

Pulmonary and Critical Care In-Service Training Examination Score as a Predictor of Board Certification Examination Performance

Robert R. Kempainen¹, Brian J. Hess², Doreen J. Addrizzo-Harris³, Douglas C. Schaad⁴, Craig S. Scott⁴, Brian W. Carlin⁵, Robert C. Shaw, Jr.⁶, Lauren Duhigg⁷, and Rebecca S. Lipner⁷

¹Division of Pulmonary and Critical Care, Hennepin County Medical Center, University of Minnesota School of Medicine, Minneapolis, Minnesota; ²Hess Consulting, St. Nicolas, Quebec, Canada; ³Division of Pulmonary, Critical Care and Sleep Medicine, New York University School of Medicine, New York, New York; ⁴Department of Biomedical Informatics and Medical Education, University of Washington, Seattle, Washington; ⁵Division of Pulmonary and Critical Care, Drexel University, Pittsburgh, Pennsylvania; ⁶Applied Measurement Professionals, Inc., Olathe, Kansas; and ⁷American Board of Internal Medicine, Philadelphia, Pennsylvania

Abstract

Rationale: Most trainees in combined pulmonary and critical care medicine fellowship programs complete in-service training examinations (ITEs) that test knowledge in both disciplines. Whether ITE scores predict performance on the American Board of Internal Medicine Pulmonary Disease Certification Examination and Critical Care Medicine Certification Examination is unknown.

Objectives: To determine whether pulmonary and critical care medicine ITE scores predict performance on subspecialty board certification examinations independently of trainee demographics, program director competency ratings, fellowship program characteristics, and prior medical knowledge assessments.

Methods: First- and second-year fellows who were enrolled in the study between 2008 and 2012 completed a questionnaire encompassing demographics and fellowship training characteristics. These data and ITE scores were matched to fellows' subsequent scores on subspecialty certification examinations, program director ratings, and previous scores on their American Board of Internal Medicine Internal Medicine Certification Examination. Multiple linear regression and logistic regression were used to identify independent predictors of subspecialty certification examination scores and likelihood of passing the examinations, respectively.

Measurements and Main Results: Of eligible fellows, 82.4% enrolled in the study. The ITE score for second-year fellows was matched to their certification examination scores, which yielded 1,484 physicians for pulmonary disease and 1,331 for critical care medicine. Second-year fellows' ITE scores ($\beta = 0.24$, $P < 0.001$) and Internal Medicine Certification Examination scores ($\beta = 0.49$, $P < 0.001$) were the strongest predictors of Pulmonary Disease Certification Examination scores, and were the only significant predictors of passing the examination (ITE odds ratio, 1.12 [95% confidence interval, 1.07–1.16]; Internal Medicine Certification Examination odds ratio, 1.01 [95% confidence interval, 1.01–1.02]). Similar results were obtained for predicting Critical Care Medicine Certification Examination scores and for passing the examination. The predictive value of ITE scores among first-year fellows on the subspecialty certification examinations was comparable to second-year fellows' ITE scores.

Conclusions: The Pulmonary and Critical Care Medicine ITE score is an independent, and stronger, predictor of subspecialty certification examination performance than fellow demographics, program director competency ratings, and fellowship characteristics. These findings support the use of the ITE to identify the learning needs of fellows as they work toward subspecialty board certification.

Keywords: education; specialty boards; curriculum

(Received in original form January 7, 2016; accepted in final form February 2, 2016)

Supported by the Association of Pulmonary and Critical Care Medicine Program Directors (APCCMPD).

Author Contributions: R.R.K.: study conception and design; data acquisition, analysis, and interpretation; drafting, revision, and final approval of the manuscript; and takes responsibility for manuscript content; B.J.H.: study conception and design; data acquisition, analysis, and interpretation; revision and final approval of manuscript; and takes responsibility for statistical analysis; D.J.A.-H.: study conception and design, data acquisition and interpretation, and revision and final approval of the manuscript; D.C.S.: study conception and design; data acquisition, analysis, and interpretation; and revision and final approval of the manuscript; C.S.S.: study conception and design; data acquisition, analysis, and interpretation; and revision and final approval of the manuscript; B.W.C.: study conception and design, data acquisition and interpretation, and revision and final approval of the manuscript; R.C.S.: study conception and design, data acquisition and interpretation, and revision and final approval of the manuscript; L.D.: study design; data acquisition, analysis, and interpretation; revision and final approval of manuscript; and takes responsibility for statistical analysis; and R.S.L.: study conception and design, data acquisition and interpretation, and revision and final approval of the manuscript.

Correspondence and requests for reprints should be addressed to Robert R. Kempainen, M.D., Division of Pulmonary and Critical Care Medicine, Hennepin County Medical Center, University of Minnesota School of Medicine, G5, 701 Park Avenue, Minneapolis, MN 55415. E-mail: kempa001@umn.edu

This article has an online supplement, which is accessible from this issue's table of contents at www.atsjournals.org

Ann Am Thorac Soc Vol 13, No 4, pp 481–488, Apr 2016

Copyright © 2016 by the American Thoracic Society

DOI: 10.1513/AnnalsATS.201601-015OC

Internet address: www.atsjournals.org

The use of in-service training examinations (ITEs) to assess knowledge in anticipation of subsequent board certification examinations dates back to the 1960s (1). The ability of ITE scores to predict certification examination performance is well established for a variety of medical specialties, including internal medicine, neurology, surgery, family medicine, and pediatrics, among others (2–8). Although for decades participation in the internal medicine ITE has been nearly universal among internal medicine residency programs, internal medicine subspecialties only recently have developed ITEs for their fellowship programs. Recent studies indicate second-year fellows' scores on ITEs developed by the Infectious Diseases Society of America and American College of Rheumatology are useful predictors of subsequent American Board of Internal Medicine (ABIM) certification examination outcomes (9, 10).

The vast majority of fellows pursuing careers in pulmonary disease or critical care medicine enroll in combined programs leading to board eligibility in both disciplines. A near-universal goal of trainees in combined fellowship programs is to become ABIM certified in both pulmonary disease and critical care medicine. Combined pulmonary and critical care medicine fellowships are of 3 years' duration, with some fellows opting for additional years of research training. Entering trainees have completed medical school and a residency in internal medicine. The ABIM Internal Medicine Certification Examination is typically then completed early in fellowship. Before taking their two subspecialty board certification examinations, fellows must pass their ABIM Internal Medicine Certification Examination and receive satisfactory or higher performance ratings from their fellowship program directors on multiple aspects of patient care inclusive of the six Accreditation Council of Graduate Medical Education (ACGME) core competencies instituted in 2002.

Since 2007, the Association of Pulmonary and Critical Care Medicine Program Directors (APCCMPD) has provided fellows with an ITE that tests both pulmonary disease and critical care medicine knowledge (referred to hereinafter as *pulmonary and critical care ITE*). Currently, approximately 85% of trainees in U.S. combined pulmonary and critical care

medicine fellowship programs complete the pulmonary and critical care ITE in their first and/or second year of fellowship, but whether pulmonary and critical care ITE scores are useful for identifying trainees at risk of failing their ABIM subspecialty board certification examinations is unknown.

The aim of the present study was to determine the extent to which pulmonary and critical care ITE scores predict performance on the two ABIM certification examinations compared with other potential predictors of certification examination performance, including trainee demographics, prior performance on the ABIM Internal Medicine Certification Examination, and ACGME fellowship program directors' ratings of clinical competency. Unlike previous studies of internal medicine subspecialty ITEs, in this study we also assessed the predictive value of first-year fellows' ITE scores and fellowship training program characteristics for ABIM subspecialty board certification examination outcomes. Some of the results of this study were previously reported in the form of an abstract (11).

Methods

Study Participants

From 2008 to 2012, fellows in their first and second year of combined pulmonary disease and critical care medicine fellowship training were invited to participate in the study at the time they sat for the pulmonary and critical care ITE. Participating fellows then completed a brief questionnaire for collection of demographic information, characteristics of their fellowship program, and cumulative clinical experiences acquired during fellowship up to the time of taking the pulmonary and critical care ITE. These data were merged with ABIM assessments of the fellows, including scores on the ABIM Internal Medicine Certification Examination, ABIM Pulmonary Disease subspecialty board certification examination, ABIM Critical Care Medicine subspecialty board certification examination, as well as fellowship program director competency ratings. The analysis included results of participants' first attempt at the certification examinations.

American Board of Internal Medicine Internal Medicine Certification Examination

As noted above, fellows typically complete the Internal Medicine Certification Examination shortly after completing their internal medicine residency. The examination is comprised of 240 single best multiple-choice questions. The examination content blueprint is available on the ABIM website (<https://www.abim.org/~media/ABIM%20Public/Files/pdf/exam-blueprints/certification/internal-medicine.pdf>). Scores are based on responses to 200 of the questions; the remaining 40 questions are for field testing only. The examination is scored using item response theory, and overall scores are equated and standardized (mean, 500 [SD, 100]). The reliability coefficient for the scores is greater than or equal to 0.90 across administrations used in this study.

Fellowship Program Director Ratings

During the study years, program directors completed a standardized evaluation form for each fellow on an annual basis to qualitatively rate the fellow's performance on six ACGME competencies plus a rating of overall clinical competence. Ratings are based on a 9-point scale that includes three categories: unsatisfactory (1 – 3), satisfactory (4 – 6), and superior (7 – 9). At the end of fellowship, fellows must receive a rating of 4 or higher in all competencies to be eligible to take the ABIM subspecialty board certification examinations. In this study, analyses included ratings of three competencies: medical knowledge, overall clinical competency, and professionalism. Ratings for these competency areas ranged from 4 to 9, with the majority falling between 7 and 9. The other competencies, such as procedural skills, practice-based learning and improvement, and systems-based practice, were deemed less directly relevant to ABIM certification examination performance and therefore were not included in the analyses.

Pulmonary and Critical Care In-Service Training Examination

The pulmonary and critical care ITE is a computer-based examination developed by the APCCMPD with the financial support of the American College of Chest Physicians and the American Thoracic Society. APCCMPD contracted with Applied

Measurement Professionals, Inc. (Olathe, KS), for services that included content development facilitation, test form assembly, item-scoring validation, delivery of test scores, and technical documentation. A new version of the ITE is administered each year. The test is administered online to fellows in their academic institutions and is proctored by representatives at each institution. The examination includes 125 single best multiple-choice questions written by the APCCMPD In-service Committee, which is comprised of physicians, the majority of whom are program directors.

The pulmonary and critical care ITE content domain is described by a detailed list of learning objectives that pertain to 28 content areas relevant to pulmonary disease and critical care medicine (see online supplement). The 28 content areas fall into five categories (subscales): pulmonary diseases and disorders (40% of questions); critical care (40% of questions); general care (4% of questions); practice skills (12% of questions); and biology, physiology, and research (4% of questions). The pulmonary and critical care ITE content domain is based on the ABIM Pulmonary Disease and the ABIM Critical Care Medicine subspecialty board certification

examination blueprints. The reliability coefficient for the scores is greater than or equal to 0.80 across administrations. The APCCMPD In-service Committee reviews a preliminary item analysis each year of item responses that are inconsistent with total test scores and/or are challenged by examinees. Changes in scoring are made as directed by the committee.

American Board of Internal Medicine Pulmonary Disease and Critical Care Medicine Certification Examinations

Typically fellows complete one of these subspecialty board certification examinations late in the third year of fellowship training and the other certification examination approximately one year later. The examinations are secure, computer-based, and each is comprised of 240 single-best-answer multiple-choice questions. Scores are based on responses to 200 of the questions; the remaining 40 questions are for field testing only. The content blueprints for the examinations are available on the ABIM website (<http://www.abim.org/about/exam-information/exam-blueprints.aspx>). The examinations are scored using item response theory, and overall scores are equated and standardized (mean, 500 [SD, 100]). The

reliability coefficient for the scores is greater than or equal to 0.90 across administrations.

Statistical Analyses

For each of the three ABIM certification examinations, scores were equated and standardized to allow scores to be compared across years of administration for statistical analyses; the pulmonary and critical care ITE overall raw scores were equated and standardized using the linear equating method so that these scores could also be compared across the study years (12). Multiple regression analysis was used to identify predictors of ABIM subspecialty board certification examination scores. Variables in the analyses included pulmonary and critical care ITE score, ABIM Internal Medicine Certification Examination score, age, sex, country of birth (United States/Canada vs. international), medical school graduation country (United States/Canada vs. international), fellowship program director ratings, and fellowship training experiences. Specific training experiences included cumulative months spent on critical care and pulmonary consult rotations at the time of sitting for the pulmonary and critical care ITE, as well as anticipated average number of half days in an outpatient clinic and months spent on

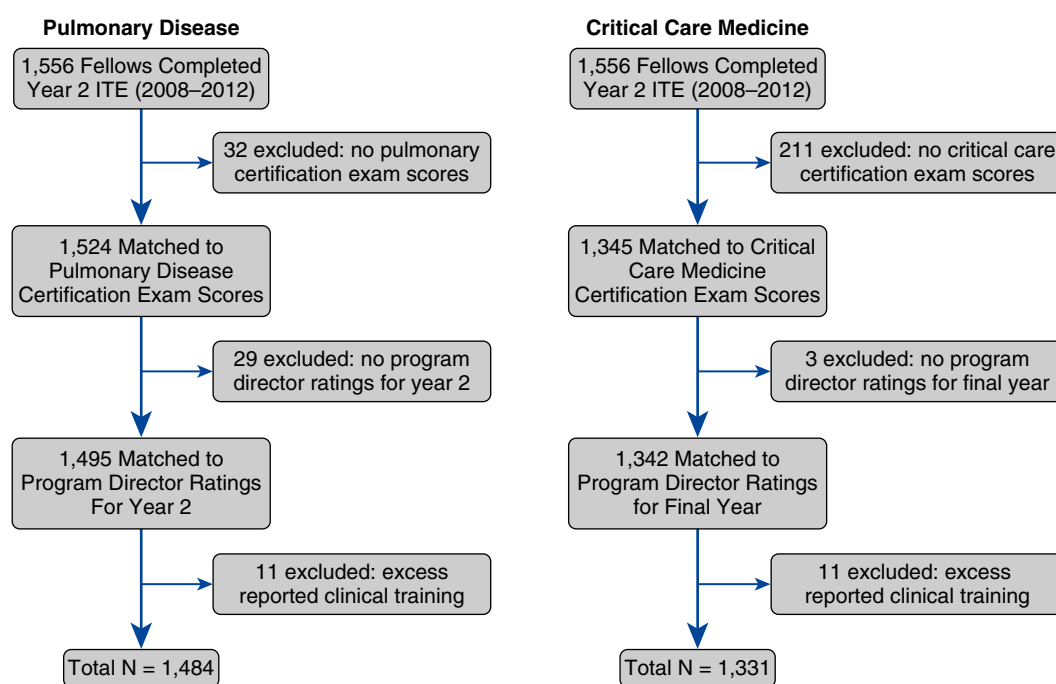


Figure 1. Trial profile for second-year fellows enrolled in the study. ITE = in-service training examination.

nonclinical/research activity during the first 3 years of fellowship.

For the primary analysis, we kept months of nonclinical/research activity as a continuous variable, but as a follow-up we dichotomized the variable (i.e., ≥ 12 mo [research-intensive program] vs. < 12 mo [clinical-focused program]) to determine if the results were similar when comparing these two types of fellowships. In addition, approximately 50–70% of the program directors' ratings were 8 or higher, depending on the competency area. Therefore, we chose to convert the ratings to dichotomous variables wherein ratings of 8 and 9 were considered "high" and ratings of 4–7 were considered "low." To determine which variables to retain as predictors in the regression models, the all-possible-regressions procedure was used, selecting the model with the highest R^2 value as our final model. The adjusted R^2 was calculated to measure how well the data fit the model and is interpreted as the percentage variance in ABIM subspecialty board certification examination scores that is accounted for by the set of predictors selected. We also calculated the variance inflation factor index to determine if multicollinearity (i.e., high correlations among the predictors) was a problem in the models. Standardized regression coefficients (β) were reported to describe and compare the relative magnitude of the predictive variables.

Logistic regression analysis was used to predict the probability of passing each of the two ABIM subspecialty board certification examinations using the same set of covariates. In addition, a restricted logistic model was fitted to generate the conditional probability of passing each of the ABIM certification examinations given only the pulmonary and critical care ITE score.

Statistical significance was assessed using an α level less than 0.05. All analyses were conducted using SAS 9.3 software (SAS Institute Inc., Cary, NC). The University of Minnesota Institutional Review Board approved the study. All data were deidentified and analyzed in the aggregate.

Results

A total of 82.4% of fellows taking the pulmonary and critical care ITE between 2008 and 2012 were enrolled in the study. Sex, age, and site of medical school

graduation (United States/Canada vs. international) did not differ significantly between participants and nonparticipants. Figure 1 summarizes the eligibility of enrolled second-year fellows for the analyses. Less than 5% of fellows were excluded from the analysis of predictors of performance on the ABIM Pulmonary Disease Certification Examination, whereas this value was 14% for the ABIM Critical Care Medicine Certification Examination, largely because insufficient time had elapsed from the 2012 pulmonary and critical care ITE for participants to have completed their ABIM Critical Care Medicine Certification Examination at the time this study was conducted.

Table 1 contains second-year fellows' demographic information as well as

characteristics of their fellowship training. Approximately 70% of participants were male, and about 53% had graduated from a U.S. or Canadian medical school. Second-year fellows' performance on the pulmonary and critical care ITE, ABIM Internal Medicine Certification Examination, and ABIM subspecialty board certification examinations are also summarized in Table 1. Of the 1,484 fellows who completed the ABIM Pulmonary Disease Certification Examination, 80 (5.39%) failed the examination on the first attempt. Of the 1,331 fellows who completed the ABIM Critical Care Medicine Certification Examination, 67 (5.03%) failed the examination on the first attempt.

Table 2 presents the results of the multiple regression analysis using the

Table 1. Demographics, training characteristics, and examination performance of second-year fellows in pulmonary disease and critical care medicine

	Pulmonary Disease* (n = 1,484)	Critical Care Medicine† (n = 1,331)
Sex, n (%)		
Male	1,040 (70.08%)	944 (70.92%)
Female	444 (29.92%)	387 (29.08%)
Age‡, yr, mean (SD)	34.1 (3.26)	35.1 (3.20)
Birth country, n (%)		
United States/Canada	775 (52.22%)	680 (51.09%)
International	709 (47.78%)	651 (48.91%)
Medical school graduation country, n (%)		
United States/Canada	797 (53.71%)	700 (52.59%)
International	687 (46.29%)	631 (47.41%)
Fellowship training, mean (SD)		
Months of nonclinical/research	9.58 (5.78)	9.41 (5.65)
Average half days in clinic	1.58 (0.74)	1.59 (0.75)
Months of critical care	7.83 (3.10)	7.85 (3.05)
Months of pulmonary consults	6.50 (3.12)	6.55 (3.15)
Program director ratings§, mean (SD)		
Medical knowledge	7.40 (0.99)	7.71 (1.01)
Professionalism	7.73 (0.97)	7.94 (1.01)
Overall clinical competence	7.51 (0.90)	7.80 (0.95)
Pulmonary and critical care in-service training examinations and certification examination (CE) scores, mean (SD)		
In-service training examination: first-year fellows	89.46 (7.95)	89.62 (7.88)
In-service training examination: second-year fellows	90.66 (7.62)	90.82 (7.52)
Internal medicine CE score	528.69 (68.03)	530.69 (66.11)
Pulmonary disease CE score	525.58 (80.87)	
Critical care medicine CE score		500.85 (77.80)

*Missing data: months of nonclinical/research, 4; average half days in clinic, 6; months of critical care, 6; months of pulmonary consults, 6; and program director ratings, 26.

†Missing data: months of nonclinical/research, 4; average half days in clinic, 6; months of critical care, 7; months of pulmonary consults, 6; program director ratings, 1; critical care medicine certification examination score, 1.

‡Age at time of taking the certification examination.

§On a 1–9 rating scale.

all-possible-regressions procedure to select the variables to be included in the final models. Multicollinearity was not judged to be a problem (variance inflation factor values were between 1 and 2 for both the pulmonary disease and critical care medicine models). The strongest independent predictors of scores on the ABIM Pulmonary Disease Certification Examination were second-year fellows' scores on the pulmonary and critical care ITE ($\beta = 0.24$, $P < 0.001$) and the ABIM Internal Medicine Certification Examination ($\beta = 0.49$, $P < 0.001$). The pulmonary and critical care ITE and ABIM Internal Medicine Certification Examination were similarly predictive of scores on the ABIM Critical Care Medicine Certification Examination (Table 2).

Male sex, graduation from a U.S./Canadian medical school, and fellowship program director ratings of medical knowledge were weak predictors of higher scores on the ABIM Pulmonary Disease Certification Examination. Male sex, birth country, medical school graduation country, fewer average number of half days in an

outpatient clinic, and program director ratings of medical knowledge were weak predictors of higher scores on the ABIM Critical Care Medicine Certification Examination. Number of months spent on nonclinical/research activity was not a significant predictor, regardless of whether it was a continuous or a dichotomous variable. The adjusted R^2 values were 0.48 and 0.45 for the pulmonary disease and critical care medicine models, respectively. That is, the set of predictors explained 48% and 45%, respectively, of the variance in scores on the two ABIM subspecialty board certification examinations.

Of note, the predictive value of pulmonary and critical care ITE scores among first-year fellows on subsequent ABIM subspecialty board certification examinations was comparable to second-year fellows' scores ($\beta = 0.26$ [$P < 0.001$] for the ABIM Pulmonary Disease Certification Examination [$n = 826$ scores]; $\beta = 0.24$ [$P < 0.001$] for the ABIM Critical Care Medicine Certification Examination [$n = 734$ scores]). As expected, fewer participants who completed the pulmonary and critical care ITE as first-year fellows had gone on to take

their ABIM certification examinations at the time of the analysis.

Table 3 summarizes the final logistic regression model of predictors of second-year fellows' passing each of the two ABIM certification examinations. Scores on the pulmonary and critical care ITE and ABIM Internal Medicine Certification Examination were significant predictors of passing each ABIM certification examination. Figure 2 demonstrates the conditional probability of second-year fellows' passing the two ABIM certification examinations based solely on their pulmonary and critical care ITE scores. The estimated probability of passing both ABIM certification examinations is approximately 95% for fellows achieving the mean equated score on the pulmonary and critical care ITE during these study years. Furthermore, fewer average number of half days in an outpatient clinic predicted a higher likelihood of passing the ABIM Critical Care Medicine Certification Examination, but with a wide confidence interval. Other fellowship training activities, including types of clinical rotations, nonclinical/research time, program director ratings, and fellow demographics

Table 2. Multiple regression analysis results predicting scores on the American Board of Internal Medicine Pulmonary Disease Certification Examination and Critical Care Medicine Certification Examination

Predictor	Unstandardized Coefficient		Standardized Coefficient	t Value	P Value
	β -Value	Standard Error of β -Value	β -Value		
Pulmonary Disease Certification Examination score of second-year fellows, $n = 1,472^*$					
ABIM Internal Medicine Certification examination score	0.58	0.03	0.49	21.79	<0.001
In-service training examination score	2.53	0.24	0.24	10.39	<0.001
Sex, 1 = male vs. 0 = female	12.06	3.39	0.07	3.55	<0.001
Medical school graduation country, 1 = United States/Canada vs. 0 = international	14.36	3.29	0.09	4.37	<0.001
Months of nonclinical/research	-0.49	0.29	-0.04	-1.73	0.08
Months of pulmonary consults	-0.82	0.51	-0.03	-1.61	0.11
Medical knowledge [†] , 1 = high vs. 0 = low	15.82	3.19	0.10	4.95	<0.001
Critical Care Medicine Certification Examination score of second-year fellows, $n = 1,292^\ddagger$					
ABIM Internal Medicine Certification examination score	0.51	0.03	0.44	17.71	<0.001
In-service training examination score	2.53	0.26	0.24	9.78	<0.001
Sex, 1 = male vs. 0 = female	25.93	3.60	0.15	7.20	<0.001
Birth country, 1 = United States/Canada vs. 0 = international	10.54	4.45	0.07	2.37	0.02
Medical school graduation country, 1 = United States/Canada vs. 0 = international	16.90	4.59	0.11	3.68	<0.001
Months of nonclinical/research	0.53	0.30	0.04	1.74	0.08
Average half days in clinic	-13.11	2.17	-0.13	-6.03	<0.001
Medical knowledge [†] , 1 = high vs. 0 = low	9.15	3.43	0.06	2.67	0.01

Definition of abbreviation: ABIM = American Board of Internal Medicine.

*Twelve observations were excluded from the regression analysis because of missing values.

[†]High is defined as a program director rating of 8 or 9; low is defined as a program director rating of 7 or less.

[‡]Thirty-nine observations were excluded from the regression analysis because of missing values.

Table 3. Logistic regression analysis results predicting passing status on the American Board of Internal Medicine Pulmonary Disease Certification Examination and Critical Care Medicine Certification Examination

Predictor	β -Value	Standard Error of β -Value	P Value	OR (95% CI)
Pulmonary Disease Certification Examination of second-year fellows (N = 1,472)*				
ABIM Internal Medicine Certification examination score	0.01	0.00	<0.001	1.01 (1.01–1.02)
In-service training examination score	0.11	0.02	<0.001	1.12 (1.07–1.16)
Sex, 1 = male vs. 0 = female	–0.19	0.30	0.52	0.83 (0.46–1.48)
Birth country, 1 = United States/Canada vs. 0 = international	0.27	0.32	0.39	1.31 (0.71–2.44)
Medical school graduation country, 1 = United States/Canada vs. 0 = international	0.26	0.33	0.43	1.30 (0.68–2.46)
Months of nonclinical/research	–0.01	0.02	0.74	0.99 (0.95–1.04)
Months of pulmonary consults	–0.04	0.04	0.29	0.96 (0.88–1.04)
Average half days in clinic	0.11	0.18	0.54	1.12 (0.79–1.58)
Medical knowledge [†] , 1 = high vs. 0 = low	0.45	0.48	0.34	1.58 (0.62–4.00)
Professionalism [†] , 1 = high vs. 0 = low	0.12	0.33	0.73	1.12 (0.58–2.16)
Overall clinical competence [†] , 1 = high vs. 0 = low	0.40	0.49	0.41	1.49 (0.57–3.88)
Critical Care Medicine Certification Examination of second-year fellows, n = 1,293[‡]				
ABIM Internal Medicine Certification examination score	0.01	0.00	<0.001	1.01 (1.01–1.02)
In-service training examination score	0.07	0.02	<0.001	1.08 (1.03–1.12)
Sex, 1 = male vs. 0 = female	0.46	0.29	0.12	1.58 (0.89–2.81)
Birth country, 1 = United States/Canada vs. 0 = international	0.60	0.34	0.08	1.81 (0.93–3.53)
Medical school graduation country, 1 = United States/Canada vs. 0 = international	0.24	0.35	0.49	1.27 (0.65–2.50)
Months of nonclinical/research	0.00	0.03	0.93	1.00 (0.95–1.06)
Months of critical care	0.02	0.04	0.68	1.02 (0.94–1.11)
Average half days in clinic	–0.43	0.16	0.01	0.65 (0.47–0.90)
Medical knowledge [†] , 1 = high vs. 0 = low	0.45	0.42	0.29	1.57 (0.68–3.59)
Professionalism [†] , 1 = high vs. 0 = low	–0.30	0.39	0.45	0.74 (0.34–1.61)
Overall clinical competence [†] , 1 = high vs. 0 = low	0.19	0.48	0.69	1.21 (0.48–3.07)

Definition of abbreviations: ABIM = American Board of Internal Medicine; CI = confidence interval; OR = odds ratio.

*Twelve observations were excluded from the regression analysis because of missing values.

[†]High is defined as a program director rating of 8 or 9; low is defined as a program director rating of 7 or less.

[‡]Thirty-eight observations were excluded from the regression analysis because of missing values.

were not predictive of passing the ABIM certification examinations.

Finally, among first-year fellows, performance on the pulmonary and critical care ITE and ABIM Internal Medicine Certification Examination were also independently predictive of passing each of the two subsequent ABIM certification examinations (odds ratios, 1.10 and 1.02, respectively, for both the ABIM Pulmonary Disease Certification Examination and the ABIM Critical Care Medicine Certification Examination).

Discussion

This national study confirms that a widely used ITE that tests both pulmonary disease and critical care medicine content is a valid measure of medical knowledge for fellows preparing for their subsequent ABIM subspecialty board certification

examinations. Scores on the pulmonary and critical care ITE independently predicted performance on subsequent ABIM certification examinations in these two internal medicine subspecialties, even after controlling for prior assessments of medical knowledge, fellow demographics, and fellowship program characteristics. Scores on the pulmonary and critical care ITE and ABIM Internal Medicine Certification Examination were the strongest predictors of scores on, as well as passing of, the ABIM certification examinations. Of note, the scores of first- and second-year fellows on the pulmonary and critical care ITE were equally predictive. A subset of demographic variables, fellowship program director ratings, and characteristics of fellowship programs were weaker independent predictors of subsequent performance on the ABIM certification examinations. Together, these findings suggest that the pulmonary

and critical care ITE serves as a valuable formative tool for assessing fellows' readiness for future ABIM certification examinations.

The findings of the present study are consistent with those of prior studies demonstrating a positive relationship between results on ITEs and subsequent performance on certification examinations in a variety of medical fields (2–8) as well as internal medicine subspecialties (9,10). Unlike the aforementioned analyses of ITEs in infectious disease and rheumatology, in the present study we assessed the predictive value of first-year fellows' ITE scores on ABIM subspecialty board certification examination performance. Despite first-year fellows' having completed less of fellowship training, the predictive value of their scores matched that of second-year fellows. As previously noted, scores on the ABIM Internal Medicine Certification Examination also strongly predicted ABIM certification examination

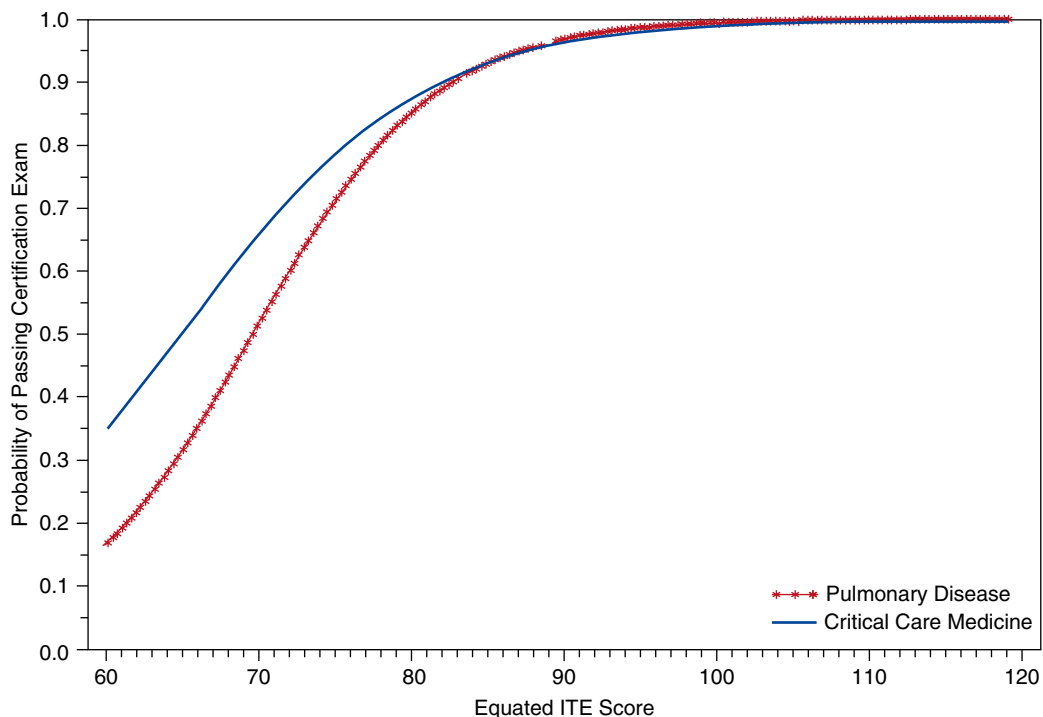


Figure 2. Conditional probability of passing the American Board of Internal Medicine Pulmonary Disease Certification Examination and the Critical Care Medicine Certification Examination based on Pulmonary and Critical Care In-Training Examination score. The x axis denotes scores on the in-service training examination that were equated across the study years. ITE = in-service training.

performance. Similarly, studies of infectious disease and rheumatology fellows revealed scores on the ABIM Internal Medicine Certification Examination and on step 1 of the United States Medical Licensing Examination predicted fellows' performance on their subspecialty board certification examinations (9, 10).

To our knowledge, this study is the first exploration of whether specific characteristics of fellowship training programs predict performance on subsequent ABIM certification examinations. The number of months that second-year fellows had spent on inpatient pulmonary consult services versus intensive care services was similar at the time they completed the pulmonary and critical care ITE, and neither was predictive of subsequent performance on the ABIM certification examinations. More time dedicated to an outpatient clinic was associated with worse performance on the ABIM Critical Care Medicine Certification Examination but did not predict results on the ABIM Pulmonary Disease Certification Examination.

However, the effect of clinic time on ABIM Critical Care Medicine Certification Examination scores was modest and must be viewed in light of fellows' need to gain

expertise in evaluating patients with respiratory disease in an ambulatory setting. Relative to fellows at clinical fellowship programs, trainees at research-intensive programs dedicate fewer months to patient care. However, this difference in clinical training did not translate to worse performance on the ABIM certification examinations, as the number of months spent on nonclinical/research activity was not significantly predictive of examination scores or of passing the examinations. This finding is consistent with a previous study in which researchers found that the performance of internal medicine residents in research-intensive tracks on the ABIM Internal Medicine Certification Examination was similar to that of their peers in traditional tracks, despite their spending fewer months in clinical training (13).

Scores on the pulmonary and critical care ITE and ABIM Internal Medicine Certification Examination predicted performance on ABIM subspecialty board certification examinations, but various measures of clinical experience did not, which raises the possibility that test-taking ability is the more important determinant of success on certification examinations. This is unlikely to be the case. First-year pulmonary

and critical care medicine fellows have already accumulated a wealth of knowledge and experience relevant to their specialties during medical school, internal medicine residency, and the first 9 months of fellowship by the time they take the pulmonary and critical care ITE. The incremental gain in knowledge acquired during the remainder of a fellowship is relatively small, and this likely explains why the ITE scores of first- and second-year fellows were equally predictive of ABIM certification examination performance, whereas the number of months fellows spent on various types of clinical rotations and the type of fellowship (research focus vs. clinical focus) were not predictive. Furthermore, the predictive value of prior examination performance is limited in that the statistical models in the present study and previous studies accounted for only about 50% of the variance in test scores.

Previous studies from a variety of medical fields showed a correlation between program director ratings of resident clinical competence and subsequent performance on board certification examinations (14, 15). However, in the present study we found that fellowship director ratings of medical knowledge did not independently predict

passing the two ABIM certification examinations and only weakly predicted scores on these certification examinations. In prior studies, rheumatology and infectious disease fellowship director ratings similarly did not predict passing and only weakly predicted scores on certification examinations (9, 10). We found that director ratings of overall clinical competence and professionalism did not independently predict scores or passing the ABIM certification examinations.

The limited range of values for some variables may account for their reduced capacity to predict performance. Program director competency ratings were uniformly high, especially for professionalism; about 70% of critical care medicine fellows received a rating of 8 or higher for professionalism. In 2015, assessments of fellows transitioned from the previous rating system to the ACGME Next Accreditation System, which requires programs to report whether fellows have achieved competency-based milestones as well as the program's board certification examination pass rates. Additional research is needed to determine whether ITE performance at the fellowship level can be used by

program directors to identify targets for curriculum improvement to meet these goals (16).

A limitation of our study is that not all fellows taking the pulmonary and critical care ITE were enrolled in the study. However, the results are likely generalizable because over 80% of eligible fellows participated, and, for the vast majority, their scores could be matched to subsequent performance on the two ABIM subspecialty board certification examinations. Furthermore, demographic characteristics of participants and nonparticipants did not differ significantly. Participants self-reported their clinical and nonclinical activities, and the accuracy of this information was not verified with individual programs.

However, there is reason to believe fellows would be able to recall this information with reasonable accuracy, given that it was collected at the time of pulmonary and critical care ITE administration. There are other factors that could affect ABIM certification examination performance that were not measured and included in the study (e.g., examination preparation method). Also, the model used to predict a passing outcome on each of the ABIM certification examinations based on

fellows' pulmonary and critical care ITE scores should be interpreted with caution because the predictive relationships were not very strong; in addition, the results cannot be directly generalizable to future examinees, because we equated ITE scores across the study years and the ABIM certification examination content and minimum passing scores can change over time. A recent study highlighted the pitfalls of excessive emphasis on general surgery ITEs (17).

In conclusion, this study establishes the pulmonary and critical care ITE as an independent predictor of fellows' performance on the ABIM subspecialty board certification examinations, which is consistent with previous evaluations of ITEs. The pulmonary and critical care ITE is a stronger predictor of subsequent ABIM certification examination scores and pass/fail status than fellow demographics, program director competency ratings, and fellowship program characteristics. These findings support the use of the pulmonary and critical care ITE to identify the learning needs of fellows as they work toward ABIM board certification. ■

Author disclosures are available with the text of this article at www.atsjournals.org.

References

- Hubbard JP, Furlow LT, Matson DD. An in-training examination for residents as a guide to learning. *N Engl J Med* 1967;276:448–451.
- Grossman RS, Fincher RM, Layne RD, Seelig CB, Berkowitz LR, Levine MA. Validity of the in-training examination for predicting American Board of Internal Medicine certifying examination scores. *J Gen Intern Med* 1992;7:63–67.
- Waxman H, Braunstein G, Dantzker D, Goldberg S, Lefrak S, Lichstein E, Ratzan K, Schiffman F. Performance on the internal medicine second-year residency in-training examination predicts the outcome of the ABIM certifying examination. *J Gen Intern Med* 1994;9:692–694.
- Babbott SF, Beasley BW, Hinchey KT, Blotzer JW, Holmboe ES. The predictive validity of the internal medicine in-training examination. *Am J Med* 2007;120:735–740.
- Juul D, Flynn FG, Gutmann L, Pascuzzi RM, Webb L, Massey JM, Dekosky ST, Foertsch M, Faulkner LR. Association between performance on Neurology In-Training and Certification Examinations. *Neurology* 2013;80:206–209.
- de Virgilio C, Yaghoubian A, Kaji A, Collins JC, Deveney K, Dolich M, Easter D, Hines OJ, Katz S, Liu T, et al. Predicting performance on the American Board of Surgery qualifying and certifying examinations: a multi-institutional study. *Arch Surg* 2010;145:852–856.
- O'Neill TR, Li Z, Peabody MR, Lybarger M, Royal K, Puffer JC. The predictive validity of the ABFM's in-training examination. *Fam Med* 2015;47:349–356.
- Althouse LA, McGuinness GA. The in-training examination: an analysis of its predictive value on performance on the general pediatrics certification examination. *J Pediatr* 2008;153:425–428.
- Grabovsky I, Hess BJ, Haist SA, Lipner RS, Hawley JL, Woodward S, Engleberg NC. The relationship between performance on the infectious diseases in-training and certification examinations. *Clin Infect Dis* 2015;60:677–683.
- Lohr KM, Clauser A, Hess BJ, Gelber AC, Valeriano-Marcet J, Lipner RS, Haist SA, Hawley JL, Zirkle S, Bolster MB; American College of Rheumatology Committee on Rheumatology Training and Workforce Issues. Performance on the adult rheumatology in-training examination and relationship to outcomes on the rheumatology certification examination. *Arthritis Rheumatol* 2015;67:3082–3090.
- Kempainen R, Hess B, Addrizzo-Harris DJ, Schaad D, Scott C, Carlin BW, Duhigg L, Lipner R. In-service training examination score as a predictor of performance on pulmonary and critical care board certification examinations [abstract]. *Am J Respir Crit Care Med* 2015;191:A1001.
- Kolen MJ, Brennan RL. Test equating, scaling, and linking: methods and practices, 2nd ed. New York: Springer; 2004.
- Lipner RS, Lelieveld C, Holmboe ES. Performance of physicians trained through the research pathway in internal medicine. *Acad Med* 2012;87:1594–1599.
- Shea JA, Norcini JJ, Kimball HR. Relationships of ratings of clinical competence and ABIM scores to certification status. *Acad Med* 1993;68(10 Suppl):S22–S24.
- Langenau EE, Pugliano G, Roberts WL. Relationships between high-stakes clinical skills exam scores and program director global competency ratings of first-year pediatric residents. *Med Educ Online* 2011;16:7362.
- Fessler HE, Addrizzo-Harris D, Beck JM, Buckley JD, Pastores SM, Piquette CA, Rowley JA, Spevetz A. Entrustable professional activities and curricular milestones for fellowship training in pulmonary and critical care medicine: report of a multisociety working group. *Chest* 2014;146:813–834.
- Jones AT, Biester TW, Buyske J, Lewis FR, Malangoni MA. Using the American Board of Surgery In-Training Examination to predict board certification: a cautionary study. *J Surg Educ* 2014;71:e144–e148.