Clin Orthop Relat Res (2021) 00:1-4
DOI 10.1097/CORR.0000000000001891

Clinical Orthopaedics and Related Research®
A Publication of The Association of Bone and Joint Surgeons®

Not the Last Word

Published online: 28 June 2021

Not the Last Word: Masks and the Veil of Ignorance

Joseph Bernstein MD[1]

There was a time not long ago when the debate over mask-wearing for COVID-19 was grounded in science. Thus, when the "Danish mask study" [4]—a randomized controlled trial of 4800 subjects—appeared in the *Annals of Internal Medicine*, it was met with great expectations.

Before the study was published, I thought the results could go either way. On the one hand, the pores of surgical masks are much larger than the virus—so much so, that one might expect masks to filter SARS-CoV-2 no better than a chain link fence might impede the passage of a house fly. On the other hand, masks block spittle and sneezes, trap larger virus-containing droplets, and limit finger-to-face contact, perhaps enough to have a small clinical effect. When I read that the study reported no efficacy seen, I came away unmoved.

Then the smart people went to work. Mask advocates were quick to criticize the methodology of the study. This was not an analysis of *using* masks, they said, but a study of *recommending* masks, with no determination of whether masks were actually worn. "Excellent job, mask-lovers," I thought to myself, "you came up with something I overlooked."

The opposition was rapid with its retort. They argued that the authors had merely employed a routine "intention-to-treat" analysis, one in which subjects are assessed according to "the groups to which they were randomly assigned … regardless of the treatment they actually received" [10]. With an intention-to-treat analysis, subjects randomized to wear masks have their outcomes logged in the mask-wearers' column independent of their actions, just as subjects randomized to operative care are evaluated with that cohort, even if they ultimately decline surgery [16]. "Excellent job, mask-loathers," I thought to myself, "you, too, came up with something I overlooked."

There are at least two reasons why these individuals came up with arguments that I did not. The simplest is that they are just more astute. But from the perspective I prefer, these commentators may have been more motivated, eager to devise arguments to defend their a priori beliefs.

Biases concentrate the mind wonderfully. For example, reviewers will more likely discover flaws in works whose conclusions they dislike. Emerson et al. [6] demonstrated this in an elegant experiment that studied positive-outcome bias during peer review. They created two nearly identical versions of a manuscript, differing only in the direction of the findings (positive/negative). The papers were then sent to 238 blinded reviewers. Emerson et al. found that these reviewers were more likely to reject the negative version of the test manuscript. Even more interestingly, reviewers who received the negative version scrutinized the submission more closely, evidently, for they detected more of the mathematical and citation errors that Emerson et al. had intentionally planted in the manuscript [6].

"Reason is … the slave of the passions," [15] as the philosopher, David Hume, put it. As such, peer reviewers might scrutinize papers not to discover their true worth, but to discover plausible justifications for their gut instincts.

A *note from the Editor-in-Chief: We are pleased to present to readers of* Clinical Orthopaedics and Related Research® *the next* Not the Last Word. *The goal of this column is to explore timely and controversial issues that affect how orthopaedic surgery is taught, learned, and practiced. We welcome reader feedback on all of our columns and articles; please send your comments to eic@clinorthop.org.*

The author certifies that there are no funding or commercial associations (consultancies, stock ownership, equity interest, patent/licensing arrangements, etc.) that might pose a conflict of interest in connection with the submitted article related to the author or any immediate family members.
All ICMJE Conflict of Interest Forms for authors and *Clinical Orthopaedics and Related Research®* editors and board members are on file with the publication and can be viewed on request.
The opinions expressed are those of the writer, and do not reflect the opinion or policy of Clinical Orthopaedics and Related Research® or The Association of Bone and Joint Surgeons®.

J. Bernstein ✉, University of Pennsylvania, 424 Stemmler Hall, Philadelphia, PA 19104, USA, Email: orthodoc@uphs.upenn.edu

[1]Department of Orthopaedic Surgery, University of Pennsylvania, Philadelphia, PA, USA

Wolters Kluwer

Not the Last Word

If biases are fundamental to human thinking, then it would be futile to attempt to try to eradicate them. Rather, editors should try to corral "motivated vigilance" for the benefit of the readers. One way to do this, as suggested by Mirkin and Bach [14], is a two-stage process. Reviewers would first examine a manuscript with a background and methods section but "stripped of actual results." Only if the paper passes this screening test is the final version sent to the reviewers for further assessment.

To more fully harness the power of blinded review, journals have been established for "methods only" manuscripts: *International Journal of Surgery: Protocols*, for example. Placing peer reviewers behind a veil of ignorance—evaluating the methods with no idea about the results—might help suppress biases. Moreover, when authors commit to a given analytic plan in advance, it is harder for them to engage in after-the-fact "significance hunting, data dredging, post hoc hypothesis testing," [13] and other unpleasantries.

Furthermore, when researchers declare that they have studied a question, their failure to publish results produces a signal that improves our understanding of the published data [13]. Because of positive-outcome bias, unpublished studies are more likely to have negative results. In turn, meta-analyses limited to published studies alone will tend to inflate the measured effects [17].

We need to push back on this.

Granted, it's been more than 10 years since Mirkin and Bach proposed two-stage review, and I know of no journal that has adopted it. Likewise, *International Journal of Surgery: Protocols* has been published for 5 years, and I was not aware of it until recently. That lack of adoption is understandable. More work for journals and additional constraints on researchers is hardly appealing to either party. Yet the rest of us—readers, researchers, and patients, current and future—all stand to gain.

One means to harvest these gains at low cost would be to compel researchers to pre-register their studies: articulating the research questions, defining the methods to answer those questions, and, critically, posting a document with that information at sites such as The Research Registry [18].

Currently, most journals require preregistration for prospective studies of living patients. Broadening that requirement for all types of research, even retrospective or laboratory studies, should not be too burdensome.

There was a time not long ago when a suggestion that all studies should be preregistered could be dismissed by saying "that's just not how we do things." In the past year, however, we have out of necessity learned how to overcome habit, tradition, and inertia. I hope we can meet that challenge here.

Eugene J. Carragee MD

Editor-in-Chief, Emeritus, The Spine Journal

Professor, Department of Orthopaedic Surgery, Stanford University School of Medicine

Dr. Joseph Bernstein presents the case that many studies are methodologically flawed, that these flaws are often missed by reviewers, and that some biases, such as those toward positive vs negative results are not occurring at random. He is concerned about bias in our publications, rightfully so. He suggests we revisit the strategy of a two-phase review, as first advocated by Mirkin and Bach in 2011 [14], to include the full scope of orthopaedic surgery research.

I am skeptical. Since the mid-1990s, the problem of publication bias, hidden conflict of interests, and frankly, misleading data selection has become national news, with Senate Investigations, whistle-blower suits, and innumerable Department of Justice indictments. All with little effect. Many manuscripts [5] are still written as marketing ads for devices and drugs under the FDA radar prohibiting promotion of off-label use. Dr. Bernstein, like Ahab, has in his sights the great white whale which menaces journal integrity; are we up to it?

First are practical considerations. As Editor-in-Chief of *The Spine Journal*, I found the proliferation of journals, papers submitted, and use of citations "with an agenda" has stretched manpower to extremes and problematical "too good to be true" submission especially burdensome. Particularly with manufacturer-sponsored studies, I found what everyone already seemed to know. When authors are asked for details, datasets, and corroboration, we found the data and analytic methods were "proprietary" and unreviewable, the authors conflicted, and the time required to tease the mess apart was staggering.

The most important Level I and II studies already require the two-phase approach whether through FDA filings, NIH or other grant proposal documents, or institutionally required review and clinicaltrials.gov registration. But results are mixed [2, 22]. Requiring that an editorial staff carefully review phase 1 and phase 2 publications and reconciliation of the two documents in every study will, I suspect, be met with resistance or workflow bottlenecks. It is extraordinarily uncommon for either reviewers or section editors to check between the two documents when readily available now.

What's more, we've got bigger problems that everybody knows. The editor of *The Lancet*, Richard Horton,

Not the Last Word

lamented that in publishing industry-sponsored studies, the systematic bias to favorable publications corrupted journals of best intent. He felt that "[j]ournals have devolved into information laundering operations for … industry" [19], despite all industry trials having both FDA and clinicaltrials.gov prepublications methods and designs available.

And the industry is a behemoth, financially dwarfing both publications and medical societies one thousand-fold (Pfizer's last 5 years of revenue: > USD 200 billion [21]; Medtronic's 5-year revenue: USD 140 billion [20]). With widening wealth disparity, the Citizens United Verdict granting "person" status to corporations, and the enormous political leverage that follows comes the impression of being beyond regulation. Intense lobbying has decreased FDA, CDC, and other regulatory funding and oversight. Conversely, the marketing and legal resources of the medical industry are larger than the national budgets of medium-sized countries. And the line between marketing and publication support has never been less transparent—if it still exists at all [3, 9, 19].

Any fix would need to be a fundamental change by serious journals and their relationship with industry [3, 12]; but against all good intention cometh the leviathan.

James L. Carey MD, MPH

Director, Penn Center for Advanced Cartilage Repair

Perelman School of Medicine at the University of Pennsylvania

Bias has been defined as sources of variation that distort the study findings in one direction, resulting in systematic error [11]. As a critical reviewer, this definition of bias most closely reflects what I am typically trying to recognize in the design, conduct, and analysis of a study. Dozens of sources of this kind of bias within a study have been named, including some common ones like observer bias, selection bias, and nonrespondent bias.

In Dr. Bernstein's column, he thoughtfully addressed a kind of bias outside of the study—the prejudice of commentators and reviewers who may be motivated to "defend their a priori beliefs" and "to discover plausible justifications for their gut instincts." I agree that this kind of bias may lead to some unfair scrutiny. However, from a practical standpoint, "if biases are fundamental to human thinking," as Dr. Bernstein proposes, then perhaps we should try to value the critical ideas produced by this "motivated vigilance" at the right time.

Dr. Bernstein outlined how the "Danish Mask Study" reported no significant improvement in infection rates following the addition of a mask recommendation to other public health measures [4]. The "mask-lovers" criticized the methodology of the study, especially the absence of any determination of whether masks were actually worn. In contrast, the "mask-loathers" supported that methodology and argued that the study design employed a standard intention-to-treat analysis. The motivations of the "mask-lovers" and "mask-loathers" may have been grounded in defense of their beliefs. Regardless, they have both provided valuable insights and critical direction. In fact, the addition of an "as treated" data presentation (as implied by the "mask-lovers") would have meaningfully complemented the intention-to-treat analysis (supported by the "mask-loathers").

The "Swedish ACL study" [8] is a parallel example of a "negative" study in orthopaedic surgery. In the randomized controlled trial of young, active adults, 62 patients were assigned to early ACL reconstruction and 59 were assigned to the option of having a delayed ACL reconstruction if needed. The authors concluded that "a strategy of rehabilitation plus early ACL reconstruction was not superior to a strategy of rehabilitation plus optional delayed ACL reconstruction" [8]. Fortunately, the authors presented the data in sets of parallel tables—as "intention-to-treat" groups and "as treated" groups. The "as treated" presentation of data in supplementary tables illustrated the decreased laxity in the ACL reconstruction group and the increased nonrepairable meniscus tears in the delayed ACL reconstruction group. Interestingly, this "negative" study supported the AAOS Clinical Practice Guideline recommendation that "Moderate evidence supports surgical reconstruction in active young adult (ages 18-35) patients with an ACL tear" [1] and informed its rationale that ACL reconstruction decreases pathologic laxity and reduces the incidence of meniscal tears.

Did the mask study have the potential to be as influential as the ACL study? Did the critical direction from the "mask-lovers" and "mask-loathers" come too late? The British statistician R.A. Fisher said, "To consult the statistician after an experiment is finished is often merely to ask him to conduct a post-mortem examination. He can perhaps say what the experiment died of" [7]. Similarly, feedback related to study design is most helpful when it occurs early. Perhaps the real value of a two-stage review process or a journal of "methods only" manuscripts is to provide early focused feedback on the study design—prior to study conduct and analysis. The critical insights from "motivated vigilance" will

contribute more to science when they come earlier from biased reviewers (evaluating work prior to publication) rather than later from biased commentators (evaluating work after publication). That is, in my opinion, the timing and content of critical reviews are more important than their motivations.

References

1. American Academy of Orthopaedic Surgeons. Management of anterior cruciate ligament injuries evidence-based clinical practice guideline. Available at: https://www.aaos.org/globalassets/quality-and-practice-resources/anterior-cruciate-ligament-injuries/anterior-cruciate-ligament-injuries-clinical-practice-guideline-4-24-19.pdf. Accessed June 8, 2021.
2. Birkmeyer NJ, Weinstein JN, Tosteson AN, et al. Design of the Spine Patient outcomes Research Trial (SPORT). *Spine (Phila Pa 1976)*. 2002;27:1361-1372.
3. Boetto E, Golinelli D, Carullo G, Fantini MP. Frauds in scientific research and how to possibly overcome them. *J Med Ethics*. Published online October 6, 2020. DOI: 10.1136/medethics-2020-106639.
4. Bundgaard H, Bundgaard JS, Raaschou-Pedersen DET, et al. Effectiveness of adding a mask recommendation to other public health measures to prevent SARS-CoV-2 infection in Danish mask wearers: a randomized controlled trial. *Ann Intern Med*. 2021;174:335-343.
5. Carragee EJ, Ghanayem AJ, Weiner BK, Rothman DJ, Bono CM. A challenge to integrity in spine publications: years of living dangerously with the promotion of bone growth factors. *Spine J*. 2011;11:463-468.
6. Emerson GB, Warme WJ, Wolf FM, Heckman JD, Brand RA, Leopold SS. Testing for the presence of positive-outcome bias in peer review: a randomized controlled trial. *Arch Intern Med*. 2010;170:1934-1939.
7. Fisher RA. Presidential address to the first Indian statistical congress. *Sankhya*. 1938;4:14-17.
8. Frobell RB, Roos EM, Roos HP, Ranstam J, Lohmander LS. A randomized trial of treatment for acute anterior cruciate ligament tears. *N Engl J Med*. 2010;363:331-342.
9. Gelberman RH, Samson D, Mirza SK, Callaghan JJ, Pellegrini VD Jr. Orthopaedic surgeons and the medical device industry: the threat to scientific integrity and the public trust. *J Bone Joint Surg Am*. 2010;92:765-777.
10. Gupta SK. Intention-to-treat concept: a review. *Perspect Clin Res*. 2011;2:109-112.
11. Hulley SB, Cummings SR, Browner WS, Grady DG, Newman TB. *Designing Clinical Research*. 3rd ed. Lippincott Williams and Wilkins; 2007.
12. Keefe PR. The family that built an empire of pain. Available at: https://www.newyorker.com/magazine/2017/10/30/the-family-that-built-an-empire-of-pain. Accessed June 8, 2021.
13. Leopold SS. Editorial: No-difference studies make a big difference. *Clin Orthop Relat Res*. 2015;473:3329-3331.
14. Mirkin JN, Bach PB. Outcome-blinded peer review. *Arch Intern Med*. 2011;171:1213-1214.
15. Online Library of Liberty. A Treatise of Human Nature. Available at: https://oll.libertyfund.org/title/bigge-a-treatise-of-human-nature. Accessed May 25, 2021.
16. Qvist AH, Væsel MT, Jensen CM, Jakobsen T, Jensen SL. Minimal pain decrease between 2 and 4 weeks after nonoperative management of a displaced midshaft clavicle fracture is associated with a high risk of symptomatic nonunion. *Clin Orthop Relat Res*. 2021;479:129-138
17. Reddy AK, Anderson JM, Gray HM, Fishbeck K, Vassar M. Clinical trial registry use in orthopaedic surgery systematic reviews. *J Bone Joint Surg Am*. 2021;103:e41.
18. Registry Research. About the research registry. Available at: https://www.researchregistry.com/about. Accessed June 1, 2021.
19. Smith R. Medical journals are an extension of the marketing arm of pharmaceutical companies. *PLoS Med*. 2005;2:e138.
20. Statista. Medtronic's total revenue from 2006 to 2020. Available at: https://www.statista.com/statistics/241515/total-revenues-of-medtronic-since-2006/. Accessed June 8, 2021.
21. Statista. Pfizer's total revenue from 2006 to 2020. Available at: https://www.statista.com/statistics/266171/revenue-of-pfizer-since-2006/. Accessed June 8, 2021.
22. Weinstein JN, Tosteson TD, Lurie JD, et al. Surgical vs nonoperative treatment for lumbar disk herniation: the Spine Patient Outcomes Research Trial (SPORT): a randomized trial. *JAMA*. 2006;296:2441-2450.