Editor's Spotlight/Take 5

Editor's Spotlight/Take 5: Small Improvements in Mechanical Axis Alignment Achieved With MRI versus CT-based Patient-specific Instruments in TKA: A Randomized Clinical Trial

Seth S. Leopold MD

Even if you have not mixed methacrylate cement since residency, some elements of the randomized trial by Pfitzner and colleagues, and the interview that follows this spotlight, should influence how you think about new technology in your practice.

The study evaluated the performance of patient-specific cutting blocks in TKA (custom made based either on preoperative MRI or CT images) against that of conventional cutting guides. If this technique has barely penetrated the subspecialty of

arthroplasty, why should someone who does not “do knees” be interested?

Several reasons. First, it raises important issues about the adoption of new, expensive technology. The authors found the newer guiding systems demonstrated better alignment and faster operative times, yet they do not recommend the widespread use of these tools; their reasons for this modesty have broad applicability across our specialty and are worthy of our respect and emulation. Second, the study points us to the important differences between what we can measure

easily and what we really want to know, which is a theme we see repeated in all fields and merits discussion. Finally, it is a well done randomized trial. If you are reading this, some part of you must appreciate an elegant experiment for its own sake, almost regardless of the topic.

The paper was a multicenter collaboration among Charité–Universitätsmedizin Berlin, the Mayo Clinic in Rochester, MN, USA, and Hospital Maerkisch Oderland in Wriezen, Germany. The authors found CT- and MRI-based patient-specific guides provided improvements in almost every measurable TKA alignment parameter. But most of the observed alignment differences were quite small. In fact, the differences were so small that the numbers probably would not have made the typical surgeon take note, even before the earlier Mayo study [1] that caused a buzz about whether alignment even matters. And alignment in the control group missed the mark by more than one might have expected in a high-volume setting, tending to make the intervention group

Note from the Editor-In-Chief: In “Editor's Spotlight,” one of our editors provides brief commentary on a paper we believe is especially important and worthy of general interest. Following the explanation of our choice, we present “Take Five,” in which the editor goes behind the discovery with a one-on-one interview with an author of the article featured in “Editor's Spotlight.”

The authors certify that they, or any members of their immediate family, have no 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 comment refers to the article available at: [10.1007/s11999-014-3784-6](https://doi.org/10.1007/s11999-014-3784-6).

S. S. Leopold MD (✉)
Clinical Orthopaedics and Related
Research, 1600 Spruce Street,
Philadelphia, PA 19013, USA
e-mail: sleopold@clinOrthop.org

Editor's Spotlight/Take 5

look better. But even with that, the interventions themselves only improved things by 1° to 3° or so. Yes, the p values were tiny, so the results were “significant,” but will patients perceive such small alignment differences? Will any benefits there offset the increased costs of the new technology, or the risks associated with it? The authors, in a move we do not see every day, raise many of these questions themselves.

Patients do not ask whether we can get them a mechanical axis of 2° instead of 3.5°, just as they do not ask whether some new tool for carpal tunnel surgery can help us finish the operation in 11 minutes instead of 14. But operative time – like tibial alignment – is easy to measure, report, and compare statistically. While the newer technique studied by Pfitzner and colleagues saved about 15 minutes on a 75-minute operation, it added preoperative time, cost, and an additional imaging study, and the planned component was not even usable in 17% to 30% of the operations in the custom groups. Looking at the p values alone would give a different impression than actually looking at the data and the sizes of the observed effects. What we really want to know is whether patients notice implant-performance differences in knees inserted using one or another of these approaches; the answer appears to be no, at least in the

short term. We also want to know whether implants inserted this way will last longer; obviously this one will take decades to answer, and there is no evidence now to think this will be the case.

The trialists did the work to get this right. Randomization was properly done – they had not one, but two treatment groups of interest. The authors also performed a rigorous assessment of their endpoints. Their self-assessment of the meaning of their results was nuanced, focusing appropriately on effect size rather than “statistical significance.” For that, more than anything else, they earn my compliments, and they should get your attention.

But most importantly, if you are not a knee surgeon, retain its take-home message: If the p value in a study is not low enough to be convincing, then any “differences” may well have been the result of chance alone. But, as in Pfitzner et al., when the p value is small, be sure also to look at the data. Are the authors measuring endpoints that matter? And if so, is the effect size large enough to pay for? Remember, “paying” may be something done in dollars (direct costs of implementing new technology), in errors (surmounting the learning curve of a new technology), or in uncertainty (needing to use a particular implant whose performance over the longer run is

unknown). Effect sizes on endpoints that matter, rather than p values, should drive our therapeutic decisions.

Please join me as I go behind the discovery in the Take 5 Interview that follows with Matthew P. Abdel MD, the corresponding author of this paper, and talk about adopting new technologies, evaluating clinical research, and this paper itself.

Take Five Interview with Matthew P. Abdel MD, corresponding author of “Small Improvements in Mechanical Axis Alignment Achieved With MRI versus CT-based Patient-specific Instruments in TKA: A Randomized Clinical Trial”

Seth S. Leopold MD: *Congratulations on a well-done randomized trial. Before getting into your study itself, I think readers might benefit from your expertise as a reader of this kind of work. As an accomplished clinician-scientist, when you read a paper about a new technology, what elements of a well-designed trial do you focus in on to see whether the authors have convinced you a technology is worth adopting?*

Matthew P. Abdel MD: Thank you for the kind comments, Dr. Leopold, and for highlighting our randomized trial. When I read an article about new technology, there are three main areas I focus on to determine if it is worth adopting.

Editor's Spotlight/Take 5

Matthew P. Abdel MD

Foremost, the technology must directly improve patient care. While that can be short-term outcomes such as length of stay or blood loss, it also includes longer-term results such as functional gains and enhanced survivorship. Second, the technology must be cost-effective. In 2014, technologies that only offer incremental benefits to patients, but at a large cost both to patients and the healthcare infrastructure, are not viable. Finally, the technology must be applicable in practice. New technologies that require intensive specialized training and substantial additional equipment only available at a handful of large institutions may be impractical for efficient patient care.

Dr. Leopold: *More specifically, how should the reader evaluate effect size in contrast to p value?*

Dr. Abdel: Great question. There often is confusion between the meanings of p value and effect size. In reality, both are essential in studies such as ours. Foremost, it is important to remember that there are two major types of errors that can occur, statistically speaking. Type I (α) error is the rejection of the null hypothesis when it is true. Type II (β) error is the failure to find a difference when indeed one exists. In addition, there is a third issue, which you raised in your commentary: That of effect size. With that in mind, again, p value is simply the estimated probability of falsely rejecting the null hypothesis, or mistakenly concluding that a difference is present when in truth there is no such difference. In most orthopedic investigations, the α -error rate is set at 0.05, meaning they accept a 1 in 20 chance that this might occur. On the other hand, effect size is a quantitative measure of how large the treatment effect is. In essence, it is the strength of an observed phenomenon. As an easy example, consider an imaginary study on length of stay. This study might point to a statistically significant difference in the length of a hospital stay between two interventions, and offer p values below 0.05 to show that the observed difference is not likely related to chance. In a large

enough study, very small differences – say, something on the order of 0.1 days – might be detectable as significant at that level. But is a 2-hour difference in a hospital stay clinically relevant for a patient who is in the hospital for several days? While the p value tells one story, the effect size tells quite another.

Dr. Leopold: *I thought you and your colleagues displayed tremendous integrity in this work – your p values, interpreted superficially, might have been construed as making a case for widespread use of a new technology. Yet, you stopped well short of recommending that, and in fact recommended against dissemination of the new approach until it is validated by further study. How did those conversations play out in your group as you were crafting the message of this report?*

Dr. Abdel: We were very fortunate in this international group to have investigators who are all dedicated to the science of orthopaedic surgery, and the accurate dissemination of information, particularly in its direct relation to patient care. Under the leadership of Dr. Pfitzner, we were all aware of the “statistically significant” findings highlighted in the Results section. However, the key to this investigation was in the Discussion section. This is where we had to avoid the temptation

Editor's Spotlight/Take 5

to recommend patient-specific instrumentation just because there was a statistically significant difference. In fact, as you pointed out, we agreed that the differences were minimal, and likely not large enough to be clinically important.

Dr. Leopold: *It seems to me that the questions of interest for this or any other new technology might involve longer-term performance and durability. Yet, those are fiendishly difficult to answer, and if those were the only standards, new technologies might only rarely be tested and never introduced outside of study protocols. Would that be a good thing or a bad thing for patients, and why?*

Dr. Abdel: Innovation in orthopaedic surgery is essential for the advancement of patient care. However, it must have the potential to help our patients,

be based on previous literature, and undergo rigorous in vitro investigation prior to widespread dissemination. As a specialty, we have not always been as careful as we could have been in this regard; dual-modular femoral necks for THA, metal-on-metal bearings, and recombinant human BMPs are several examples that come to mind. In addition, with any new technology, we must promote and analyze joint replacement registries to identify early failures and analyze long-term successes.

Dr. Leopold: *One specific question about the technology you actually tested: Patient-specific guides. As I am sure you will agree, we cannot draw inferences about long-term success from your study; perhaps the small alignment differences will prove important in terms of durability, some*

years from now. So, look 5 years down the road with me: What kinds of studies will help us to know whether patient-specific guides are worth the costs and risks associated with it?

Dr. Abdel: Another great question. As you astutely pointed out, our study does not speak to the mid- or long-term reliability or durability of patient-specific instrumentation. In my opinion, before we can design studies that focus on hitting that target, future studies must first focus on what the “target” should be for individual patients.

Reference

1. Parratte S, Pagnano MW, Trousdale RT, Berry DJ. Effect of postoperative mechanical axis alignment on the fifteen-year survival of modern, cemented total knee replacements. *J Bone Joint Surg Am.* 2010;92:2143–9.