

Developing an Assessment Framework for Essential Internal Medicine Subspecialty Topics

Natasha Chida, MD, MSPH
Christopher Brown, MD
Jyoti Mathad, MD
Kelly Carpenter, MD
George Nelson, MD
Marcos C. Schechter, MD
Paulina A. Rebolledo, MD
Valeria Fabre, MD
Diana Silva Cantillo, MD

Sarah Longworth, MD
Valerianna Amorosa, MD
Christian Petruskis, MD
Catherine Boulanger, MD
Natalie Cain, BS
Amita Gupta, MD
Jane McKenzie-White, MEd
Robert Bollinger, MD, MPH
Michael Melia, MD

ABSTRACT

Background Assessing residents by direct observation is the preferred assessment method for infrequently encountered subspecialty topics, but this is logistically challenging.

Objective We developed an assessment framework for internal medicine (IM) residents in subspecialty topics, using tuberculosis diagnosis for proof of concept.

Methods We used a 4-step process at 8 academic medical centers that entailed (1) creating a 10-item knowledge assessment tool; (2) pilot testing on a sample of 129 IM residents and infectious disease fellow volunteers to evaluate validity evidence; (3) implementing the final tool among 886 resident volunteers; and (4) assessing outcomes via retrospective chart review. Outcomes included tool score, item performance, and rates of obtaining recommended diagnostics.

Results Following tool development, 10 infectious disease experts provided content validity. Pilot testing showed higher mean scores for fellows compared with residents (7 [SD = 1.8] versus 3.8 [SD = 1.7], respectively, $P < .001$) and a satisfactory Kuder-Richardson Formula 20 (0.72). Implementation of the tool revealed a 14-minute (SD = 2.0) mean completion time, 61% (541 of 886) response rate, 4.4 (SD = 1.6) mean score, and $\leq 57\%$ correct response rate for 9 of 10 items. On chart review ($n = 343$), the rate of obtaining each recommended test was $\leq 43\%$ (113 of 261), except for chest x-rays (96%, 328 of 343).

Conclusions Our assessment framework revealed knowledge and practice gaps in tuberculosis diagnosis in IM residents. Adopting this approach may help ensure assessment is not limited to frequently encountered topics.

Introduction

Assessment using the Accreditation Council for Graduate Medical Education (ACGME) competencies is critical to determining resident readiness for unsupervised practice.^{1,2} Unfortunately, commonly used assessment methods may lack validity evidence and have other limitations.¹⁻⁷ The American Board of Internal Medicine (ABIM) Certification Examination and the American College of Physicians Internal Medicine In-Training Examination (IM-ITE) have validity evidence, but they occur annually and may not provide for timely identification of deficits. Direct observation is optimal but can be logistically challenging, particularly for core internal medicine (IM) subspecialty topics, with which residents have few encounters.^{1,4,8} As internists should be well-versed in these topics, additional assessment strategies are needed.

The ABIM Certification Examination blueprint identifies tuberculosis (TB) as a core infectious disease (ID) subject, yet both the ABIM examination and the IM-ITE contain few TB-focused questions.^{9,10} We reviewed PubMed and MedEdPORTAL and could not locate TB diagnosis assessment tools. Because patients with TB often present to internists, and 1 untreated patient may infect up to 15 people, a missed diagnosis can have significant consequences.¹¹⁻¹³ Using pulmonary TB diagnosis as a proof of concept, we created and tested an assessment framework for IM subspecialty topics.

Methods

We developed our framework through a 4-step process: We (1) created a TB diagnosis knowledge assessment tool; (2) evaluated the tool's evidence of validity; (3) distributed the tool to IM residents; and (4) assessed resident practice by reviewing charts of inpatients evaluated for TB. We included 972

DOI: <http://dx.doi.org/10.4300/JGME-D-17-00377.1>

TABLE 1

Performance by Item on Knowledge Assessment Tool Among Sites and Respondents^a

Item	Content Area ^b	Correct Response, n (%)
1	Identify TB disease risk factors	201 (37)
2	Interpret TST results in a patient with abnormal chest imaging	311 (57)
3	Interpret IGRA results in a patient with TB symptoms	297 (55)
4	Identify correct microbiological workup of pulmonary TB	208 (38)
5	Identify indication for NAAT in diagnostic workup of pulmonary TB	141 (26)
6	Identify correct sputum type for diagnostic workup of pulmonary TB	208 (38)
7	Identify number and timing of sputum samples needed for diagnostic workup of pulmonary TB	411 (76)
8	Interpret TB diagnostic results in a patient at risk for active TB	235 (43)
9	Identify criteria for noninfectiousness in a patient with smear-positive TB	240 (44)
10	Identify discharge criteria in a patient with smear-positive TB	112 (21)

Abbreviations: TB, tuberculosis; TST, tuberculin skin test; IGRA, interferon gamma release assay; NAAT, nucleic acid amplification test.

^a N = 541 residents.^b Based on US Centers for Disease Control and Prevention Core Curriculum on TB.¹⁴

residents from 8 IM residency programs. The study was conducted between May and September 2015.

Development of Assessment Tool and Evaluation of Validity Evidence

Using established question-writing guidelines, we created 10 multiple-choice items based on the Centers for Disease Control and Prevention (CDC) TB Core Curriculum (TABLE 1).^{14,15} Ten ID attending physicians provided content validity through question review. Using Qualtrics software (Provo, Utah), we pilot tested the tool with 86 residents—with equal representation from each postgraduate year (PGY)—and a convenience sample of 43 ID fellows. Participation was voluntary and responses were anonymous. Respondents received a \$10 gift card. We used the American Association for Public Opinion Research type 1 response rate, which includes only fully completed tools.¹⁶

We expressed site scores as means of individual results, compared resident and fellow scores using Student's *t* tests, and evaluated institutional differences using 1-way analysis of variance. Hypothesis testing was done using 2-tailed tests; significance was determined at $P < .05$. We performed item analyses and discrimination and rewrote items with correct response percentages of $< 25\%$ or $> 95\%$ or with point-biserial indices of < 0.20 .¹⁷ We assessed internal consistency via the Kuder-Richardson Formula 20.¹⁸ Analyses were done using STATA version 13.0 (StataCorp LLC, College Station, TX).

Tool Distribution

We assessed the remaining 886 residents in the 7 programs using the refined tool and the same methodology as the pilot. We captured resident demographics

and self-reported experience caring for patients with TB, and we used 1-way analysis of variance to compare mean scores by PGY level and TB experience.

Chart Review of IM Resident Practice

To assess resident practice beyond knowledge, we retrospectively reviewed the charts of 343 patients randomly selected from 2136 inpatients assessed for pulmonary TB across the sites in 2014. A uniform abstraction instrument was used to record whether residents obtained tests the CDC recommend for every patient evaluated for pulmonary TB¹⁴: 3 sputum specimens, a nucleic acid amplification test, a chest x-ray, and a latent TB infection test. Sample size was based on the assumption each test would be performed in 70% of patients with a 95% confidence interval and $\pm 5\%$ error margin. The number of charts reviewed per site was proportional to the site's contribution to the 2136 patients. We included adults admitted to an IM or ID resident team with a respiratory sample sent for acid-fast bacilli smear and culture. Exclusion criteria included death during hospitalization and evaluation for nontuberculous mycobacteria. Resident and fellow study team members performed the chart reviews to determine the feasibility of trainees completing this task.

The study was approved by each site's Institutional Review Board.

Results

Development of Assessment Tool and Evaluation of Validity Evidence

Question development required less than 10 hours, while each expert review required less than 1 hour. The response rate for the pilot was 57% (74 of 129),

TABLE 2

Characteristics, Tuberculosis Experiences, and Performance Among Knowledge Assessment Tool Respondents^a

Characteristic	Respondents, n (%)
Resident level ^b	
PGY-1	195 (36)
PGY-2	174 (32)
PGY-3	161 (30)
PGY-4	11 (2)
Residency program	
Site 1 (n = 31)	17 (3)
Site 2 (n = 131)	63 (12)
Site 3 (n = 169)	112 (21)
Site 4 (n = 110)	76 (14)
Site 5 (n = 131)	52 (10)
Site 6 (n = 171)	123 (23)
Site 7 (n = 143)	98 (18)
Tuberculosis patients within previous 12 mo	
0	237 (44)
1–5	278 (51)
6–10	19 (4)
> 10	7 (1)

Abbreviation: PGY, postgraduate year.

^a N = 886 residents.

^b No difference in response rate by PGY level was seen ($F_{3,537} = 0.23$, $P = .90$). PGY-4 were fourth-year medicine-pediatrics residents (n = 20).

with 53% (46 of 86) of residents and 65% (28 of 43) of fellows responding. Mean tool completion time was 14 minutes (SD = 2.0). Fellows scored higher than residents, confirming criterion validity (7.0 of 10.0 [SD = 1.8], median score 7 [interquartile range {IQR} 6–8], versus 3.8 of 10 [SD = 1.7], median score 4 [IQR 2.5–5], respectively; $P < .001$). Scores did not differ by site. Item analyses and discrimination yielded 1 item with a point-biserial index of 0.01 and another with a 14% correct response percentage (18 of 129); both were revised. The Kuder-Richardson Formula 20 coefficient was 0.72, indicating satisfactory reliability.

Tool Distribution

The response rate for the final assessment was 61% (541 of 886), with no difference by PGY level ($F_{3,537} = 0.23$, $P = .90$; TABLE 2). Mean tool completion time was 14 minutes (SD = 3.0); the overall mean score was 4.4 (SD = 1.6). Site 5 performed better than other sites, with a mean score of 4.8 ($F_{6,534} = 2.46$, $P = .021$). There was no difference in mean score among other sites ($F_{5,483} = 2.04$, $P = .07$), by PGY ($F_{3,485} = 0.20$, $P = .90$), or by TB experience ($F_{3,537} = 1.10$, $P = .35$). The correct response percentage was $\leq 57\%$ (311 of 541) for 9 of the items (TABLE 1).

Chart Review of IM Resident Practice

Data abstraction required approximately 20 minutes per chart. Eighteen percent (62 of 343) of patients had a nucleic acid amplification test, 35% (121 of 343) had latent TB infection testing, and 96% (328 of 343) had a chest x-ray. Sputa were obtained in 261 patients; 43% (113 of 261) submitted 3 specimens, and 22% (57 of 261) submitted 1 specimen. The remaining 82 patients underwent bronchoscopy. Therefore, the rate of obtaining each recommended test was $\leq 43\%$, except for chest x-rays.

Discussion

We developed an assessment framework for an infrequently encountered IM subspecialty topic. Implementation revealed concordance between our knowledge assessment tool and chart review, indicating gaps in resident knowledge and practice in TB diagnosis.

Prior work has identified IM resident knowledge gaps in commonly encountered subspecialty topics such as chronic kidney disease and asthma.^{19,20} To our knowledge this is the first framework to assess an infrequently encountered subspecialty topic. Programs may be able to utilize our framework to understand opportunities for improvement at individual and program levels. For example, programs could use our framework to explore deficiencies consistently identified by the ITE, while residents could perform the chart review on their own patients under the purview of quality improvement-oriented faculty; the latter would help residents analyze their practice, as recommended by the ACGME.²¹

A limitation of our study was the retrospective nature of our chart review. The small number of TB evaluations made linking individual residents' tool performances to their patients impractical, and we used review as a surrogate for practice. Another limitation was the time required for framework development, given program faculty often have many demands on their time.^{3,22,23} While it likely is not feasible for individual programs to generate this type of framework for each subspecialty topic, programs may build frameworks for topic(s) with which their residents have the most difficulty, and the resulting tools could be shared via open access journals such as MedEdPORTAL.

Future research should investigate the feasibility of implementing our framework for other subspecialty topics and its impact on resident knowledge and skills.

Conclusion

We developed an assessment framework that revealed resident knowledge and practice gaps in TB diagnosis.

Incorporating our framework into existing assessment methods can help ensure that timely individualized assessment is not limited to frequently encountered topics.

References

- Potts JR. Assessment of competence: the Accreditation Council for Graduate Medical Education/residency review committee perspective. *Surg Clin North Am*. 2016;96(1):15–24.
- Accreditation Council for Graduate Medical Education; American Board of Internal Medicine. The Internal Medicine Milestone Project. https://www.acgme.org/portals/0/pdfs/milestones/internalmedicine_milestones.pdf. Accessed April 24, 2018.
- Kogan JR, Holmboe ES, Hauer KE. Tools for direct observation and assessment of clinical skills of medical trainees: a systematic review. *JAMA*. 2009;302(12):1316–1326.
- Holt KD, Miller RS, Nasca TJ. Residency programs' evaluations of the competencies: data provided to the ACGME about types of assessments used by programs. *J Grad Med Educ*. 2010;2(4):649–655.
- Hauer KE, Chesluk B, Lobst W, et al. Reviewing residents' competence: a qualitative study of the role of clinical competency committees in performance assessment. *Acad Med*. 2015;90(8):1084–1092.
- Davis DA, Mazmanian PE, Fordis M, et al. Accuracy of physician self-assessment compared with observed measures of competence: a systematic review. *JAMA*. 2006;296(9):1094–1102.
- Holmboe ES, Sherbino J, Long DM, et al. The role of assessment in competency-based medical education. *Med Teach*. 2010;32(8):676–682.
- Weinberger SE, Pereira AG, Lobst WF, et al. Competency-based education and training in internal medicine. *Ann Intern Med*. 2010;153(11):751–756.
- American Board of Internal Medicine. Internal Medicine Certification Exam. <http://www.abim.org/exam/certification/internal-medicine.aspx#content>. Accessed April 24, 2018.
- American College of Physicians. IM-ITE. <https://www.acponline.org/featured-products/medical-educator-resources/im-ite>. Accessed April 24, 2018.
- Taylor Z, Marks SM, Rios Burrows NM, et al. Causes and costs of hospitalization of tuberculosis patients in the United States. *Int J Tuberc Lung Dis*. 2000;4(10):931–939.
- World Health Organization. Tuberculosis: key facts. <http://www.who.int/mediacentre/factsheets/fs104/en>. Accessed April 24, 2018.
- Karakousis PC, Sifakis FG, de Oca RM, et al. US medical resident familiarity with national tuberculosis guidelines. *BMC Infect Dis*. 2007;7:89.
- Centers for Disease Control and Prevention. Core curriculum on tuberculosis: what the clinician should know. <http://www.cdc.gov/tb/education/corecurr>. Accessed April 24, 2018.
- Case SM, Swanson DB. *Constructing Written Test Questions for the Basic and Clinical Sciences*. 3rd ed. Philadelphia, PA: National Board of Medical Examiners; 2002.
- American Association for Public Opinion Research. Response rates—an overview. <http://www.aapor.org/Education-Resources/For-Researchers/Poll-Survey-FAQ/Response-Rates-An-Overview.aspx>. Accessed April 24, 2018.
- De Champlain AF. A primer on classical test theory and item response theory for assessments in medical education. *Med Educ*. 2010;44(1):109–117.
- Streiner DL. Starting at the beginning: an introduction to coefficient alpha and internal consistency. *J Pers Assess*. 2003;80(1):99–103.
- Hemnes AR, Bertram A, Sisson SD. Impact of medical residency on knowledge of asthma. *J Asthma*. 2009;46(1):36–40.
- Estrella MM, Sisson SD, Roth J, et al. Efficacy of an internet-based tool for improving physician knowledge of chronic kidney disease: an observational study. *BMC Nephrol*. 2012;13:126.
- Accreditation Council for Graduate Medical Education. Common program requirements. <http://www.acgme.org/What-We-Do/Accreditation/Common-Program-Requirements>. Accessed April 24, 2018.
- Malik MU, Diaz Voss Varela DA, Stewart CM, et al. Barriers to implementing the ACGME Outcome Project: a systematic review of program director surveys. *J Grad Med Educ*. 2012;4(4):425–433.
- Swing SR, Clyman SG, Holmboe ES, et al. Advancing resident assessment in graduate medical education. *J Grad Med Educ*. 2009;1(2):278–286.



Natasha Chida, MD, MSPH, is Assistant Professor, Johns Hopkins Center for Clinical Global Health Education and Department of Medicine, Johns Hopkins University School of Medicine; **Christopher Brown, MD**, is Internal Medicine Resident, Department of Medicine, Weill Cornell Medicine; **Jyoti Mathad, MD**, is Assistant Professor, Department of Medicine, Weill Cornell Medicine; **Kelly Carpenter, MD**, is Internal Medicine Resident, Department of Medicine, Vanderbilt University School of Medicine; **George Nelson, MD**, is Assistant Professor, Department of Medicine, Vanderbilt University School of Medicine; **Marcos C. Schechter, MD**, is Assistant Professor, Department of Medicine, Division of Infectious Diseases, Emory University School of Medicine; **Paulina A. Rebollo, MD**, is Assistant Professor, Department of Medicine, Emory University School of Medicine, and Hubert Department of Global Health, Rollins School of Public Health, Emory University; **Valeria Fabre, MD**, is Instructor, Department of Medicine, Johns Hopkins University School of Medicine; **Diana Silva Cantillo, MD**, is Internal Medicine Resident, Department of Medicine, Alpert

Medical School of Brown University; **Sarah Longworth, MD**, is Infectious Disease Fellow, Department of Medicine, Perelman School of Medicine, University of Pennsylvania; **Valerianna Amorosa, MD**, is Associate Professor, Department of Medicine, Perelman School of Medicine, University of Pennsylvania; **Christian Petruskis, MD**, is Internal Medicine Resident, Department of Medicine, University of Miami Miller School of Medicine; **Catherine Boulanger, MD**, is Associate Professor, Department of Medicine, University of Miami Miller School of Medicine; **Natalie Cain, BS**, is a Medical Student, University of Miami Miller School of Medicine; **Amita Gupta, MD**, is Associate Professor, Johns Hopkins Center for Clinical Global Health Education and Department of Medicine, Johns Hopkins University School of Medicine; **Jane McKenzie-White, MEd**, is Managing Director, Johns Hopkins Center for Clinical Global Health Education, Johns Hopkins University School of Medicine; **Robert Bollinger, MD, MPH**, is Professor, Johns Hopkins Center for Clinical Global Health Education and Department of Medicine, Johns Hopkins University School of Medicine; and **Michael Melia,**

MD, is Associate Professor, Department of Medicine, Johns Hopkins University School of Medicine.

Funding: This study was supported by National Institutes of Health and Ujala Foundation grants.

Conflict of interest: The authors declare they have no competing interests.

This brief report was presented as an abstract at IDWeek, San Diego, California, October 7, 2015.

The authors would like to thank Katie McIntire for her careful review of this work.

Corresponding author: Natasha Chida, MD, MSPH, Johns Hopkins School of Medicine, 600 North Wolfe Street, Baltimore, MD 21287, 407.927.8817, nchida1@jhmi.edu

Received May 20, 2017; revisions received November 3, 2017, and December 27, 2017; accepted January 8, 2018.