



Not the Last Word

Not the Last Word: Ending The Residency Application Arms Race—Starting with the USMLE

Joseph Bernstein MD

Two campers are sitting around the fire. Off in the distance, the sound of a large bear trampling through the woods is heard. One of the campers quickly removes his boots, and begins to lace up a pair of running shoes. The other camper looks at him in disbelief and says, “What are you doing wasting your time? You can’t outrun a bear!” The first camper continues to put on his shoes, and replies, “I

don’t have to outrun the bear—I only have to outrun you.”

That’s the nature of most competitions. One does not have to beat some objective standard, but merely outpace the other contestants. This concept is seen in athletic events like road races, of course, but it also pertains to “competitions” of everyday life. The nicest house in town can be owned by a man of ordinary wealth, as long as he is richer than all his neighbors. And so on.

Earning a residency spot in orthopaedic surgery is a lot like an untimed road race. Each year, there are about 2000 students vying for 700 spots. Earning a position is not assured by mastering a given amount of material, demonstrating a specific set of skills, or attaining a prescribed level of merit. Rather, it is simply a matter of crossing

the finish line of the “residency desirability race” among the top 700, ahead of the 1300 also-rans who must find their happiness in other specialties.

The residency desirability race has at least two elements that are less like a road race and more like an arms race. An arms race—for those too young to remember—is a competition where countries strived to amass a greater stockpile of weapons, not so much to use them, but simply to have more than the other.

The first arms-race is the quest to perform more and more visiting rotations. Just as two atomic bombs are enough if the enemy has only one, performing two visiting rotations confers an advantage to an applicant when everybody else is performing one or none. On the other hand, if the mean number of visiting rotations is three, then it takes four to benefit.

A second arms-race results from the emphasis programs place on the United States Medical Licensing Exam (USMLE). Here, too, the only standard is attaining a higher score than the other applicants. In my experience as a faculty advisor over the last 10 years, I have seen the de facto passing score

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 section 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 he, or any 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 *Clinical Orthopaedics and Related Research®* or The Association of Bone and Joint Surgeons®.

J. Bernstein MD (✉)
Department of Orthopaedic Surgery,
University of Pennsylvania, 424
Stemmler Hall, Philadelphia, PA 19104,
USA
e-mail: orthodoc@uphs.upenn.edu

Not the Last Word

migrate from south of 230 to north of 240, though the students seem no smarter.

Everybody would be better off, as I have argued elsewhere [2], if residency positions were just sold at auction directly, rather than through this proxy system. At least with an auction, the seller receives the selling price. In the visiting rotations arms race, most of the money is spent on Amtrak, airfare, and accommodations. As for the USMLE, the time dedicated to memorizing the Krebs Cycle, the ratio of FSH to LH throughout the menstrual cycle, or the differences between schizoid and schizotypal personality disorders is just wasted, as this information has scant relevance to orthopaedic surgery.

It's fair to say, however, a residency auction is not coming any time soon. Even students who would benefit from an explicit market for residency positions find the notion repugnant [7]: some things, they tell me, just should not be for sale. Beyond that, it would be impractical to implement an auction. If nothing else, the coordination required might not withstand legal scrutiny [3]. Thus, efforts should be applied at fixing the problems at hand.

I'll revisit the possible solutions to the visiting rotation problem in a later column ... if I can think of any. I do, however, have an idea for reducing the undue emphasis placed on the USMLE

by orthopaedic surgery residency program directors.

The USMLE is favored by program directors because it provides objective data that facilitate the evaluation of applicants. Grading schemes vary from school to schooling, and of course the schools themselves differ in important ways. Further, some distinctions suggested by letters of recommendation are almost meaningless. Sure, a student touted as "magnificent" is better than one described "eukaryotic and possibly diploid" [8], but the middle can be a muddle. Only to those in the know is "outstanding" better than "excellent", for example. Compared to metrics like grades, school pedigree, and letters of reference, USMLE scores are starkly clear. There is simply no doubt that a 242 on the examination is higher than a 241—not necessarily "better", but "higher" it certainly is.

The need for an objective residency selection is based on the numbers. Students apply to dozens of programs, and in turn, programs receive dozens of applications for every interview slot they can offer. A screening criterion that can be applied by nonexperts—or, better still, by a computer algorithm—is accordingly appealing. Also, a higher score on any multiple choice test typically reflects better test-taking abilities. That is a desirable attribute, as residency programs themselves are evaluated on the rate with which their

graduates pass the American Board of Orthopedic Surgery Part 1 exam, itself a multiple choice test [1].

The basic problem with the USMLE examination is that one can study for it. Ordinarily, that a test rewards studying is a feature, not a bug—but not here. Because studying can be effective, an applicant facing this high-stakes examination would be wise to study more. Yet that engenders an arms race. With no natural limit to how much time should be spent memorizing esoterica that may nudge one's score up and point or two, larger and larger blocks of time are wasted. (At my institution, students routinely devote 2 months for study—more than double what I recall a decade ago—simply to keep up with the aforementioned 10-point escalation in the "passing" score).

For better or worse, the USMLE can't be buried; it can only be replaced. The benefits to programs from having an objective and easily-compared standardized test are so great that USMLEs will be used unless and until something better comes along—an objective and easily-compared standardized test that assesses things relevant to orthopaedics, not the minutiae of glucose metabolism, obstetrics, or psychiatry.

Basing residency selection on a relevant test will be a boon to the applicants, but also to the programs

Not the Last Word

themselves. In particular, if an examination were to be constructed that concentrated on musculoskeletal anatomy, bone and cartilage physiology, and other germane basic sciences, medical school graduates entering residency will have mastered these topics. In turn, residency programs can shed their obligation to teach these subjects in the limited time they have, liberating time and resources for more clinical education.

The test I propose might be constructed under the auspices of the National Board of Medical Examiners (the creator of the USMLE, but also, interestingly enough, a participant in writing the American Board of Orthopaedic surgery certification exam [9]). There is a mechanism in place for the National Board of Medical Examiners to create, and medical schools to offer, subject examinations in many areas [5]. Indeed, such a musculoskeletal medicine exam was written years ago, and may be a worthy foundation on which a new test can be built. The missing ingredient is not the question bank itself, but rather the assurance that a score will have currency among applicants and their evaluators.

Replacing the USMLE with a more on-point examination will take effort, but this effort will be rewarded across the board. Applicants will waste less time and programs will recruit better-prepared candidates. The alternative is

perpetuating the arms-race, and an arms race—for those too young to remember—is not guaranteed to end happily.

Joseph Zuckerman MD

Professor and Chairman, Department of Orthopaedic Surgery

New York University Langone Medical Center

I appreciate the opportunity to comment on Dr. Bernstein's provocative piece that challenges the value of the USMLE for selecting orthopaedic surgery residents. Dr. Bernstein has been a leader in orthopaedic education at all levels for many years and I applaud his efforts to bring attention to the overemphasis placed on the USMLE scores. His analogy to an arms race is well understood by this writer who can still remember "air raid drills" in elementary school in the 1960s, which were conducted in preparation for the possibility of nuclear attack. Fortunately, the arms race Dr. Bernstein describes is far less dangerous.

The thrust of Dr. Bernstein's opinion is that the undue emphasis on the USMLE can be fixed by replacing it with "an objective and easily compared standardized test that assesses things relevant to orthopaedics, not the minutiae of glucose metabolism, obstetrics or psychiatry." He further proposes that this exam could be

similar to the "shelf exams" frequently utilized at the end of required clerkships like pediatrics, internal medicine, and obstetrics and gynecology.

I certainly agree with Dr. Bernstein about the excessive importance placed on USMLE scores. However, the answer is not a new standardized exam. The problem is not the exam; the problem is how it is used. I have often heard program directors state, "we do not interview any applicants with a USMLE Step 1 score below 240". I see this approach as simply a short cut—a way to avoid reviewing the hundreds of applications. We all recognize that the USMLE score is certainly not a reliable predictor of success as an orthopaedic resident. Rather, it is one of many factors that should be utilized in selecting our residents.

Although I do agree with Dr. Bernstein that letters of recommendation can be less than helpful and often misleading, I do not agree that they are "almost meaningless." They are also part of the multifactorial equation that is designed to help us select residents.

For the past 7 years, our program has used a scoring system for each applicant that includes different components of the application including USMLE Step 1 and 2 scores, AOA and Gold Humanism honor society membership (see <http://www.gold-foundation.org/ghhs/> for more information about the Arnold P. Gold foundation, which does

Not the Last Word

wonderful philanthropic work around the world), Dean's letters, grades in required clerkships, faculty letters of recommendation, research productivity, extra-curricular activities/life experiences, and a personal statement. Each application is reviewed by two members of the selection committee. The scores are totaled and we select our applicants to interview based upon the total scores beginning with the highest scores. Recognizing that a score of "44" may not be much different than a score of "42," when we get to the last 10 interview spots, our Program Director re-reviews the next 20 or 30 applications to determine the final selections. Is this system perfect? Of course not; is it better than using a single test score as a "do not pass go" factor? Absolutely. It gives each applicant the opportunity to have his or her application reviewed. And, as importantly, it provides an opportunity to have a USLME score of 220 balanced by an impressive Dean's letter, a life experience that speaks strongly about the individual, or outstanding research credentials.

The answer to the USMLE arms race is not to replace it with another exam that focuses on orthopaedic knowledge. Students already are specializing too early. Instead of going to medical school to become a physician first and an orthopaedic surgeon second, many seem to bypass the physician part. Dr. Bernstein's "solution" could make this

even worse. Program Directors and Department Chairs need to take a step back and reexamine the selection process. There are no short cuts to selecting residents for our programs. It is a challenging process in which we continue to refine our approach in an effort to select residents who will be successful, skilled, and caring orthopaedic surgeons. In fact, recently we have utilized personality assessments for all of the students we interview as yet another tool to add to the equation.

As students become smarter and more accomplished, the impact of a multifactorial evaluation becomes even more important. The students we select will be the residents we train and, ultimately, the orthopaedic surgeons who provide care. It is our responsibility to devote the time and effort needed to select the most qualified applicants. The way to accomplish this requires more than another standardized exam.

S. Elizabeth Ames MD

Program Director, Department of Orthopaedics & Rehabilitation, University of Vermont

Chair, Council of Orthopaedic Residency Directors

It is time to face the uncomfortable work of creating a better system

to select the orthopaedic surgeons of the future. The required skills, both in education and clinical practice, have evolved. The current population of orthopaedic residents is changing. We need to identify the candidates best suited to succeed in the new world, and our current tools do not do that. We need to define the parameters we consider important to measure, measure those parameters accurately, and select candidates based on the results of this process in order to ensure the success of our specialty and the patients we treat. Currently, we have no such processes in place, and creating one is challenging because of changes in both the structure of our current residencies and in the candidates themselves.

What has changed? Meeting educational mandates makes it nearly impossible to manage both a clinical practice and a directorship—the appeal of easily available metrics is clear when there isn't time to thoroughly read every application. Program directors report satisfaction with available resident performance metrics, but currently more time is consumed addressing resident satisfaction and the remediation of a struggling resident than on improving a selection process that badly needs our attention. The ability to prepare for an USMLE exam can result in outstanding scores but does not reflect the individual's daily abilities, technical

Not the Last Word

acumen, or work ethic. We need to decide: Should we continue to use old standards, or take the time to develop the necessary tools or new measurements?

I propose that we choose the latter for several reasons.

First, our curriculum is broadening. We cover an expanding technical skill set and musculoskeletal knowledge base in 60 months of education, despite being limited by mandates like training hours. Interpersonal skills, teamwork, and leadership acumen are increasingly important and we are working to integrate them into residency training.

Second, we have a distinct lack of diversity; we do not reflect the general medical school population in terms of gender or race. It is fair to ask whether we will continue to inspire the strongest medical students of the next decade.

Finally, there is the generational divide—our current residents have a different vision of what a quality life looks like compared with their predecessors or their faculty.

The decision to act requires us to investigate the ways in which these factors affect our residencies and the application process. It would be easier to maintain the status quo. We could accept the argument that we are doing well enough using the available metrics. In fact, demand is so high that a

program can offer interviews to 10% to 15% of applicants and still create a match list with quality and depth. There is no question, though, that each of us would prefer to match those individuals who will be outstanding rather than just solid performers. At a program level, we have an arms race of our own. We want those that will excel every day in their self-education, technical skills, and patient care—not just skills measured by examinations. It may not be worth it to us to merely define a “qualified” applicant yet. I believe it is worth it to at least start with defining the exceptional or fully-qualified individual. What we learn will inform the future.

William C. McGaghie PhD

Professor, Medical Education and Preventive Medicine

Northwestern University Feinberg School of Medicine

Dr. Bernstein is right. We should not continue the practice of using USMLE Step 1 and 2 scores as screening metrics for postgraduate residency selection decisions in orthopaedic surgery or any other medical specialty. This is not a new idea. Medical education scholars including myself [4], and leading academic physicians tied to the National Board

of Medical Examiners [6] have pointed out for decades that USMLE Step 1 and 2 test scores are intended to inform medical licensure decisions, not decisions about readiness for postgraduate medical education. These decisions are quite different.

The controversy about using USMLE Step 1 and 2 scores for postgraduate residency screening and selection decisions has two sources: (1) Academic and professional discourse about the valid use of test scores to make accurate decisions about medical persons in specific situations, under known conditions, that have real consequences; and (2) the convenience and safety of residency program directors. The first source of controversy addresses scholarly, evidence-based medical education. The second source addresses administrative expedience.

To illustrate, using test scores to make accurate decisions about doctors at all levels is not easy. Five years ago, my colleagues and I reported a research synthesis [4] and argued that, “The validity argument about using USMLE Step 1 and 2 scores for postgraduate residency selection decisions is neither structured, coherent, nor evidence-based.” We concluded that, “Continued use of USMLE Step 1 and 2 scores for postgraduate medical residency decisions is discouraged” [4].

Not the Last Word

Dr. Jeffrey G. Wong, senior associate dean for medical education and professor of internal medicine at the Medical University of South Carolina endorsed this conclusion in an editorial. Wong argued, “It may force [residency programs] to identify those traits, skill sets, and attitudes that best predict excellence in their particular specialties rather than simply focusing on a number” [10].

Residency program directors continue to rely on USMLE Step 1 and 2 scores for resident screening because the scores provide a standard metric to rank residency candidates who graduate from different medical schools. The scores are available at no cost and have the aura of scientific objectivity, even though the tests were designed for a purpose completely different from residency selection. In addition, resident acceptance can ride on a difference of only several score points despite knowledge of measurement error in all test scores. These may be differences between candidates without distinctions.

Dr. Bernstein is also right when he calls for “an objective and easily compared standardized test that

assesses things relevant to orthopaedics [or any other medical specialty], not the minutiae of glucose metabolism, obstetrics, or psychiatry.” Other leading academic physicians have argued, “Like the use of an off-label drug for purposes that have not been well studied, it is ill advised to use USMLE examination scores for a purpose for which the test was not developed and has not been directly validated” [6].

Residency program directors in orthopaedics and other medical specialties should heed this sage advice.

References

1. Bernstein J. Harvard beats Yale and other fallacies. *Clin Orthop Relat Res.* 2013;471:3745–3749.
2. Bernstein J. Sell those residency slots. Available at: <http://www.aaos.org/AAOSNow/2012/Feb/clinical/clinical5/?ssopc=1>. Accessed August 31, 2016.
3. DePalma A. Ivy universities deny price fixing but agree to avoid it in the future. Available at: <http://www.nytimes.com/1991/05/23/us/ivy-universities-deny-price-fixing-but-agree-to-avoid-it-in-the-future.html>. Accessed August 31, 2016.
4. McGaghie WC, Cohen ER, Wayne DB. Are United States Medical Licensing Exam Step 1 and 2 scores valid measures for postgraduate medical residency selection decisions? *Acad Med.* 2011;86:48–52.
5. National Board of Medical Examiners. NBME subject examinations. Available at: <http://www.nbme.org/students/Subject-Exams/subexams.html>. Accessed September 26, 2016.
6. Prober CG, Kolars JC, First LR, Melnick DE. A plea to reassess the role of United States Medical Licensing Examination Step 1 scores in residency selection. *Acad Med.* 2016;91:12–15.
7. Roth AE. Repugnance as a constraint on markets. *J Econ Perspect.* 2007;21:37–58.
8. Schneiderman H. Le mot juste. *JAMA.* 1988;259:87.
9. The American Board of Orthopaedic Surgery. Overview of board certification. Available at: <https://www.abos.org/certification/board-certification.aspx>. Accessed August 31, 2016.
10. Wong JG. To the editor. *Acad Med.* 2011;86:793–794.