;171(18):1675-1676.
6. Ioannidis JP. Contradicted and initially stronger effects in highly cited clinical research. *JAMA.* 2005;294(2):218-228.
7. Prasad V, Cifu A, Ioannidis JP. Reversals of established medical practices: evidence to abandon ship. *JAMA.* 2012;307(1):37-38.
8. Prasad V, Rho J, Cifu A. The diagnosis and treatment of pulmonary embolism: a metaphor for medicine in the evidence-based medicine era. *Arch Intern Med.* 2012;172(12):955-958.
9. Prasad V, Vandross A. Cardiovascular primary prevention: how high should we set the bar? *Arch Intern Med.* 2012;172(8):656-659.
10. Hennekens CH, Demets D. The need for large-scale randomized evidence without undue emphasis on small trials, meta-analyses, or subgroup analyses. *JAMA.* 2009;302(21):2361-2362.
11. D'Agostino RB Sr. Changing end points in breast-cancer drug approval--the Avastin story. *N Engl J Med.* 2011;365(2):e2.
12. Krumholz HM. Documenting the methods history: would it improve the interpretability of studies? *Circ Cardiovasc Qual Outcomes.* 2012;5(4):418-419.
13. Ioannidis JP. The importance of potential studies that have not existed and registration of observational data sets. *JAMA.* 2012;308(6):575-576.
14. Ioannidis JP. Are medical conferences useful? And for whom? *JAMA.* 2012;307(12):1257-1258.

Chetan Huded, MD, is a resident physician in internal medicine at Northwestern Memorial Hospital in Chicago. He was last year's winner of the Rambach Award, given to the most meritorious resident of the intern class. He is interested in cardiology.

Jill Rosno, MD, is a resident physician in internal medicine at Northwestern Memorial Hospital in Chicago. A graduate of Dartmouth Medical School, she is interested in general internal medicine.

Vinay Prasad, MD, is chief fellow in medical oncology at the National Cancer Institute in Bethesda, Maryland. Dr. Prasad coined the term “medical reversal” and has published on the subject in the *Journal of the American Medical Association* and *Archives of Internal Medicine*, among others. He is interested in the adoption of rational methods in medical practice.

Related in VM

[A Call to Integrate Ethics and Evidence-Based Medicine](#), January 2013

[Teaching Critical Appraisal of Medical Evidence](#), January 2013

[Rating Evidence in Medical Literature](#), January 2011

The viewpoints expressed on this site are those of the authors and do not necessarily reflect the views and policies of the AMA.

Copyright 2013 American Medical Association. All rights reserved.