Do Scores of the USMLE Step 1 and OITE Correlate with the ABOS Part I Certifying Examination?

A Multicenter Study

Paul J. Dougherty MD, Norman Walter MD,
Peter Schilling MD, Soheil Najibi MD, PhD,
Harry Herkowitz MD

Received: 6 September 2009 / Accepted: 16 March 2010 / Published online: 30 March 2010
© The Association of Bone and Joint Surgeons® 2010

Abstract

Background The US Medical Licensing Examination (USMLE) and Orthopaedic In-Training Examination (OITE) are commonly used to select medical students or residents, respectively. Knowing how well these examinations predict performance on the American Board of Orthopaedic Surgery (ABOS) Part I certifying examination is important to provide evaluations for medical students and residents. Previous studies comparing the OITE scores with the ABOS Part I scores have been limited to one program.

Questions/purposes Therefore, we compared the scores on the USMLE Step 1 and OITE examinations with those on the ABOS Part I certifying examination using data from four ACGME approved residency programs.

One or more of the authors (HH and NW) has received funding from Stryker, Zimmer, and Biomet corporations.

Each author certifies that his or her institution has approved the human protocol for this investigation and that all investigations were conducted in conformity with ethical principles of research.

P. J. Dougherty (✉), P. Schilling
Department of Orthopaedic Surgery, University of Michigan,
1500 E Medical Center Drive, Ann Arbor, MI 48109, USA
e-mail: pauldoug@med.umich.edu

N. Walter
McLaren Regional Medical Center, 401 South Ballenger
Highway, Flint, MI 48532, USA

S. Najibi
Henry Ford Hospital, 2799 West Grand Blvd, Detroit, MI 48202,
USA

H. Herkowitz
William Beaumont Hospital, 3535 W. Thirteen Mile Road,
Royal Oak, MI 48073, USA

Methods We reviewed 202 resident files from 1996 to 2008 from four programs in the same geographic region. Of those, 181 (90%) had complete records. De-identified data were used to compare USMLE Step 1 scores, OITE percentile rank scores, and ABOS Part I percentile rank scores. Pearson coefficients and receiver operator curves were calculated to assess the relationships between tests.

Results We found a correlation of 0.53 between the USMLE Step 1 and ABOS Part I, and an average correlation of 0.50 for postgraduate years (PGY) 2 through 5 OITE scores and ABOS Part I. There was a stepwise increase in correlation from PGY 2 through PGY 5 between the OITE scores and ABOS scores. Those who averaged in the 27th percentile or lower on the OITE had a 57% chance of failing the ABOS Part I examination.

Conclusion USMLE Step 1 scores correlated with ABOS Part I certifying examination scores, and we therefore believe it may be used as one factor in resident selection. Use of the OITE scores in guiding education and feedback appears to be justified.

Introduction

A major goal of all orthopaedic surgery residency programs is to ensure their graduating residents are capable of becoming board-certified by passing the ABOS certifying examinations [1–6, 8–17]. A board certification rate greater than 75% is required to maintain Residency Review Committee Accreditation.

Residency programs commonly use scores on two examinations to predict or assess resident performance: the USMLE Step 1 and the OITE [2–6, 8, 10, 13–16]. Residency programs have an interest in determining how well performance on these examinations correlates with resident

performance and, ultimately, the ABOS Part I certifying examination. Medical students take the USMLE Step 1 after the second year of medical school. Correlating the USMLE Step 1 and OITE scores with passing the ABOS Part I is important for program directors to help guide resident selection [5–8, 12–14, 16–18] and resident education [13–15].

Correlation of USMLE Step 1 scores and the ABOS examination is not clear. Case and Swanson [5], Herndon et al. [13], and Swanson et al. [16] all observed a correlation of the USMLE and ABOS Part I examinations. Klein et al., however, found low correlation in evaluating one program [14]. Herndon et al. [13], Swanson et al. [16], and Klein et al. [14] all reported varying degrees of correlation at their respective programs with the OITE percentile ranking scores and ABOS Part I Examination results.

As part of the outcomes project for the Accreditation Council for Graduate Medical Education (ACGME), it is important to understand how accurate the OITE is for assessment of residents. To better define the limitations of the OITE in correlation with ABOS Part I results, more studies, with perhaps differing methods, are needed to aid in this goal.

Therefore, we examined (1) the relationship between the USMLE Step 1 scores and the ABOS Part I percentile ranking scores; (2) whether there is an association between certain threshold USMLE Step 1 scores and passing the ABOS Part I examination; (3) the relationship between OITE scores (percentile rankings for each year an individual takes the examination) and ABOS Part I percentile rank scores; and (4) the mean OITE scores for PGYs 2–5) and ABOS Part I percentile rank scores.

Materials and Methods

We pooled and analyzed the scores from the USMLE Step 1, OITE (PGY 2 through PGY 5), and the first attempt at the ABOS Part I examination for graduates from four residency programs in one geographic area from 1996 to 2009. For the USMLE we used the raw scores. For the OITE and ABOS Part I examination results, we used the percentile rankings for our calculations. Four orthopaedic surgery programs contributed the resident data from 1996 to 2008. Two programs have six residents per year each. One program has two residents per year. The fourth program expanded from three to five residents per year during the study period. Of the four programs, one is university based, two are academic nonuniversity based, and the fourth is a community program. All programs were ACGME accredited during the period that the records were studied. The total number of potential resident scores available was 202. Twenty-three residents had incomplete

scores for the USMLE Step 1, and between 17 (PGY 2) and six (PGY 5) residents had incomplete OITE scores (Table 1). For each category with incomplete data, we looked at mean ABOS percentile and the number of failures. For the USMLE alone, the mean ABOS percentile was 55, similar to the mean of 57 for all 202 reported ABOS Part I examination scores (Table 1). The PGYs 2–5 average score (Fig. 1) is an average across all years, regardless of the missing values for one year.

The study was approved by the Institutional Review Boards at the authors' institutions.

The USMLE Step 1 has been a computer-based examination since 1999; it is administered at Prometric centers daily throughout the year. Test questions are prepared by examination committees composed of basic scientists and

Table 1. USMLE Step 1, OITE, and ABOS Part I results

Examination	Number of residents	Mean score	Standard deviation	Range of scores
USMLE	179	227.5	15.1	182–227
OITE PGY 2	185	54.2	22.7	1–99
OITE PGY 3	187	59.2	23.11	1–99
OITE PGY 4	195	62.3	23.7	1–99
OITE PGY 5	196	66.5	21.9	3–100
OITE (mean PGY 2 to PGY 5)	202	61.1	19.3	2.75–99.25
ABOS Part I	202	57.0	25.9	0–99

USMLE Step 1 data are raw scores; OITE and ABOS Part I data are percentile rankings.

Fig. 1 A ROC curve comparing the mean percentile rankings of the OITE examination scores for PGY 2 through PGY 5 with the ABOS Part I certifying examination scores is shown. If the mean OITE scores were in the 27th percentile or less, 57% failed the ABOS examination. If the scores were in the 37th percentile or less, the failure rate was 32%.

clinicians who are faculty members, teachers, and investigators with recognized prominence in their respective fields. Committee members are selected to provide broad representation from the academic, practice, and licensing communities across the United States and Canada. The test is designed to measure application of basic science knowledge.

The OITE is a proctored paper and pencil examination given annually on the second Saturday of each November. The test is given in only one sitting and the materials must be gathered and submitted for scoring within 2 days of taking the examination. Questions used for this examination are submitted by a group of orthopaedic surgeons from different subspecialties; the OITE examination committee selects the questions to be included on the actual examination [3, 5, 13–16].

The ABOS Part I examination is a proctored paper and pencil examination given annually in July (beginning in July 2009, the ABOS Part I became a 1-day computer-based examination administered at Prometric test centers). All test questions included on the Part I examination have been through a rigorous review process, including multiple reviews and pilot testing by board-certified orthopaedic surgeons. The test is built to an established blueprint that specifies the exact numbers of items to be included in each major content area and statistical criteria to be used. Items selected for inclusion in the test must be appropriate in terms of content and psychometrics (ability to test the area that pertains to the question).

During the period of review, the National Board of Medical Examiners (NBME) examination changed from part examinations to USMLE step examinations. We made a simple conversion to convert the NBME part scores to comparable USMLE step scores [5, 14]. There is a simple linear relationship between the USMLE and NBME scores as follows: $USMLE = 200 + \{[NBME - 500]/5\}$. We determined correlation coefficients between the OITE scores and the ABOS Part I percentiles using Microsoft Excel software (Microsoft, Redmond, WA). Similarly, the USMLE Step 1 scores were compared with the ABOS Part I scores.

Scatterplots were made to visually inspect the relationship between ABOS Part I and Step 1 percentile and scores on previous standardized tests. The Pearson correlation coefficient was used to measure the degree of linear correlation among ABOS Part I percentiles, USMLE Step 1 scores, and each of the OITE percentiles, including PGYs 2, 3, 4, and 5 examinations and trainees' mean OITE percentile at PGY 5.

In our analysis, we used receiver-operated characteristic curves (ROC) to identify threshold scores and percentiles on the USMLE and OITE, respectively, that maximized the identification of those destined to fail ABOS Part I (true positives), yet minimized the identification of those who

went on to pass (false positives). ROC curves are graphic representations of this concept. ROC curves are, in essence, plots of right versus wrong answers as a detection threshold or the criterion is varied across its range. Comparison of the ABOS Part I and OITE percentile rankings were calculated at various thresholds for passing and calculated as a ROC. The results were tabulated looking at particular points on the curve and looking for differences between a true-positive, meaning at a particular point a person was receiving the predicted outcome based on the OITE percentile ranking scores for the ABOS Part I examination.

The conceptual framework for the remainder of our analysis treated the USMLE Step 1 scores and OITE percentile rankings as screening tests designed to identify trainees at risk for ABOS Part I examination failure. As such, we characterized the ability of these tests to accurately predict failure on ABOS Part I by using ROC curves, which are graphic depictions of a screening test's sensitivity (1-specificity) for detection of an outcome (ie, board examination failure) because its discrimination threshold (USMLE score or OITE percentile) is varied. OITE percentile rank scores for PGY 5 were compared with the ABOS Part I examination regarding passing. We visually inspected the ROC curves for each examination to identify threshold scores and percentiles that maximized the identification of those destined to fail the ABOS Part I (true-positives), yet minimized the identification of those who went on to pass (false-positives). Finally, we calculated the relative risk of ABOS Part I failure with its 95% confidence interval for each threshold Step 1 score and OITE percentile ranking (average for PGYs 2–5 and PGY 5 separately). Calculations were made using SAS software (Version 9.2; SAS Institute, Cary, NC).

Results

We found a wide range of scores for all three examinations (Table 1). Twelve of the 202 residents (5.9%) failed the ABOS Part I certifying examination.

We found a correlation between the USMLE Step 1 scores and the ABOS Part I percentile rank scores. In comparing USMLE Step 1 scores with ABOS Part I scores (Fig. 2), 33% of those who scored 209 or less on Step 1 failed the ABOS Part I. For a threshold of 223 or less, 16% failed the ABOS Part I. All candidates who scored above 227 on the Step 1 passed the ABOS Part I examination.

There was a stepwise increase in correlation between the OITE percentile ranking scores from PGY 2 to PGY 5 and the ABOS Part I scores (Table 2). For the residents who failed the ABOS Part I, 10 of 12 had average (PGYs 2 to 5) OITE percentile ranking scores of 28 or less for those years (Fig. 1).

Fig. 2 A ROC curve comparing USMLE scores with ABOS Part I certifying examination scores is shown. For residents scoring 209 or less on the USMLE, 33% failed the ABOS examination.

Table 2. Correlations of USMLE Step 1 and percentile rankings of OITE scores with percentile rankings of the ABOS score

Examination percentile ranking	Correlation coefficient
USMLE	0.53
OITE 1 percentile rank	0.39
OITE 2 percentile rank	0.42
OITE 3 percentile rank	0.47
OITE 4 percentile rank	0.52
Mean OITE percentile rank	0.50
ABOS percentile rank	1.00

All coefficients $p < 0.0001$.

Fifty-four percent of the PGY 5 trainees who were below the 22nd percentile on the OITE failed the examination (relative risk, 19.5; CI, 7.2–53.0) (Fig. 3). Similarly, for the threshold at or below the 40th percentile, 32% failed the examination (relative risk, 17.8; CI, 5.1–61.7). Of those who were at or above the 71st percentile, all passed the ABOS Part I examination. OITE percentile ranking scores below the 30th percentile year-in-training were found in 10 of 12 of those who failed the ABOS Part I certifying examination.

Discussion

The ability to predict how well a medical student will do on the ABOS Part I certifying examination using the USMLE Step 1 scores is important for orthopaedic surgery residency selection committees [2, 5, 13, 14, 16]. Additionally, the usefulness of the OITE individual percentile rank scores for predicting outcome on the ABOS Part I

Fig. 3 A ROC curve comparing the OITE percentile ranking of PGY 5 residents with the ABOS Part I certifying examination scores is shown. If the scores were at or below the 22nd percentile, 54% of the candidates failed the ABOS examination. If their scores were at or below the 40th percentile, 32% failed the examination.

certifying examination is useful for residents and faculty to assess resident performance. Several studies report variable correlations of USMLE Step 1 scores with the ABOS Part I examination scores (Table 3). Herndon et al. [13], Swanson et al. [16], and Case and Swanson [5] found correlation of the USMLE (Case and Swanson [5] used the NBME) Step 1 examination scores and passing the ABOS Part I certifying examination. Klein et al. [14], however, found a weak correlation. Except for the study by Case and Swanson [5], all of the studies were confined to one program over a period of time. OITE scores (percentile ranking) also correlated with passing the ABOS Part I examination in two studies. To confirm correlations based on a wider distribution of programs, we compared the scores on the USMLE Step 1 and OITE examinations with those on the ABOS Part I certifying examination using data from four ACGME approved residency programs.

The main limitation of our study is the missing USMLE Step 1 scores for 23 residents, and between six to 17 OITE examination scores in PGYs 2 to 5 for the 202 graduating residents included in this study. For each examination, we calculated the mean ABOS percentile ranking scores, and the number of failures for each group. The mean ABOS scores were similar for the groups. USMLE scores were missing 23 results, yet the mean ABOS score was 55. (It was 57 for all 202 residents in this study.) Second, we cannot ensure the scores from the four programs in our study reflect those in programs throughout the United States. However, the overall pass rate reported in this study is similar to that reported by the American Board of Orthopaedic Surgery. Additionally, Swanson et al. [16]

Table 3. Comparison of USMLE and OITE with ABOS scores

Study	USMLE	OITE	Number of programs	Results	Number of residents
Herndon et al. [13]	Yes	Percentile rankings PGY 5 and PGYs 2–5	1	Found USMLE scores and OITE percentile rankings predictive of success on the ABOS Part I examination	161
Swanson et al. [16]	Yes	No	?	Significant but moderately strong relationship between USMLE Step 1, USMLE Step 2, and combined.	2654 pooled data from ABOS Part I certifying examination
Klein et al. [14]	Yes	Yes	1	Found moderate but significant correlation with ABOS Part I scores and OITE percentile rank in PGY 3 (0.52) and PGY 5 (0.69). Weak correlation between USMLE and ABOS Part I scores	64
Risner et al. [15]	No	Yes	1	Longitudinal descriptive score of residents and department. Can use scores to help drive study efforts in areas that are weak	16
Case and Swanson [5]	Yes	No	ABOS examinees	Found a strong relationship between pass/fail on ABOS Part I examination and the NBME Step 1 examination	Pooled ABOS data of 656 examinees (85% of total)
Black et al. [3]	Yes	Yes	1	Comparison of USMLE and OITE percentile ranking scores. Did not compare with ABOS Part I scores. Found moderate but significant correlation between USMLE Step 2 scores and OITE for each year group. Step 1 scores were found to weakly correlate, except for the PGY 3	Complete scores for 20 of 31 residents; used individual scores for each year
Carmichael et al. [4]	Yes	Yes	2	Residency selection factors including USMLE, AOA status, marital status, gender, age at the beginning of residency, and research were compared with average OITE scores. Found correlation of OITE scores to USMLE scores and marital status	60
Current study	Yes	Yes	4	Found correlation of USMLE Step 1 and OITE percentile rankings with ABOS Part I certifying examination results. OITE correlation improved stepwise from PGY 2 to PGY 5	202 ABOS Part I

reported recent pooled data from 2002–2006 of 2654 candidates who are graduates from American medical schools and observed similar overall pass rates to our study and similar to USMLE Step 1 results.

Our data suggest a correlation between the USMLE Step 1 scores and ABOS Part I percentile scores (Fig. 2). At or below a threshold value of 209, there was 33% failure of the ABOS examination. Swanson et al. [16] evaluated the results of 2654 graduates who took the ABOS examinations from 2002 to 2006, and found a 60% pass rate with a USMLE score of 180 and a 90% pass rate with a USMLE score of 220. Klein et al. [14], reporting on 64 graduates from one program, observed a low correlation between USMLE Step 1 and the ABOS Part I scores. Case and Swanson [5] reviewed 621 candidates taking the ABOS Part I examination in 1991 and compared an individual's scores with his or her own scores on NBME Parts 1 and 2. They found a correlation with NBME part scores and

ABOS Part I scores. Of those failing the ABOS Part I, 38% with an NBME Part 1 score below 400 failed, whereas only 1.6% of those above 600 failed.

Ten of the 12 residents in our study who failed the ABOS Part I had average OITE percentile rank scores of 28 or less for PGYs 2 to 5 (Fig. 1). Additionally, 10 of 12 had OITE percentile scores at or below 30 in PGY 5. Of the residents who passed the examination, none had an average below 28 percentile for PGYs 2 to 5.

In our study, comparison of the OITE scores with the ABOS Part I scores revealed the strongest correlation in PGY 4 and PGY 5. There was also a stepwise increase in correlation from PGY 2 through PGY 5. The average of percentile rank scores for all 4 years (PGYs 2–5) also correlated. Herndon et al. [13] retrospectively reviewed 15 years of residents taking the ABOS Parts I and II and compared OITE scores for those who did not pass with scores of those who passed. Among the 161 residents, they

observed a difference in the mean percentile OITE scores for PGY 5 year alone and an average of the PGYs 2–5 in the ABOS pass and failure groups. Klein et al. [14] found the correlation of OITE and ABOS Part I scores highest in PGY 5. In contrast to our study, Klein et al. [14] found the next highest correlation with ABOS Part I examination results to the PGY 3. Although not stated by Klein et al. [14], this may be the result of the rotation schedule rather than a generalization about the examination. We reached similar conclusions using data from four programs in the same geographic region.

Correlation between the USMLE Step 1 and ABOS Part I suggest that the USMLE Step 1 scores can be used as a factor with resident selection committees. USMLE Step 1 scores at or less than 209 were associated with a failure rate of 33% on the ABOS Part I examination. Our data suggest correlation between the OITE and passing the ABOS Part I examination, supporting the findings of Klein et al. [14] that residents whose OITE scores are below the 30th percentile average are at risk for failure of the ABOS Part I. Because of this, residents with scores at or below this level should receive extra attention from the program, and residents with scores at or below this threshold should devote extra time to study.

References

1. ACGME. V.C.3. ACGME Program Requirements for Graduate Medical Education in Orthopaedic Surgery. Available at: http://www.acgme.org/acWebsite/downloads/RRC_progReq/260orthopaedicsurgery07012007.pdf. Accessed February 3, 2010.
2. Bernstein AD, Jazrawi LM, Elbeshbesy B, Della Valle CJ, Zuckerman JD. Orthopaedic resident-selection criteria. *J Bone Joint Surg Am.* 2002;84:2090–2096.
3. Black KP, Abzug JM, Chinchilli VM. Orthopaedic in-training examination scores: a correlation with USMLE results. *J Bone Joint Surg Am.* 2006;88:671–676.
4. Carmichael KD, Westmoreland JB, Thomas JA, Patterson RM. Relation of residency selection factors to subsequent orthopaedic in-training examination performance. *South Med J.* 2005;98:528–532.
5. Case SM, Swanson DB. Validity of NBME Part I and Part II scores for selection of residents in orthopaedic surgery, dermatology, and preventive medicine. *Acad Med.* 1993;68(2 suppl):S51–S56.
6. Dirschl DR. Scoring of orthopaedic residency applicants: is a scoring system reliable? *Clin Orthop Relat Res.* 2002;399:260–264.
7. Dirschl DR, Adams GL. Reliability in evaluating letters of recommendation. *Acad Med.* 2000;75:1029.
8. Dirschl DR, Campion ER, Gilliam K. Resident selection and predictors of performance: can we be evidence based? *Clin Orthop Relat Res.* 2006;449:44–49.
9. Dirschl DR, Dahners LE, Adams GL, Crouch JH, Wilson FC. Correlating selection criteria with subsequent performance as residents. *Clin Orthop Relat Res.* 2002;399:265–271.
10. Evarts CM. Resident selection: a key to the future of orthopaedics. *Clin Orthop Relat Res.* 2006;449:39–43.
11. Godellas CCV, Hauge LLS, Huang RR. Factors affecting improvement on the American Board of Surgery In-Training Exam (ABSITE). *J Surg Res.* 2000;91:1–4.
12. Green M, Jones P, Thomas JX Jr. Selection criteria for residency: results of a national program directors survey. *Acad Med.* 2009;84:362–367.
13. Herndon JH, Allan BJ, Dyer G, Jawa A, Zurakowski D. Predictors of success on the American Board of Orthopaedic Surgery examination. *Clin Orthop Relat Res.* 2009;467:2436–2445.
14. Klein GR, Austin MS, Randolph S, Sharkey PF, Hilibrand AS. Passing the Boards: can USMLE and Orthopaedic in-Training Examination scores predict passage of the ABOS Part-I examination? *J Bone Joint Surg Am.* 2004;86:1092–1095.
15. Risner B, Nyland J, Crawford CH 3rd, Roberts CS, Johnson JR. Orthopaedic in-training examination performance: a nine-year review of a residency program database. *South Med J.* 2008;101:791–796.
16. Swanson DB, Sawhill A, Holtzman KZ, Bucak SD, Morrison C, Hurwitz S, DeRosa GP. Relationship between performance on part I of the American Board of Orthopaedic Surgery certifying examination and scores on USMLE Steps 1 and 2. *Acad Med.* 2009;84(10 suppl):S21–S24.
17. Thordarson DB, Ebramzadeh E, Sangiorgio SN, Schnall SB, Patzakis MJ. Resident selection: how we are doing and why? *Clin Orthop Relat Res.* 2007;459:255–259.
18. Turner NS, Shaughnessy WJ, Berg EJ, Larson DR, Hanssen AD. A quantitative composite scoring tool for orthopaedic residency screening and selection. *Clin Orthop Relat Res.* 2006;449:50–55.