


© Health Research and Educational Trust
DOI: 10.1111/1475-6773.12800
RESEARCH ARTICLE

Implicit Review Instrument to Evaluate Quality of Care Delivered by Physicians to Children in Emergency Departments

James P. Marcin , Patrick S. Romano, Madan Dharmar, James M. Chamberlain, Nanette Dudley, Charles G. Macias, Lise E. Nigrovic, Elizabeth C. Powell, Alexander J. Rogers, Meridith Sonnett, Leah Tzimenatos, Elizabeth R. Alpern, Rebecca Andrews-Dickert, Dominic A. Borgialli, Erika Sidney, Charlie Casper, Jonathan Michael Dean, and Nathan Kuppermann, for the Pediatric Emergency Care Applied Research Network

Objective. To evaluate the consistency, reliability, and validity of an implicit review instrument that measures the quality of care provided to children in the emergency department (ED).

Data Sources/Study Setting. Medical records of randomly selected children from 12 EDs in the Pediatric Emergency Care Applied Research Network (PECARN).

Study Design. Eight pediatric emergency medicine physicians applied the instrument to 620 medical records.

Data Collection/Extraction Methods. We determined internal consistency using Cronbach's alpha and inter-rater reliability using the intraclass correlation coefficient (ICC). We evaluated the validity of the instrument by correlating scores with four condition-specific explicit review instruments.

Principal Findings. Individual reviewers' Cronbach's alpha had a mean of 0.85 with a range of 0.76–0.97; overall Cronbach's alpha was 0.90. The ICC was 0.49 for the summary score with a range from 0.40 to 0.46. Correlations between the quality of care score and the four condition-specific explicit review scores ranged from 0.24 to 0.38.

Conclusions. The quality of care instrument demonstrated good internal consistency, moderate inter-rater reliability, high inter-rater agreement, and evidence supporting validity. The instrument could be useful for systems' assessment and research in evaluating the care delivered to children in the ED.

Key Words. Pediatrics, quality, emergency department

It has been nearly one decade since the Institute of Medicine's (IOM; now the National Academy of Medicine) series of reports on the status and future of emergency care (IOM 2007a,b,c). In the pediatric report, "Emergency Care for Children: Growing Pains," the IOM described the "fragmented" system of emergency care, noting in particular the "uneven" nature of emergency care for children (IOM 2007b). Efforts to improve the quality of care delivered to acutely ill and injured children in emergency departments (EDs) depend in part on the availability of reliable and valid measures of quality of care and determinations of clinical outcomes. Valid instruments that could identify care processes or other factors associated with quality of care are needed to facilitate improvements in care delivery and policies.

Currently, few instruments and measures of quality can be applied to all children treated in EDs. Outcome measures such as mortality, length of stay,

Address correspondence to James P. Marcin, M.D., M.P.H., Department of Pediatrics, University of California, Davis School of Medicine, 2516 Stockton Blvd, Sacramento, CA 95817; e-mail: jpmarcin@ucdavis.edu. Patrick S. Romano, M.D., M.P.H., Madan Dhanmar, M.B.B.S., Ph.D., and Nathan Kuppermann, M.D., M.P.H., are with the Department of Pediatrics, University of California, Davis School of Medicine, Sacramento, CA. Patrick S. Romano, M.D., M.P.H., is also with the Department of Internal Medicine, University of California, Davis School of Medicine, Sacramento, CA. Nathan Kuppermann, M.D., M.P.H., is also with the Department of Emergency Medicine, University of California, Davis School of Medicine, Sacramento, CA. James M. Chamberlain, M.D., is with the Division of Emergency Medicine, Children's National Health System, Washington, DC. Nanette Dudley, M.D., is with the Department of Pediatrics, University of Utah School of Medicine, Salt Lake City, UT. Charles G. Macias, M.D., M.P.H., is with the Department of Pediatrics and Center for Clinical Effectiveness, Baylor College of Medicine, Houston, TX. Lise E. Nigrovic, M.D., M.P.H., is with the Division of Emergency Medicine, Boston Children's Hospital, Boston, MA. Elizabeth C. Powell, M.D., M.P.H., is with the Department of Pediatrics, Northwestern University's Feinberg School of Medicine, Chicago, IL. Alexander J. Rogers, M.D., is with the Departments of Emergency Medicine and Pediatrics, University of Michigan, Ann Arbor, MI. Meridith Sonnett, M.D., is with the Department of Pediatrics, Columbia University Medical Center, Columbia University College of Physicians and Surgeons, New York, NY. Leah Tzimenatos, M.D., and Nathan Kuppermann, M.D., M.P.H., are with the Department of Emergency Medicine, University of California, Davis School of Medicine, Sacramento, CA. Elizabeth R. Alpern, M.D., M.S.C.E., is with the Department of Pediatrics, The Children's Hospital of Philadelphia, University of Pennsylvania School of Medicine, Philadelphia, PA; Present address: Department of Pediatrics, Ann & Robert H. Lurie Children's Hospital, Northwestern University Feinberg School of Medicine, Chicago, IL. Rebecca Andrews-Dickert, M.D., is with the Department of Emergency Medicine, DeVos Children's Hospital, Michigan State University College of Human Medicine, Grand Rapids, MI. Dominic A. Borgjalli, D.O., M.P.H., is with the Department of Emergency Medicine, Hurley Medical Center and University of Michigan, Flint, MI. Erika Sidney, M.D., is with the Division of Emergency Medicine, Children's Hospital of Colorado, University of Colorado, Aurora, CO. Charlie Casper, Ph.D., and Jonathan Michael Dean, M.D., are with the Department of Pediatrics, University of Utah and PECARN Data Coordinating Center, Salt Lake City, UT.

recidivism, appropriateness of admission, health-related quality of life, and patient and/or parent satisfaction may not be reliable if the outcomes are rare or insufficiently sensitive to the processes of care we are trying to improve. Many of these measures are also subject to confounding factors, such as patient age, severity of illness, insurance status, and others. Therefore, more general and valid instruments are needed to assess quality of care that can be applied to all children receiving care in the ED.

Peer review plays an important role in the ascertainment of quality of care both at the individual provider and at the system levels (Dans, Weiner, and Otter 1985; Goldman 1994; Hofer et al. 2004). Implicit review is a type of peer review where assessments of quality are based on expert reviewers' judgment of care (Donabedian 1988) and has been used in both outpatient (Hulka et al. 1979) and inpatient care settings (Kahn et al. 1990; Rubenstein et al. 1990). Structured review of medical records to assess the implicit quality of care has been shown to have high face validity (Goldman 1994) and offers better inter-rater reliability (Goldman 1992, 1994) than unstructured review (Goldman 1992).

In 2007, our group developed and tested a five-item, structured quality of care implicit review instrument for children receiving care in the ED using medical records from four rural hospitals. This peer review instrument in which assessments of quality are based on expert reviewers' judgment encompasses four dimensions of care that had been previously developed and validated (Rubenstein et al. 1990; Rubin 1990). These include the physician's initial data gathering, integration of information and development of appropriate diagnoses, initial treatment plans and physician orders, and plan for disposition and follow-up. In addition, there is one item assessing the overall quality of care (Dharmar et al. 2007). This implicit review instrument was shown to have moderate reliability, internal consistency, and validity. The evaluation, however, was only conducted on 178 pediatric patients from four rural EDs in northern California and was tested by only two pediatric emergency medicine experts. Although the instrument demonstrated significant correlations with medication errors assessed by pediatric pharmacists blinded to clinical assessments, additional evaluations of the instrument's construct or predictive validity are needed (Dharmar et al. 2007).

We performed this study to further evaluate and validate this structured quality of care implicit review instrument in a multicenter cohort of children presenting to EDs with acute illnesses and injuries. We measured the consistency, reliability, and validity of the instrument using a sample of EDs participating in the Pediatric Emergency Care Applied Research Network (PECARN) (Alpern et al. 2006; Tzimenatos, Kim, and Kuppermann 2015) to

determine whether the review instrument has sufficient reliability and validity for use as a quality measurement tool.

METHODS

Study Design

We conducted a retrospective, observational cohort study of children presenting with acute illnesses and injuries to 12 EDs participating in PECARN. At the time of the study, there were four “nodes” participating in PECARN, and each node included between five and six EDs for a total of 22 PECARN EDs. For the purposes of this study, we included three EDs from each of the nodes for equal nodal representation. The three EDs were specifically selected to maximize clinician and patient diversity with regard to hospital size, treating physicians (i.e., general vs. pediatric emergency medicine), and patient populations. We applied the quality of care implicit review instrument (Dharmar et al. 2007) to the medical records of a random sample of children presenting to the 12 EDs selected and evaluated the instrument using standard assessment and validation techniques. This study was approved by the institutional review board at each participating hospital.

Study Setting and Population

Children younger than 18 years who presented to any of the participating EDs for evaluation from January through December 2011 were eligible for inclusion. We randomly sampled patient visits from the ED logs at each of the participating hospitals using a two-stage date/patient sampling scheme generated by the PECARN Data Coordinating Center. First, we used this sampling scheme to select patients presenting with any medical condition and did not use stratification because we wanted the cohort to be representative of the ED population as a whole. Then, for the purpose of validation, a second random sample was generated using the same methods to select patients from the ED logs with one of four medical conditions (asthma, febrile seizure, diarrhea and dehydration, and head trauma) for which there were existing condition-specific, criterion-based explicit quality of care instruments (Gausche-Hill 2007). We excluded medical records of children who were seen in the ED for scheduled procedures (e.g., suture removal), those evaluated for nonmedical complaints (e.g., abandonment or endangerment), those transiently evaluated in the ED in the process of direct admission to the hospital, and those who left

the ED without being seen by the attending physician. In the event that a patient medical record met exclusion criteria, the reviewer skipped to the next randomly sampled patient.

Study Protocol

After removing all patient, hospital, and physician identifiers, the research coordinator at each participating hospital photocopied medical records of sampled patients, including all ED treatment sheets, reports of radiologic examinations and ED procedures, laboratory test results, consultations, and discharge instructions. The research coordinator abstracted all relevant patient data from each medical record after which the de-identified medical record was uploaded to a secure server at the PECARN Data Coordinating Center for review.

The research coordinator abstracted patient age, sex, race (American Indian or Alaska Native, Asian, African American, Native Hawaiian or Other Pacific Islander, White, Multiracial, and unknown), ethnicity (Hispanic or Latino, not Hispanic or Latino, and unknown), arrival by emergency medical services (yes/no), triage category (nonurgent, urgent, emergent, unknown), and disposition of care (discharged home, observation unit—transitional care unit, admitted to floor, admitted to intensive care unit, transferred, and other).

Measures

Structured Quality of Care Implicit Review Instrument. The quality of care provided to each child in the ED was assessed using the previously published implicit review instrument (see Appendix SA2) (Dharmar et al. 2007). This five-item instrument includes four items assessing different dimensions of care and one item assessing the overall quality of care. The four dimension-specific items focus on processes of care and include the initial data gathering about acute problems; the integration of information and development of appropriate diagnoses; the initial treatment plan and orders; and the plan for disposition and follow-up. All five items were assessed on a seven-point ordered adjectival scale: extremely inappropriate; very inappropriate; somewhat inappropriate; intermediate; somewhat appropriate; very appropriate; and extremely appropriate (Dharmar et al. 2007). We also calculated a summary quality of care score (Rubin 1990), which was the sum of the five item-specific

scores from each record, resulting in a score ranging from 5 to 35 for each patient (Dharmar et al. 2007).

To assess the construct validity of the instrument, the reviewer also answered the following question: “What is the likelihood that you would want this physician taking care of your (own) child in the Emergency Department?” (Kahn et al. 1990; Rubenstein et al. 1990). This was similarly measured using a seven-point ordered adjectival scale from extremely unlikely to extremely likely. Reviewers were not aware that this question would be used to evaluate construct validity.

To apply the quality of care instrument, we selected two expert physician reviewers board certified in pediatric emergency medicine from each of the four participating PECARN nodes for a total of eight reviewers. Prior to reviewing the medical records, all of the expert reviewers met for a one-day, in-person training session to review the manual of operations. The group discussed general principles of structured implicit review, how the instrument should be applied, outlined anchors for the adjectival scale, and reviewed several sample medical records both individually and as a group. The anchors that were developed for each item and described in the manual of operations

Table 1: Guidelines for Anchors Used in the Manual of Operations for Individual Quality of Care Adjectival Scale

<i>Score</i>	<i>Category of Care</i>	<i>Description of Care</i>
7	Extremely Appropriate	Extremely appropriate medical care is the best you can imagine in the average U.S. hospital ED. Extremely appropriate care minimizes the risk of complications, maximizes the likelihood of a good outcome, and maximizes humane care and respect for patients/parent’s wishes
6	Very Appropriate	Medical care that is slightly less in quality than “Extremely Appropriate”
5	Somewhat Appropriate	Medical care that is slightly better in quality than “Intermediate”
4	Intermediate	Intermediate care is acceptable but just minimally so
3	Somewhat Inappropriate	Medical care that is slightly less in quality than “Intermediate”
2	Very Inappropriate	Medical care that is slightly better in quality than “Extremely Inappropriate”
1	Extremely Inappropriate	Extremely inappropriate care is malpractice. This care has egregious errors and is likely to result in more harm than benefit or is likely to cause more harm or less benefit than alternative approaches available at the average U.S. hospital ED.

followed the original RAND definitions and guidelines for a seven-point adjectival scale (Table 1) (Rubenstein et al. 1990; Rubenstein 1991). We conducted regular electronic meetings to address issues as they arose, but we did not try to recalibrate or sanction individual raters.

Each de-identified medical record was randomly assigned to four expert reviewers for independent assessments of quality (Fitch 2001; Dharmar et al. 2007). This number was chosen to balance reliability and susceptibility to outliers against efficiency, based on prior literature (Hofer et al. 2004; Hayward et al. 2007). Each expert reviewer was assigned an equal number of medical records and did not review records from his or her own institution. Reviewers were only provided the test results and ED records that would have been available to the ED physician during the patient visit, effectively blinding them from the ultimate outcome of the patient after discharge or admission. Reviewers were also advised to inform the study coordinator of extreme circumstances such as unreadable records or serious patient safety concerns, but no such issues were identified.

Explicit Quality Instrument. To further validate our quality of care implicit review instrument, we selected four medical conditions for which explicit quality measures existed for comparison. We then simultaneously applied our implicit review instrument and the four condition-specific, criterion-based, explicit quality instruments to cohorts of patient visits with the selected medical conditions: asthma, febrile seizures, diarrhea and dehydration, and head trauma. The condition-specific instruments were previously created for assessing the quality of care children received in the ED, using evidence-based guidelines and national expert consensus panels (Gausche-Hill 2007). Three lead investigators reviewed the literature for relevant data published after the creation of the original condition-specific instruments (Kuppermann et al. 2009; Subcommittee on Febrile Seizures and American Academy of Pediatrics 2011; Agency for Healthcare Research and Quality 2014), which resulted in minor updates to the explicit quality instruments (see Appendix SA3). The updated condition-specific instruments were then reviewed by all study emergency medicine investigators-reviewers, and final modifications were made.

The maximum attainable scores for the four condition-specific explicit review instruments were as follows: 100 for asthma; 90 for febrile seizure; 75 for diarrhea and dehydration; and 100 for head trauma. Each instrument had explicit criteria for each condition regarding the quality of triage, history and physical examination, diagnostic evaluation, therapeutic and management plan, and appropriate patient disposition. One pediatric critical care nurse

reviewer with expertise in both pediatric emergency and critical care medicine applied the condition-specific explicit review instruments to all four cohorts of patient medical records and scored them using the guidelines accompanying the instruments. Prior to medical record abstraction, the nurse reviewer was trained during a one-day session by lead investigators and was blinded to the implicit review instrument scores.

Data Analysis

To evaluate the quality of care implicit review instrument, we analyzed internal consistency, inter-rater reliability, and construct validity, using each of the five item-specific scores as well as the summary quality score from each reviewer. While there are limitations to different statistical methods used to measure correlations (Sijtsma 2009), we assessed internal consistency using the Cronbach's alpha coefficient, corrected item-total correlations (Spearman), and the change in Cronbach's alpha after removal of each single item from the tool (Cronbach 1951; Nunnally and Bernstein 1994). We measured inter-rater reliability with the intraclass correlation coefficient (ICC) (Fleiss 1981) using the 1, k method, where $k = 4$ (Shrout and Fleiss 1979; SAS Institute Inc. 2013), representing the reliability of the mean of four reviewers based on the categorization recommended by Landis and Koch (Landis and Koch 1977). As an additional method of measuring inter-rater reliability, we calculated the inter-rater agreement as the proportion of records for which all four reviewers' scores fell within a three-point range for each of the five item-specific scores (Fitch 2001). We also evaluated inter-rater disagreement, defined as the proportion of records for which at least one reviewer ranked quality in the lowest range of scores (1–3) and at least one reviewer ranked quality in the highest range of scores (5–7), for each of the five item-specific scores (Fitch 2001).

To evaluate the construct validity of the instrument, we determined the correlation between the summary quality of care scores and the construct validity question score using the Pearson correlation coefficient. We also estimated correlations between the summary quality of care score and the four explicit, condition-specific scores, also using the Pearson correlation coefficient. Interpretation of the correlation coefficients was based on the categorization recommended by Cohen (1988). Last, because of an anticipated "ceiling effect" with the explicit condition-specific scores, we compared the mean summary quality of care scores among those medical records that were scored the maximum condition-specific scores to those that were not scored the maximum condition-specific scores using Student's t -test. p -values $<.05$

were considered to be significant. All analyses were performed using SAS version 9.4 (SAS Institute, Cary, NC).

RESULTS

Table 2 shows the characteristics of the 620 ED encounters included in the study. Approximately 50 medical records (range: 47–55) were reviewed from each of the 12 participating EDs. The median age in years was 2.8 with an interquartile range of 1.2 to 7.9; 21.6 percent of the patients were infants. Most patients ($N = 528$, 85.2 percent) were discharged home from the ED; 77 (12.4

Table 2: Patient Characteristics, $N = 620$

Age in years, mean (SD)	5.2 (5.2)
Gender, n (%)	
Male	344 (55.5)
Female	276 (44.5)
Race, n (%)	
American Indian or Alaska Native	5 (0.8)
Asian	7 (1.1)
Black or African American	181 (29.0)
Native Hawaiian or other Pacific Islander	4 (0.7)
White	298 (48.1)
Multiracial	5 (0.8)
Unknown	121 (19.5)
Ethnicity, n (%)	
Hispanic or Latino	159 (25.7)
Not Hispanic or Latino	348 (56.1)
Unknown	113 (18.2)
Triage category*, n (%)	
Nonurgent	38 (6.1)
Urgent	437 (70.5)
Emergent	144 (23.2)
Unknown	1 (0.2)
Arrived by emergency medical service, n (%)	
Yes	99 (16.0)
No	521 (84.0)
Discharge disposition, n (%)	
Discharged home	528 (85.2)
Observation unit—transitional care unit	11 (1.8)
Admitted to floor	65 (10.5)
Admitted to pediatric intensive care unit	12 (1.9)
Transferred	2 (0.3)
Other	2 (0.3)

*Emergent = Emergency Severity Index levels 1 and 2; Urgent = levels 3 and 4; Nonurgent = level 5.

percent) patients were hospitalized, 11 (1.8 percent) were transferred to an observation unit, and two (0.3 percent) were transferred to another facility.

The eight expert reviewers each reviewed approximately 300 medical records (range: 292 to 317). Their mean Cronbach's alpha was 0.85 with a range across reviewers from 0.76 to 0.97. The overall Cronbach's alpha for all eight reviewers was 0.90. The overall item-total correlation for the item-specific overall quality of care score was 0.89, and the change in Cronbach's alpha after removal of the item-specific total quality of care score was 0.85. These data, in addition to the item-total correlation and change in Cronbach's alpha for each item in the quality of care instrument for each reviewer, are provided in detail (see Appendix SA4).

Table 3 shows the means, medians, and inter-rater reliability measures for each of the item-specific quality of care scores and the summary quality of care score. The mean item-specific quality of care scores ranged from 6.03 to 6.17, with a mean overall quality of care score of 6.03 (SD=0.52). The mean summary quality of care score was 30.6 (SD=2.7). In 99 of the 620 reviews (16 percent), the reviewers rated all four of the dimension-specific items and the overall quality with the maximum score of 7. The ICC's for each of the five item-specific quality of care scores demonstrated moderate agreement, with a range from 0.40 to 0.46 and an ICC of 0.49 for the summary quality of care score. The proportion of the five item-specific quality of care scores for which all four reviewers reached agreement varied between 95.2 percent and 99.4 percent within a three-point range, between 85.0 percent and 96.8 percent within a two-point range, and between 55.6 percent and 75.3 percent within a one-point range (see Appendix SA5). The proportion of the five item-specific quality of care scores for which at least one reviewer rated quality as low and at least one reviewer rated quality as high ranged from 1.6 percent to 7.1 percent. The frequency with which individual reviewers rated item-specific quality scores with the maximum score of 7 or with very low scores (1-3) is provided in Appendix SA6.

The Pearson correlation coefficients comparing the quality of care scores for each of the four dimensions of care with the construct validity question score ranged between 0.65 and 0.79 ($p < .001$ for all). The correlation coefficient was 0.91 between the overall quality of care score and the validation question score, and 0.91 between the summary quality of care score and the validation question score (Cohen 1988).

The correlation between the summary quality of care scores and the explicit, condition-specific scores ranged from 0.24 to 0.38 (Table 4). Because of the "ceiling effect" noted with the condition-specific scores, we compared

Table 3: Implicit Review Quality of Care Score Metrics and Inter-Rater Reliability

<i>Items in Implicit Instrument</i>	<i>Implicit Quality Score</i>		<i>Intraclass Correlation Coefficient (95% CI)</i>	<i>Percent Agreement*</i>	<i>Percent Disagreement†</i>
	<i>Mean (SD)</i>	<i>Median (Q1, Q3)</i>			
Initial data gathering by physician about acute problems	6.17 (0.43)	6.25 (6.00, 6.50)	0.43 (0.35, 0.50)	616/620 = 99.4%	10/620 = 1.6%
Physician's integration of information and development of appropriate diagnoses	6.20 (0.48)	6.25 (6.00, 6.50)	0.42 (0.34, 0.49)	609/620 = 98.2%	16/620 = 2.6%
Physician's initial treatment plan and initial orders	6.14 (0.57)	6.25 (5.75, 6.50)	0.46 (0.39, 0.53)	590/620 = 95.2%	44/620 = 7.1%
Physician's plan for disposition and follow-up	6.06 (0.56)	6.25 (5.75, 6.50)	0.40 (0.32, 0.48)	594/620 = 95.8%	42/620 = 6.8%
Assess the overall quality of care provided to the patient	6.03 (0.52)	6.00 (5.75, 6.50)	0.43 (0.35, 0.50)	597/620 = 96.3%	37/620 = 6.0%
Summary Quality of Care Score	30.6 (2.17)	30.8 (29.3, 32.3)	0.49 (0.42, 0.55)		

*Percent agreement calculated as the proportion of medical records where all four reviewers' scores fell within a three-point range for each of the five item-specific scores.

†Percent disagreement calculated as the proportion of medical records for which at least one reviewer rated quality of care as low (scores 1 to 3) and at least one reviewer rated quality of care as high (scores 5–7).

Table 4: Measures of Validity: Correlation of Implicit Review Quality of Care Scores with Explicit, Evidence-Based Condition-Specific Quality of Care Scores

<i>Disease groups (N = 271)</i>	<i>Total Summed Implicit Quality Score*</i>		<i>Condition-Specific Explicit Quality Score†</i>		<i>Pearson Correlation Coefficient (p)</i>
	<i>Mean (SD)</i>	<i>Median (Q1, Q3)</i>	<i>Mean (SD)</i>	<i>Median (Q1, Q3)</i>	
Asthma, N = 69 (11.1%)	31.1 (1.66)	31.3 (30.0, 32.5)	89.3 (9.14)	92.0 (84.0, 92.0)	0.32 (0.008)
Closed head injury, N = 68 (11.0%)	31.1 (2.15)	31.6 (29.6, 32.9)	95.3 (6.28)	98.0 (93, 100)	0.38 (0.002)
Diarrhea and dehydration, N = 68 (11.0%)	30.4 (1.82)	30.5 (29.5, 31.5)	61.5 (8.40)	61.0 (57.0, 68.0)	0.29 (0.015)
Febrile seizures, N = 66 (10.7%)	30.1 (2.16)	30.0 (29.3, 31.3)	87.7 (6.56)	90.0 (90.0, 90.0)	0.24 (0.055)

*Max score = 35.

†Max score: Asthma and closed head injury = 100; Gastroenteritis and dehydration = 75; Febrile seizures = 90.

the mean total summary quality of care scores between medical records that were scored with the maximum condition-specific scores and medical records that were not scored with the maximum condition-specific scores. We found a significantly higher mean total summary quality of care score, 31.5 ($N = 46$), for records that received the maximum condition-specific score than for records that did not receive the maximum condition-specific score, 30.5 ($N = 225$), $p = .001$.

DISCUSSION

Our quality of care implicit review instrument had excellent internal consistency, moderate inter-rater reliability, and high inter-rater agreement when applied to a diverse cohort of acutely ill and injured children receiving care in a large sample of EDs. The instrument also had acceptable construct validity, with statistically significant correlations between the summary quality of care scores and the explicit, condition-specific scores. Our findings demonstrate that our implicit review instrument performs well when measuring the overall quality of care delivered to children in the ED. This instrument could be used in a variety of settings to evaluate

quality of care and changes in quality following changes in care delivery (Ashton et al. 1999; Mistry et al. 2014).

Peer review has played an important role in the ascertainment of quality of care, both at the individual provider level and at the health system level (Dans, Weiner, and Otter 1985; Rubin 1990; Goldman 1992, 1994; Smith et al. 1997). The use of implicit review, or peer review in which assessments of quality are based on expert reviewers' judgment, when applied in a structured or systematic manner, has high face validity and excellent inter-rater reliability (Dans, Weiner, and Otter 1985; Goldman 1992, 1994). Furthermore, structured implicit review correlates well with criterion-based measures of processes of care and health care outcomes (Ashton et al. 1999). While the Centers for Medicare & Medicaid Services, the Agency for Healthcare Research and Quality, and other organizations involved in pediatric health care quality measurement traditionally prefer tools or instruments that can be applied using explicit methods or evidence-based guidelines (Mangione-Smith, Schiff, and Dougherty 2011; Mistry et al. 2014), the key advantage of implicit review is that it can be applied to all children presenting to EDs, regardless of underlying age and diagnoses, and can be applied in circumstances where evidence-based guidelines or explicit measures of quality do not exist. While there are several other methods proposed to evaluate quality of the care delivery processes (Olsen, Grossmann, and McGinnis 2011; Alper, Sanders, and Saunders 2013), structured implicit review can also provide a composite measurement of quality of care for either an individual patient or cohort of patients, which cannot be achieved using administrative databases or individual measures focused on specific conditions.

We previously developed and published the structured implicit review instrument evaluated here as a new means of measuring the quality of care provided to acutely ill and injured children in EDs (Dharmar et al. 2007). This instrument was developed according to implicit review guidelines (Rubin 1990) by modifying similar implicit review instruments previously published and validated by the RAND Corporation (Rubenstein et al. 1990; Rubenstein 1991). In the original study, the instrument was applied by two pediatric emergency medicine physicians to a cohort of 178 acutely ill and injured children seen at four rural EDs. The results of the original study were consistent with the results we obtained from the current study. Regarding internal consistency, for example, the overall Cronbach's alpha was very similar: 0.90 in this study and 0.91 in our previous study. With respect to inter-rater reliability, the ICC for the mean total summary score was lower at 0.49 in the current study compared to 0.65 in our previous study. This difference could reflect several

factors. In our original evaluation, only two pediatric emergency medicine physicians from the same institution applied the instrument to the records of children seen in four rural EDs in the highest triage category, while in this current study, eight reviewers from eight different institutions reviewed the records of more heterogeneous patients from 12 institutions triaged at all acuity levels. Despite this difference in ICC between the studies, we believe that the implicit review instrument has sufficient validity and reliability to justify its use for quality assessment.

LIMITATIONS

There are several limitations to our study. First, while we applied the implicit review instrument to the medical records of patients of varying ages and severities of illness, all of the records came from EDs that participate in PECARN. These EDs are typically large, have expertise in pediatric emergency medicine, are affiliated with academic institutions, and are located in urban settings. In addition, our scores were skewed toward higher quality than would be expected if patient care was provided by nonexperts in pediatric emergency medicine. Our results may not be similar if the instrument were applied in smaller, rural, and/or nonacademic EDs. Second, because there are relatively few criterion-based, explicit measures of quality of care in the pediatric ED setting, our ability to validate the instrument against reference standards is limited, as is our ability to determine or propose clinically meaningful differences in quality. While we found significant score correlations between the implicit review instrument and four explicit process-of-care review instruments for specific medical conditions, this association was limited by ceiling effects, could only be evaluated for four conditions, and warrants further evaluation. Third, because the implicit review instrument focuses on physician quality of care, our instrument does not capture other domains of health care quality such as timeliness, patient centeredness, or equity. Fourth, quality of care review based on retrospective review of medical records can be time intensive and costly and is limited by the completeness and accuracy of the source documents. Finally, this and other implicit review instruments are meant to measure quality of care delivered to large patient cohorts and are not well suited or intended for identifying instances of low or high quality of care among individual patients and/or physicians; rather this method of evaluating quality of care is generally limited to research or review conducted by internal or external oversight entities.

CONCLUSIONS

We evaluated and validated a structured, quality of care implicit review instrument in a diverse cohort of acutely ill and injured children receiving care in a large sample of EDs. The instrument had excellent reliability and acceptable validity when compared to samples of records from cohorts of patients for whom there are existing explicit review instruments based on evidence-based practice guidelines. Future studies should evaluate the performance of this instrument on more diverse pediatric populations receiving care in a variety of EDs. Although a tool of this type is less actionable for quality improvement than specific process-of-care measures, this type of tool could be more reliable and sensitive to changes in true quality of care than risk-adjusted outcome measures such as mortality and adverse events (e.g., the Agency for Healthcare Research and Quality's Pediatric Quality Indicators) (Bardach, Chien, and Dudley 2010; Bardach et al. 2013). Accordingly, the implicit review instrument could be used in future investigations to identify factors associated with high-quality care delivered to children in the ED, such that disparities in care can be identified and corrected. It could also be used to assess the overall impact of changes in care delivery and systems of care (Dharmar et al. 2008).

ACKNOWLEDGMENTS

Joint Acknowledgment/Disclosure Statement: We acknowledge the efforts of the following individuals participating in PECARN at the time this study was initiated:

PECARN Steering Committee: N. Kuppermann, Chair; E. Alpern, D. Borgialli, J. Callahan, J. Chamberlain, P. Dayan, J. M. Dean, M. Gerardi, M. Gorelick, J. Hoyle, E. Jacobs, D. Jaffe, R. Lichenstein, K. Lillis, P. Mahajan, R. Maio, F. Moler, D. Monroe, R. Ruddy, R. Stanley, M. Tunik, A. Walker.

MCHB/EMSC liaisons: D. Kavanaugh, H. Park.

Data Coordinating Center (DCC): M. Dean, R. Holubkov, S. Knight, A. Donaldson, S. Zupan, M. Miskin, J. Wade, A. Jones, M. Fjelstad.

Feasibility and Budget Subcommittee (FABS): T. Singh, Chair; A. Dronowski, L. Fukushima, E. Kim, D. Monroe, G. O'Gara, H. Rincon, M. Tunik, S. Zupan.

Grants and Publications Subcommittee (GAPS): M. Gorelick, Chair; E. Alpern, D. Borgialli, K. Brown, L. Babcock, A. Donaldson, G. Foltin, F. Moler, S. Teach.

Protocol Concept Review and Development Subcommittee (PCRADs): D. Jaffe, Chair; J. Chamberlain, A. Cooper, P. Dayan, J. M. Dean, R. Holubkov, P. Mahajan, R. Maio, N. C. Mann, K. Shaw, A. Walker.

Quality Assurance Subcommittee (QAS): R. Stanley, Chair; P. Ehrlich, R. Enriquez, M. Gerardi, R. Holubkov, E. Jacobs, R. Lichenstein, K. Lillis, B. Millar, R. Ruddy, M. Shults.

Safety and Regulatory Affairs Subcommittee (SRAS): W. Schalick, J. Callahan, Co-Chairs; S. Atabaki, J. Burr, K. Call, J. Hoyle, R. Ruddy, J. Suha-jda, N. Schamban.

Funding Sources: 1R01HS019712 Agency for Healthcare Quality and Research.

This project was also supported in part by the Health Resources and Services Administration (HRSA), Maternal and Child Health Bureau (MCHB), Emergency Medical Services for Children (EMSC) Network Development Demonstration Program under cooperative agreements U03MC00008, U03MC00001, U03MC00003, U03MC00006, U03MC00007, U03MC22684, and U03MC22685. This information or content and conclusions are those of the author and should not be construed as the official position or policy of, nor should any endorsements be inferred by HRSA, HHS, or the U.S. Government. The authors have no financial relationships relevant to this article to disclose.

Disclosure: None.

Disclaimer: None.

REFERENCES

- Agency for Healthcare Research and Quality. 2014. "National Guideline Clearinghouse" [accessed on June 16, 2014]. Available at <http://www.guideline.gov/>
- Alper, J., J. Sanders, R. S. Saunders, and Institute of Medicine. 2013. *Roundtable on Value and Science-Driven Health Care*. Core Measurement Needs for Better Care, Better Health, and Lower Costs: Counting What Counts: Workshop Summary. Washington, DC: National Academies Press.
- Alpern, E. R., R. M. Stanley, M. H. Gorelick, A. Donaldson, S. Knight, S. J. Teach, T. Singh, P. Mahajan, J. G. Goepf, N. Kuppermann, J. M. Dean, and J. M. Chamberlain, and Pediatric Emergency Care Applied Research. 2006. "Epidemiology of a Pediatric Emergency Medicine Research Network: The PECARN Core Data Project." *Pediatric Emergency Care* 22 (10): 689–99.
- Ashton, C. M., D. H. Kuykendall, M. L. Johnson, and N. P. Wray. 1999. "An Empirical Assessment of the Validity of Explicit and Implicit Process-of-Care Criteria for Quality Assessment." *Medical Care* 37 (8): 798–808.

- Bardach, N. S., A. T. Chien, and R. A. Dudley. 2010. "Small Numbers Limit the Use of the Inpatient Pediatric Quality Indicators for Hospital Comparison." *Academic Pediatrics* 10 (4): 266–73.
- Bardach, N. S., E. Vittinghoff, R. Asteria-Penaloza, J. D. Edwards, J. Yazdany, H. C. Lee, W. J. Boscardin, M. D. Cabana, and R. A. Dudley. 2013. "Measuring Hospital Quality Using Pediatric Readmission and Revisit Rates." *Pediatrics* 132 (3): 429–36.
- Cohen, J. 1988. *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale: L. Erlbaum Associates.
- Cronbach, L. J. 1951. "Coefficient Alpha and the Internal Structure of Tests." *Psychometrika* 16 (3): 297–334.
- Dans, P. E., J. P. Weiner, and S. E. Otter. 1985. "Peer Review Organizations. Promises and Potential Pitfalls." *New England Journal of Medicine* 313 (18): 1131–7.
- Dharmar, M., J. P. Marcin, N. Kuppermann, E. R. Andrada, S. Cole, D. J. Harvey, and P. S. Romano. 2007. "A New Implicit Review Instrument for Measuring Quality of Care Delivered to Pediatric Patients in the Emergency Department." *BMC Emergency Medicine* 7: 13.
- Dharmar, M., J. P. Marcin, P. S. Romano, E. R. Andrada, F. Overly, J. H. Valente, D. J. Harvey, S. L. Cole, and N. Kuppermann. 2008. "Quality of Care of Children in the Emergency Department: Association With Hospital Setting and Physician Training." *Journal of Pediatrics* 153 (6): 783–9.
- Donabedian, A. 1988. "The Quality of Care. How Can It Be Assessed?" *Journal of the American Medical Association* 260 (12): 1743–8.
- Fitch, K. 2001. *The Rand/UCLA Appropriateness Method User's Manual*. Santa Monica, CA: Rand.
- Fleiss, J. L. 1981. *Statistical Methods for Rates and Proportions*. New York: Wiley.
- Gausche-Hill, M. 2007. "Implementation and Evaluation of Care of Children in the Emergency Department: Guidelines for Preparedness." Presented at the American Academy of Pediatrics Emergency Department Preparedness Guidelines Advisory Council; National EMSC Resource Center; Silver Springs, MD; May 2–4, 2007.
- Goldman, R. L. 1992. "The Reliability of Peer Assessments of Quality of Care." *Journal of the American Medical Association* 267 (7): 958–60.
- . 1994. "The Reliability of Peer Assessments. A Meta-Analysis." *Evaluation and the Health Professions* 17 (1): 3–21.
- Hayward, R. A., M. Heisler, J. Adams, R. A. Dudley, and T. P. Hofer. 2007. "Overestimating Outcome Rates: Statistical Estimation When Reliability Is Suboptimal." *Health Services Research* 42 (4): 1718–38.
- Hofer, T. P., S. M. Asch, R. A. Hayward, L. V. Rubenstein, M. M. Hogan, J. Adams, and E. A. Kerr. 2004. "Profiling Quality of Care: Is There a Role for Peer Review?" *BMC Health Services Research* 4 (1): 9.
- Hulka, B. S., F. J. Romm, G. R. Parkerson Jr, I. T. Russell, N. E. Clapp, and F. S. Johnson. 1979. "Peer Review in Ambulatory Care: Use of Explicit Criteria and Implicit Judgments." *Medical Care* 17 (3 Suppl): i-vi, 1–73.

- IOM – Institute of Medicine. Committee on the Future of Emergency Care in the United States Health System. 2007a. *Emergency Medical Services at the Crossroads*. Washington, DC: National Academies Press.
- IOM – Institute of Medicine. Committee on the Future of Emergency Care in the United States Health System. 2007b. *Hospital-Based Emergency Care: At the Breaking Point*. Washington, DC: National Academies Press.
- IOM – Institute of Medicine. Committee on the Future of Emergency Care in the United States Health System. 2007c. *Emergency Care for Children: Growing Pains*. Washington, DC: National Academies Press.
- Kahn, K. L., W. H. Rogers, L. V. Rubenstein, M. J. Sherwood, E. J. Reinisch, E. B. Keebler, D. Draper, J. Kosecoff, and R. H. Brook. 1990. “Measuring Quality of Care With Explicit Process Criteria before and after Implementation of the DRG-Based Prospective Payment System.” *Journal of the American Medical Association* 264 (15): 1969–73.
- Kuppermann, N., J. F. Holmes, P. S. Dayan, J. D. Hoyle Jr, S. M. Atabaki, R. Holubkov, F. M. Nadel, D. Monroe, R. M. Stanley, D. A. Borgialli, M. K. Badawy, J. E. Schunk, K. S. Quayle, P. Mahajan, R. Lichenstein, K. A. Lillis, M. G. Tunik, E. S. Jacobs, J. M. Callahan, M. H. Gorelick, T. F. Glass, L. K. Lee, M. C. Bachman, A. Cooper, E. C. Powell, M. J. Gerardi, K. A. Melville, J. P. Muizelaar, D. H. Wisner, S. J. Zuspan, J. M. Dean, S. L. Wootton-Gorges, and Pediatric Emergency Care Applied Research. 2009. “Identification of Children at Very Low Risk of Clinically-Important Brain Injuries after Head Trauma: A Prospective Cohort Study.” *Lancet* 374 (9696): 1160–70.
- Landis, J. R., and G. G. Koch. 1977. “The Measurement of Observer Agreement for Categorical Data.” *Biometrics* 33 (1): 159–74.
- Mangione-Smith, R., J. Schiff, and D. Dougherty. 2011. “Identifying Children’s Health Care Quality Measures for Medicaid and CHIP: An Evidence-Informed, Publicly Transparent Expert Process.” *Academic Pediatrics* 11 (3 Suppl): S11–S21.
- Mistry, K. B., F. Chesley, K. Llanos, and D. Dougherty. 2014. “Advancing Children’s Health Care and Outcomes Through the Pediatric Quality Measures Program.” *Academic Pediatrics* 14 (5 Suppl): S19–S26.
- Nunnally, J. C., and I. H. Bernstein. 1994. *Psychometric Theory*. New York: McGraw-Hill.
- Olsen, L., C. Grossmann, J. M. McGinnis, and Institute of Medicine. Roundtable on Value & Science-Driven Health Care. 2011. *Learning What Works: Infrastructure Required for Comparative Effectiveness Research: Workshop Summary*. Washington, DC: National Academies Press.
- Rubenstein, L. V. 1991. *Structured Implicit Review of the Medical Record: A Method for Measuring the Quality of Inhospital Medical Care and a Summary of Quality Changes Following Implementation of the Medicare Prospective Payment System*. Santa Monica, CA: Rand.
- Rubenstein, L. V., K. L. Kahn, E. J. Reinisch, M. J. Sherwood, W. H. Rogers, C. Kamberg, D. Draper, and R. H. Brook. 1990. “Changes in Quality of Care for Five Diseases Measured by Implicit Review, 1981 to 1986.” *Journal of the American Medical Association* 264 (15): 1974–9.

- Rubin, H. R., and Rand Corporation. 1990. *Guidelines for Structured Implicit Review of the Quality of Hospital Care for Diverse Medical and Surgical Conditions*. Santa Monica, CA: Rand.
- SAS Institute Inc. 2013. [accessed on January 1, 2016]. Available at <http://support.sas.com/kb/25/031.html>
- Shrout, P. E., and J. L. Fleiss. 1979. "Intraclass Correlations: Uses in Assessing Rater Reliability." *Psychological Bulletin* 86 (2): 420–8.
- Sijtsma, K. 2009. "On the Use, the Misuse, and the Very Limited Usefulness of Cronbach's Alpha." *Psychometrika* 74 (1): 107–20.
- Smith, M. A., A. J. Atherly, R. L. Kane, and J. T. Pacala. 1997. "Peer Review of the Quality of Care. Reliability and Sources of Variability for Outcome and Process Assessments." *Journal of the American Medical Association* 278 (19): 1573–8.
- Subcommittee on Febrile Seizures and American Academy of Pediatrics. 2011. "Neurodiagnostic Evaluation of the Child with a Simple Febrile Seizure." *Pediatrics* 127 (2): 389–94.
- Tzimenatos, L., E. Kim, and N. Kuppermann. 2015. "The Pediatric Emergency Care Applied Research Network: A History of Multicenter Collaboration in the United States." *Pediatric Emergency Care* 31 (1): 70–6.

SUPPORTING INFORMATION

Additional supporting information may be found online in the supporting information tab for this article:

Appendix SA1: Author Matrix.

Appendix SA2: Quality of Care Instrument.

Appendix SA3: Pediatric ED Condition-Specific Quality of Care Instruments.

Appendix SA4: Cronbach's Alpha, Item-Total Correlation and Cronbach's Alpha without Item for Individual Reviewers.

Appendix SA5: Proportion of the Five Item-Specific Quality of Care Scores for Which All Four Reviewers Reached Agreement within a Two-Point and One-Point Range.

Appendix SA6: Extreme Scores by Individual Reviewers.