

Predicting Performance on the American Board of Surgery Qualifying and Certifying Examinations

A Multi-institutional Study

Christian de Virgilio, MD; Arezou Yaghoubian, MD; Amy Kaji, MD, PhD; J. Craig Collins, MD, MBA; Karen Deveney, MD; Matthew Dolich, MD; David Easter, MD; O. Joe Hines, MD; Steven Katz, MD; Terrence Liu, MD, PhD; Ahmed Mahmoud, MD; Marc L. Melcher, MD, PhD; Steven Parks, MD; Mark Reeves, MD; Ali Salim, MD; Lynette Scherer, MD; Danny Takanishi, MD; Kenneth Waxman, MD

Background: We sought to determine whether US Medical Licensing Examination (USMLE) Step 1 score, American Board of Surgery (ABS) In-Training Examination (ABSITE) score, and other variables are associated with failing the ABS qualifying and certifying examinations. Identifying such factors may assist in the early implementation of an academic intervention for at-risk residents.

Design: Retrospective review.

Setting: Seventeen general surgery training programs in the western United States.

Participants: Six hundred seven residents who graduated in 2000-2007.

Main Outcome Measures: First-time pass rates on the qualifying and certifying examinations, US vs non-US medical school graduation, USMLE Steps 1 and 2 scores, ABSITE scores, operative case volume, fellowship training, residency program type, and mandatory research.

Results: The first-time qualifying and certifying exami-

nation pass rates for the 607 graduating residents were 78% and 74%, respectively. On multivariable analysis, scoring below the 35th percentile on the ABSITE at any time during residency was associated with an increased risk of failing both examinations (odds ratio, 0.23 [95% confidence interval, 0.08-0.68] for the qualifying examination and 0.35 [0.20-0.61] for the certifying examination), as was scoring less than 200 on the USMLE Step 1 (0.36 [0.21-0.62] for the qualifying examination and 0.62 [0.42-0.93] for the certifying examination). A mandatory research year was associated with an increased likelihood of passing the certifying examination (odds ratio, 3.3 [95% confidence interval, 1.6-6.8]).

Conclusions: Residents who are more likely to fail the ABS qualifying and certifying examinations can be identified by a low USMLE Step 1 score and by poor performance on the ABSITE at any time during residency. These findings support the use of the USMLE Step 1 score in the surgical residency selection process and a formal academic intervention for residents who perform poorly on the ABSITE.

Arch Surg. 2010;145(9):852-856

FOLLOWING SUCCESSFUL completion of a general surgery residency, board certification is a 2-step process that requires passing a written qualifying examination followed by an oral certifying examination administered

See Invited Critique at end of article

by the American Board of Surgery (ABS). In preparation for these examinations, the ABS administers an annual in-training examination (ABSITE) to all residents. Recently, the ABS began publishing online the 5-year first-time pass rates for all surgical residency programs in the United States. First-time pass rates are used by the

Accreditation Council of Graduate Medical Education (ACGME) as one of the measures of the educational success of a surgical residency. It is a requirement that at least 65% of surgical residents in a residency program pass both the certifying and qualifying examinations the first time. Given the importance of these examinations for surgical residency programs, the purpose of this multi-institutional study was to determine whether residents at risk of failing could be identified early, thus allowing for early academic intervention.

METHODS

The study was approved by the Human Subjects Committee of the Los Angeles Biomedical Research Institute at Harbor-UCLA Medi-

Author Affiliations are listed at the end of this article.

cal Center. Program directors from 17 residency programs in the western United States were contacted and surveyed. Variables for graduates from 2000-2007 were collected, including medical education (US vs non-US medical school), US vs non-US citizens, US Medical Licensing Examination (USMLE) Steps 1 and 2 scores, ABSITE scores (reported as national percentiles) for each postgraduate year (PGY), ABSITE score during the chief year (the score residents achieved in their final year of residency), final total major operative case volume, whether fellowship training was pursued (ACGME-accredited or non-ACGME-accredited post-general surgery residency graduate education), and type of fellowship. Self-reported variables collected from each program director included surgical residency program type (university program [defined by whether the program was at a university], community program [having no university affiliation], or hybrid program [a community-based program with a university affiliation]), whether the program held specific preparatory courses for the qualifying and certifying examinations, whether the program had a remediation program for residents with a low ABSITE score, whether the program had a minimal threshold performance score for the ABSITE, and whether the program included at least 1 year of mandatory research.

Resident data were collected in a database (Excel; Microsoft Corporation, Redmond, Washington) and translated into a native SAS format using DBMS/Copy software (Dataflux Corporation, Cary, North Carolina). Analyses were conducted using SAS statistical software, version 9.1 (SAS Institute Inc, Cary, North Carolina). Descriptive statistics were calculated for all variables. When appropriate, numerical variables were compared using the nonparametric Wilcoxon rank sum test and are reported as medians with interquartile ranges. Categorical or nominal variables were compared using the χ^2 test or Fisher exact test, as appropriate. Univariate predictors that were found to be statistically significantly associated (defined by $P < .05$) with passing the qualifying and certifying examinations were selected for multivariable logistic regression models. The ABSITE and USMLE Step 1 threshold cutoff scores that best predicted passing the qualifying and certifying examinations were investigated by comparing the χ^2 likelihood ratios of multivariable logistic regression models using different cutoff points for both the ABSITE and USMLE Step 1 scores (ABSITE scores were categorized as <25 , <30 , <35 , <40 , etc, and USMLE Step 1 scores were categorized as <190 , <200 , <210 , <220 , etc).

To account for potential correlations among residents at the same training program, a cluster analysis was performed using generalized estimating equations and an exchangeable covariance matrix.

A post-hoc sensitivity analysis was performed comparing models that contained the variable "presence of a mandatory research year" because this was affirmative in only 3 of the 17 training programs and because this variable was defined at the program level rather than at the individual resident level. For example, residents who trained at programs where research was not mandatory may have performed research; however, they would have been classified as having trained at a program *without* a mandatory research year.

RESULTS

RESIDENT CHARACTERISTICS

A total of 607 residents graduated from the 17 programs. The program and resident characteristics are given in **Table 1**. Ninety percent of the residents were US citizens, and 88% were graduates of US medical schools. The

Table 1. Descriptive Analysis of Entire Data Set

Characteristic	No. (%)
Program Characteristics (n = 17)	
Mandatory research year by program	
Yes	3 (18)
No	14 (82)
Program type	
Community	7 (41)
University	7 (41)
Hybrid	3 (18)
Remediation program for low ABSITE score	11 (65)
Formal course for written examination	9 (53)
Formal course for oral examination	12 (71)
Resident Characteristics (n = 607)	
Fellowship	
Yes	362 (60)
No	184 (30)
Unknown	62 (10)
Graduate of US medical school	536 (88)
Median USMLE score, No. (IQR)	
Step 1	219 (206-232)
Step 2	211 (191-226)
Median operative case volume	988 (832-1133)
Median ABSITE score by PGY, No. (IQR)	
PGY 1	67 (48-86)
PGY 2	57 (35-80)
PGY 3	57 (33-80)
PGY 4	56 (32-79)
PGY 5	56 (32-78)
PGY 6	57 (33-82)
PGY 7	59 (41-86)
First-time written examination pass rate	475 (78)
First-time oral examination pass rate	447 (74)

Abbreviations: ABSITE, American Board of Surgery In-Training Examination; IQR, interquartile range; PGY, postgraduate year; USMLE, US Medical Licensing Examination.

median USMLE Step 1 score was 219 (83% of Step 1 scores were available), and the median Step 2 score was 211 (only 38% of Step 2 scores were available). The median ABSITE national percentiles for each PGY level were as follows: 67th percentile for the PGY 1 level, 57th percentile for both the PGY 2 and PGY 3 levels, and 56th percentile for both the PGY 4 and PGY 5 levels. The median operative case volume was 988. Three hundred sixty-two residents (60%) pursued fellowship training, with vascular surgery (chosen by 10% of residents) and cardiothoracic surgery (chosen by 9%) as the most frequent choices. The first-time pass rate was 78% for the qualifying examination and 74% for the certifying examination.

UNIVARIATE ANALYSIS FOR QUALIFYING EXAMINATION

Factors associated with a higher likelihood of a first-time pass for the qualifying examination (**Table 2** and **Table 3**) included USMLE Step 1 score (median score of 222 for those who passed vs 204 for those who did not; $P < .001$), USMLE Step 2 score (median score of 213 for those who passed vs 193 for those who did not; $P < .001$), ABSITE national percentile scores in PGY 1 through PGY 4 and the chief year (all $P < .001$), the pres-

Table 2. Univariate Analysis for Categorical Variables and Qualifying Examination vs Certifying Examination

Variable	Qualifying Examination		Certifying Examination	
	Pass Rate, %	P Value	Pass Rate, %	P Value
Program type				
Community	84	.82	85	.30
University	89		90	
Hybrid	85		81	
Written board preparation				
Yes	86	.33	86	.50
No	89		89	
Oral board preparation				
Yes	89	.20	88	.70
No	84		87	
Remediation program for low ABSITE score				
Yes	93	<.001	89	.20
No	79		85	
Medical school training				
US graduate	87	>.99	88	>.99
Non-US graduate	86		89	
Mandatory research year				
Yes	93	.03	92	.08
No	85		86	
Postgraduate fellowship				
Yes	88	.30	88	.80
No	85		88	
Unknown	86		71	

Abbreviation: ABSITE, American Board of Surgery In-Training Examination.

Table 3. Univariate Analysis for Numerical Variables and Qualifying Examination vs Certifying Examination^a

Variable	Qualifying Examination			Certifying Examination		
	Fail	Pass	P Value	Fail	Pass	P Value
ABSITE percentile						
PGY 1	45	70	<.001	63	69	.10
PGY 2	34	61	<.001	61	59	.40
PGY 3	32	62	<.001	40	62	.001
PGY 4	24	60	<.001	51	69	.02
Chief year ^b	72	83	<.001	45	59	.003
Operative case volume	911	1000	.05	976	1000	.40
USMLE score						
Step 1	204	222	<.001	212	221	.001
Step 2	193	213	<.001	209	213	.20

Abbreviations: ABSITE, American Board of Surgery In-Training Examination; PGY, postgraduate year; USMLE, US Medical Licensing Examination.

^aFail and pass data are given as median values.

^bFor residents undertaking more than 5 years of training, the chief year was defined as the last year of residency.

ence of a remediation program for a low ABSITE score (93% pass rate with such a program vs 79% without; $P < .001$), and a mandatory research year during residency (93% pass rate with research vs 85% without; $P = .03$). Factors not associated with an improved qualifying examination performance included program type, having a formal program for qualifying and certifying examination preparation, citizenship status, being a graduate of a US medical school, and fellowship status.

UNIVARIATE ANALYSIS FOR CERTIFYING EXAMINATION

Factors associated with a higher likelihood of a first-time pass for the certifying examination included USMLE

Step 1 score (median score of 221 for those who passed vs 212 for those who did not; $P = .001$) and ABSITE national percentile scores in PGY 3 ($P = .001$), PGY 4 ($P = .02$), and the chief year ($P = .003$).

MULTIVARIABLE ANALYSIS

After adjusting for the potential for clustering among residents at each of the programs and for covariates that were found to be significant in the univariate analysis (USMLE Step 1 scores, ever scoring below a certain ABSITE score, the program type, and whether there was a remediation program, an oral board preparation program, a written board preparation program, and a mandatory research year), we found the following 2 factors to be consis-

Table 4. Multivariable Models for Passing the ABS Qualifying and Certifying Examinations

Variable	Qualifying Examination ^a	P Value	Certifying Examination ^a	P Value
Mandatory research	Not predictive		3.3 (1.6-6.8)	<.001
ABSITE score <35th percentile ^b	0.23 (0.08-0.68)	.008	0.35 (0.20-0.61)	<.001
USMLE Step 1 score <200	0.36 (0.21-0.62)	<.001	0.62 (0.42-0.93)	.02

Abbreviations: ABS, American Board of Surgery; ABSITE, American Board of Surgery In-Training Examination; USMLE, US Medical Licensing Examination.

^aData are given as odds ratio (95% confidence interval).

^bAt any time during residency.

tently predictive of failing the qualifying and certifying examinations: scoring below the 35th percentile on the ABSITE at any time during residency ($P=.008$ for the qualifying examination and $P<.001$ for the certifying examination) and scoring less than 200 on the USMLE Step 1 ($P<.001$ for the qualifying examination and $P=.02$ for the certifying examination). A mandatory research year was associated with an increased likelihood of passing the certifying examination ($P<.001$) (**Table 4**).

SENSITIVITY ANALYSIS

When we removed the program-level predictor of a mandatory research year, the coefficient estimates for the ABSITE and USMLE scores hardly changed, indicating the stability and strength of the association between these 2 predictors and pass rates for individual residents. In contrast, the coefficients for other program-level variables, such as the presence of a remediation program, significantly changed when predicting passage of the qualifying examination, demonstrating near statistical significance (odds ratio, 3.1; $P=.053$).

COMMENT

The present multi-institutional study of 17 general surgery residency programs on the West Coast found that the USMLE Step 1 score and ABSITE performance at any PGY level predicted performance on the ABS qualifying and certifying examinations. In addition, programs with mandatory research had a higher pass rate on the certifying examination. On multivariable analysis, a USMLE Step 1 score below 200 and scoring below the 35th percentile on the ABSITE at any PGY level predicted failing both the qualifying examination and the certifying examination. A mandatory research year was associated with increased likelihood of passing the certifying examination. It is unclear why a mandatory research year improves the certifying examination pass rate; it may simply be a reflection of the resident having an extra year of experience. This latter finding should be viewed with caution because only 3 training programs had a mandatory research year.

The USMLE is one of several criteria used by surgery programs for the residency selection process. However, there are few data regarding the utility of the USMLE for predicting passage of the ABS examinations. Osland and colleagues¹ recently reported their single-institution experience. Similar to the present study, they noted a Step 1 threshold score of 200 and a Step 2 threshold score of

186 for passing both the qualifying and certifying examinations the first time. Residents scoring 200 or higher on Step 1 had an 86% pass rate vs only a 60% pass rate for those with a score of less than 200 ($P=.02$). For Step 2, residents scoring 186 or higher had an 80% pass rate on both examinations vs only a 40% pass rate for those below this threshold ($P=.045$). Thus, the present study validates the utility of the USMLE as one criterion in the residency selection process. Of note, a task force was recently commissioned to evaluate whether the USMLE was meeting its original objectives. Recent suggestions include a recommendation of making the USMLE a pass/fail test.² Although the USMLE should not be used as a sole criterion for residency selection, in light of the present findings we believe that surgery residency programs should lobby to ensure that USMLE scores remain available.

The correlation between ABSITE performance in the chief resident year and success on the ABS qualifying examination has consistently been strong; however, few data are available regarding an association between ABSITE scores in earlier postgraduate years and qualifying examination performance or such an association with the certifying examination.^{3,4} In a previous single-institution study, de Virgilio et al⁵ demonstrated that scoring below the 30th percentile at any time during residency predicted subsequent failure on the qualifying examination. Osland and colleagues¹ noted that an ABSITE score above the 50th percentile for PGYs 1 and 3 and above the 33rd percentile for PGYs 4 and 5 were predictive of passing a subsequent qualifying examination. The present study confirmed an association between ABSITE performance and passing the qualifying examination at each PGY level. In addition, ABSITE performance at PGYs 3 and 4 predicted passing the certifying examination.

The association between ABSITE performance early in residency and subsequent qualifying and certifying examination performance creates opportunities for meaningful educational endeavors. Several studies have shown that a program of weekly reading combined with examinations can significantly improve ABSITE scores.⁵⁻⁸ Other investigators have shown that specifically targeting residents who have a low ABSITE score via a remediation program likewise improves ABSITE performance.⁹ Specific ABSITE preparation programs have recently been shown to pay dividends with respect to ABS pass rates. In one recent study,¹⁰ pass rates on the qualifying and certifying examinations increased after instituting such a program. The authors noted that, for each year of exposure to the educational program, the odds ratio for pass-

ing the combined qualifying and certifying examinations was 2.2 ($P=.04$). Borman¹⁰ demonstrated that remediation for a low ABSITE score resulted in improved performance on the qualifying examination. In the present study, a preparation course for the certifying and qualifying examinations was not associated with improvement in performance on either test. However, a remediation program for a low ABSITE score was associated with an increased likelihood of passing the qualifying examination. It is our opinion that the key element for success in these endeavors is not so much the content of the preparatory courses, per se, but rather stimulating the resident to read regularly and instilling a lifelong habit of self-study. Although an ABSITE remediation program was not found to be predictive of passing the ABS examinations in the present study, this may also represent a misclassification. Residents who do poorly on the ABSITE may have been assigned to a self-study program or a written board preparatory course by their program director, which would not have been considered an “official” remediation.

There is a paucity of data regarding other variables that affect ABS qualifying and certifying examination pass rates. Osland et al¹ noted that Alpha Omega Alpha honors, honors in medical school on a surgery rotation, interview scores, resident rank list number, and faculty evaluations were not predictive of success. The present study is noteworthy because of other variables that did not predict pass rates: having US citizenship status and whether the resident was a US or non-US medical school graduate did not affect pass rates. The program type—whether it was a university, community, or hybrid program—also did not affect pass rates.

In summary, we demonstrated that ABSITE and USMLE scores are predictive of success on the ABS qualifying and certifying examinations. The findings of the study create an opportunity for early identification of residents who are at increased risk of failing these examinations, with the potential for remediation. Furthermore, these results lend support to using the USMLE score as one tool in the surgical resident selection process.

Accepted for Publication: April 9, 2010.

Author Affiliations: Departments of Surgery, Harbor-UCLA Medical Center, Torrance (Drs de Virgilio, Yaghoubian, and Kaji), Kaiser Permanente Los Angeles Medical Center, Los Angeles (Dr Collins), University of California, Irvine (Dr Dolich), University of California, San Diego (Dr Easter), UCLA (University of California, Los Angeles) David Geffen School of Medicine, Los Angeles (Dr Hines), Huntington Memorial Hospital, Pasadena (Dr Katz), University of California, San Francisco (UCSF) East Bay, Oakland (Dr Liu), San Joaquin Community Hospital, Stockton (Dr Mahmoud), Stanford University, Palo Alto (Dr Melcher), UCSF, Fresno (Dr Parks), Loma Linda University, Loma Linda (Dr Reeves), Cedars-Sinai Medical Center, Los Angeles (Dr Salim), University of Cali-

fornia Davis Medical Center, Davis (Dr Scherer), and Santa Barbara Cottage Hospital, Santa Barbara (Dr Waxman), California; Oregon Health and Science University, Portland (Dr Deveney); and University of Hawaii, Honolulu (Dr Takanishi).

Correspondence: Christian de Virgilio, MD, Division of Vascular Surgery, Harbor-UCLA Medical Center, 1000 W Carson St, Campus Box 25, Bldg 1-E, Torrance, CA 90509 (cdevirgilio@labiomed.org).

Author Contributions: *Study concept and design:* de Virgilio, Dolich, Hines, and Waxman. *Acquisition of data:* de Virgilio, Yaghoubian, Collins, Deveney, Dolich, Easter, Hines, Katz, Liu, Mahmoud, Melcher, Parks, Reeves, Salim, Scherer, Takanishi, and Waxan. *Analysis and interpretation of data:* de Virgilio and Kaji. *Drafting of the manuscript:* de Virgilio, Yaghoubian, and Takanishi. *Critical revision of the manuscript for important intellectual content:* de Virgilio, Kaji, Collins, Deveney, Dolich, Easter, Hines, Katz, Liu, Mahmoud, Melcher, Parks, Reeves, Salim, Scherer, and Waxman. *Statistical analysis:* de Virgilio, Yaghoubian, and Kaji. *Administrative, technical, and material support:* Easter, Hines, Mahmoud, Parks, Reeves, Scherer, Takanishi, and Waxman. *Study supervision:* de Virgilio.

Financial Disclosure: None reported.

Previous Presentation: This paper was presented at the 81st Annual Meeting of the Pacific Coast Surgical Association; February 16, 2010; Kapalua, Hawaii; and is published after peer review and revision.

REFERENCES

- Osland JS, Shellito JL, Helmer SD, Chang FC. American board of surgery examinations: can we identify surgery residency applicants who will pass the examinations on the first attempt? *J Surg Educ*. 2008;65(1):17. doi:10.1016/j.jsurg.2008.01.019.
- Committee to Evaluate the USMLE Program (CEUP). Comprehensive review of the USMLE: summary of the final report and recommendations. http://www.usmle.org/general_information/CEUP-Summary-Report-June2008.pdf. Accessed April 2009.
- Shetler PL. Observations on the American Board of Surgery In-Training examination, board results, and conference attendance. *Am J Surg*. 1982;144(3):292-294.
- Garvin PJ, Kaminski DL. Significance of the In-Training Examination in a surgical residency program. *Surgery*. 1984;96(1):109-113.
- de Virgilio C, Chan T, Kaji A, Miller K. Weekly assigned reading and examinations during residency, ABSITE performance, and improved pass rates on the American Board of Surgery Examinations. *J Surg Educ*. 2008;65(6):499-503.
- Mahmoud A, Andrus CH, Matolo NM, Ward CC. Directed postgraduate study result in quantitative improvement in American Board of Surgery In-Training Exam scores. *Am J Surg*. 2006;191(6):812-816.
- de Virgilio C, Stabile BE. Weekly reading assignments and examinations result in sustained improvement in American Board of Surgery In-Training Examination (ABSITE) scores. *Am Surg*. 2005;71(10):830-832.
- de Virgilio C, Stabile BE, Lewis RJ, Brayack C. Significantly improved American Board of Surgery In-Training Examination scores associated with weekly assigned reading and preparatory examinations. *Arch Surg*. 2003;138(11):1195-1197.
- Harthun NL, Schirmer BD, Sanfey H. Remediation of low ABSITE scores. *Curr Surg*. 2005;62(5):539-542.
- Borman KR. Does academic intervention impact ABS qualifying examination results? *Curr Surg*. 2006;63(6):367-372.