What Do Certification Examinations Tell Us About Quality?

OVER THE PAST DECADE, EFFORTS TO MEASURE, PUBLICLY REPORT, AND REWARD THE PERFORMANCE OF INDIVIDUAL PHYSICIANS HAVE BECOME COMMONPLACE. STIMULATED BY INFLUENTIAL REPORTS FROM THE INSTITUTE OF MEDICINE IN 1999 AND 2001 AND A 2003 REPORT BY McGlynn et al highlighting the problems with quality of care in the United States, these performance measurement activities are increasingly being used to assess the competence of individual physicians and physician groups, with a focus on both the quality and the efficiency of care.

Yet, ready availability of accurate and easily interpretable information on the quality of care provided by individual physicians remains an elusive goal. Such information could be of substantial use both to patients in choosing high-quality physicians and to physicians and provider systems in implementing quality improvement and monitoring programs. To date, however, clinical and administrative measures often cover a small proportion of care and are not representative of the often complex decisions that physicians must make. Therefore, they are not optimal for comparative decision making.

See also page 1396

Board certification can be considered an alternative to these traditional discrete measures of quality that is designed to provide an overall assessment of physician competence. Certification by a medical specialty board is meant to indicate that a physician has the knowledge, experience, and skills for providing quality health care within a given specialty. Current data suggest that the vast majority of physicians in the United States are board certified in at least 1 specialty and that many physicians have gone on to receive subspecialty certification as well. Because almost all physicians are board certified, specialty certification in and of itself is not useful for distinguishing between physicians in a field.

Before 2001, physicians were granted permanent board certification on the basis of a 1-time certifying examination that was typically taken shortly after the completion of formal training. This system was modified to reflect the rapid changes in medical knowledge. During the period from 2001 to 2005, all of the major medical specialty boards phased in programs that require physicians to update their certification every 6 to 10 years through maintenance of certification (MOC) programs of the various specialties. Each of these programs requires physicians to demonstrate evidence of lifelong learning and self-assessment, to maintain a valid unrestricted license to practice, and to requalify through repeated certifying examinations. Also, physicians are required to evaluate and compare their current clinical performance with accepted guidelines of care through assessments of their own clinical practice and to apply the best practices for improving care. Notably, physicians who were granted board certification before a specified date were grandfathered out of these new requirements.

Despite these changes, unresolved questions remain about whether scores on credentialing examinations can predict clinical skill and quality of daily practice. Currently, there is little research relating performance on written cognitive examinations with actual clinical practice, so it is not clear to what extent performance on such examinations is reflective of individual physician practice. Furthermore, much of the extant literature relies on the simple pass/fail determination without using the entire range of scores available. Nevertheless, the meaning of specific knowledge deficits on examinations is unclear, because, in practice, many physicians have available professional colleagues and consultants as well as additional information sources that they can use in real time to help them evaluate patients and formulate treatment plans. Therefore, knowledge deficits in the artificial setting of an examination do not imply that physicians will make mistakes in caring for patients.

In this issue of the Archives, Holmboe et al attempt to address this question by examining the relationship between scores on the American Board of Internal Medicine MOC examination and subsequent quality of care. They linked physicians who completed the MOC examination between 2000 and 2005 to Medicare patients seen in 2002 or 2003 and examined performance on 5 claims-based process-of-care measures, including annual hemoglobin A1c testing, lipid profiles, and retinal examinations for patients with diabetes, mammography screening, and lipid profiles for patients with cardiovascular disease. Patients were linked to the primary care physician with whom they had the most contact but to be included had to be linked to the same physician in both years. The authors studied approximately 3600 physicians linked to more than 220 000 Medicare beneficiaries. All of the physicians completed the MOC pro-
cess and had between 30 and 800 linked Medicare patients.

Holmboe and colleagues classified physicians as being in the top quartile, middle half, or bottom quartile of performance and then examined the relationship with measured quality using hierarchical models to account for the clustering of patients within physician. They found that scores on the MOC examination were correlated with quality performance for the diabetes and mammography measures but not for the cardiovascular disease measure. They evaluated 3 models (unadjusted, adjusted for patient characteristics, and adjusted for patient and physician characteristics), and each model showed consistent results, although there was some attenuation of the relationship with the addition of patient and physician characteristics as control variables. The bivariate results also provide additional evidence of the construct validity of the examination. Consistent with prior research, foreign medical graduates, older physicians, and solo practitioners had both lower examination scores and lower-quality scores.\(^7\)\(^8\) Patients with more frequent visits and with specialists involved in their care had higher-quality scores for diabetes and cardiovascular disease but not for mammography screening.

Although Holmboe and coauthors’ article represents an important contribution, there are several methodological limitations to their study that are worthy of mention. First, with regard to the quality-of-care measures, the authors were limited to administrative data for deriving their measures; therefore, those measures may not be representative of the higher cognitive processes that are assessed through the certifying examination. None of the measures examines a complex cognitive task, and there is no way to evaluate whether the response to an abnormal screening test result was appropriate without detailed chart review data, which were not available for Holmboe and colleagues’ study. Nonetheless, they are the only measures that were available at the time, and all of them are valid and have been endorsed by national organizations such as the National Quality Forum.

The study also examined several models but did not include one that controls for patient demographic and clinical comorbidities without other control variables. Instead, when the authors controlled for patient characteristics, they also controlled for “access to care” as measured by the number of physician visits and use of specialists. Prior research has shown that more visits are associated with higher quality; the same might be argued for involvement of a pertinent specialist. To the extent that the number of visits and the involvement of a specialist are decisions that physicians make in the course of providing care, controlling for these factors might not be appropriate. Similarly, Holmboe and coworkers included measure-specific patient volume and location of medical school (eg, United States or Canada vs other) as physician predictors. One could reasonably argue that each of these characteristics is associated with both quality and examination scores but that examination scores might be a mediator of the relationship between specific physician characteristics and quality, rather than a true confounder. Finally, although the analysis included physicians who “failed” the MOC examination (they are among the lowest quartile), it does not replicate the pass/fail nature of board certification based on the test. Such information would have been useful for judging the adequacy of recertification as a measure of physician competence.

It should also be recognized that performance on measures of the type included in this study can also be influenced by the systems of care in which a physician practices. For instance, electronic medical records often include reminder systems or prompts that identify patients for whom specific tests or treatments are indicated. Moreover, such systems can also be used to identify patients who are in need of specific tests and to proactively contact them to arrange for testing. Effective systems of care might also use other health care professionals to treat patients along with physicians to ensure that recommended screening and monitoring tests such as those assessed in Holmboe and coauthors’ study are performed systematically, rather than relying on the physician. Therefore, the extent to which performance on these quality measures is determined by systems of care is not clear, nor is the extent to which performance on the cognitive examination is associated with the sorting of physicians into practices with better or worse systems of care.

In summary, Holmboe and colleagues provide evidence that supports the notion that performance on a written cognitive examination that evaluates knowledge and clinical judgment in areas in which an internist should demonstrate competency is associated with quality of care. These results are welcome and demonstrate that board certification (or, in this case, MOC) is potentially a meaningful measure of physician competency. It is not clear, however, whether the simple pass/fail designation used to determine board certification would have the same discriminatory power. More importantly, as noted above, the quality measures that are assessed in this study are not indicative of the higher cognitive processes that are assessed in the cognitive examination. Measures of this type are readily available for health plans and are becoming increasingly available at the level of individual physicians or physician groups. What continues to be lacking is a method for measuring how physicians perform when delivering care reflective of higher cognitive abilities and processes. Cognitive examinations such as the MOC hold the potential to accurately measure this aspect of physician performance, but such a link has yet to be established empirically. If and when this relationship is established empirically, performance on cognitive examinations like the MOC will become an important tool for assessing physician competence.

Bruce E. Landon, MD, MBA, MSc

Correspondence: Dr Landon, Department of Health Care Policy, Harvard Medical School, 180 Longwood Ave, Boston, MA 02115 (landon@hcp.med.harvard.edu).

Financial Disclosure: Dr Landon has received consulting funds from the American Board of Internal Medicine.
Additional Contributions: Barbara J. McNeil, MD, PhD, provided comments on an earlier version of this article, and Emily Corcoran provided editorial assistance.

REFERENCES