
Using Patient Care Quality Measures to Assess Educational Outcomes

Susan R. Swing, PhD, Sandra Schneider, MD, Ken Bizovi, MD, Dane Chapman, MD, PhD, Louis G. Graff, MD, Cherri Hobgood, MD, Thomas Lukens, MD, PhD, Martha J. Radford, MD, Arthur Sanders, MD, Rebecca Smith-Coggins, MD, Linda Spillane, MD, Laura Hruska, MEd, Robert L. Wears, MD

Abstract

Objectives: To report the results of a project designed to develop and implement a prototype methodology for identifying candidate patient care quality measures for potential use in assessing the outcomes and effectiveness of graduate medical education in emergency medicine.

Methods: A workgroup composed of experts in emergency medicine residency education and patient care quality measurement was convened. Workgroup members performed a modified Delphi process that included iterative review of potential measures; individual expert rating of the measures on four dimensions, including measures quality of care and educational effectiveness; development of consensus on measures to be retained; external stakeholder rating of measures followed by a final workgroup review; and a post hoc stratification of measures. The workgroup completed a structured exercise to examine the linkage of patient care process and outcome measures to educational effectiveness.

Results: The workgroup selected 62 measures for inclusion in its final set, including 43 measures for 21 clinical conditions, eight medication measures, seven measures for procedures, and four measures for department efficiency. Twenty-six measures met the more stringent criteria applied post hoc to further stratify and prioritize measures for development. Nineteen of these measures received high ratings from 75% of the workgroup and external stakeholder raters on importance for care in the ED, measures quality of care, and measures educational effectiveness; the majority of the raters considered these indicators feasible to measure. The workgroup utilized a simple framework for exploring the relationship of residency program educational activities, competencies from the six Accreditation Council for Graduate Medical Education general competency domains, patient care quality measures, and external factors that could intervene to affect care quality.

Conclusions: Numerous patient care quality measures have potential for use in assessing the educational effectiveness and performance of graduate medical education programs in emergency medicine. The measures identified in this report can be used as a starter set for further development, implementation, and study. Implementation of the measures, especially for high-stakes use, will require resolution of significant measurement issues.

ACADEMIC EMERGENCY MEDICINE 2007; 14:463–473 © 2007 by the Society for Academic Emergency Medicine

Keywords: outcome and process assessment (health care), quality indicators, educational measurement, internship and residency, program evaluation, emergency medicine

From the Department of Research and Education, Accreditation Council for Graduate Medical Education (SRS, LH), Chicago, IL; Department of Emergency Medicine, University of Rochester School of Medicine and Dentistry (SS, LS), Rochester, NY; Emergency Department, Providence St. Vincent Health Center (KB), Portland, OR; Emergency Department, Sanpete Valley Hospital/Intermountain Health Care (DC), Mt. Pleasant, UT; Departments of Emergency Medicine and Internal Medicine, University of Connecticut School of Medicine (LGG), Farmington, CT; Department of Emergency Medicine, New Britain General Hospital (LGG), New Britain, CT; Office of Educational Development, Department of Emergency Medicine, University of North Carolina School of Medicine (CH), Chapel Hill, NC; Department of Emergency Medicine, MetroHealth Medical Center (TL), Cleveland, OH;

Department of Medicine, New York University School of Medicine (MJR), New York, NY; Department of Emergency Medicine, Arizona Health Sciences Center (AS), Tucson, AZ; Department of Emergency Medicine, Stanford University Medical Center (RS-C), Palo Alto, CA; and Department of Emergency Medicine, University of Florida College of Medicine (RLW), Jacksonville, FL. Received October 2, 2006; revision received November 28, 2006; accepted December 3, 2006.

Supported by grant 034768 from the Robert Wood Johnson Foundation (to SRS).

This work represents the perspectives of the authors and not those of their affiliated organizations.

Contact for correspondence and reprints: Susan R. Swing, PhD; e-mail: srs@acgme.org.

Graduate medical education (GME) programs are expected to graduate residents who can practice competently and independently.¹ Ideally, the newly graduated, competent physician will be able to provide quality care: care that is effective, safe, efficient, timely, equitable, and patient centered.² An assumption of the Accreditation Council for Graduate Medical Education's (ACGME's) general competency and outcome assessment initiative is that resident physician competence results when GME programs provide learning opportunities that foster residents' development in the six general competency domains established by the ACGME³ and the American Board of Medical Specialties.

Patient care settings are a primary venue for resident learning. Acquisition of competency occurs as residents care for patients with the assistance of more experienced physician teachers. This includes applying input and feedback from their teachers and modeling their teachers' care processes. Therefore, quality of care for patients treated and managed in learning environments is directly attributable, at least in part, to the capabilities and competence of residents and their teachers and is indirectly attributable to other features of the educational program that contribute to learning.

Hospital and practicing physician performance are already being assessed using quality-of-care measures, such as desired patient outcomes and condition-specific care processes associated with desired outcomes.^{4,5} Similar measures, selected or adjusted for use in educational environments, could function as educational outcomes. These indices would directly measure the extent to which residents have learned to provide quality care and indicate the educational effectiveness of the program. These measures could add value by indicating specific ways patient care performance needs to change. This type of feedback is not an inherent quality of the current, most commonly used methods for assessing resident learning and performance, that is, clinical performance ratings and written examinations. The patient care quality measures could also function as indicators of the educational potential of the patient care and learning environment.

Use of patient care process and outcome measures for assessment by residency programs would align with the ACGME's phase 3 implementation guideline for the Outcome Project.⁶ The phase 3 goal is to integrate the general competencies and patient care and to begin using external measures, such as quality-of-care indicators, to assess program performance. Associating competencies with quality-of-care measures and linking competencies with educational experiences whereby they are fostered could help elucidate ways to improve education, resident performance, and patient care.

Candidate measures for assessing emergency department (ED) care quality have been presented in three recently published reports.⁷⁻⁹ They include some of the disease- and condition-specific measures currently used at a national level for hospital performance assessment and improvement. To the best of our knowledge, however, no one has examined whether these or other patient care quality measures would be appropriate or useful for assessing emergency medicine (EM) residency education. This article reports the results of a project designed to develop and implement a prototype methodology for

identifying and evaluating candidate patient care quality measures for potential use in assessing the outcomes and effectiveness of GME in EM.

METHODS

The measure identification and evaluation activity took place through the following activities: 1) construction and orientation of the GME and Patient Care Quality Workgroup that functioned as the expert panel; 2) performance of a six-phase modified Delphi process, involving the workgroup and external stakeholders as raters of the candidate measures; and 3) construction and application of a framework for examining the validity of the measures for assessing residency educational effectiveness. Figure 1 presents a more detailed overview of the steps.

Construction and Orientation of the Workgroup

The GME and Patient Care Quality Workgroup was the primary development group. The main selection criteria for group members was expertise in residency education and/or quality measurement. A criteria for the overall group composition was representation of the major stakeholder groups in EM: the Residency Review Committee (RRC), American Board of Emergency Medicine, American College of Emergency Physicians, Society for Academic Emergency Medicine, and Council of Emergency Medicine Residency Directors. Potential members were identified through peer nominations, publication records, their involvement in high-profile activities in residency education, or physician performance measurement. Members were invited to participate by the workgroup chair.

The workgroup exhibited the following characteristics. There were four members of the RRC from three appointing bodies. Six of the workgroup members had one or more primary organizational affiliations, as determined by board or committee membership within the organization; the other members were not actively engaged in EM organizations. Including the RRC members, organizational representation in the workgroup was as follows: American College of Emergency Physicians ($n = 3$), American Board of Emergency Medicine ($n = 1$), Society for Academic Emergency Medicine ($n = 6$), and Council of Emergency Medicine Residency Directors ($n = 3$). Among the Council of Emergency Medicine Residency Directors members were a current program director, an associate program director, and a distinguished educator. An American Board of Emergency Medicine executive staff member attended and observed the workgroup meetings. All workgroup members had expertise in quality and performance measurement, residency education, or both, as evidenced by records of scholarly publication and positions held (e.g., residency program director, chief of hospital quality, or representative to the American Medical Association's Consortium on Physician Performance Improvement [$n = 2$]). One of the quality experts was a cardiologist. Nine of the workgroup members (all physicians) participated in all aspects of the measure identification and discussion as described in the following text. The remaining members participated in a subset of the processes.

Orientation of the workgroup consisted of presentation of the project aims and the rationale for considering patient care quality and outcome measures for assessing

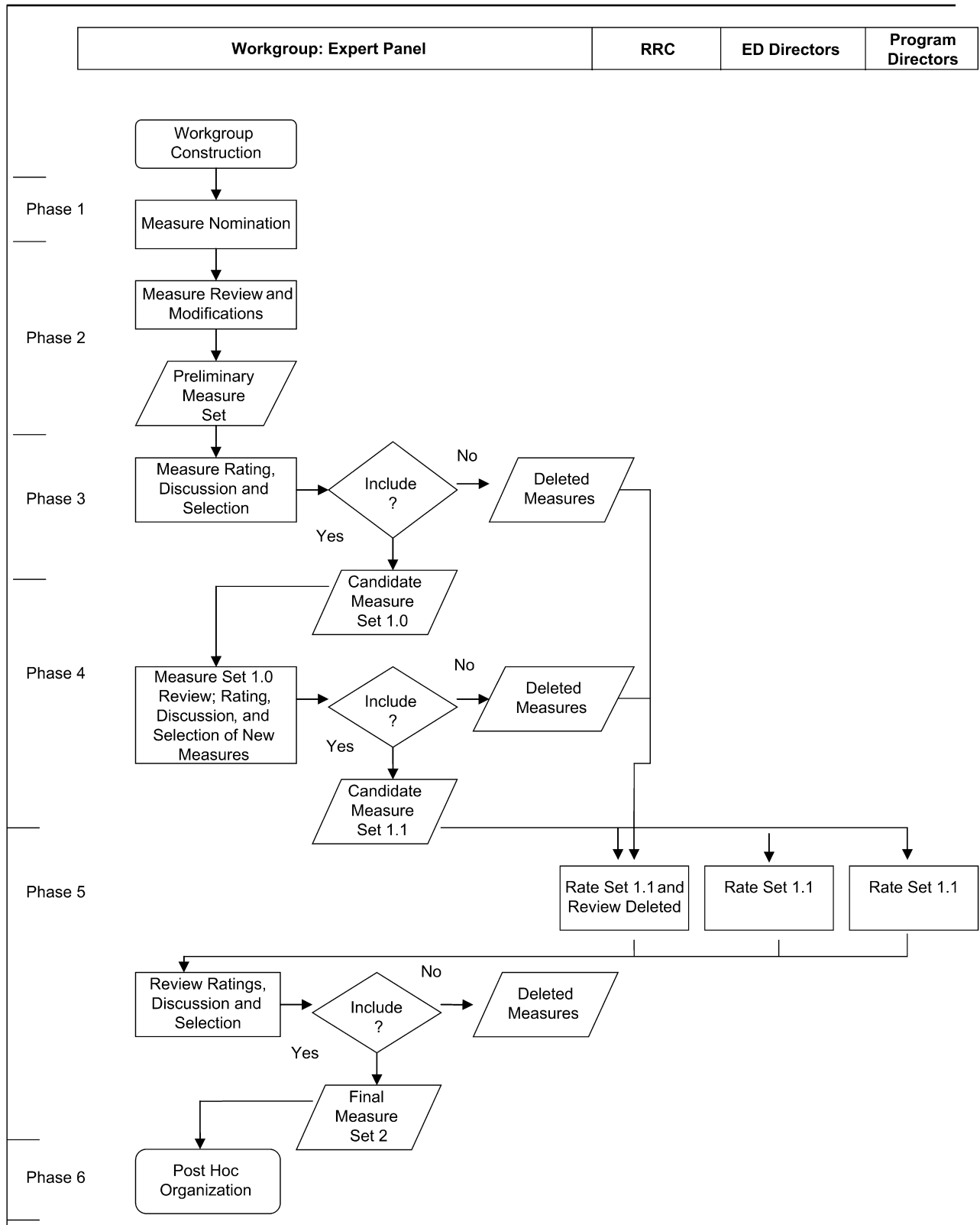


Figure 1. Modified Delphi process flow. RRC = Residency Review Committee Members.

resident and residency program performance. This was followed by a structured exercise during which workgroup members generated and discussed factors that could account for good and poor patient care process

and outcome measures in an ED where resident physicians were learning and providing care. Five readings that discussed quality of care measurement in EM were provided in advance of the meeting.

Modified Delphi Process

The workgroup adapted the modified Delphi methodology used by Lindsay et al.⁷ to this project's unique aim of exploring the link of the patient care quality measures to residency education. A modified Delphi methodology was selected because it allows a group to develop consensus by systematically assessing an expert panel's agreement or disagreement on complex issues. Two or more rounds of voting on issues are conducted, and areas of disagreement are resolved by discussion within the expert group.^{10–12} This study's approach also included features from the RAND appropriateness methodology (RAM), specifically, a relatively small Delphi panel of nine members and the RAM quantitative definition and criterion for establishing agreement.¹³

Phase 1: Nomination of Measures for a Preliminary Set

The first phase of the workgroup's activity was to construct a list of potential measures appropriate for assessing quality of care provided by resident physicians in the ED. Individual workgroup members submitted their recommended measures. These measures were compiled into a preliminary list. During a conference call, the workgroup reviewed this compilation of potential measures and made additional suggestions. No items were removed from consideration at this phase of the activity. The list was then further refined by linking measures to clinical conditions (where appropriate) and by organizing the measures into four categories: clinical conditions, medications, tasks and procedures, and departmental efficiency.

Phase 2: Review of the Preliminary Measure Set against Criteria and Refinement of the Preliminary Measure Set

In phase 2 of measure development, during a second conference call, workgroup members reviewed the preliminary list of measures again to determine whether 1) the measures were representative of the spectrum of ED clinical conditions for patients of various ages and clinical acuity and 2) the clinical conditions identified were common reasons for which emergency care is sought and treated in most EDs. Measures were refined during the course of the group discussion, and gaps were identified. Individuals generated additional measures after the meeting to fill the gaps in accordance with assignments made during the conference call. During the conference call, the workgroup also identified the three critical dimensions of an appropriate measure: 1) importance, 2) measures quality of care, and 3) measures educational effectiveness. The degree to which a measure fit the dimensions was used as the basis for including or excluding individual measures in the next phase of the Delphi process. The group identified a fourth dimension, "feasible to measure," to collect input on the probability that a measure could be implemented.

Phase 3: Workgroup Ratings, Discussion, and Selection of Candidate Measures for Set 1.0

In the third major phase of the measure identification process, workgroup members individually rated each of the conditions, procedures, and specific measures. Each condition, procedure, and departmental efficiency mea-

sure was rated from 1 (not important) to 9 (very important) on the importance dimension. This dimension indicated high prevalence in the ED. Specific measures were rated on "measures quality of care" and "feasible to measure" using a scale of 1 (strongly disagree) to 9 (strongly agree). Response options for "measures educational effectiveness" ranged from 1 (not at all) to 9 (to a great extent). "Measures educational effectiveness" was defined as the extent to which the measure is attributable to effectiveness of teaching and learning and clinical performance within the residency (and not external factors).

The ratings were aggregated and provided to the workgroup members at a face-to-face meeting where the results were reviewed and discussed. Each member also received his or her own ratings. A mean score of 5 on the three dimensions of importance, quality of care, and educational effectiveness was set as a screening criterion for measure retention. The workgroup agreed to discuss measures with borderline mean scores with the understanding that criterion-based decisions could be overridden by a consensus of the group. Feasibility was not considered for purposes of measure selection and retention, because the aim was to identify measures that were substantively appropriate. In addition, the workgroup believed that feasibility would depend on local resources. As a result of this review and the accompanying discussion, 40 measures were dropped and 50 measures were retained, including three new measures defined and voted on during the meeting.

Phase 4: Review of Candidate Measure Set 1.0 for Representativeness; Rating, Discussion, and Selection of New Measures; and Location of Evidence

Two workgroup members conducted a postmeeting review of the measures for representativeness against the Model of the Clinical Practice of EM.¹⁴ The measures identified as a result of the review along with other previously identified but unrated measures were scored and aggregated and later reviewed and discussed utilizing the same approach described previously. Two new clinical conditions and 15 measures were retained. Two members of the workgroup compiled external evidence for the measures as measures of patient care quality. The search for evidence was limited to evidence-based reviews and documented development, use, or endorsement of the measures by major medical or quality improvement organizations.

Phase 5: External Stakeholder Ratings, Workgroup Discussion, and Construction of Final Candidate Measure Set

Because the project goals were novel, external validation of the workgroup's ratings and selections was sought. Thirty-four individuals from three stakeholder groups were invited to participate by rating the candidate set of measures. The 20 individuals who accepted the invitation and completed the ratings were seven RRC members, five ED directors, and eight program directors. The RRC members were those who had not participated in the workgroup; the ED directors were volunteers from a larger group of 14 who were invited because of their participation in a focus group convened by the American College of Emergency Physicians to discuss recent

graduates' performance. The program directors were from a larger group of 13 nominated by workgroup members. The external stakeholder raters composed a convenience sample associated with major stakeholder groups in EM. None of the participants had seen results from earlier phases of the measure identification process.

Each of the external stakeholder group members individually rated the phase 4 candidate measure set 1.1 on the four dimensions. The RRC group was asked to review the measures that had been dropped in the preceding phases and to identify any that should be put back. Three previously dropped measures were recommended for re-inclusion. All raters were also asked to suggest additional measures.

Mean ratings were calculated for each group separately. At its final meeting, the workgroup reviewed and compared the mean ratings from each of the three stakeholder groups and the mean across all three groups with the workgroup's own mean ratings and the previously defined criteria. As a result of the consensus discussion, three measures were dropped. Six measures suggested by the stakeholders were added to a list of new measures for future consideration.

Phase 6: Post Hoc Analysis and Stratification of the Measures

After the workgroup had completed its decision making, the measures were organized post hoc into four groups based on strength of support for the measures overall across the dimensions of importance and measures quality of care and educational effectiveness. The purpose of the post hoc analysis was to better prioritize measures for future development. The post hoc groupings were made based on the classic definition of agreement or disagreement from the RAM.¹³ According to this approach, agreement occurs when approximately 67% of the ratings fall into the same three-point range on a nine-point Likert scale (either 1–3, 4–6, or 7–9) as the median of the ratings. Replicability of results across rating groups is expected when this definition is used.

In this study, a measure was classified as a priority for future development when raters agreed that it is important, measures quality of care, and measures educational effectiveness. Agreement was indicated when at least 67% of the ratings for each of the three dimensions across all raters from the workgroup and external stakeholder groups were in the 7–9 point range on the scale. For the practical purpose of further distinguishing the most strongly supported measures, those measures receiving ratings of 7–9 by at least 75% of raters on all three dimensions were classified into a high agreement group. Measures were included in an "uncertain" group if the agreement criteria was not reached for one or more dimensions and ratings on the other dimensions displayed uncertainty rather than disagreement when the RAND definition was applied. Measures meeting the RAND disagreement definition on one or more dimensions were put into the disagreement group.

Structured Exercise for Exploring the Linkage of Education, Competencies, and Patient Care Quality
After constructing the final version of the preliminary set of measures, the workgroup performed a structured

exercise to explore linkages among education, competencies, and patient care quality. Establishment of causal relationships is a necessary step for demonstrating the validity of the measures for assessing educational outcomes. The exercise consisted of selecting a sample of measures and identifying for each of them: 1) specific competencies (knowledge and skills from the six general competency domains) needed to successfully treat the condition or perform the procedure being assessed using the measure, 2) educational activities likely to occur in residency programs to foster development of the competencies, and 3) factors extraneous to the educational program that might intervene to affect patient care and the associated quality-of-care measures.

RESULTS

A set of 62 measures in four categories was identified through the workgroup and external stakeholder ratings and selection process. They included 43 measures for 21 clinical conditions; eight medication measures, including four specific high-priority drug interactions; seven measures for six tasks or procedures; and four measures of department efficiency.

Twenty-six measures met the stricter quantitative criteria for agreement applied post hoc using the RAM. These measures are presented in the high and moderate columns in Table 1 and the Data Supplement under "Agree" (available as an online Data Supplement at <http://www.aemj.org/cgi/content/full/j.aem.2006.12.011/DC1>). For these measures, a minimum of 67% of raters provided ratings of 7–9 on the scale for each of the dimensions. For the 19 measures in the high agree column, a minimum of 75% of ratings were in the 7–9 point range on the scale. Because the raters agreed that the measures rate highly on the dimensions, these measures can be considered the most appropriate for further development. From 26% to 93% of the workgroup and external stakeholder group members rated the measures between 7 and 9 on the "feasible to measure" dimension. Fourteen of these met the RAM criteria for agreement. These results are presented in Table 1 and the online Data Supplement. Among the measures rated most difficult to measure were the following: for deep vein thrombosis or pulmonary embolism, measuring whether pretest probability was assessed; for headache, percent of subarachnoid hemorrhage diagnosis missed (first 72 hours); and for C-spine, conformance with Canadian C-spine or National Emergency X-Radiography Utilization Study (NEXUS) rules.

All but four of the remaining measures were classified in the uncertain category. These 32 measures received less than 67% of ratings in the 7–9 point range for at least one of the three dimensions. For nine of these, the ratings were below the agreement criteria only for the educational effectiveness dimension. There was disagreement across raters on all four departmental efficiency and effectiveness measures. Six additional measures suggested by members of the external stakeholder groups but not rated during the course of the project were retained for future consideration. These are presented in Table 2.

Documentation supporting use of 15 measures associated with six clinical conditions and one procedure was

Table 1
Summary of Measure Ratings across Critical Dimensions

	Agree	
	High	Moderate
Clinical condition		
Acute myocardial infarction		Percent administered aspirin within 24 hours*
Pneumonia	Appropriate initial antibiotic Percent high risk admitted (Pneumonia Severity Index class 4 or 5)	
Asthma	Percent administered anti-inflammatory drugs (corticosteroids)*	
Abdominal pain	Percent administered relievers* Unscheduled return with ruptured ectopic pregnancy within 72 hours	
Headache		Percent subarachnoid hemorrhage diagnosis missed (first 72 hours)
Syncope/dizzy/shortness of breath		Electrocardiography for patients older than 50 years
Deep vein thrombosis/pulmonary embolism	Percent of patients with deep vein thrombosis/pulmonary embolism receiving anticoagulation in the ED* Pretest probability assessed	
C-spine		Conformance with Canadian C-spine or NEXUS rules
Meningitis	Time to antibiotics in documented meningitis*	
Pregnancy	Rh screening done on threatened abortion and trauma with pregnancy*	
Seizures		Percent head computed tomographic scan for seizure patients (first-time seizure) excluding febrile seizure
Toxicology: unknown ingestion	Acetaminophen level* Pregnancy test if patient is a female of childbearing age*	ASA level*
Pediatrics: fever in an infant younger than 1 month old	Documentation of suicidality Lumbar puncture with cerebrospinal fluid culture and Gram stain* Urinalysis and urine culture* Blood culture* Antibiotics administered in the ED*	
Medication	Medication orders that are contraindicated due to patient allergy	
Procedures		
Intubation	Successful endotracheal intubation*	
Central lines		Complication of central lines
Sedation	Pre-sedation airway assessment in conscious sedation	

A complete version of this table is provided as an online Data Supplement at <http://www.aemj.org/cgi/content/full/j.aem.2006.12.011/DC1>. Agree means that at least 67% of raters provided ratings on the measure in the 7–9 point range on the nine-point Likert scale for the importance/relevance, quality of care, and educational effectiveness dimensions. Classification as “high agree” required 75% of ratings in the 7–9 range on the Likert scale for the three dimensions.
NEXUS = National Emergency X-Radiography Utilization Study; ASA = acetylsalicylic acid.
* At least 67% of the ratings for “feasible to measure” were between 7 and 9 on the scale.

located (see Table 3). All measures are derived from expert consensus or scientific studies.^{15–21} Six measures for two conditions currently are among the performance measures used in national hospital reporting and quality improvement initiatives, and five more are candidate measures.^{15,16} Three measures are included in guidelines

developed by the EM community.^{19–21} Seven measures in the groups designated as appropriate for further development are supported by this evidence as quality-of-care measures.

A sample of results from the structured exercise designed to explore the linkage of educational activities

and competencies to patient care process measures is presented in Table 4. For each of the measures, competencies from four to six of the general competency domains were identified as the knowledge and skills needed to provide quality patient care. Also, for each measure, numerous factors were identified that could intervene to influence patient care and associated quality-of-care measures. Typically, these factors were related to the system. They included resource (equipment, drug, and staff) availability, protocols and policies, patient mix, ED crowding, and hospital volume. The results illustrate that quality patient care (measured by the indicators identified in this study) could be a result of educational activities and residents' acquisition and performance of essential competencies but that intervening variables will need to be ruled out as causal factors.

DISCUSSION

A reliable level of agreement among raters was attained for 26 measures that received high ratings on the importance, quality of care, and educational effectiveness dimensions. These results support the conclusion that there are patient care quality measures that are appropriate for assessing the educational effectiveness of GME in EM. As measures of educational effectiveness and patient care quality, they would indicate whether patient care provided by ED residents and faculty involved appropriate diagnostic testing and treatment processes, correct diagnoses, and successfully performed procedures. Ratings for feasibility of measurement for these 26 indicators suggest that many programs should be able to collect these performance data.

Defining quality indicators using the best available evidence is a goal of this and any performance measurement initiative. Some of the indicators identified in this project were derived previously by others following systematic study of the evidence. Even so, not everyone agrees with these measures. Evolution and refinement of these measures are expected as further research is conducted.

It is appropriate in consensus studies to set selection criteria at whatever level best suits the purpose of the study.¹¹ This study was an initial inquiry into the appropriateness of using patient care quality measures to assess the effectiveness of GME. Relaxed criteria were used initially to enable a broad set of measures to be identified. The application of the stricter criteria post hoc enabled identification of the most strongly supported measures for future development.

The 19 highest rated measures (i.e., those in the high agree category) could be used as the focus of next development steps involving collection and use of these measures in residency programs. Later, the seven other measures in the agree category could be added to make the set of measures more representative of care in the ED.

The measures in this set already being collected for national performance measurement initiatives (i.e., those related to pneumonia and asthma) will require limited, if any, additional development before collection in the ED. Further research and development are needed before use of the other measures. This might include 1) identifi-

Table 2
Additional Measures Recommended for Inclusion

Condition	Measure
Asthma	Percent discharged with inhaled corticosteroids
Extremity injuries	Documentation of distal N/V examination
Productivity	Patients per hour, RVU per patient, RVU per hour
Testicular torsion	Documentation of genitourinary examination
Vital signs abnormalities	Documentation of reassessment or rationale for patient release
Wound repair	Documentation of tetanus status

RVU = relative value unit; N/V = neuro-vascular.

cation of clinical cases that should be excluded from the measures, 2) study of the reliability and validity of the measures, 3) development of data collection instruments, and 4) study of the evidence base. For all measures, it will be important to further investigate effects of contextual variables that are not elements of the educational program and to develop measurement approaches that adjust or control for these intervening variables.

The initial recommended use for the measures, following essential development activities, is for quality measurement and improvement at the residency program level. Program-level patient care process data indicating, for example, that low percentages of patients with asthma were administered relievers, or low percentages of patients with deep vein thrombosis or pulmonary embolism received anticoagulation therapy, or patients suspected of ingesting toxic substances were not tested for acetaminophen, could indicate deficits in local knowledge about current guidelines or standards of care, inadequacies in the transmission of this knowledge to residents, or inadequate supervision. The performance data would be useful in alerting both residents and ED faculty of the gaps and in signaling that changes in both clinical performance and educational processes are needed.

When collected before and after an educational intervention designed to improve care, the measures would provide evidence simultaneously of whether patient care improved and whether the education intervention was effective. Studies in practice settings have shown that providing feedback on patient care performance to providers can contribute to improved care of patients with acute myocardial infarction and pneumonia²² and that quality-of-care measures (for asthma) are sensitive to pre-post change following interventions that include education of health care providers.²³⁻²⁵

Eventually, though, it will be desirable to use the measures to assess the educational effectiveness of GME programs by considering how well residents collectively perform on these measures. This use is consistent with

Table 3
Evidence and Support for Patient Care Quality Measures

Clinical Conditions	Support for Use	
	Used Nationally for Hospital Performance Measurement	Published Literature Review
Acute myocardial infarction		
Percent administered aspirin within 24 hours	CMS, HQA, JCAHO, APU	
Percent administered beta-blockers within 24 hours	CMS, HQA, JCAHO, APU	
Percent administered thrombolytics within half an hour	CMS, HQA, JCAHO	
Percent undergoing percutaneous coronary intervention within 90 minutes	CMS, JCAHO, HQA (120 minutes)	
Pneumonia		Mandell et al. ¹⁷
Appropriate initial antibiotic	CMS, JCAHO, HQA	
Time to antibiotic (percent less than four hours)	CMS, JCAHO, HQA, APU	
Percent high risk admitted (Pneumonia Severity Index class 4 or 5)		
Asthma		Williams et al. ¹⁸
Percent administered anti-inflammatory drugs (corticosteroids)	JCAHO candidate measure	
Percent administered relievers	JCAHO candidate measure	
Percent measured lung function (peak flow, forced expiratory volume in 1 second)		
Percent return within seven days following ED or observational visit (children)	JCAHO candidate measure	
Head injury		Jagoda et al. ¹⁹
CT scan of the head conforming with NEXUS II head CT or Canadian rules		
Pregnancy		Clinical policy ²⁰
Rh screening performed on threatened abortion and trauma with pregnancy		
Seizures		Practice parameter ²¹
Percent undergoing CT scan of the head for seizure patients (first-time seizure) excluding febrile seizure		
Tasks/procedures		
Complication of central lines	JCAHO candidate measure	

CMS = Centers for Medicare & Medicaid Services; HQA = Hospital Quality Alliance; JCAHO = Joint Commission on Accreditation of Healthcare Organizations; APU = Reporting Hospital Quality Data for Annual Payment Update; CT = computed tomography; NEXUS II = National Emergency X-Radiography Utilization Study II.

the aim of GME, to prepare new physicians to provide high-quality patient care, and with the goal of phase 3 of the ACGME's Outcome Project. Measurement strategies that control for patient mix and other system variables are required before high-stakes use of the data or across-program comparisons, however. Furthermore, programs will require assistance putting into place data collection mechanisms.

Using the measures to assess individual resident performance is desirable but presents additional measurement challenges and considerations. In addition to benefits already mentioned, the use of patient care quality measures potentially would result in more precise measures of residents' ability to provide quality care than those currently obtained based on global ratings

or focused observations of resident-patient encounters that lack agreed upon performance standards. Second, as illustrated by the results of the structured exercise, the patient care quality measures could serve as indicators that essential competencies have been acquired, integrated, and applied. Last, assessment using the measures will better prepare residents for practice settings where similar measures are or will be used.

Appropriate use of the measures will require thoughtful interpretation of the results because of the mediating variables the workgroup identified. For example, to conclude that high performance on acute myocardial infarction is due to educational effectiveness, the program will need to rule out high levels of external contribution by specialized units. A conclusion of educational

Table 4
Sample Educational Processes, Competencies, and Intervening Variables that Contribute to Patient Care Quality

Condition: Measure	Educational Processes	Resident Competencies	Intervening Variables
Acute myocardial infarction: percent missing diagnosis of acute myocardial infarction (first 72 hours)	Didactic sessions	Information gathering from patient (PC and ICS)	Inadequate number of monitored or observational beds
	Bedside teaching Analysis of practice patterns	Knowledge of guidelines, indications, and contraindications (MK)	Triage (or mistriage) ED protocol
	Patient follow-up	Ability to accurately interpret electrocardiogram (PC and MK)	Practice patterns (regional variations in care regarding whether an electrocardiogram is obtained)
	Independent reading	Decision-making and judgmental bias toward diagnosis of myocardial infarction (PC) Ability to recognize atypical presentation (PC and MK) Knowledge of testing limits (MK) Coordination of care in ED and with consultants (ICS and SBP)	Resource availability (chest pain unit, stress testing, and imaging)
Otitis media: correct antibiotic prescribed	Didactic sessions	Knowledge of local flora (MK)	Equipment availability (otoscope and insufflator)
	Bedside teaching Analysis of practice patterns	Diagnostic skill (PC and MK) Skill with insufflator (PC and MK)	Formulary (drug availability) Patient mix (socioeconomic status, cultural norms, and relationship to patient preferences)
	Patient follow-up	Knowledge of guidelines (MK)	Patients' likelihood and ability to comply
	Independent reading	Consideration of costs vis-à-vis patient (PC and SBP) Therapeutic relationship (ICS) Counseling/education (PC and ICS) Analysis of practice patterns (PBLI)	
Intubation: successful endotracheal intubation	Didactic sessions	Knowledge of drugs used for rapid sequence induction (MK)	ED protocol (preprinted drug list)
	Bedside teaching		Patient mix (e.g., patients with head and neck cancer, trauma)
	Analysis of practice patterns Patient follow-up	Knowledge of difficult airway algorithms (MK) Recognition of indications and contraindications (PC and MK)	Resource and staff availability Equipment availability and location
	Independent reading	Prior experience resulting in procedural skill (PC and MK)	Hospital volume (opportunity to practice)
	Simulation (models, animal labs, cadavers) Analysis of practice and improvement projects	Team coordination (ICS and SBP) Knowledge and skills related to practice improvement (PBLI)	ED crowding
Departmental efficiency/effectiveness: patient length of stay in the ED			

Table 4
(Continued)

Condition: Measure	Educational Processes	Resident Competencies	Intervening Variables
	Participation on hospital committees	Ability to work with others to improve care (ICS and SBP)	Resource availability (ED and hospital staffing levels, trauma or other specialized centers, diagnostic test availability, on-call consultant availability, clinic and subspecialists' appointments, hospital equipment)
	Case reviews of outliers (i.e., patients with especially long stays)	Willingness to take on care improvement activities (P)	Patient mix (elders) Diversion policy Hospital flow Hospital financing Community resources availability (home visit nurses, social services, hospice, emergency housing, and beds in shelters)
<p>Bedside teaching: case presentation and resident/attending physician discussion; attending physician confirms residents' observations through patient interview/examination and provides feedback to the resident. Analysis of practice patterns: examination of a sample of cases related to the same symptoms or condition for care processes and outcomes.</p> <p>Accreditation Council for Graduate Medical Education general competencies: PC = patient care; ICS = interpersonal and communication skills; MK = medical knowledge; SBP = systems-based practice; PBLI = practice-based learning and improvement; P = professionalism.</p>			

effectiveness based on high success rates of resident-performed intubations may need to be qualified if residents treat only selected and uncomplicated patients. Obtaining a large enough sample of performance for each resident for each measure and separating team and system effects from individual performance through exclusions or adjustments are the major measurement challenges.²⁶

Relying exclusively on patient care quality measures to assess resident competence is not desirable, because not all competencies are assessed using these measures. One example is the extent to which care is patient centered, compassionate, and respectful. Furthermore, observing and assessing individual competencies during the initial learning stages is a more direct way of ascertaining the extent of attainment of individual competencies that comprise good patient care and of identifying additional improvements needed in fundamental skills and knowledge.

Timeliness and efficiency are among the dimensions of quality of care identified by the Institute of Medicine.¹ In this study, measures related to timeliness of care for individual patients received high ratings on quality of care, but measures of departmental efficiency received low ratings. However, all of these were among the measures rated lowest as indicators of educational effectiveness. Factors external to the ED will significantly affect these measures. Nonetheless, it is important to retain these measures for consideration. Failure to utilize these measures could perpetuate "normalized deviance,"²⁷ whereby residents learn to accept overcrowding, inefficiencies, and care delivered too late to be of optimal benefit to the patient. Instead, demonstrated improvements in these mea-

asures could be the basis for special commendation for excellence in systems-based practice, assuming significant resident involvement in or leadership of multidisciplinary improvement initiatives that produce increases in timeliness and efficiency.

LIMITATIONS

The modified Delphi approach used in this study departed from recommendations in two ways. First, the mean rather than the median rating was used as the initial screening criteria. A post hoc examination revealed that use of the mean or median produced comparable decisions. Second, the workgroup did not complete a second round of voting. The external stakeholder group ratings functionally served instead as the second-round vote. A comparison of final results showed no difference in selection decisions based on combined workgroup and external stakeholder ratings versus external stakeholder ratings alone. Last, given the large number of potential patient care quality measures, different measures could be identified in the initial measure nomination phase of future studies. This would not invalidate the current findings, but rather expand the potential measure set.

CONCLUSIONS

Patient care quality measures, when carefully developed and collected, provide direct measures of the desired outcomes of education: provision of high-quality care. Thus, they have the potential to increase the validity of inferences made about the educational effectiveness of

GME. Further activities to develop and test these measures should be undertaken. The measures identified in this article can be used as a starter set for further development, implementation, and study. Implementation of the measures, especially for high-stakes use, will require resolution of significant measurement issues.

References

1. Accreditation Council for Graduate Medical Education (ACGME). Common Program Requirements. Available at: http://www.acgme.org/acWebsite//dutyHours/dh_dutyHoursCommonPR.pdf. Accessed Nov 23, 2006.
2. Institute of Medicine. Crossing the Quality Chasm: A New Health System for the 21st Century. Washington, DC: National Academy Press, 2001.
3. The ACGME Outcome Project. The General Competencies. Available at: <http://www.acgme.org/outcome/comp/compFull.asp>. Accessed Nov 23, 2006.
4. National Quality Forum. Available at: <http://www.qualityforum.org/mission/home.htm>. Accessed Jan 23, 2006.
5. Centers for Medicaid and Medicare Services. Hospital Quality Initiatives. Available at: http://new.cms.hhs.gov/HospitalQualityInits/01_Overview.asp#TopOfPage. Accessed Jan 23, 2006.
6. ACGME Outcome Project Timeline. Available at: http://www.acgme.org/outcome/project/timeline/TIMELINE_index_frame.htm. Accessed Jan 23, 2006.
7. Lindsay P, Schull M, Bronskill S, Anderson G. The development of indicators to measure the quality of clinical care in emergency departments following a modified-Delphi approach. *Acad Emerg Med*. 2002; 9:1131-9.
8. Beattie E, Mackway-Jones K. A Delphi study to identify performance indicators for emergency medicine. *Emerg Med J*. 2004; 21:47-50.
9. Graff L, Stevens C, Spaite D, Foody J. Measuring and improving quality in emergency medicine. *Acad Emerg Med*. 2002; 9:1091-107.
10. Coben JH. Measuring the quality of hospital-based domestic violence programs. *Acad Emerg Med*. 2002; 9:1176-83.
11. Fink A, Kosecoff J, Chassin M, Brook RH. Consensus methods: characteristics and guidelines for use. *Am J Public Health*. 1984; 74:979-83.
12. Jones J, Hunter D. Consensus methods for medical and health services research. *BMJ*. 1995; 311:376-80.
13. Fitch, K, Bernstein SJ, Aguilar MD, et al. The RAND/UCLA Appropriateness Method User's Manual. Available at: <http://www.rand.org/health/tools/appropriateness.html>. Accessed Apr 22, 2004.
14. Hockberger RS, Binder LS, Graber MA, et al. The model of the clinical practice of emergency medicine. *Ann Emerg Med*. 2001; 37:745-70.
15. Joint Commission on Accreditation of Healthcare Organizations (JCAHO). Performance Measurement: Core Measures. Available at: <http://www.jcaho.org/pms/core+measures/candidate+core+measure+sets.htm>. Accessed Jan 24, 2006.
16. Centers for Medicaid and Medicare Services. The Hospital Quality Alliance (HQA) Ten Measure "Starter Set." Available at: http://new.cms.hhs.gov/HospitalQualityInits/01_Overview.asp#TopOfPage. Accessed Jan 23, 2006.
17. Mandell LA, Bartlett JG, Dowell SF, File TM Jr, Musher DM, Whitney C. Update of practice guidelines for the management of community-acquired pneumonia in immunocompetent adults. *Clin Infect Dis*. 2003; 37:1405-33.
18. Williams SG, Schmidt DK, Redd SC, Storms W. Key clinical activities for quality asthma care. Recommendations of the National Asthma Education and Prevention Program. *MMWR Recomm Rep*. 2003; 52:1-8.
19. Jagoda AS, Cantrill SV, Wears RL, et al. Clinical policy: neuroimaging and decisionmaking in adult mild traumatic brain injury in the acute setting. *Ann Emerg Med*. 2002; 40:231-49.
20. American College of Emergency Physicians. Clinical policy: critical issues in the initial evaluation and management of patients presenting to the emergency department in early pregnancy. *Ann Emerg Med*. 2003; 41:123-33.
21. American College of Emergency Physicians, American Academy of Neurology, American Association of Neurological Surgeons, American Society of Neuroradiology. Practice parameter: neuroimaging in the emergency patient presenting with seizure (summary statement). *Ann Emerg Med*. 1996; 28:114-8.
22. Schade CP, Cochran BF, Stephens MK. Using statewide audit and feedback to improve hospital care in West Virginia. *Jt Comm J Qual Saf*. 2004; 30:143-51.
23. Boychuk RB, Demesa CJ, Kiyabu KM, et al. Change in approach and delivery of medical care in children with asthma: results from a multicenter emergency department educational asthma management program. *Pediatrics*. 2006; 117:S145-51.
24. Cabana MD, Slish KK, Evans D, et al. Impact of physician asthma care education on patient outcomes. *Pediatrics*. 2006; 117:2149-57.
25. Schonlau M, Mangione-Smith R, Chan KS, et al. Evaluation of a quality improvement collaborative in asthma care: does it improve processes and outcomes of care? *Ann Fam Med*. 2005; 3:200-8.
26. Kirk SA, Campbell SM, Kendell-Webb S, et al. Assessing the quality of care in multiple conditions in general practice: practical and methodological problems. *Qual Saf Health Care*. 2003; 12:421-7.
27. Croskerry P, Chisholm C, Vinen J, Perina D. Quality and education. *Acad Emerg Med*. 2002; 9:1108-15.