

The Association Between Physicians' Cognitive Skills and Quality of Diabetes Care

Brian J. Hess, PhD, Weifeng Weng, PhD, Eric S. Holmboe, MD, and Rebecca S. Lipner, PhD

Abstract

Purpose

To examine the association between physicians' cognitive skills and their performance on a composite measure of diabetes care that included process, outcome, and patient experience measures.

Method

The sample was 676 physicians from the United States with time-limited certification in general internal medicine between 2005 and 2009. Scores from the American Board of Internal Medicine (ABIM) internal medicine maintenance of certification (MOC) examination were used to measure practicing physicians' cognitive skills (scores reflect fund of

medical knowledge, diagnostic acumen, and clinical judgment). Practice performance was assessed using a diabetes composite measure aggregated from clinical and patient experience measures obtained from the ABIM Diabetes Practice Improvement Module.

Results

Using multiple regression analyses and controlling for physician and patient characteristics, MOC examination scores were significantly associated with the diabetes composite scores ($\beta = .22, P < .001$). The association was particularly stronger with intermediate outcomes than with process and patient experience measures. Performance in the endocrine

disease content domain of the examination was more strongly associated with the diabetes composite scores ($\beta = .19, P < .001$) than the performance in other medical content domains ($\beta = .06-.14$).

Conclusions

Physicians' cognitive skills significantly relate to their performance on a comprehensive composite measure for diabetes care. Although significant, the modest association suggests that there are unique aspects of physician competence captured by each assessment alone and that both must be considered when assessing a physician's ability to provide high-quality care.

Substantial deficiencies in the quality of health care delivered to patients in the United States are well documented.^{1,2} Effective microsystems must have competent physicians who possess sufficient knowledge and cognitive skills to make the correct diagnosis and who exercise informed decision making.³ The link between what physicians "know" and what physicians "do" (e.g., performance) in clinical practice is critical to measuring quality patient care⁴ because effective

clinical performance requires the integration of medical knowledge with other important competencies. In Miller's pyramidal hierarchy of testing clinical competence and performance,⁵ testing of knowledge forms the foundation, followed by application of knowledge. Therefore, a secure examination of cognitive skills is an important component of specialty board certification and maintenance of certification (MOC) programs. The examination is an effective and efficient way of testing physicians' knowledge of the breadth of a medical field. Its purpose is to assess physicians' ability to incorporate and synthesize new knowledge to arrive at a correct diagnosis or treatment, and it is used to ensure the public that physicians are held accountable to meeting a minimum level of competence of cognitive ability.⁶

Studies have shown that physicians with stronger cognitive skills, as measured by a secure examination, provide better quality of care.⁷⁻¹⁰ For practicing physicians, higher MOC examination scores have been associated with higher rates of processes of care (e.g., mammography screening) for Medicare beneficiaries.⁷ Although the link between

physicians' cognitive performance and quality of care assessed through individual process and patient outcome measures has been demonstrated, previous research has not examined this relationship using a composite measure of performance for a specific chronic medical condition (e.g., diabetes) and for a broad range of patients. Composite measures comprising intermediate outcome, process, and patient experience measures that are psychometrically robust can comprehensively represent the overall clinical care that physicians provide to patients with a specific condition,¹¹⁻¹³ provided that they include valid measures that are evidence based, relevant to practice, and statistically sound.¹⁴ Research examining the relationship between cognitive skills and clinical performance in the context of a specific clinical domain is negligible—for instance, do practicing physicians with greater knowledge and skill in endocrine disease provide better care to diabetic patients? Furthermore, prior research examining the relationship between cognitive performance and quality care has focused on initial certification examination performance; only one study has focused on MOC examination scores,⁷ and, as a result, others have called

Dr. Hess is director of research analysis, American Board of Internal Medicine, Philadelphia, Pennsylvania.

Dr. Weng is health services researcher, American Board of Internal Medicine, Philadelphia, Pennsylvania.

Dr. Holmboe is senior vice president and chief medical officer, American Board of Internal Medicine, Philadelphia, Pennsylvania.

Dr. Lipner is vice president of psychometrics and research analysis, American Board of Internal Medicine, Philadelphia, Pennsylvania.

Correspondence should be addressed to Dr. Hess, American Board of Internal Medicine, 510 Walnut St., Suite 1700, Philadelphia, PA 19106; telephone: (215) 446-3611; e-mail: bhess@abim.org.

Acad Med. 2012;87:157-163.
First published online December 20, 2011
doi: 10.1097/ACM.0b013e31823f3a57

for additional studies to better elucidate the value of MOC programs.¹⁵ Therefore, the purpose of our study is to examine the relationship between cognitive skills and care of diabetic patients using internal medicine MOC examination scores and diabetes composite scores.

Method

Instruments

We used the American Board of Internal Medicine (ABIM) internal medicine MOC examination, a secure, computer-based examination comprising 180 patient vignettes that require a single-best-answer response, to measure practicing physicians' cognitive skills (scores reflect fund of medical knowledge, diagnostic acumen, and clinical judgment in general internal medicine). Physicians were expected to integrate information, prioritize alternatives, and/or use clinical judgment to reach an appropriate decision about a course of action in each question. Table 1 displays the medical content domains and percentage of questions in that domain as defined by the blueprint (the expanded blueprint is available at http://www.abim.org/pdf/blueprint/im_moc.pdf).

We used clinical performance data from the ABIM Diabetes Practice Improvement Module (Diabetes PIM), a Web-based self-evaluation tool that guides physicians through collecting data

from their own practice, using medical chart reviews, patient surveys, and a practice system survey—all of which form a comprehensive performance assessment.¹⁶ We encouraged physicians to abstract 25 patient charts and distribute 25 patient surveys using a retrospective or prospective sequential sample, or a systematic random sample, with a minimum of 10 charts and 10 surveys required. Patients were eligible if they had type 1 or type 2 diabetes, were between 18 and 75 years old, and received care from the physician for at least 12 months (including at least one visit within the past 12 months), with diabetes care management decisions made primarily by that physician. To acknowledge the difficulty of scheduling office visits, we gave a grace period of one to three months, depending on the recommended interval, to periodic measures (e.g., retinal exam).

The development of the diabetes composite measure has been previously described.¹³ Briefly, measures used in the diabetes composite are shown in Table 2. The intermediate outcome and process measures, originally developed by the National Committee for Quality Assurance in partnership with the American Diabetes Association, use guidelines that are evidence based describing ideal diabetes care.¹⁷ We defined performance on intermediate outcome measures as the percentage of a

physician's patient panel that met the recommended performance level. For process measures, we calculated the performance as the percentage of a physician's patient panel that received the test/exam or counseling. The two patient experience measures, created using specific patient survey questions (see the footnote under Table 2), were included because they underscore the importance of patient-centered care.

Physician sample

We obtained data from a retrospective cohort of 676 physicians certified only in general internal medicine between 1990 and 2002, were enrolled in MOC (2% were enrolled for the second time), and completed the MOC examination and the Diabetes PIM between 2005 and 2009 to satisfy MOC requirements. Certificates have a 10-year time limit, and physicians certified since 1990 must complete MOC every 10 years to maintain certification.

Study variables

We used physicians' scores from their first attempt on the internal medicine MOC examination. Because physicians took the examination at different times, overall scores were equated and reported on a standardized score scale (mean = 500, SD = 100).¹⁸ Equated scores were not available for each medical content domain (e.g., endocrine disease); therefore, percentiles were used to measure physicians' relative ability in each domain. Percentiles are more stable than raw scores because the average ability and distribution of first takers (as measured by the overall equated score) was not significantly different across administrations (2005 through 2009). The average reliability (alpha) coefficients across administrations for individual medical content domains ranged from .32 to .65 (Table 1); coefficients for individual domains were generally similar across administrations. Alpha coefficients for overall scores ranged from .89 to .91 across administrations.

To compute physicians' scores on the diabetes composite, each physician's actual performance rate for an individual measure was multiplied by a point value (or weight) assigned to it by an expert panel (Table 2).¹³ For example, if a physician completed a foot examination for 50% of his or her patients and its weight was four points, then the

Table 1
Descriptive Statistics for Each Medical Content Domain of the Internal Medicine Maintenance of Certification (MOC) Examination for 676 Physicians, 2005–2009

Content domain	No. (%) of exam questions*	Percentile, mean (SD) [†]	Average reliability (alpha) [‡]
Cardiovascular disease	26 (14)	45.3 (26.8)	.56
Endocrine disease	14 (8)	43.3 (27.9)	.50
Gastrointestinal disease	16 (9)	43.8 (27.1)	.45
Hematologic disease	11 (6)	39.4 (26.8)	.44
Infectious disease	16 (9)	42.9 (28.3)	.44
Oncologic disease	11 (6)	39.8 (27.2)	.33
Renal disease	11 (6)	39.2 (26.3)	.32
Respiratory disease	16 (9)	42.4 (27.5)	.44
Rheumatologic disease	14 (8)	43.8 (27.1)	.42
Other disease conditions (e.g., allergy and immunology, psychiatry)	45 (25)	48.0 (27.9)	.65

* The number of questions per domain was very similar across administration forms.
[†] Percentiles were computed from physicians' first attempt of the MOC examination.
[‡] The reliability for each individual medical content domain is the average reliability coefficient across administrations of the examination (2005–2009).

Table 2

Measures From the American Board of Internal Medicine (ABIM) Diabetes Practice Improvement Model (PIM) Used to Compute the Diabetes Composite Scores for 676 Physicians, 2005–2009

Measure	Definition	Physician performance mean, % (SD) [§]	Threshold [¶]	Point value (weight) ^{**}
Process measures				
Retinal exam	Completed	60.1 (29.2)	28.8%	9
Nephropathy assessment	Completed	91.3 (12.6)	73.1%	10
Foot exam	Completed	68.3 (32.6)	35.6%	4
Smoking status and cessation advice/treatment	Completed	96.5 (7.2)	67.5%	7
Intermediate outcome measures				
A1C poor control*	>9.0%	16.3 (18.1)	N/A	10
A1C at goal [†]	<8.0% or <7.0% (based on the patient)	61.3 (20.0)	N/A	7
Blood pressure poor control*	≥140/90	26.6 (16.2)	N/A	10
Blood pressure superior control	<130/80	37.4 (16.7)	N/A	9
LDL poor control*	≥130 mg/dL	25.9 (19.4)	N/A	10
LDL superior control	<100 mg/dL	52.5 (20.0)	N/A	8
Patient experience measures				
Overall diabetes care satisfaction [‡]	Excellent or very good responses	73.5 (16.0)	N/A	7
Patient self-care support [‡]	Excellent or very good responses	77.8 (12.2)	N/A	9

* For the poor control measures, points were awarded to physicians when their patients did *not* meet the definition. For example, if 80% of a physician's patients did *not* have poor control of their blood pressure, then the physician would earn 8 points for that measure ($0.80 \times 10 = 8$).

[†] The definition for A1C at goal is <8.0% for patients aged 65 and over, with coronary heart disease, cerebrovascular disease, peripheral artery disease, end-stage renal disease, or significant loss of vision or blindness; the level is <7.0% for other patients.

[‡] Overall diabetes care satisfaction was defined as the percentage of patients in a physician's panel who rated their overall diabetes care "excellent" or "very good" based on one question using a five-point Likert scale. Patient self-care support was defined as the percentage of responses per physician that were "excellent" or "very good" across seven questions about showing understanding of living with diabetes, encouraging questions and answering them clearly, providing information on taking medications properly, providing information on side effects of medications, teaching foot care, providing information on proper diabetic diet, and teaching home blood glucose monitoring.

[§] The performance mean is the average proportion of a physician's patients meeting the measure across the sample of 676 physicians.

[¶] Because physicians typically have direct control of processes, if a physician's performance rate did not meet or exceed the threshold for a process measure, then the physician earned 0 points for that measure.

^{**} Points for each measure were determined through a consensus-based, standard-setting process described in a previous study.¹⁹

physician would receive two points for that measure ($.50 \times 4 = 2$). Because physicians typically have direct control of processes of care, if a physician's performance rate did not meet or exceed the threshold for a process measure (Table 2), then the physician earned zero points for that measure. We determined process measure thresholds, as well as points for each measure, through a consensus-based, standard-setting process previously described.¹⁹ Points earned for individual measures were

summed to yield a total score between 0 and 100 points. Psychometric evidence supports the composite score; the reliability coefficient, estimated using a bootstrap (resampling) method that takes into account the effect of nesting patients within physicians, is high—approximately .92.¹³ To examine associations with the types of measures that make up the composite, we also computed separate subcomposite scores for the intermediate outcome, process, and patient experience measures by

summing the points earned for those measure sets.

Statistical analyses

We used multiple regression analysis to examine the relationship between overall internal medicine MOC examination scores and the diabetes composite scores, controlling for physician and patient characteristics (at the physician level). We also examined regression models for each individual measure and for the intermediate outcome, process, and patient experience subcomposites separately. Physician characteristics obtained from the ABIM database of MOC enrollees were age, gender, type of practice (solo versus nonsolo), average percent of time practicing in an ambulatory setting, and birth country (United States/Canada versus international) and medical school graduation country (United States/Canada versus international), which were combined and dummy coded so that each group was compared with U.S./Canadian born and internationally trained (reference group). Research has shown that U.S. citizens who graduated from international medical schools perform lower on quality-of-care measures compared with physicians who graduated from international medical schools and were not U.S. citizens.²⁰ Patient characteristics were age and gender averaged at the physician level. Interactions between explanatory variables were also tested. Stepwise regression dictated variable inclusion in the model, starting with the full model and ending when the minimum Akaike information criterion value was found among the series of models. To ensure that all other explanatory variables were not highly correlated in each regression model, we examined bivariate correlations and collinearity diagnostic statistics (i.e., variance inflation factor and tolerance indexes). Next, we used a series of similar models to examine the association between percentile scores from *each* individual medical content domain and the diabetes composite scores, again controlling for physician and patient characteristics. Because individual medical content domain scores were less reliable than overall examination scores, we adjusted the coefficient associated with the medical content domain performance to correct for the attenuation in reliability (i.e., assume perfect reliability).²¹ In each of

these models, we controlled for overall examination performance using pass/fail status instead of overall examination scores because of the high correlation between the domain scores and overall examination scores. Analyses were performed using SPSS, version 12.0 (SPSS Inc., Chicago, Illinois). Ethical approval for our study was given by the Essex Institutional Review Board, Inc.

Results

Of the 676 physicians in our sample, 189 (28%) were in solo practice and 451 (67%) were men. Two hundred seventy (40%) were born and graduated from medical school in the United States or Canada, whereas 41 (6%) were born in the United States or Canada but graduated from an international medical school; conversely, 304 (45%) were internationally born and graduated from an international medical school, whereas 61 (9%) were internationally born but graduated from a U.S. or Canadian medical school. Most of physicians' time in clinical practice was spent in an ambulatory setting (mean = 77%, SD = 20%); mean age was 46 (SD = 6 years). This sample was similar to the population

of all general internists with time-limited certificates enrolled in MOC except for a slightly higher percentage of solo practitioners. It is not representative of older physicians with time-unlimited certificates not enrolled in MOC.

The mean number of charts abstracted per physician was 21 (SD = 7.1), which yielded 14,095 patient charts. The mean patient age for the charts was 58.9 (SD = 10.7), and 7,188 (51%) patients were male. The mean number of patient surveys per physician was 19.7 (SD = 7.2), which yielded 15,267 patient surveys. The mean patient age for the surveys was 58.9 (SD = 10.7), and 7,634 (50%) were male. Patient age and gender were similar for the two data sources (chart and patient survey).

Physicians performed consistently across the medical content domains (Table 1) and tended to score slightly lower than the average physician who took the examination; the mean equated overall MOC examination score was 487 (SD = 92), and 566 (84%) passed the examination on the first attempt across the administrations used in our study. The mean diabetes composite score

was 70.12 (SD = 11.96; range = 18.72–93.24). Table 2 shows that physicians performed somewhat better on the process and patient experience measures compared with intermediate outcomes.

Table 3 presents the results of the multiple regression analysis associating overall MOC examination scores with the diabetes composite scores, controlling for physician and patient characteristics; bivariate correlations and collinearity statistics indicated that the variables in the model were not highly correlated. Overall examination scores were positively associated with the diabetes composite scores ($\beta = .22, P < .001$). The β coefficients are standardized and reflect the relative importance of each explanatory variable; both overall MOC examination scores and patient age ($\beta = .23, P < .001$) contributed the most to the model (e.g., physicians with older patient panels tended to provide better care). Physicians who were internationally born and trained had significantly higher diabetes composite scores than physicians born in the United States/Canada but were internationally trained ($\beta = .09, P = .03$). Physicians had significantly higher diabetes composite scores if they were women, spent more time in ambulatory practice, and treated a higher percentage of male patients. The adjusted R^2 for the model was 13.0%; interaction terms were not statistically significant, and adding solo practice status and physician age to the model did not substantially improve the goodness-of-fit, nor did it substantially change the regression coefficients for other variables in the model (results not shown).

Results of follow-up multiple regression analyses associating MOC examination scores with performance on each individual measure and on the intermediate outcome, process, and patient experience subcomposites are presented in the Appendix. Associations were stronger with the intermediate outcome subcomposite ($\beta = .23, P < .001$), specifically for poor LDL control ($\beta = .23, P < .001$) and superior LDL control ($\beta = .25, P < .001$), compared with the association observed with the process subcomposite ($\beta = .13, P = .01$). Association with the patient experience subcomposite was the weakest ($\beta = .08, P = .04$).

Table 3
Multiple Regression Analysis Results Showing the Relationship Between Internal Medicine Maintenance of Certification (MOC) Examination Scores, Physician and Patient Characteristics, and the Diabetes Composite Scores for 676 Physicians, 2005–2009

Explanatory variables	b value	95% CI for b	β coefficient	t	P
Examination performance					
Equated overall MOC examination scores (first attempt)	0.03	0.02 to 0.04	.22	5.86	<.001
Physician characteristics					
Birth country and medical school country*					
<i>Internationally born and trained</i>	1.78	0.20 to 3.25	.09	2.22	.03
<i>Internationally born but U.S./Canadian trained</i>	0.13	-2.23 to 2.48	.01	0.011	.92
<i>U.S./Canadian born and trained</i>	-0.52	-2.09 to 1.04	-.03	-0.66	.51
Physician gender [†]	3.11	1.08 to 5.14	.12	3.00	.003
% Time in ambulatory or office setting	0.06	0.02 to 0.11	.10	2.73	.007
Patient characteristics[‡]					
Patient age	0.70	0.48 to 0.92	.23	6.27	<.001
% Male patients	6.63	1.64 to 11.63	.11	2.61	.008

* Birth country (U.S./Canada versus international) and medical school country (U.S./Canada versus international) were combined and dummy coded so that each group was compared with U.S./Canadian born and internationally trained (reference group).

[†] Physician gender (0 = men, 1 = women).

[‡] Patient age and percent male patients were averaged at the physician level.

Table 4

Association Between Performance in Each Medical Content Domain of the Internal Medicine Maintenance of Certification (MOC) Examination and Diabetes Composite Scores for 676 Physicians, 2005–2009

Medical content domain	Unadjusted		Adjusted	
	β coefficient*	P value	β coefficient*	P value
Cardiovascular disease	.06	.11	.09	.04
Endocrine disease	.13	.001	.19	<.001
Gastrointestinal disease	.09	.04	.14	<.001
Hematologic disease	.04	.31	.06	.11
Infectious disease	.08	.04	.13	.001
Oncologic disease	.08	.04	.14	<.001
Renal disease	.06	.11	.11	.01
Respiratory disease	.07	.09	.11	.01
Rheumatologic disease	.05	.22	.09	.04
Other disease conditions (e.g., allergy and immunology, psychiatry)	.09	.04	.11	.01

* Coefficients were obtained from 10 separate multiple regression models (i.e., a separate analysis was conducted for each individual medical content domain). Unadjusted coefficients reflect the association between each medical content domain score and the diabetes composite scores after controlling for examination outcome (pass versus fail) and physician and patient characteristics. Adjusted coefficients reflect the same associations but correct for the attenuation of reliability in the medical content domain scores.

Table 4 presents the unadjusted and adjusted standardized coefficients when associating each medical content domain with the overall diabetes composite scores, controlling for MOC examination outcome (fail versus pass) and physician and patient characteristics in each separate regression model. Based on the adjusted β coefficients, the endocrine disease content domain percentiles were significantly (but modestly) associated with the diabetes composite scores ($\beta = .19, P < .001$). This relationship was stronger than the relationship exhibited by each of the other medical content domains ($\beta = .06$ – $.14$). Finally, the endocrine disease content domain percentiles were also more strongly associated with the intermediate outcome subcomposite ($\beta = .20, P < .001$) than with the process measure subcomposite ($\beta = .11, P = .01$) and patient experience subcomposite ($\beta = .06, P = .11$).

Discussion

Our findings demonstrate that physicians' cognitive skills, as measured by internal medicine MOC examination scores, are related to a comprehensive measure of diabetes care based on real practice data. We also found that physicians' performance on the endocrine disease domain questions was more strongly associated with better diabetes care compared with other

general internal medicine content domains. When we examined the specific types of measures that make up the composite, overall examination scores and particularly the endocrine disease domain scores yielded stronger relationships with the intermediate outcomes than with process and patient experience measures. Our study is consistent with previous findings from Holmboe and colleagues,⁷ which showed a positive relationship between ABIM MOC examination scores and quality of care on a set of process measures for Medicare beneficiaries. However, our study extends the understanding of this relationship by including outcome and patient experience measures for a specific chronic condition from a broader sample of patients to create a more comprehensive composite performance measure.

Although the associations that we observed were statistically significant, they were modest, and not surprising given the complexity of clinical practice. This complexity is evidenced by the description of the six competencies that are necessary for providing high-quality patient care.²² That is, other competencies affect clinical performance, such as interprofessional communication skills, practice-based learning and improvement, and systems-based practice. For example, physician empathy, an element of good

communication skills, is associated with positive clinical outcomes for diabetic patients.²³ Future research should include additional measures of these competencies as well as other personal characteristics (e.g., physicians' participatory decision making) to understand the relative contribution of cognitive skills for explaining performance on quality measures.

We do, however, offer some explanations for the positive relationships observed. Physicians' cognitive skills are an essential element in making an accurate diagnosis and in executing informed decision making,³ and therefore these skills are foundational and necessary but not sufficient in providing high-quality patient care. One expects that physicians with stronger cognitive skills may be more effective at “doing the correct thing” by demonstrating a greater likelihood that appropriate care processes are performed, better intermediate outcomes are achieved, and patient self-care programs are supported. When we examined specific measures in the composite, we found stronger associations between examination scores and intermediate outcomes, specifically LDL control. One explanation is that because process measures are widely accepted and can be implemented by nonphysician staff in a well-functioning office practice, it is logical that a weaker association with process measures would be seen. On the other hand, having patients take control of outcomes, like their LDL, is more complicated, which might require higher cognitive skills. For example, lowering LDL does require a degree of physician decision making and recognition. Many decisions are packed into a short patient visit, and thus the physician may be more likely to either miss or delay starting therapies if the physician uses a “physician-centric” approach to care (i.e., does not engage other providers within the clinic to help make decisions and start needed therapies). Most process measures really do not need much physician decision—they can be automated in effective practices. Conversely, many physician offices still lack effective systems to ensure processes of care, and these systems may not be directly under the physician's control (e.g., although physicians might know the “correct thing,” their office staff may not know or do not comply). Furthermore, examination scores yielded weaker associations with the patient experience

measures, which is expected because other foundational competencies such as communication and interpersonal skills would more likely exhibit stronger relationships with patient experience measures.

The modest associations between the cognitive examination and the performance composites suggest that there are unique aspects of physician competence captured by each assessment alone and that both must be considered when assessing a physician's ability to provide high-quality care. For example, our diabetes composite is condition-specific and does not measure general diagnostic skill. Thus, if the diagnosis is wrong, it does not matter that "appropriate" processes of care are being performed or that the patient is being educated to take care of a condition he or she might not have. The cognitive examination, on the other hand, does assess diagnostic skill across a broad spectrum of conditions. Moreover, the ABIM MOC examination and the Diabetes PIM are part of the ABIM's professional self-regulatory requirements to demonstrate to the public physicians' commitment to maintaining competence in medical knowledge and patient care, respectively; together, these capture different levels of Miller's⁵ competency hierarchy by providing an assessment of what physicians "know" and "do" in practice.

The role of physician and patient characteristics in explaining variation in the diabetes composite scores was also consistent with our expectations. Physicians caring for older diabetic patients demonstrated a higher quality of care, perhaps because older patients may be seen more often or are more likely to comply with physician recommendations.²⁴ Physicians who practiced largely in an ambulatory setting also demonstrated higher-quality care, possibly because they have a systematic approach to managing diabetic patients,²⁵ whereas those practicing largely in an inpatient setting typically spend less time managing diabetic patients. Female internists also tended to demonstrate higher-quality care. This finding is somewhat inconsistent with a recent study²⁶ showing that male and female physicians tend to provide similar-quality diabetes care (however, those findings showed that women performed somewhat better on A1C and LDL

measures). The inconsistent findings may be due to the use of individual measures in the latter study, which are less reliable than composite measures. Finally, consistent with another study,²⁰ we found that physicians who were born in the United States/Canada but who had graduated from international medical schools provided poorer diabetes care than physicians who were internationally born and trained.

Our study has limitations. First, the associations between medical content domain performance and the diabetes composites should be interpreted with care because the observed reliability of each medical content domain score was modest and because correcting for attenuation provides only estimates of effects. Second, we controlled for differences in patient age and gender at the physician level (and did not adjust for patient case mix at the patient level) because risk adjustment is less relevant for process measures and because the A1C-at-goal measure by definition accounts for some differences in patients. Kaplan and colleagues¹¹ have shown that adjusting for patient case mix for a similar set of diabetes measures did not impact how physicians were ranked. Third, physicians selected the Diabetes PIM to earn credit toward satisfying the MOC practice performance requirement, thus limiting the generalizability of our findings to other physicians. Fourth, our sample scored slightly below average on the MOC examination, and therefore may not generalize to physicians in MOC with average cognitive skills. Fifth, some physicians may not have adhered to the PIM sampling instructions, and therefore their performance on individual measures might be inflated. However, because there was no consequence for performing poorly on the Diabetes PIM, there is no reason to believe that physicians "cherry-picked" patients; previous work has also confirmed the accuracy of physician-reported data in the Diabetes PIM.¹⁶

Our findings deepen the understanding of physicians' cognitive skills and their relationship to condition-specific quality of care by using a psychometrically robust composite measure of diabetes care composed of evidence-based clinical and patient experience measures instead of less reliable, individual clinical measures. The one previous study known

to us that found an association between MOC examination scores and quality of care⁷ used only relatively few process measures obtained from Medicare claims. We thus extend our understanding of this relationship by using medical record data to capture performance on outcomes and processes from a broader sample of patients. Our findings suggest that cognitive skills are an important foundational competency to facilitate other patient-care activities, even when measuring the quality of care comprehensively through composites. This should appeal to both the policy and professional communities that have been asking for more evidence on the validity of cognitive testing as part of MOC.¹⁵ More research is needed to examine the relationship between cognitive skills and other specific clinical abilities, such as accuracy of diagnosis and treatment decisions. Notwithstanding, it should be reassuring to patients that clinical knowledge, an essential competency for clinical practice, matters.

Funding/Support: None.

Other disclosures: All authors are employed by the ABIM. Drs. Hess, Weng, and Lipner are coinventors of a business method invention describing the application of a standard-setting method to practicing physicians. The invention is patent pending. Dr. Holmboe received honoraria for teaching about clinical assessment from the Uniformed Services University of the Health Sciences, the University of Kansas, and the Harvard-Macy Systems Assessment Course. Dr. Holmboe receives royalties for a textbook on assessment published by Mosby-Elsevier.

Ethical approval: Essex Institutional Review Board, Inc., approved this study.

Previous presentations: This study was presented at the annual meeting of the American Educational Research Association, May 3, 2010, Denver, Colorado.

References

- 1 Institute of Medicine. Improving the Quality of Health Care of Mental and Substance-Use Conditions. Washington, DC: National Academy Press; 2006.
- 2 McGlynn EA, Asch SM, Adams J, et al. The quality of health care delivered to adults in the United States. *N Engl J Med*. 2003;348:2635–2645.
- 3 Gruppen LD, Frohna AZ. Clinical reasoning. In: Norman GR, van der Vleuten CP, Newble DJ, eds. *International Handbook of Research in Medical Education*. Dordrecht, Netherlands: Kluwer Academic; 2002:205–230.
- 4 Holmboe ES, Lipner RS, Greiner A. Assessing quality of care: Knowledge matters. *JAMA*. 2008;299:338–340.

- 5 Miller GE. The assessment of clinical skills/competence/performance. *Acad Med.* 1990; 65(9 suppl):S63–S67.
- 6 Brennan TA, Horwitz RI, Duffy FD, Cassel CK, Goode LD, Lipner RS. The role of physician specialty board certification status in the quality movement. *JAMA.* 2004;292:1038–1043.
- 7 Holmboe ES, Wang Y, Meehan TP, et al. Association between maintenance of certification examination scores and quality of care for medical beneficiaries. *Arch Intern Med.* 2008;168:1396–1403.
- 8 Norcini JJ, Lipner RS, Kimball HR. Certifying examination performance and patient outcomes following acute myocardial infarction. *Med Educ.* 2002;36:853–859.
- 9 Tamblyn R, Abrahamowicz M, Dauphinee WD, et al. Association between licensure examination scores and practice in primary care. *JAMA.* 2002;288:3019–3026.
- 10 Wenghofer E, Klass D, Abrahamowicz M, et al. Doctor scores on national qualifying examinations predict quality of care in future practice. *Med Educ.* 2009;43:1166–1173.
- 11 Kaplan SH, Griffith JL, Price LL, Pawlson LG, Greenfield S. Improving the reliability of physician performance assessment: Identifying the “physician effect” on quality and creating composite measures. *Med Care.* 2009;47:378–387.
- 12 Lipner RS, Weng W, Arnold GK, Duffy FD, Lynn LA, Holmboe ES. A three-part model for measuring diabetes care in physician practice. *Acad Med.* 2007;82(10 suppl):S48–S52.
- 13 Weng W, Hess BJ, Lynn LA, Holmboe ES, Lipner RS. Measuring physicians’ performance in clinical practice: Reliability, classification accuracy, and validity. *Eval Health Prof.* 2010;33:302–320.
- 14 Landon BE, Normand SL, Blumenthal D, Daley J. Physician clinical performance assessment: Prospects and barriers. *JAMA.* 2003;290:1183–1189.
- 15 Goldman L, Goroll AH, Kessler B. American Board of Internal Medicine maintenance of certification program. *N Engl J Med.* 2010; 362:948–952.
- 16 Holmboe ES, Meehan TP, Lynn L, Doyle P, Sherwin T, Duffy FD. Promoting physicians’ self-assessment and quality improvement: The ABIM Diabetes Practice Improvement Module. *J Contin Educ Health Prof.* 2006;26:109–119.
- 17 American Diabetes Association. Standards of medical care in diabetes—2009. *Diabetes Care.* 2009;32(suppl 1):S13–S61.
- 18 Holland PW, Dorans NJ. Linking and equating. In: Brennan RL, ed. *Educational Measurement.* 4th ed. Westport, Conn: Praeger Publishers; 2006:187–220.
- 19 Hess BJ, Weng W, Lynn LA, Holmboe ES, Lipner RS. Setting a fair performance standard for physicians’ quality of patient care. *J Gen Intern Med.* 2011;26:467–473.
- 20 Norcini JJ, Boulet JR, Dauphinee WD, Opalek A, Krantz ID, Anderson ST. Evaluating the quality of care provided by graduates of international medical schools. *Health Aff (Millwood).* 2010;29:1461–1468.
- 21 Bohrnstedt GW. Measurement. In: Rossi PH, Wright JD, Anderson AB, eds. *Handbook of Survey Research.* New York, NY: Academic Press; 1983:69–121.
- 22 Accreditation Council for Graduate Medical Education. Competencies and Outcome Assessment. <http://www.acgme.org/outcome/comp/compHome.asp>. Accessed October 8, 2011.
- 23 Hojat M, Louis DZ, Markham FW, Wender R, Rabinowitz C, Gonnella JS. Physicians’ empathy and clinical outcomes for diabetic patients. *Acad Med.* 2011;86:359–364.
- 24 Holmboe ES, Wang Y, Tate JP, Meehan TP. The effects of patient volume on the quality of diabetic care for Medicare beneficiaries. *Med Care.* 2006;44:1073–1077.
- 25 Kitahata MM, Koepsell TD, Deyo RA, Maxwell CL, Dodge WT, Wagner EH. Physicians’ experience with the acquired immunodeficiency syndrome as a factor in patients’ survival. *N Engl J Med.* 1996;334: 701–706.
- 26 Kim C, McEwen LN, Gerzoff RB, et al. Is physician gender associated with the quality of diabetes care? *Diabetes Care.* 2005;28: 1594–1598.

Appendix

The Association Between Overall Internal Medicine Maintenance of Certification (MOC) Examination Scores and Performance on Each Individual Diabetes Measure and on the Three Diabetes Subcomposite Measures for 676 Physicians, 2005–2009

Diabetes measure	β coefficient*	P value
Process measures		
Retinal exam	.10	.01
Nephropathy assessment	.10	.01
Foot exam	.07	.08
Smoking status and cessation advice / treatment	.08	.04
Process subcomposite	.13	.01
Intermediate outcome measures[†]		
A1C poor control	.16	<.001
A1C at goal	.11	.01
Blood pressure poor control	.16	<.001
Blood pressure superior control	.10	.01
LDL poor control	.23	<.001
LDL superior control	.25	<.001
Intermediate outcome subcomposite	.23	<.001
Patient experience measures[†]		
Overall diabetes care satisfaction	.11	.01
Patient self-care support	.04	.29
Patient experience subcomposite	.08	.04

* Coefficients were obtained from 15 separate multiple regression models (i.e., one for each individual measure and for the three subcomposites). Coefficients are standardized and reflect the association between overall MOC examination scores and the diabetes measure after controlling for the physician and patient characteristics.

[†] See Table 2 footnotes for definitions of the intermediate outcome measures and the patient experience measures.