Assessing the Quality of Osteoporosis Care in Practice

Weifeng Weng, PhD[1], Brian J. Hess, PhD[2], Lorna A. Lynn, MD[1], and Rebecca S. Lipner, PhD[1]

[1]American Board of Internal Medicine, Philadelphia, PA, USA; [2]Hess Consulting, St-Nicolas, Québec, Canada.

BACKGROUND: Patients with osteoporosis can sustain fractures following falls or other minimal trauma. This risk of fracture can be reduced through appropriate diagnostic testing, pharmacologic therapy, and other readily measured standards of care.

OBJECTIVES: Our aim was to develop a credible clinical performance assessment to measure physicians' quality of osteoporosis care, and determine reasonable performance standards for both competent and excellent care.

DESIGN: This was a retrospective cohort study.

PARTICIPANTS: Three hundred and eighty one general internists and subspecialists with time-limited board certification were included in the study.

MAIN MEASURES: Performance rates on eight evidence-based measures were obtained from the American Board of Internal Medicine (ABIM) Osteoporosis Practice Improvement Module® (PIM), a web-based tool that uses medical chart reviews to help physicians assess and improve care. We applied a patented methodology, using an adaptation of the Angoff standard-setting method and the Dunn-Rankin method, with an expert panel skilled in osteoporosis care to form a composite and establish standards for both competent and excellent care. Physician and practice characteristics, including a practice infrastructure score based on the *Physician Practice Connections Readiness Survey (PPC-RS)*, were used to examine the validity of the inferences made from the composite scores.

KEY RESULTS: The mean composite score was 67.54 out of 100 maximum points with a reliability of 0.92. The standard for competent care was 46.87, and for excellent care it was 83.58. Both standards had high classification accuracies (0.95). Sixteen percent of physicians performed below the competent care standard, while 22 % met the excellent care standard. Specialists scored higher than generalists, and better practice infrastructure was associated with higher composite scores, providing some validity evidence.

CONCLUSIONS: We developed a rigorous methodology for assessing physicians' osteoporosis care. Clinical performance feedback relative to absolute standards of care provides physicians with a meaningful approach to self-evaluation to improve patient care.

KEY WORDS: osteoporosis; quality assessment; performance measurement.

INTRODUCTION

Patients with osteoporosis can sustain fractures following falls or other minimal trauma, often leading to loss of function and independence, with annual direct medical costs estimated at $17 billion.[1] Increasingly, attention has focused on identifying physicians who deliver high-quality care, especially for common medical conditions like osteoporosis, where evidence-based care can improve outcomes and/or reduce costs.[2,3] Through appropriate risk assessment, diagnosis, and treatment,[4,5] the risk of osteoporosis-related fractures can be reduced, alleviating consequent dependency and economic burden.

Meaningful clinical performance assessment includes clinically important evidence-based measures, feasible data collection, and a psychometrically robust assessment.[2,6] Recent research on composites drawn from evidence-based measures has advanced the science of quality assessment.[7–11] Composites are comprehensive measure sets that provide a more reliable assessment than individual measures alone.[8,9] Using a composite to make a judgment about an individual physician's performance in practice requires a defensible absolute standard of performance on the composite. An absolute standard that reflects the level of proficiency required for practice in the profession is a widely-accepted requirement for certification testing of physicians' knowledge.[12] We have previously shown that reasonable absolute standards can be set for composites in diabetes and preventive cardiology care.[10,11] Multiple individual osteoporosis measures are currently available in National Quality Measures Clearinghouse;[13] we believe this is the first study to address the development and evaluation of a composite for osteoporosis care.

In this study, we addressed measuring osteoporosis care quality using evidence-based guidelines.[14,15] We applied a patented methodology[16] to assess physicians' clinical performance. We determined performance standards for both competent care, defined as the standard of care patients should expect to receive from all certified physicians, and excellent care, defined as the quality of care patients can expect to receive from a certified physician who has the knowledge and skills needed to deliver recommended care at a higher achievable standard. We then conducted analyses to

understand the stability and meaningfulness of these standards in measuring osteoporosis care.

METHODS

Instrument and Sample

We used data from the American Board of Internal Medicine (ABIM) Osteoporosis Practice Improvement Module® (PIM), augmented by data on physician and practice characteristics drawn from ABIM's registration and practice characteristics survey. The PIM is a web-based self-evaluation tool that uses medical chart reviews to help physicians improve clinical care. Physicians who make management decisions about patients with or at high risk of osteoporosis can choose to complete the Osteoporosis PIM as part of ABIM's Maintenance of Certification (MOC) program.[17]

To complete the PIM, individual physicians abstracted 25 patient charts using a retrospective or prospective sequential sample, or a systematic random sample. Eligible patients included patients aged 18 years and older with a diagnosis of osteoporosis, osteopenia, or a prior low impact fracture; women aged 65 years and older; and men aged 70 years and older. Patients must have received care from the physician's practice for at least 12 months (with at least one visit within the past 12 months). Physicians also completed a web-based adaptation of the *Physician Practice Connections® Readiness Survey* (PPC-RS) that assesses the physician's practice infrastructure. The PPC-RS is a precursor to the *Physician Practice Connections®–Patient-Centered Medical Home* assessment survey developed by the National Committee for Quality Assurance (NCQA).[18] It asks physicians to report on whether the following are present in their practice: structured quality measurement and improvement; patient data tracking systems; processes for care management; proactive management of important conditions; patient-centered self-care support and education; and technology infrastructure. The PPC-RS score is on a scale of 0 to 100 points, with 100 being the maximum possible score.

We obtained data from a retrospective cohort of 463 physicians who completed the osteoporosis PIM between November 2011 and January 2013.

Performance Measures

We used eight evidence-based process measures (Table 1; online Appendix 1 for specification) derived from evidence-based guidelines on the prevention, diagnosis, and treatment of osteoporosis.[4,5] Eighty-two physicians were excluded because of less than five eligible patients for some measures, yielding 381 physicians. Performance on each measure was defined as the percentage of a physician's patient panel receiving the service.

Composite and Standard-Setting Methodology

The methodology has been previously described[9–11] and patented.[16] For each measure, we established a threshold for delivering both competent and excellent care using an expert panel and an adaptation of the Angoff standard-setting method.[19] A panel of 12 physicians was selected through a call for nominations to represent essential perspectives of clinical practice around osteoporosis care (online Appendix 2). All were certified in internal medicine and/or rheumatology, endocrinology, and geriatric medicine. A majority spent at least 50 % of their time in clinical practice. The panel reached a shared understanding of the characteristics of a hypothetical physician whose care for patients with or at risk for osteoporosis would be just at the threshold for competent care (i.e., a "marginally competent" physician). This physician would have a basic understanding of osteoporosis care, but would not always deliver appropriate care to all patients.

Next, thresholds for *each* measure were estimated for how the hypothetical "marginally competent" physician would perform (online Appendix 3). For example, to determine the threshold for bone density testing, each panelist independently answered the question "What percentage of eligible patients seen by a 'marginally competent' physician would have had bone density testing?" Statistics describing the characteristics of patients seen by the 381 physicians were presented. After panelists shared their initial estimates with the other panel members, actual results for each measure were presented as a "reality check." Panelists discussed differences and were given a chance to change their estimates. Final estimates were averaged across panelists to determine each measure's threshold for competent care.

The Dunn-Rankin method was used to determine the point value for each measure used in the composite score (online Appendix 3).[20] Panelists independently rated each measure's importance to delivering quality osteoporosis care using an 11-point Likert scale (0=Not at all important to 10=Very important). Panelists' ratings of importance were then used to derive point values for each measure. The points across all eight measures were scaled to sum to 100.

To determine the competent care standard (online Appendix 3), the threshold for each measure was multiplied by its point value (Table 1). For example, the threshold of 44.6 % for calcium intake assessment and counseling was multiplied by its point value (i.e., importance weight) of 13 (i.e., 0.446×13=5.80). The products for all measures were summed to yield the "standard" for competent care. Likewise, to determine the excellent care standard, the panel repeated the steps above (Table 1). This time, the panel considered a hypothetical "marginally excellent" physician whose care would be just at the threshold for excellent care.

Computing a Physician's Performance Score

A physician's performance on each measure was multiplied by its point value, but no points were awarded unless the threshold for competent care was met. Points earned by a physician for all measures were summed to yield a composite score that ranged between 0 and 100 points (online Appendix 3).

Table 1. Computation of the Competent and Excellent Care Standard Based on Measures from the ABIM Osteoporosis PIM

Measure	Physician Performance Mean (SD)[*]	Denominator Size Mean (SD)[†]	Reliability Coefficient	Points[‡]	Competent Care Standard		Excellent Care Standard	
					Threshold[§]	Threshold X Points[‖]	Threshold[§]	Threshold X Points[‖]
Bone Density Testing in Older Adults	86 % (19 %)	20 (4.9)[¶]	0.83	17	56.7 %	9.64	86.7 %	14.74
Pharmacologic Therapy	65 % (26 %)	13 (6.1)[¶]	0.70	15	41.9 %	6.29	75.0 %	11.25
Falls Risk Management: Falls Risk Screening, Complete Assessment, and Plan of Care	36 % (37 %)	25 (2.0)	0.97	11	13.3 %	1.46	72.9 %	8.02
Current Level of Physical Activity Documentation and Exercise Advice	59 % (32 %)	25 (2.0)	0.92	12	38.6 %	4.63	80.8 %	9.70
Current Level of Alcohol Use and Advice on Potentially Hazardous Drinking Prevention	82 % (26 %)	25 (2.0)	0.89	10	62.1 %	6.21	90.0 %	9.00
Smoking Status Documentation and Cessation Support	97 % (8 %)	25 (2.0)	0.69	9	77.1 %	6.94	93.8 %	8.44
Calcium Intake Assessment and Counseling	72 % (29 %)	25 (2.0)	0.91	13	44.6 %	5.80	86.3 %	11.22
Vitamin D Intake Assessment and Counseling	76 % (24 %)	25 (2.0)	0.86	13	45.4 %	5.90	86.3 %	11.22
Standard (= Sum of Threshold x Points)						46.87		83.58

[*] *The physician performance mean is the average proportion of patients meeting the measure across the sample of 381 physicians*
[†] *Denominator size is the number of eligible patients for a measure for a given physician*
[‡] *Points: Scoring weight for each measure. The Dunn-Rankin scaling method was used to calculate the point value for each measure, incorporating both the average importance rating and maximum rating provided by each panelist so that the sum of the points ranged from 0 to 100*
[§] *Threshold: Each panelist estimated the percent of patients seen by a "marginally competent"/"marginally excellent" physician that would meet the measure. The estimates were averaged across the panelists to determine each measure's threshold for competent/excellent care*
[‖] *A physician must earn at least the threshold for competent care standard to be awarded any points*
[¶] *The minimum number of eligible patients for the first two measures is five; all patients are eligible for the rest of the measures*
ABIM American Board of Internal Medicine; PIM Practice Improvement Module

Estimating Reliability and Classification Accuracy

We used the bootstrap sampling method (1,000 samples per physician) to estimate the reliability of the composite[21] and the classification accuracy of the two standards.[22,23] We estimated the standard error of measurement through the bootstrap samples, and then derived the reliabilities using the classical true score model. Classification accuracy describes the accuracy of decisions made and how well they minimize false classifications. For both standards, each physician was first classified as meeting or not meeting the standard based on the observed sample. Second, a composite was calculated for each physician's bootstrap sample and a classification was made. If classification decisions were the same for the bootstrap and observed samples, the decision was deemed accurate. The proportion of accurate classifications over all samples for each physician was calculated. These were then averaged across physicians to compute the classification accuracy index, which can range from 0 to 1.

Statistical Analyses

Chi-square tests and *t*-tests were used to compare the differences between our study sample and those who completed other available ABIM PIMs. We computed coefficient of variation and intraclass correlation coefficient to assess inter-rater agreement in panelists' final estimates of each measure's threshold. To examine the validity of the assessment, we used multivariate regression to determine if physician characteristics (e.g., specialty), practice characteristics (e.g., practice type), and practice infrastructure level (PPC-RS score) were meaningfully associated with the composite. We utilized the same set of variables in logistic regression models to examine whether they predicted meeting competent and excellent care standards. We conducted a sensitivity analysis to evaluate the effect of patient risk group (e.g., having osteoporosis versus being at risk due to age) on the composite. Statistical analyses were conducted using SAS Version 9.3. All data were Health Insurance Portability and Accountability Act (HIPAA) compliant and data were reported only in aggregate. Permission to use data for research purposes was granted by physicians upon enrollment in MOC.[24] The ABIM Privacy Policy can be found at www.abim.org/privacy.aspx. We were blinded to physicians' identities, and we viewed and analyzed the data in aggregate. Essex Institutional Review Board, Inc. approved this study.

RESULTS

Most of the 381 physicians used retrospective sequential (54.3 %) or systematic random (33.6 %) samples. Table 2

Table 2. Physician and Practice Characteristics of the Study Sample and All Physicians Who Completed Any One of the ABIM PIMs

Characteristics	Study Sample (N=381)	Initial sample (N=463) [†]	All Physicians (N=11,813)	p value* (study sample versus all
Age [mean (SD)]	46.7 (6.5)	47.0 (6.7)	47.3 (6.9)	0.09
Female (%)	50.1	49.2	33.3	< 0.001
Equated IM certification exam score [mean (SD)] [‡]	490 (88.5)	490 (89.7)	495 (87.1)	0.25
Birth/Training (%)				0.005
U.S. born/U.S. trained	58.8	60.0	53.2	
Foreign born/US trained	12.6	12.5	10.6	
Foreign born/foreign trained	21.5	21.0	29.9	
U.S. born/foreign trained	7.1	6.5	6.3	
Specialty (%)				< 0.001
General internist	55.6	57.0	41.3	
Rheumatologist	19.2	16.6	3.2	
Endocrinologist	10.0	9.3	3.0	
Geriatrician	7.9	9.5	4.1	
Other subspecialist [§]	7.4	7.6	48.4	
% patient care in office setting	83.7 (20.5)	83.0 (21.7)	53.6 (35.0)	< 0.001
PPC-RS score [mean (SD)]	59.5 (19.1)	59.3 (19.2)		
% Medicaid/dual eligible patients	18.5 (19.7)	19.2 (20.3)	24.5 (21.0)	< 0.001
Practice Type (%)				< 0.001
Solo	13.4	13.2	10.2	
Group, 2–10	33.9	32.2	23.7	
Group, ≥ 11	23.4	22.0	18.1	
Academic/hospital	23.1	24.6	29.8	
Other	6.3	8.0	18.3	

* t-tests were conducted for age, percent of patient care time in office setting, equated internal medicine (IM) certification exam scores, and percent of Medicaid/dual eligible patients in the practice; chi-squared tests for the categorical variables
† Initial sample includes physicians in the study sample and the 82 physicians who were dropped due to fewer than five eligible patients for pharmacologic therapy measure. THere were no statistically significant differences between our study and initial sample across all physician and practice characteristics and PPC-RS score (p values ranged from 0.45 to 0.95)
‡ Scores were statistically equated to be comparable over time and scaled to have a mean=500 and SD=100. Exam scores were available for 364 physicians in our study sample, for 441 physicians in the initial osteoporosis completer sample, and for 11,179 physicians who completed any one of the available ABIM PIMs
§ Thirty-five other subspecialists in the initial sample are certified in hematology, nephrology, and other subspecialties within the IM field
ABIM American Board of Internal Medicine; PIMs Practice Improvement Modules; U.S. United States; IM Internal Medicine; PPC-RS Physician Practice Connections Readiness Survey

compares physician and practice characteristics of our study sample with the initial 463 osteoporosis PIM completers and with 11,813 physicians who completed any ABIM PIM during the same time period. No statistically significant differences were observed between our study sample and the initial sample. Compared with all physicians completing any PIM, our study sample had similar age and internal medicine certification exam scores, but differed in other characteristics. Our sample had a larger proportion of general internists (55.6 %) and subspecialists whose training includes emphasis on osteoporosis: rheumatologists (19.2 %) and endocrinologists (10.0 %).

The mean number of charts abstracted was 25.36 (SD=2.0; range: 24-50). The mean patient age was 72.21 years (SD=10.48); 89.4 % were female; 39.8 % had osteoporosis, 28.5 % had osteopenia and 3.1 % had a prior low-impact fracture; 23.6 % were women 65 years or older and 5.0 % were men 70 years or older without these diagnoses. Table 1 shows physicians' mean performance on each measure. The reliability coefficients of individual measures ranged from 0.69 to 0.97. The estimated reliability of the composite was 0.92, indicating that 92 % of the measured performance reflects true ability, not random error. The mean composite was 67.54 out of a possible 100 points (SD=18.22).

Table 1 presents the thresholds and number of points assigned to each measure and the standards for both competent and excellent care. The absolute percentage point differences between the panel's overall initial and final threshold estimates were small for most measures (2.3 % to 6.7 % for competent care; 2.1 % to 4.6 % for excellent care), except for the falls risk management competent care threshold (12.8 %). The variability in panelists' final threshold ratings was small (coefficients of variation: 0.06 to 0.28 for competent care; 0.02 to 0.13 for excellent care). The inter-rater agreement (0.81 for competent care; 0.68 for excellent care) indicated substantial and moderate agreement in panelists' final judgments. Panelists showed consistency in their ratings of the importance of individual measures; variability in points assigned was small, ranging from 0.4 to 2.1 points on average.

The standard for competent care was 46.87 points (Table 1). Figure 1 displays a histogram of physician composite scores. Three hundred and twenty-one physicians (84.3 %) met that standard. Those not meeting the standard performed statistically differently across all measures compared with those meeting it. Rheumatologists (94.5 %) and endocrinologists (92.1 %) had a higher proportion meeting the standard than general internists (81.1 %) and geriatricians (76.7 %). The classification accuracy index was high (0.95), meaning that,
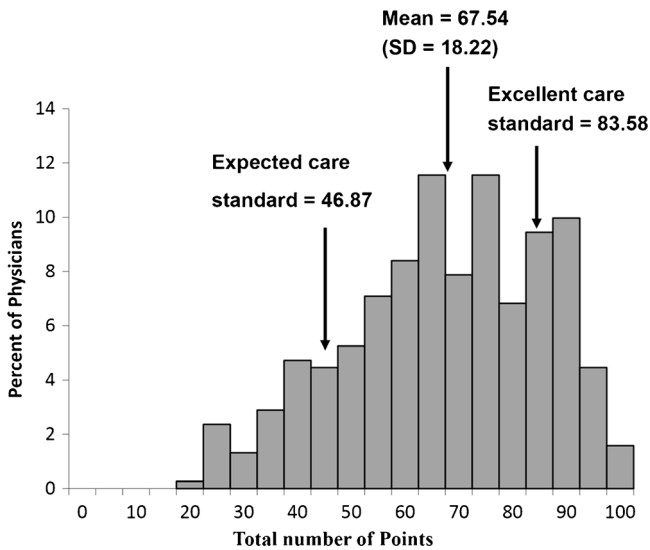
Figure 1. Distribution of composite scores with standards for performance (N=381 physicians).

with repeated sampling, the same classification result (meeting the standard or not) would occur 95 % of the time.

The standard for excellent care was 83.58 points (Table 1) and was achieved by 85 physicians (22.3 %). Rheumatologists (38.4 %) and endocrinologists (34.2 %) had a higher proportion meeting the excellent standard than general internists (16.0 %) and geriatricians (26.7 %). The classification accuracy index was again high (0.95).

Table 3. Results of Multivariate Regression Analysis Associating Physician and Patient Characteristics with Composite Scores (N=364)[*]

Characteristics	b	Standard error of b	p value
Physician Characteristics			
Rheumatologist[†]	11.21	2.40	< 0.001
Endocrinologist[†]	9.65	3.14	0.00
Geriatrician[†]	2.39	4.30	0.58
Other specialist[†]	1.00	3.68	0.79
Equated IM certification exam score [‡]	0.00	0.01	0.83
Female	-1.83	1.86	0.33
Age	0.13	0.15	0.40
% patient care in office setting	0.07	0.05	0.13
Practice/Patient Characteristics			
PPC-RS score [§]	0.21	0.05	<0.001
Solo practice	4.46	2.65	0.09
Average patient age	-0.49	0.22	0.03
% female patients	0.22	0.07	0.00
Medicaid/dual eligible patients (%) [‖]			
6–25 %	1.22	1.98	0.54
>25 %	-0.90	2.46	0.71

[*] *17 observations were dropped due to missing data; they did not have equated exam scores since they took the exam prior to 1986 and scores are not comparable*
[†] *Reference group: General internist*
[‡] *Scores were from the physicians' first attempt, and were statistically equated to be comparable over time and scaled to have a mean of 500 and SD of 100*
[§] *PPC-RS scores ranged from 0 to 100 and are an assessment of practice infrastructure*
[‖] *Omitted category is≤5 %*
IM Internal Medicine; PPC-RS Physician Practice Connections Readiness Survey

Table 3 presents the association between physician, patient characteristics, practice infrastructure (PPC-RS) scores, and the composite. Twenty percent of the variance in the composite was explained by the regression model. A few variables explained most of the variances, as indicted by their squared partial correlation coefficients: PPC-RS (0.058), certification in rheumatology (0.059) or endocrinology (0.026), and proportion of female patients (0.028). Controlling for other characteristics, rheumatologists and endocrinologists were estimated to have a higher composite score (11.21 and 9.65 points, respectively) than general internists; a physician with a PPC-RS score of 73.3 (75th percentile) would have a predicted composite score 3.05 points higher than one with a score of 59 (50th percentile). The average patient age in a physician's sample and the physician's composite score were negatively related; proportion of female patients was positively related.

Table 4 presents the logistic regression results for the association between physician and practice characteristics and meeting the two care standards. Controlling for other characteristics, rheumatologists were more likely to meet both competent (adjusted odds ratio [AOR], 7.41; 95 % CI, 1.67 to 32.84) and excellent (AOR, 3.56; CI, 1.79 to 7.09) care standards than general internists. Endocrinologists were more likely to meet the excellent care standard (AOR, 3.24; CI, 1.34 to 7.85) than general internists; and also tended to be more likely to meet the competent care standard compared to general internists; however, the result was not statistically significant (AOR=3.46; CI, 0.70 to 17.2). For a one-point increase in PPC-RS score out of a possible 100 points, there was a 2 % and 3 % increase in the likelihood of meeting competent (AOR, 1.02; CI, 1.01 to 1.04) and excellent (AOR, 1.03; CI, 1.01 to 1.04) care standards. For a one-year increase in sampled patients' average age, there was a 10 % reduction in the likelihood of meeting competent care standard (AOR, 0.90; CI, 0.83 to 0.98). For a 1 % increase in the proportion of sampled female patients, there was a 4 % increase in the likelihood of meeting the competent care standard (AOR, 1.04; CI, 1.02 to 1.06).

We compared the results including only patients with osteoporosis with the original results, which included patients with other risk factors. Two hundred and seventy-two physicians had at least five eligible patients for each measure. There was an average increase of 4.1 points in composite scores. Physician classification changed for 43 physicians (15.9 %), with 21 now meeting the competent care standard, 18 now meeting the excellent care standard, and two physicians now falling below each standard.

DISCUSSION

We evaluated *individual* physicians' osteoporosis care by creating a robust composite using clinical data and a

Table 4. Odds of Meeting the Competent Care and Excellent Care Standard From a Single Multivariable Logistic Regression Model Including All Listed Characteristics (N=364)[*]

Characteristics	Meeting the competent care standard		Meeting the excellent care standard	
	OR (95 % CI)	p value	OR (95 % CI)	p value
Physician Characteristics				
Rheumatologist[†]	7.41 (1.67–32.84)	0.01	3.56 (1.79–7.09)	< 0.001
Endocrinologist[†]	3.46 (0.70–17.2)	0.13	3.24 (1.34–7.85)	0.01
Geriatrician[†]	1.75 (0.46–6.65)	0.41	2.57 (0.68–9.71)	0.17
Other specialist[†]	0.56 (0.18–1.76)	0.32	0.54 (0.11–2.66)	0.45
Equated IM certification exam score [‡]	1.00 (1.00–1.00)	0.82	1.00 (1.00–1.00)	0.31
Female	0.53 (0.27–1.07)	0.08	0.90 (0.50–1.62)	0.72
Age	1.02 (0.96–1.07)	0.58	1.01 (0.96–1.06)	0.67
% Patient care in office setting	1.00 (0.99–1.02)	0.78	1.02 (1.00–1.04)	0.04
Practice/Patient Characteristics				
PPC-RS score [§]	1.02 (1.01–1.04)	0.01	1.03 (1.01–1.04)	< 0.001
Solo practice	0.83 (0.32–2.14)	0.70	1.67 (0.77–3.62)	0.19
Average patient age	0.90 (0.83–0.98)	0.01	0.97 (0.91–1.04)	0.41
% female patients	1.04 (1.02–1.06)	<0.001	1.02 (0.99–1.04)	0.24
Medicaid/dual eligible patients (%) [‖]				
6–25 %	0.67 (0.32–1.4)	0.28	1.62 (0.87–3.00)	0.13
> 25 %	0.42 (0.18–1.00)	0.05	1.61 (0.76–3.38)	0.21

[*] *17 observations were dropped due to missing data; they did not have equated exam scores since they took the exam prior to 1986 and scores are not comparable*
[†] *Reference group: General internist*
[‡] *Scores were from the physicians' first attempt and were statistically equated to be comparable over time and scaled to have a mean of 500 and SD of 100*
[§] *PPC-RS scores ranged from 0 to 100 and are an assessment of practice infrastructure*
[‖] *Omitted category is≤5 %.*
IM Internal Medicine; PPC-RS Physician Practice Connections Readiness Survey

rigorous methodology for determining absolute standards for both competent and excellent care. The composites were meaningfully associated with physician and practice characteristics, supporting the validity of the methodology.[25] Subspecialists in rheumatology and endocrinology demonstrated higher quality care than general internists on our composite; this is expected, given that their training and practices are more likely to focus on care for osteoporosis patients than those of generalists who care for patients with a broad spectrum of healthcare needs. The negative association between patient age and composite scores might be due to the competing needs of the elderly patients with multiple chronic conditions. The association between patient gender and composite indicated that osteoporosis care for men lagged behind; while osteoporosis and osteoporosis-related fractures occur more often in women, the mortality rate associated with fractures is higher in men.[26] Including only patients with osteoporosis versus those at risk had an effect on physicians' composite scores.

Both competent and excellent care standards classified physicians accurately, as indicated by their high classification accuracy value. Most physicians (84 %) met the competent care standard, and about 22 % met the excellent care standard. Some characteristics were more strongly associated with meeting the competent care standard than with meeting the excellent care standard. Notably, age and gender of the patients sampled were statistically significant in predicting meeting only the competent standard. The competent care standard serves to identify those physicians whose care for osteoporosis

patients was below what patients should be able to expect; individual measures highlight specific areas for improvement. With more research on the composite's generalizability and linkage with outcomes, performance reports based on this approach could potentially inform patients' and purchasers' health care choices, reward physicians providing superior care, and guide improvement in care. We utilized the results from this study to help physicians in the ABIM MOC program to identify areas of improvement in osteoporosis care. Prior cognitive testing with physicians who completed other ABIM PIMs indicated that performance relative to their peers as well as to an absolute standard provided valuable feedback that they were not getting from any other source.[11] We created feedback reports for the ABIM Osteoporosis PIM that include histograms of the composite and individual measures showing the physician's performance compared to the two absolute performance standards and relative to their peers. Physicians see only their own individual performance and the distribution of the performance of peers. For physicians with fewer than five eligible patients for some measures, the composite score cannot be calculated. Information on performance on individual measures is provided to guide improvement activities. This study has several limitations. First, physicians voluntarily selected the osteoporosis PIM. These results may not generalize to all physicians providing osteoporosis care. Second, there was no audit process to ensure that instructions for sampling and data extraction were followed. However, prior work confirmed the accuracy of physician-reported data in a quality improvement framework,[27] and physicians had no reason to provide false results as there were no consequences for low or

high performance. Third, we based the relative importance of each measure on ratings by 12 experts; this may be better informed by further research on outcomes of care. The range of experts' experience and perspective, however, provided a balance that promoted fairness. Finally, there was no control for differences in patient panel characteristics; for process measures, formal risk adjustment is not generally necessary.

CONCLUSION

We have established an approach for assessing a physician's quality of osteoporosis care that is backed by evidence-based guidelines and empirical data. Performance on individual measures and on the composite compared to absolute standards of care may provide a more meaningful way to judge whether physicians provided evidence-based quality of care than a relative standard that simply depends on how other physicians perform. Our hope is that this type of feedback will help physicians identify areas where they can improve osteoporosis care, and decrease morbidity, dependency, and healthcare costs.

Corresponding Author: Weifeng Weng, PhD; American Board of Internal Medicine, 510 Walnut Street, Suite 1700, Philadelphia, PA 19106, USA (e-mail: wweng@abim.org).

REFERENCES

1. **Burge R, Dawson-Hughes B, Solomon D, Wong JB, King A, Tosteson A.** Incidence and economic burden of osteoporosis-related fractures in the United States, 2005-2025. J Bone Miner Res. 2007;22:465–475.

2. **Landon BE, Normand S-LT, Blumenthal D, Daley J.** Physician clinical performance assessment: Prospects and barriers. JAMA. 2003;290:1183–1189.

3. Centers for Medicare & Medicaid Services. Physician Quality Reporting System. Available at http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/PQRS/index.html. Accessed March 5, 2015.

4. National Osteoporosis Foundation. Clinician's Guide to Prevention and Treatment of Osteoporosis. Washington, DC: National Osteoporosis Foundation; 2010.

5. **Watts N, Bilezikian J, Camacho P, et al.** American association of clinical endocrinologists medical guidelines for clinical practice for the diagnosis and treatment of postmenopausal osteoporosis. Endocr Prac. 2010;16:1–37.

6. **Scholle SH, Pawlson LG, Solberg LI, et al.** Measuring practice systems for chronic illness care: Accuracy of self-reports from clinical personnel. Joint Comm J Qual Pat Saf. 2008;34:407–416.

7. **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:S48–S52.

8. **Kaplan S, Griffith J, Price L, Pawlson L, 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.

9. **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.

10. **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.

11. **Lipner RS, Weng W, Caverzagie KJ, Hess BJ.** Physician performance assessment: prevention of cardiovascular disease. Adv Health Sci Educ. 2013;18:1029–1045.

12. **Clauser BE, Case SM.** Testing for licensure and certification in the professions. In: Brennan RL, ed. Educational Measurement. 4th ed. West Port, CT: Praeger Publishers; 2006:701–731.

13. Agency for Healthcare Research and Quality. National Quality Measures Clearinghouse. Available at: http://www.qualitymeasures.ahrq.gov/browse/by-topic.aspx. Accessed March 5, 2015.

14. **Cohen K, Maier D.** Osteoporosis: Evaluation of screening patterns in a primary-care group practice. J Clin Densitom. 2008;11:498–502.

15. **Teng G, Curtis J, Saag K.** Quality health care gaps in osteoporosis: How can patients, providers, and the health system do a better job? Curr Osteoporos Rep. 2009;7:27–34.

16. **Lipner RS, Hess BJ, Weng W, Arnold A, inventors**. American Board of Internal Medicine, assignee. Method and system for determining a fair benchmark for physician's quality of patient care. United States Patent No. 08452610. May 28, 2013.

17. American Board of Internal Medicine. Maintenance of Certification Program Practice Assessment. Available at http://www.abim.org/maintenance-of-certification/earning-points/practice-assessment/default.aspx. Accessed March 5, 2015.

18. **Solberg LI, Asche SE, Pawlson LG, Scholle SH, Shih SC.** Practice systems are associated with high-quality care for diabetes. Am J Manag Care. 2008;14(2):85–92.

19. **Angoff WH.** Scales, norms, and equivalent scores. In: Thorndike RL, ed. Educational Measurement. 2nd ed. Washington, DC: American Council on Education; 1971:508–600.

20. **Dunn-Rankin P.** Scaling Methods. Hillsdale, NY: Lawrence Erlbaum; 1983.

21. **Reeves D, Campbell SM, Adams J, Shekelle PG, Kontopantelis E, Roland MO.** Combining multiple indicators of clinical quality: An evaluation of different analytic approaches. Med Care. 2007;45:489–496.

22. **Lee W-C, Brennan RL, Wan L.** Classification consistency and accuracy for complex assessments under the compound multinomial model. Appl Psychol Meas. 2009;33:374–390.

23. **Brennan RL, Wan L.** Bootstrap procedures for estimating decision consistency for single-administration complex assessments. Iowa City, IA: Center for Advanced Studies in Measurement and Assessment, The University of Iowa; 2004. Report No. 7.

24. American Board of Internal Medicine. ABIM HIPAA Business Associate Agreement. Available at: http://www.abim.org/pdf/hipaa/hipaa-compliance.pdf. Accessed March 5, 2015.

25. **Weng W, Arnold GK, Lynn LA, Lipner RS**. Estimating reliability and decision consistency of physician practice performance assessment. In: AcademyHealth Annual Research meeting; 2009.

26. **Guggenbuhl P.** Osteoporosis in males and females: Is there really a difference? Joint Bone Spine. 2009;76:595–601.

27. **Holmboe ES, Meehan TP, Lynn LA, Doyle P, Sherwin T, Duffy FD.** Promoting physicians' self-assessment and quality improvement: The ABIM diabetes practice improvement module. J Cont Educ Health Prof. 2006;26:109–119.