DOI: 10.52544/2642-7184(1)2001
MedJEM
Original Research

USMLE Scores Do Not Predict the Clinical Performance of Emergency Medicine Residents

Karima Sajadi-Ernazarova[1], Edward A Ramoska[1], Mark A Saks[1,2]

[1]Department of Emergency Medicine, Drexel University College of Medicine, Philadelphia, USA
[2]Department of Emergency Medicine, Crozer-Chester Medical Center, Philadelphia, USA

ABSTRACT

Background: Scores on "high-stakes" multiple choice exams such as the United States Medical Licensing Examination® (USMLE) are important screening and applicant ranking criteria used by residencies.

Objective: We tested the hypothesis that USMLE scores do not predict overall clinical performance of emergency medicine (EM) residents.

Methods: All graduates from our University-based EM residency between the years 2008 and 2015 were included. Residents who had incomplete USMLE records were terminated, transferred out of the program, or did not graduate within this timeframe were excluded from the analysis. Clinical performance was defined as a gestalt of the residency program's leadership and was classified into three sets: top, average, and lowest clinical performer. Dissimilarities of the initial blind rankings were adjudicated during a consensus conference.

Results: During the eight years of the study period, there were a total of 115 graduating residents: 73 men (63%) and 42 women. Nearly all of them (109; 95%) had allopathic medical degrees; the remainder had osteopathic degrees. There was not a statistically significant correlation between our ranking of clinical performance and the Step 2 Clinical Knowledge score. There was a non-significant correlation between clinical performance and the Step 1 score.

Conclusion: Neither USMLE Step 1 nor Step 2 Clinical Knowledge were good predictors of the actual clinical performance of residents during their training. We feel that their scores are overemphasized in the resident selection process.

Keywords: Emergency Medicine Education, USMLE Scores, Resident Clinical Performance, Emergency Medicine Training, Residency Recruitment, Residency Selection Criteria

INTRODUCTION

Throughout undergraduate and graduate medical education, aspiring physicians are subjected to numerous standardized tests in an attempt to determine a standard of minimal competency and to objectively measure their fund of knowledge. While individual state medical licensing boards make the ultimate determination of certification for independent medical practice, "high-stakes" multiple choice examinations such as the United States Medical Licensing Examination® (USMLE), have become widely used to gauge the mastery of applied basic and clinical science knowledge. In an attempt to recruit the best possible candidates, scores on these examinations have become important applicant screening and ranking criteria. They are used by residency programs in a wide variety of medical specialties, including Emergency

Correspondence to:

Karima R Sajadi-Ernazarova
Associate Professor of Emergency Medicine
Associate Program Director
Emergency Medicine Residency
Drexel University College of Medicine
Department of Emergency Medicine
E-mail: karimasajadi@gmail.com
245 North 15th Street, Room 2108
Philadelphia, PA 19102

Medicine, despite this not being their intended use.[1,2]

Several prior studies have examined the relationship between USMLE scores and performance on emergency medicine-specific examinations such as the in-training examination (ITE) and the American Board of Emergency Medicine® qualifying examination.[3-7] While most of these studies have demonstrated a statistically significant correlation, several other recent studies have questioned the relationship between performance on the USMLE and ITE examinations and a resident's clinical performance.[8-10] Additionally, we are not aware of any data, which shows whether USMLE scores correlate with evaluation of the actual clinical performance of emergency medicine residents at the conclusion of their residency training. In fact, after an extensive literature search, we are not aware of any study that has attempted to examine the relationship between USMLE scores and resident overall clinical performance.

This study attempts to clarify the relationship between performance on the most commonly used multiple choice USMLE examinations (Step 1, Step 2 CK) and global clinical performance of the residents at the conclusion of their training at an emergency medicine residency program. The objective of this study was to assess the association between USMLE scores and overall clinical performance of emergency medicine (EM) residents.

METHODS

All graduating residents from our university-based emergency medicine residency program between the years 2008 and 2015 were eligible for inclusion in the analysis. There were 115 study subjects screened for eligibility. Residents were excluded if they had incomplete USMLE records because they were osteopathic medical school graduates who did not take the USMLE examinations. Residents who were terminated, transferred out of the program, or did not graduate within the specified timeframe were also excluded due to incomplete information available regarding their overall clinical performance. This study was granted exempt status by our IRB.

The primary endpoint of the study is the comparison of the residents' overall clinical performance with their USMLE scores.

The resident's overall clinical performance was defined as a gestalt of the three physicians who comprised the residency program's leadership (program director, associate program director, and assistant program director) during the study period. These physicians had an intimate knowledge of all aspects of the resident's clinical performance during the entire duration of their 3 years of emergency medicine residency training, including access to all of their end-of-rotation evaluations (both on-service and off-service rotations), ITE scores, biannual formative evaluations, and performance improvement data (patient, nurse, or physician complaints/compliments, case reviews, morbidity and mortality referrals, etc.). This gestalt was based on the Accreditation Council for Graduate Medical Education's six core competencies (patient care, medical knowledge, professionalism, systems-based practice, practice-based learning and improvement, and interpersonal and communication skills)[11] that were used at that time to evaluate the residents' clinical performance. The other attending physicians' opinions were purposefully not sought due to their incomplete knowledge of all aspects of the residents' clinical performance and due to potential "popularity" and "likability" bias. This bias was believed to be minimized by the PD/APDs' complete knowledge of the residents' files. The years of training that were selected for the study included only those years when all three stated physicians were working together as PD/APDs.

These three physicians were presented with lists of graduating residents by year and asked to classify the residents into three groups: top clinical performer (top one-third), average clinical performer (middle one-third), and lowest clinical performer (bottom one-third). The grouping process was based on these three physicians' recall of these residents' overall clinical performance during their residency training at the time of graduation. The physicians were initially blinded to each other's grouping selections. Dissimilarities of the rankings were adjudicated during a consensus conference. The residents' files were then accessed and the residents' USMLE Step 1 and Step 2CK scores were obtained.

Descriptive statistics were generated. The data

from the initial physicians' grouping selection were compared and inter-rater reliability was calculated using intraclass correlation (ICC). A correlation analysis utilizing Spearman's rho (rs) was performed to uncover whether a resident's examination scores were associated with clinical performance. A p value of < 0.05 was considered significant.

RESULTS

During the eight years of the study period, we graduated 115 residents. There were 73 men (63%) and 42 women. Nearly all of them (109; 95%) had allopathic medical degrees; the remainder had osteopathic degrees. Both USMLE Step 1 and Step 2CK scores were available for 106 residents; a total of 9 had one or more score results that could not be obtained. Of these, four were osteopathic graduates who did not take the USMLE Step 2CK examination and one osteopathic medical student did not take either USMLE Step 1 or Step 2CK. In 4 cases, the scores could not be located.

Table 1 shows the distribution of the final consensus ranking of the residents. The inter-rater reliability of the initial rankings was strong with an ICC = 0.845 ($p < 0.01$).

Table 1 Final Ranking of Residents

Category	Number	Percentage
Top	38	33.0%
Middle	44	38.3%
Bottom	33	28.7%

There was a poor, but statistically significant, correlation between our ranking of overall clinical performance and the USMLE Step 2 CK score. The coefficient of determination, r^2, was 0.042 for Step 2 CK. There was not a statistically significant correlation between clinical performance and the USMLE Step 1 score (Table 2).

DISCUSSION

The recruiting season is an important part of every residency program's life cycle. Each residency program hopes to recruit outstanding applicants who will develop into excellent residents and physicians. Various criteria are used to screen candidates for an interview and to determine rank list order; USMLE scores being one of the most commonly used benchmarks.

The existing literature suggests that scores on the USMLE do correlate with scores on the ITE3 as well as with the pass or fail status on a resident's first attempt on the ABEM qualifying examination.4 It seems that a good performance on one standardized test predicts a good performance on all subsequent standardized tests, while a poor performance on one standardized test predicts a poor performance on all subsequent standardized tests. In other words, good test-takers do well on tests.

Table 2 Correlation between Clinical Performance and Examination Scores

	USMLE Step 1	USMLE Step 2 CK
Correlation Coefficient	0.067	0.205
P Value	0.49	0.04
N	109	106

Our study reveals that neither USMLE Step 1 nor Step 2CK are useful predictors of a resident's final clinical performance as judged by program leadership. The most readily available piece of information, the Step 1 score, showed no correlation with clinical performance; whereas, the Step 2 CK score showed only a poor correlation, with that score accounting for about 4% of the total variation in clinical performance. McGaghie et al have likewise "discouraged" the use of USMLE scores for postgraduate medical residency selection decisions.[9] Their research synthesis article referenced several studies that show that measures of clinical skills, as diverse as cardiac auscultation, central venous catheter insertion, thoracentesis, advanced cardiac life support scenarios, and communication skills, are not correlated or poorly correlated with USMLE Step 1 and Step 2 scores. We did not evaluate the correlation of the USMLE Step 3 or ITE scores with clinical performance since these scores are not readily available during the recruitment process.

USMLE scores are an important piece of the future resident's overall application packet, and they should neither be discarded nor ignored. However as shown in our study, since neither USMLE Step 1 nor Step 2 CK are good predictors of the actual clinical performance of the residents at the completion of their training, we believe that

their scores should not be overemphasized in the resident selection process. A resident with a higher USMLE score will not necessarily become a more effective or efficient emergency physician than one with a lower score.

LIMITATIONS

We are missing test scores for several residents, either because they could not be found or because they were osteopathic residents who did not take the USMLEs. Furthermore, since osteopathic-trained candidates represented a small subgroup of our study population, we did not include the Comprehensive Osteopathic Medical Licensing Examination® (COMLEX) scores in a separate analysis. As this examination tests different material than the USMLE, it is likely that our findings are not generalizable to this examination.

Second, we used only the "final" passing score for each examination in this analysis. It is possible that a resident may have failed an examination one or more times before obtaining this passing score. Therefore, we cannot draw conclusions as to the significance of a failure on the future clinical performance.

Third, there is no "gold standard" to assess a resident's overall clinical performance at the end of residency training. We felt that the consensus opinion of the three physicians comprising the residency leadership was the best method for achieving this ranking; and as mentioned, included all facets of a resident's performance. These rankings may be subject to recall bias since we performed all the ratings at one time and a significant amount of time had passed since the earliest classes graduated.

Finally, during our consensus conference we discussed the possibility of the "halo effect" for discordant rankings. This is a cognitive bias in which the overall impression is influenced by excellent, or poor, performance in one specific area, such as communication skills, work-load efficiency, or personality. In an effort to ameliorate the effects of social popularity, or likability bias, we specifically discussed a resident's standing in each of the six core competencies and form a unanimous opinion based on the overall clinical performance.

CONCLUSION

Neither USMLE Step 1 nor Step 2 CK are good predictors of the actual clinical performance of residents at the completion of their training. We feel that USMLE scores may be overemphasized in the resident selection process.

Conflicts of Interest: The authors declare no conflicts of interest or sources of funding.

REFERENCES

1. Green M, Jones P, Thomas JX Jr. Selection criteria for residency: Results of a national program directors survey. Acad Med. 2009;84:362–367.

2. Balentine J, Gaeta T, Spevack T. Evaluating applicants to emergency medicine residency programs. J of Emerg Med. 1999;17:131–134.

3. Thundiyil JG, Modica RF, Silvestri S, et al. Do United States Medical Licensing Examination (USMLE) scores predict in-training test performance for emergency medicine residents? J of Emerg Med. 2010; 38:65-9.

4. Harmouche E, Goyal N, Pinawin A, et al. USMLE Scores Predict Success in ABEM Initial Certification: A Multicenter Study. West J of Emerg Med. 2017; 18:544-549.

5. Wagoner NE, Suriano R. Program directors' responses to a survey on variables used to select residents in a time of change. Acad Med. 1999;74:51-58.

6. Fine PL, Hayward RA. Do the criteria of resident selection committees predict residents' performances? Acad Med. 1995;70:834-838.

7. Nelson M, Calandrella C. Does USMLE Step 1 & 2 Scores Predict Success on ITE and ABEM Qualifying Exam: A Review of an Emergency Medicine Residency Program from its Inception. Ann of Emerg Med. 2017;70:S58 - S59.

8. Blanchard J. Board scores and resident performance: Is there a link? Ann of Emerg Med. 2000;36:64-67.

9. 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.

10. Ryan JG, Barlas D, Pollack S. The Relationship Between Faculty Performance Assessment and Results on the In-Training Examination for Residents in an Emergency Medicine Training Program. J Grad Med Educ. 2013;5:582-6.

11. LaMantia J. The ACGME core competencies: getting ahead of the curve. Acad Emerg Med. 2002;9:1216-7.