

Predicting American Board of Emergency Medicine Qualifying Examination Passage Using United States Medical Licensing Examination Step Scores

Terrell Caffery, MD,¹ Jenna Fredette, MD,² Mandi W. Musso, PhD,^{1,3} Glenn N. Jones, PhD⁴

¹Emergency Medicine Residency Program, Louisiana State University Health Sciences Center School of Medicine in New Orleans, Baton Rouge, LA ²Department of Emergency Medicine, Christiana Care Health System, Newark, DE ³Division of Academic Affairs, Our Lady of the Lake Regional Medical Center, Baton Rouge, LA and Louisiana Department of Quality and Safety, Our Lady of the Lake Regional Medical Center, Baton Rouge, LA ⁴Department of Family Medicine, Louisiana State University Health Sciences Center School of Medicine in New Orleans, Baton Rouge, LA

Background: The objective of the current study was to determine whether emergency medicine residents' United States Medical Licensing Examination (USMLE) scores are significantly associated with first-attempt passage of the American Board of Emergency Medicine (ABEM) qualifying (written) examination. We hypothesized that USMLE Step 2 Clinical Knowledge (CK) scores would be useful in predicting students who passed the ABEM qualifying examination on their first attempt.

Methods: For this retrospective cohort study, we examined the data of residents who successfully completed training at two emergency medicine residency programs between the years 2002-2013. Because scores on the USMLE Step examinations varied greatly across years, we obtained means and standard deviations from the National Board of Medical Examiners. We subtracted the mean score for the year each resident took the examination from the resident's examination score, creating centered Step 1 and centered Step 2 CK scores.

Results: A multivariate logistic regression analysis indicated that centered Step 2 CK scores could be used to predict the odds of passing the ABEM qualifying examination (odds ratio = 1.05 [95% confidence interval 1.02 to 1.08, $P < 0.001$]). Using a Step 2 CK score cutoff of 7 points lower than the mean yielded 64% sensitivity and 81% specificity for predicting passage of the ABEM written examination on the first attempt.

Conclusion: Program directors and selection committees may wish to consider whether applicants' Step 2 CK scores are near the national average when making ranking decisions, as this variable is highly predictive of passing the ABEM qualifying examination on the initial attempt.

Keywords: Education–medical–graduate, emergency medicine, governing board

Address correspondence to Mandi W. Musso, PhD, Academic Research Director, Emergency Medicine Residency Program, Our Lady of the Lake Regional Medical Center, 5246 Brittany Drive, Baton Rouge, LA 70808. Tel: (225) 757-4148. Email: mmuss1@lsuhsc.edu

INTRODUCTION

In the 2017 National Resident Matching Program, a total of 2,047 emergency medicine positions were offered, and the number of applicants was 2,703.¹ This discrepancy in the number of applicants and available positions results in applicants applying to many residency programs to maximize their potential for matching. In 2013, the Accreditation Council for Graduate Medical Education (ACGME) mandated that emergency medicine residency programs demonstrate an 80% first-time passage rate on the American Board of Emergency Medicine (ABEM) qualifying (written) examination.² Consequently, selection committees want to identify applicants who will be able to pass the certifying examination on their initial attempt. Residency program directors and se-

lection committees review large numbers of applications from medical students with various backgrounds and training and select a small percentage of competitive applicants to interview. Few standardized metrics are available to help identify applicants who will be successful.

Because all medical students complete the United States Medical Licensing Examination (USMLE), residency program directors place an emphasis on these scores during the selection process, particularly when selecting applicants to invite to interview.³ Selection committees and program directors assume that USMLE scores are good predictors of resident outcomes. Evidence indicates that USMLE scores are associated with emergency medicine residency in-training examination (ITE) scores,⁴ and ITE scores correlate

with certifying examination scores. Studies from other medical specialties indicate that USMLE scores are significantly associated with performance on qualifying examinations in internal medicine,⁵ pathology,⁶ orthopedic surgery,⁷ and the American Board of Surgery.⁸ Maker et al reported that USMLE Step 2 scores are also associated with passage of the American Board of Surgery oral certifying examination.⁹ Only one study has examined the relationship between USMLE Step 1 and Step 2 scores with the ABEM qualifying examination.¹⁰ The authors reported that a Step 1 score of 227, a Step 2 Clinical Knowledge (CK) score of 225, and a composite score of 444 predicted a 95% chance of passing the ABEM qualifying and oral examinations. However, one limitation of this study is that the authors examined raw scores. The mean score for the USMLE examinations has increased steadily since its inception, limiting the utility of raw scores.

The purpose of this study was to examine whether difference from the mean USMLE scores (centered Step scores) can be used to select residents who will pass the written ABEM qualifying examination on their initial attempt. We hypothesized that residents who scored higher than the USMLE mean for the year they took the examinations would be significantly more likely to pass the initial ABEM qualifying examination. We also hypothesized that centered Step 2 CK scores would be more predictive than centered Step 1 scores because of the clinical content of the Step 2 CK examination.

METHODS

Study Design

This retrospective cohort study examined the data of residents who successfully completed training at two emergency medicine residency programs between the years 2002-2013. This study was determined to be exempt by both medical schools' institutional review boards.

Study Setting and Population

One residency program is located in the Southern Region. The program had a 3-year (postgraduate year [PGY] 1 to PGY 3) format and recruited an average of 12.6 residents per year (range, 10-15 residents) during the study period. A total of 164 residents graduated from the program within the study period (2002-2013). The first-time written ABEM qualifying passage rate was 78.3%. The other residency is located in the Mid-Atlantic Region. A total of 135 residents graduated from this program between 2002 and 2012. Data from 2013 were not available. The program also had a 3-year (PGY 1 to PGY 3) format and recruited an average of 12.3 residents per year (range, 11-13 residents) during the study period. First-time ABEM qualifying examination passage rates were available from 2007-2013 for this program and averaged 96.4%.

Study Protocol

The residency programs provided deidentified databases to the investigators. Demographic variables such as sex, age, and ethnicity were not collected because the resident classes tend to be relatively homogenous and outliers might have made these variables identifiable. Data collected included USMLE Step 1 and Step 2 CK scores and performance on the written ABEM qualifying examination (dichotomously coded as pass or fail).

Data Analysis

The national average of USMLE Step 1 and Step 2 CK scores varied greatly during the study period. The investigators contacted the National Board of Medical Examiners to obtain the mean and standard deviations for the USMLE Step 1 and Step 2 CK scores for each year. Because the standard deviations were relatively consistent across years, we calculated the distance from the mean for the year that each resident completed Step 1 and Step 2 CK, respectively. We subtracted the mean score for that year from each resident's USMLE score to create a centered Step score. For example, if a resident scored 220 in a year when the Step 2 CK mean score was 212, his or her centered Step 2 CK score would be +8. However, if a resident scored 220 in a year when the mean score was 228, his or her centered Step 2 CK score would be -8. This approach allows for an equitable comparison of Step scores across the years.

Logistic regression analyses were used to determine whether USMLE centered Step scores were predictive of passing the ABEM qualifying certification examination. First, univariate logistic regression was used to determine whether the centered Step 1 or centered Step 2 CK scores were predictive of initial ABEM qualifying examination passage separately. Then multivariate logistic regression was used to examine if the combination of the centered Step 1 and Step 2 CK scores was more predictive than the simpler univariate model. Results are presented as odds ratios (ORs) with 95% confidence intervals (CIs). Statistical significance was set at $P < 0.05$. A receiver operating characteristic (ROC) curve was constructed to obtain area under the curve (AUC) and its 95% CI. The ROC was also used to examine potential cutoff scores for centered Step 2 CK scores in predicting passage of the ABEM qualifying examination on the first attempt. Sensitivities and specificities were calculated using the ROC curve. Analyses were conducted using IBM SPSS Statistics v.22.

RESULTS

In all, 299 residents graduated from the programs within the study period. Data were available for 135 residents from the Mid-Atlantic residency program and 164 residents from the Southern residency program. Complete data were missing for 8 residents from the Southern residency program. USMLE Step 1 data were missing for 16 individuals (4 Mid-Atlantic and 12 Southern), and USMLE Step 2 CK data were missing for 68 individuals (29 Mid-Atlantic and 39 Southern). ABEM qualifying examination data were missing for 12 individuals (10 Mid-Atlantic and 2 Southern). Complete data were available for a total of 206 residents, and Step 2 CK and ABEM qualifying examination data were available for 212 residents.

In the univariate logistic regression analyses, centered Step 1 (OR = 1.05, 95% CI 1.03 to 1.08, $P < 0.001$) and centered Step 2 CK (OR = 1.06, 95% CI 1.04 to 1.09, $P < 0.001$) scores were both significant predictors of ABEM qualifying examination passage when considered separately. However, in a multivariate logistic model, the centered Step 1 score was not a significant predictor (OR = 1.02, 95% CI .99 to 1.05, $P < 0.17$), while the Step 2 CK score remained significant (OR = 1.05, 95% CI 1.02 to 1.08, $P < 0.001$). That is, the Step 1 score did not contribute uniquely to the prediction of ABEM qualifying examination passage when

the centered Step 2 CK score was in the equation. Based on these results, a more complex multivariate model is not warranted. The OR of 1.06 from the univariate model represents an increase or decrease in the odds of passing the ABEM qualifying examination with each 1-point increase or decrease in centered Step 2 CK scores, respectively. That is, the odds of passing the ABEM qualifying examination on the first attempt double with every 12-point increase in the centered Step 2 CK score. The Figure illustrates the probability of passing the ABEM qualifying examination on the first attempt as predicted by centered Step 2 CK scores.

The AUC was significant (AUC = 0.80, 95% CI 0.72 to 0.87, $P < 0.001$). The Table presents the sensitivity and specificity of various points from the mean. Overall, scores that were at the mean for that year (a centered score of 0), yielded 97% (95% CI 86.3% to 99.7%) specificity and 53% (95% CI 45.5% to 60.0%) sensitivity for predicting passage of the ABEM written examination on the first attempt. Only one resident who scored at or above the mean for his/her year failed the ABEM written examination on the first try. However, of the 180 residents in our sample who passed the ABEM written examination on their first attempt, 85 residents had Step 2 CK scores that were below the mean for the year they took the examination but still passed the ABEM written examination on their first attempt. Using a cut-off of Step 2 CK scores of 7 points lower than the mean yielded 64% (95% CI 56.7% to 70.6%) sensitivity and 81% (95% CI to 65.4% to 91.8%) specificity. Using this cutoff score, 6 individuals who had centered Step 2 CK scores between -7 and -1 failed the ABEM qualifying examination on their first attempt. However, 65 individuals had Step 2 CK

Table. Sensitivities and Specificities for Predicting American Board of Emergency Medicine Board Examination Passage on the First Attempt Based on United States Medical Licensing Examination Step 2 Clinical Knowledge (CK) Scores

Points From the Mean Step 2 CK Score	Sensitivity	Specificity
-20	0.84	0.50
-19	0.83	0.50
-18	0.81	0.50
-17	0.79	0.53
-16	0.78	0.56
-15	0.78	0.56
-14	0.76	0.59
-13	0.76	0.66
-12	0.73	0.69
-11	0.72	0.72
-10	0.68	0.72
-9	0.67	0.72
-8	0.65	0.72
-7	0.64	0.81
-6	0.62	0.84
-5	0.59	0.87
-4	0.57	0.87
-3	0.56	0.91
-2	0.56	0.91
-1	0.54	0.97
0	0.53	0.97
1	0.49	0.97
2	0.48	0.97
3	0.45	0.97
4	0.44	0.97
5	0.42	0.97
6	0.39	0.97
7	0.38	0.97
8	0.36	0.97
9	0.33	0.97
10	0.30	0.97
11	0.29	0.97
12	0.28	0.97
13	0.26	0.97
14	0.23	0.97
15	0.23	0.97
16	0.22	0.97
17	0.18	0.97
18	0.18	1.00
19	0.16	1.00
20	0.14	1.00
21	0.13	1.00

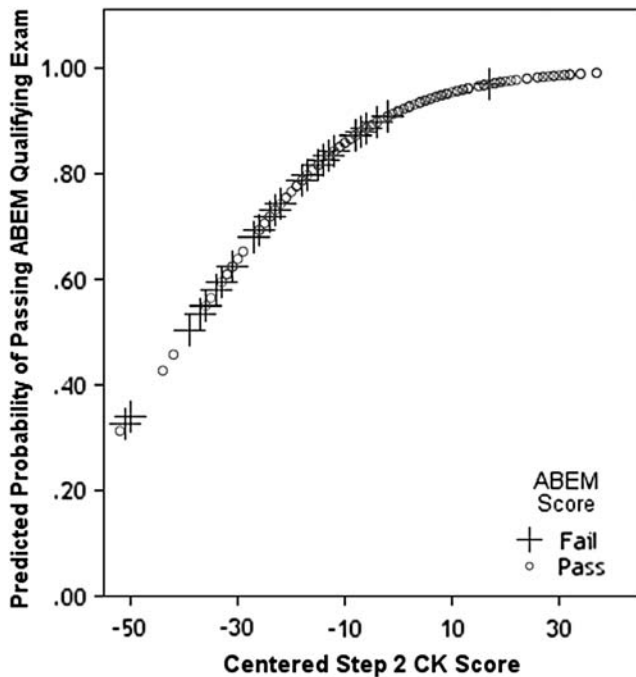


Figure. Probability of passing the American Board of Emergency Medicine (ABEM) qualifying examination on the first attempt as predicted by how many points away from the mean a person scores on the United States Medical Licensing Examination Step 2 Clinical Knowledge (CK) examination.

scores more than 7 points below the mean and still passed the ABEM qualifying examination on their first attempt.

DISCUSSION

The purpose of this study was to examine whether USMLE Step 1 and Step 2 CK scores are predictive of passing the written ABEM qualifying examination on the initial attempt. This study yielded 3 important findings. First, while Step 1 scores were predictive of initial ABEM qualifying examination passage when entered alone in a logistic regression, they were no longer significant once Step 2 CK scores were entered into the logistic regression. Step 2 CK scores were a better predictor of ABEM qualifying examination passage. Second, all but one person in our sample who had a Step 2 CK score at or above the mean in the year he/she took the examination passed the ABEM qualifying examination on the first attempt. Finally, Step 2 CK scores within 7 points of the mean yielded adequate specificity for passing the initial ABEM qualifying examination.

Previous studies have reported that USMLE Step 1 scores are valid predictors of board passage rates for other specialties.^{8,11,12} Our results are consistent with the findings of Harmouche et al¹⁰ in that Step 2 CK scores were better predictors of board passage than Step 1 scores. Step 2 CK scores are possibly more relevant than Step 1 scores in emergency medicine specifically. However, another important point is that previous retrospective studies did not account for variations in the scores across years. This difference in the study analysis may be the reason for differences between our study and prior work.

USMLE Step 2 CK scores are available for most applicants to residency programs. While the USMLE provides minimum passage scores, we found that the applicant's performance vs the national average (centered Step 2 CK scores) predicts ABEM board passage. Harmouche et al¹⁰ suggested using raw scores to predict success in passing the initial ABEM examination; however, medical students consistently perform better and better on the USMLE, raising the mean score each year. For this reason, we used a standardized approach to determining scores that would predict initial success on the ABEM and remain useful over time. As stated earlier, the ACGME now requires 80% of a program's graduates from the preceding 5 years to pass the ABEM.² Our analysis shows that a cutoff score of 7 points below the mean predicts an 80% probability of passing the ABEM. While the USMLE Step 2 CK score may be a readily available and useful tool, it should not be used alone in making ranking decisions because many individuals in our cohort who scored below the national average still passed the ABEM qualifying examination on their first attempt. Therefore, emergency medicine residency program directors and selection committees should consider other biographic and academic information prior to making final decisions about ranking applicants. Future studies could examine additional variables that may contribute to the prediction of successful residents.

Limitations of our study include its retrospective nature as well as results collected from only 2 residency programs. While our results indicate a significant relationship between USMLE Step 2 CK scores and ABEM qualifying examination passage, additional research is needed to ensure generalizability. In addition, prospective data from multiple institutions would be beneficial in validating these results and giving

program directors increased confidence in using USMLE for making decisions when ranking applicants. Examination of all factors associated with performance on the ABEM was beyond the scope of this study; however, programmatic changes during the decade of the study period may have impacted the results.

CONCLUSION

This study has important implications for emergency medicine residency program directors and selection committees, as it proposes a means of quantifying the risk of initial ABEM qualifying examination failure that program directors may be taking when selecting applicants with lower Step 2 CK scores. A centered Step 2 CK score within 7 points of the national average suggests a candidate is likely to pass the ABEM qualifying examination on the initial attempt. This knowledge can be beneficial in making selection decisions in light of ACGME requirements and the increasing applicant pool. However, other applicant data should be considered as well prior to making a final ranking decision. Further studies are needed to validate these findings; however, this study is the first to demonstrate a relationship between centered Step 2 CK scores and emergency medicine board passage.

ACKNOWLEDGMENTS

The authors have no financial or proprietary interest in the subject matter of this article.

REFERENCES

1. National Resident Matching Program. Results and Data: 2017 Main Residency Match. <http://www.nrmp.org/wp-content/uploads/2017/06/Main-Match-Results-and-Data-2017.pdf>. Published April 2017. Accessed October 6, 2017.
2. Accreditation Council for Graduate Medical Education. ACGME Program Requirements for Graduate Medical Education in Emergency Medicine. https://www.acgme.org/Portals/0/PFAssets/ProgramRequirements/110_emergency_medicine_2017-07-01.pdf. Revised effective July 1, 2017. Accessed May 9, 2018.
3. Green M, Jones P, Thomas JX Jr. Selection criteria for residency: results of a national program directors survey. *Acad Med*. 2009 Mar;84(3):362-367. doi: 10.1097/ACM.0b013e3181970c6b.
4. Thundiyil JG, Modica RF, Silvestri S, Papa L. Do United States Medical Licensing Examination (USMLE) scores predict in-training test performance for emergency medicine residents? *J Emerg Med*. 2010 Jan;38(1):65-69. doi: 10.1016/j.jemermed.2008.04.010.
5. Kay C, Jackson JL, Frank M. The relationship between internal medicine residency graduate performance on the ABIM certifying examination, yearly in-service training examinations, and the USMLE Step 1 examination. *Acad Med*. 2015 Jan;90(1):100-104. doi: 10.1097/ACM.0000000000000500.
6. Picarsic J, Raval JS, Macpherson T. United States Medical Licensing Examination step 1 two-digit score: a correlation with the American Board of Pathology first-time test taker pass/fail rate at the University of Pittsburgh Medical Center. *Arch Pathol Lab Med*. 2011 Oct;135(10):1349-1352. doi: 10.5858/arpa.2010-0240-EP.
7. Dougherty PJ, Walter N, Schilling P, Najibi S, Herkowitz H. Do scores of the USMLE Step 1 and OITE correlate with the ABOS Part I certifying examination?: a multicenter study. *Clin Orthop Relat Res*. 2010 Oct;468(10):2797-2802. doi: 10.1007/s11999-010-1327-3.

8. Sutton E, Richardson JD, Ziegler C, Bond J, Burke-Poole M, McMasters KM. Is USMLE Step 1 score a valid predictor of success in surgical residency? *Am J Surg.* 2014 Dec;208(6):1029-1034; discussion 1034. doi: 10.1016/j.amjsurg.2014.06.032.
9. Maker VK, Zahedi MM, Villines D, Maker AV. Can we predict which residents are going to pass/fail the oral boards? *J Surg Educ.* 2012 Nov-Dec;69(6):705-713. doi: 10.1016/j.jsurg.2012.08.009.
10. Harmouche E, Goyal N, Pinawin A, Nagarwala J, Bhat R. USMLE scores predict success in ABEM initial certification: a multicenter study. *West J Emerg Med.* 2017 Apr;18(3):544-549. doi: 10.5811/westjem.2016.12.32478.
11. Durning SJ, Dong T, Hemmer PA, et al. Are commonly used premedical school or medical school measures associated with board certification? *Mil Med.* 2015 Apr;180(4 Suppl):18-23. doi: 10.7205/MILMED-D-14-00569.
12. Grabovsky I, Hess BJ, Haist S, et al. The relationship between performance on the Infectious Diseases In-Training and Certification Examinations. *Clin Infect Dis.* 2015 Mar 1;60(5):677-683. doi: 10.1093/cid/ciu906.

This article meets the Accreditation Council for Graduate Medical Education and the American Board of Medical Specialties Maintenance of Certification competency for Medical Knowledge.