***CORR* Insights®: Have Levels of Evidence Improved the Quality of Orthopaedic Research?**

Vance W. Berger PhD

Received: 16 July 2013 / Accepted: 19 July 2013 / Published online: 1 August 2013
© The Association of Bone and Joint Surgeons® 2013

Where Are We Now?

Condemnation of what is generally perceived as the highest level of evidence is, by extension, condemnation of the entire system; and, to some degree, condemnation (or at least open-eyed critique) is called for. So we must first focus on randomized clinical trials, our so-called Level I evidence. Even among such trials, some are far better than others methodologically. One of the most popular assessment tools is the Jadad score [8], which ranges from zero to five. One particular trial that was rated a perfect five turned out to be so chock full of flaws that it might have served medical science better as a case study in how not to conduct clinical research [4]. Granted, this trial was in rheumatology, not orthopaedics, but the phenomenon of

the scandal of poor medical research [1] transcends any one therapeutic area. The heart of the matter is an unwavering, unquestioning, and uncritical acceptance of “reasonable assumptions,” an important one being that Level I studies of surgical therapy are reliable and important merely because they are randomized and prospective. Tenuous assumptions like this permeate and underlie almost all aspects of clinical trials research, so the entire structure is built on a house of cards. This is where we are now.

Where Do We Need To Go?

We need to move towards better and more reliable trials. The steepest descent (the most bang for the buck) may come from rooting out reliance on untested assumptions, such as assuming that certain distributions are normal when this is not even in the realm of possibility. But this is just the tip of the iceberg. We are equally uncritical of the assumption that just any old way of randomizing, including the use of permuted blocks, is valid in the context of clever investigators trying to predict future allocations [7]. We accept the assumption that a surrogate endpoint, if validated (as if a surrogate endpoint even could be validated in a meaningful way), can replace the true clinical endpoint of interest [8]. We accept the assumption that any old analysis population (such as modified intent-to-treat) is valid [3]. We allow for outright cheating in the form of using a prerandomization run-in period to screen otherwise eligible patients [6]. We adhere to the unquestioning belief that the randomized trial is beyond reproach, so it does not matter that the party with the greatest vested interest in the outcome is the one trusted with the handling of the trial conduct, and enjoying all the discretion that accrues

*This CORR Insights® is a commentary on the article “Have Levels of Evidence Improved the Quality of Orthopaedic Research?” by Cunningham and colleagues available at:
DOI: [10.1007/s11999-013-3159-4](https://doi.org/10.1007/s11999-013-3159-4).*

The author certifies that he, or a member of his immediate family, has no funding or commercial associations (eg, consultancies, stock ownership, equity interest, patent/licensing arrangements, etc.) that might pose a conflict of interest in connection with the submitted article.

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 writers, and do not reflect the opinion or policy of *CORR*® or the Association of Bone and Joint Surgeons®.

This *CORR* Insights® comment refers to the article available at DOI: [10.1007/s11999-013-3159-4](https://doi.org/10.1007/s11999-013-3159-4).

V. W. Berger (✉)

Division of Cancer Prevention, National Cancer Institute,
9609 Medical Center Drive, Rockville, MD 20850, USA
e-mail: vb78c@nih.gov; vance917@gmail.com

therefrom. Where do we need to go? As far from this Fantasyland as possible for the good of the patients who rely on us to act on their behalf.

How Do We Get There?

We get there by implementing zero tolerance. Each one of us is responsible for helping to clean the well and rid it of the toxins that have been dumped for years. When a fatally flawed method is proposed or, even worse, actually used, someone has to stand up and say no. We need to say no to permuted blocks, no to the assumption of normality, unless the analysis is sufficiently complicated that this assumption is absolutely essential (but what does it say [5] about the treatment effect when such complicated models are needed?). We must say no to excluding randomized patients from the analyses, and no to the cavalier attitude that misleading evidence is impossible.

It is possible to say yes. We must say yes to serious scrutiny, yes to an open mind to alternative explanations for what appear to be treatment effects, yes to the integrity and inner strength to take a stand on behalf of the patient constituency even when standing alone in the face of ridicule. Who can take this stand? Journal editors, regulators, investigators, funders, meta-analysts, insurance companies (who might insist on valid trials before they will add a treatment to their formulary for reimbursement), prescribing physicians, and even patients can lead the way.

The system is broken, and there is very little that any one party can do to fix it. There is plenty of incentive to keep the system broken, and little incentive to repair it. Among those few who want to do so, the realization that this effort may be futile may be enough to stymie any

action. This is exactly what the vested interests are counting on. If a treatment does not work, then it does not work, and it is wrong to use flawed research methods to make it appear otherwise; yet this is exactly what happens time and time again. We, as a profession, must be diligent and alert to this possibility, and we need to expose it whenever we see it. This notion of Level I evidence, as I have suggested, is a bit of a misnomer, given the (lack of) quality of most of the work published under this banner. By shining a bright light on the problem, perhaps we can have some hope for a future in which our shared standards ensure that rigor is not optional, and that studies called Level I are worthy of that label.

References

1. Altman DG. The scandal of poor medical research. *BMJ*. 1994;308:283–284.
2. Berger VW. Does the Prentice criterion validate surrogate endpoints? *Stat Med*. 2004;23:1571–1578.
3. Berger VW. Intent-to-treat: further insight. *Res Nurs Health*. 2011;34:4; author reply 5–6.
4. Berger VW. Is the Jadad score the proper evaluation of trials? *J Rheumatol*. 2006;33:1710–1711.
5. Berger VW. Making statistics boring again. *Stat Med*. 2010;29:1458.
6. Berger VW, Rezvani A, Makarewicz V. Direct effect on validity of response run-in selection in clinical trials. *Control Clin Trials*. 2003;24:156–166.
7. Berger VW. *Selection Bias and Covariate Imbalances in Randomized Clinical Trials*. Chichester, West Sussex, United Kingdom. John Wiley & Sons;2005:1–205.
8. Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJ, Gavaghan DJ, McQuay HJ. Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Control Clin Trials*. 1996;17:1–12.