United States Medical Licensing Examination and American Board of Pediatrics Certification Examination Results: Does the Residency Program Contribute to Trainee Achievement

Thomas R. Welch, MD[1], Brad G. Olson, MD, MS[1], Elizabeth Nelsen, MD, MA[1], Gary L. Beck Dallaghan, PhD[2], Gloria A. Kennedy, MD[1], and Ann Botash, MD[1]

Objective To determine whether training site or prior examinee performance on the US Medical Licensing Examination (USMLE) step 1 and step 2 might predict pass rates on the American Board of Pediatrics (ABP) certifying examination.

Study design Data from graduates of pediatric residency programs completing the ABP certifying examination between 2009 and 2013 were obtained. For each, results of the initial ABP certifying examination were obtained, as well as results on National Board of Medical Examiners (NBME) step 1 and step 2 examinations. Hierarchical linear modeling was used to nest first-time ABP results within training programs to isolate program contribution to ABP results while controlling for USMLE step 1 and step 2 scores. Stepwise linear regression was then used to determine which of these examinations was a better predictor of ABP results.

Results A total of 1110 graduates of 15 programs had complete testing results and were subject to analysis. Mean ABP scores for these programs ranged from 186.13 to 214.32. The hierarchical linear model suggested that the interaction of step 1 and 2 scores predicted ABP performance ($F[1,1007.70] = 6.44$, $P = .011$). By conducting a multilevel model by training program, both USMLE step examinations predicted first-time ABP results ($b = .002$, $t = 2.54$, $P = .011$). Linear regression analyses indicated that step 2 results were a better predictor of ABP performance than step 1 or a combination of the two USMLE scores.

Conclusions Performance on the USMLE examinations, especially step 2, predicts performance on the ABP certifying examination. The contribution of training site to ABP performance was statistically significant, though contributed modestly to the effect compared with prior USMLE scores. *(J Pediatr 2017;188:270-4).*

As is the case for most specialty boards, certification by the American Board of Pediatrics (ABP) requires successfully passing a secure examination designed to test clinical knowledge. Currently, a passing score is a scaled score of 180, a prerequisite for ABP certification. The range of scores is 1-300. Most recently, 86% of first time test-takers have been successful on their first attempt at this examination.[1]

Because of the importance of this examination, training programs carefully track the performance of their graduates and may include cumulative examination pass rates in their recruiting materials. Moreover, the Accreditation Council for Graduate Medical Education establishes minimal graduate pass rates, which programs must attain to maintain accreditation. For pediatrics, this rate is 70% of first-time examinees over the previous 5 years.[2]

Examination success rates may be attributed to the ability of training programs to prepare examinees for ABP certification. However, if there are factors unique to trainees themselves that are more predictive of success, it might be inappropriate to place excessive focus on training sites.

One trainee-specific factor to consider is past examination performance. A study of an orthopedic surgery training program demonstrated that performance on steps 1 and 2 of the US Medical Licensing Examination (USMLE) was predictive of performance on the American Board of Orthopedic Surgery Certifying Examination.[3] A similar study of a pediatric training program showed the predictive utility of the USMLE step 1 examination for performance on the ABP certifying examination.[4]

Both of these studies were from single institutions. Thus, neither could address the contribution of training site to exam performance. We have used data from 15 separate pediatric residencies to define further the predictive value of USMLE performance, as well as the contribution of training site to performance.

ABP	American Board of Pediatrics
USMLE	US Medical Licensing Examination

From the [1]Department of Pediatrics, State University of New York Upstate Medical University, Syracuse, NY; and [2]Office of Medical Education, University of Nebraska Medical Center, Omaha, NE

T.W. serves as an Associate Editor for *The Journal of Pediatrics*. The other authors declare no conflicts of interest.

0022-3476/$ - see front matter.
http://dx.doi.org10.1016/j.jpeds.2017.05.057

Methods

A convenience sample of 15 pediatric residency programs agreed to provide data.

Participating sites provided deidentified matching data for the first attempts at the USMLE steps 1 and 2 examinations and the ABP certifying examination for graduates taking the ABP examination between 2009 and 2013. The ABP changed the reporting scale of its certification examination from a scale of 0-800 to a scale of 0-300 in 2012. A conversion factor was obtained from the ABP to standardize the results for the entire 2009-2013 period. (Personal communication, Dr. Rachel Tan, ABP, June 18, 2015).

Graduates for whom results from one of the 3 examinations were not available were excluded. This resulted in exclusion of graduates for whom only Comprehensive Osteopathic Medical Licensing Examination results were available.

Statistical Analyses

Descriptive and inferential statistics were calculated from the data set. Continuous variables were examined to ensure normal distribution. A hierarchical linear model was used to analyze the data. Hierarchical linear modeling allows us to consider trainees nested with training programs where the program is a random variable rather than a fixed effect. This is important because strict linear models assume residual errors are independent; however, when individuals are trained in a similar context these residual errors can be correlated. This model was constructed with the first-time ABP results nested within training programs to isolate program contribution to ABP results while controlling for USMLE step 1 and 2 performance.

This analysis was followed up with a linear regression analysis. Variables were entered into the model using USMLE step 1 and 2 performance as predictors of ABP results. Analyses were conducted using each step examination as a predictor as well as a multiple linear regression analysis with both entered into the model.

Results

Complete data were available from 1110 graduates of 15 programs. **Table I** lists the deidentified programs with their mean scores on the ABP, USMLE step 1, and USMLE step 2 scores. Programs ranged in size from 28 to 176 residents. All but one was university affiliated. **Figures 1** and **2** scatterplots demonstrate the relationship between USMLE steps 1 and 2 scores with first time scores on the ABP examination.

Table II compares the examination performance of our entire group of 1100 trainees with general populations of test takers on the three examinations. The means and 25th to 75th percentile ranges are comparable between our sample and the general populations.

The relevant assumptions of hierarchical linear regression were tested.[5] An examination of correlations (data not shown) revealed that independent variables (USMLE step 1 and 2) were correlated. However, as the collinearity statistics (tolerance and variance inflation factor) were all within accepted limits, the assumption of multicollinearity was deemed to have been met.[5]

Table I. Programs and mean scores

Programs*	ABP†	USMLE step 1†	USMLE step 2†
1	214.32 (16.13)	221.09 (17.32)	237.00 (18.80)
2	206.99 (19.65)	229.96 (19.50)	238.02 (18.77)
3	206.94 (21.57)	222.17 (20.53)	234.92 (22.68)
4	206.82 (19.37)	233.09 (17.12)	241.73 (18.93)
5	205.03 (20.34)	225.31 (20.97)	235.98 (20.10)
6	201.42 (15.41)	214.6 (21.99)	224.18 (22.08)
7	199.11 (20.47)	213.97 (20.35)	226.03 (21.16)
8	198.00 (16.46)	213.39 (20.31)	222.54 (20.50)
9	196.80 (22.67)	208.03 (22.16)	220.18 (24.48)
10	194.76 (22.06)	207.64 (18.08)	208.94 (16.18)
11	193.4 (25.68)	207.52 (27.85)	219.69 (26.51)
12	192.89 (20.93)	211.3 (21.06)	219.20 (22.65)
13	191.06 (17.23)	200.67 (20.65)	212.08 (22.26)
14	187.84 (18.43)	213.05 (20.09)	213.37 (24.03)
15	186.13 (20.62)	209.18 (22.07)	211.74 (23.80)

*Listed in order of ABP examination mean score.
†Means (SD).

The interaction of USMLE step 1 and 2 scores predicted first-time ABP results, $F(1,1007.70) = 6.44$, $P = .011$. Individually, USMLE step 1 and 2 were not significant predictors of ABP performance (step 1: $F[1,1008.48] = 1.86$, $P = .173$; step 2: $F[1,1007.86] = .017$, $P = .896$). This interaction was broken down by conducting a multilevel model by residency program. This analysis indicates that performance on both USMLE step examinations predicts first-time ABP results ($b = .002$, $t = 2.54$, $P = .011$).

Based on the hierarchical linear model, the significance of the variance estimate of ABP performance by program was 9.89 ($z = 1.97$, $P = .049$). Although the level of significance is slightly significant, specific program influence on ABP results is minor.

Further analysis using linear regression was used to determine the relationship of USMLE examination scores with ABP results. **Table III** (available at www.jpeds.com) displays the USMLE step 1 and 2 and ABP results along with the results of the multiple regression analysis. For 9 programs, USMLE step 2 results were statistically significant predictors of ABP performance as evidenced by the *t* test. USMLE step 1 was ultimately not significant in the final model, indicating a lack of significance to the overall regression model. Upon further analysis of results in **Table IV** (available at www.jpeds.com), the variance of the multiple regression model which included USMLE steps 1 and 2 was not much different than for USMLE step 2 alone. For example, Dartmouth's R^2 for the multiple regression model was 0.628, step 2 alone was 0.624 and step 1 was 0.300. Five programs demonstrated a significant interaction between USMLE step 1 and 2 that contributed to the statistically significant association with ABP results. With only 1 training program lacking statistical significance, the results of the multiple regression analyses reinforce the results of the hierarchical linear model, indicating that USMLE step 2 is a better predictor for ABP performance than USMLE step 1 or a combination of the 2.

In light of the change in scoring of the ABP examination during the period studied, we reran these analyses and generated