Reliability of the Faces, Legs, Activity, Cry, and Consolability Scale in Assessing Acute Pain in the Pediatric Emergency Department

Adam Kochman, MD, John Howell, MD, Michael Sheridan, ScD, Maybelle Kou, MD, Esther Emory Shelton Ryan, BSN, Susan Lee, RN, Wendy Zettersten, RN, and Lauren Yoder, BSN

Objectives: The Faces, Legs, Activity, Cry, and Consolability (FLACC) scale is one of the most widely utilized observational pain assessment scales in clinical practice. Although designed and validated to assess postoperative pain, the tool is currently applied to assess acute pain in multiple settings, including the emergency department. Scarce literature exists evaluating the reliability of the FLACC scale in the nonsurgical population and none in the emergency department. We sought to investigate the reliability of the FLACC scale in assessing acute pain in the pediatric emergency department and to examine the sensitivity of FLACC scores after the administration of analgesia.

Methods: In phase 1 of this prospective study, a series of 2 independent evaluators, blinded to each other’s evaluations, scored 66 patients using the FLACC tool. Degree of concordance among the 6 dyads was used to measure interrater reliability. In phase 2, FLACC scores were obtained just before the administration of analgesia in 35 patients and measured at 30 and at 60 minutes after administration.

Results: Among the 6 dyads of evaluators, Kendall W demonstrated a strong concordance (27 of 30 measures; range, 0.63–1.00) for individual components of the scale and for the composite scores (range, 0.85–0.96). Significant mean reductions from preanalgesia FLACC scores [5.54; 95% confidence interval (CI), 4.79–6.30] were seen at 30 minutes (2.00; 95% CI, 1.61–2.39) and 60 minutes (1.14; 95% CI, 0.79–1.50) postanalgesia (P < 0.0001 for all comparisons).

Conclusions: The FLACC scale demonstrated high interrater reliability for both individual FLACC items and total scores in a convenience sample of patients aged 6 months to 5 years in a pediatric emergency department. It seems to be an appropriate observational tool to assess acute pain in this population.

Key Words: FLACC, pain assessment, pediatric pain assessment, observational pain scale

Although pain remains one of the most common concerns of patients presenting to the emergency department, proper assessment of this symptom has continued to trouble providers. Pediatric patients have been identified as a population at increased risk for the inappropriate recognition and treatment of acutely painful conditions. In 2001, the Joint Commission on Accreditation of Health Care Organizations mandated the assessment and documentation of pain for all health care encounters. Because of the large number of visits to the pediatric emergency department that are associated with acute pain, it is critical that pain assessment be optimized.

The American Academy of Pediatrics and the American Pain Society support the widely accepted theory that self-reported measures of this symptom are preferable to observational assessments by others. However, given the complexity and potential bias surrounding self-reported measurements, others contend that designating one “gold standard” may be overreaching. Most studies validating the use of self-report measures have identified obvious deficits in the ability to reliably assess young, preverbal, and inconsistently cooperative children. Because of the necessity of possessing basic cognitive and discriminatory abilities, the age at which children can reliably provide self-report remains controversial. Therefore, the use of observational and behavioral measures must be relied upon in these circumstances.

The Faces, Legs, Activity, Cry, and Consolability (FLACC) behavioral scale is perhaps one of the most well-recognized and widely utilized observational pain scales (Fig. 1). Developed in 1997 to evaluate postoperative pain in young children, the FLACC scale has been validated for the assessment of postoperative and postprocedural pain in this population. However, a recent and comprehensive systematic review by Crollin et al demonstrated a paucity of literature describing its application outside this setting. This review confirms the need to assess the reliability of this scale in a wider variety of environments and populations before broad utilization can be recommended, particularly in the emergency department where it is commonly used.

We set out to investigate the interrater reliability of the FLACC behavioral scale in assessing acute pain in the pediatric emergency department and to examine the sensitivity of FLACC scores at 30 and 60 minutes after administration of analgesia.

Methods

This was a prospective observational study conducted at a tertiary pediatric emergency department (Inova Children’s Hospital) with an annual census of 38,000 patients. The study was approved by the institutional review board of the Inova Health System (Falls Church, Va).

The FLACC behavioral scale is an observational assessment tool that measures pain by quantifying scores for 5 separate pain behaviors: facial expression, leg movement, activity, cry, and ability to be consoled. Each behavior is scored 0 to 2, with total scores subsequently ranging from 0 (no pain) to 10 (highest possible pain behavior). Before the commencement of data collection, each of the study enrollers (4 pediatric emergency nurses and 1 pediatric emergency medicine fellow) underwent a formal educational training session regarding proper administration of the FLACC scale.

The study was conducted in 2 phases. In phase 1, we sought to assess interrater reliability of the scale. To do this, we enrolled a convenience sample of 66 children between the ages of 6 months and 5 years who presented to the pediatric emergency department.
FIGURE 1. The FLACC behavioral pain assessment scale.

The FLACC scale was administered to each patient by 2 study enrollers. Within 2 minutes of the first enroller's administration, a second study enroller, who was blinded to the results of the first enroller's assessment, also scored the patient using the FLACC scale. The 2 results were then analyzed for degree of interrater reliability (concordance), which was calculated both for individual FLACC items and for the total FLACC score. To assess the interrater reliability across a range of conditions and levels of pain, patients were enrolled without regard to chief complaint or perceived pain. Patients in this phase of the study were enrolled only when 2 or more study enrollers were present in the department at the same time.

Patients were excluded from the study if they were (1) clinically unstable and required emergent intervention, (2) had an underlying medical disorder associated with chronic pain, (3) carried a diagnosis of developmental delay, or (4) presented with altered mental status. It has been suggested that patients in these populations do not outwardly manifest pain in the same manner as other patients.

In phase 2 of the study, we sought to investigate the change in FLACC score after analgesia administration. For this, we enrolled a convenience sample of an additional 35 patients presenting to the pediatric emergency department. Patients in this population were enrolled only if they received analgesia at the discretion of the patient's triage and clinical care were performed separately from the FLACC scores assessed for the study.

We did not obtain parental consent before enrollment of the study patients because they already had consented to pain assessment as standard of care for the emergency department visit. In addition, we felt that parental awareness of the clinician's behavioral observations might influence the outcome. Immediately after FLACC administration, parents were provided with an information sheet, which detailed the purpose of the investigation and the role of each patient, including the limited utilization of protected health information. If parents declined their child's participation, their data were immediately destroyed at that time.

Kendall coefficient of concordance (KCC) was used to measure the association of ordinal scoring made by 2 observers when assessing the same patient. Each behavior is scored as 0 to 2 with total scores subsequently ranging from 0 (no pain) to 10 (highest possible pain behavior). Kendall coefficient of concordance commonly is used in attribute agreement analysis. Unlike the kappa statistic, which measures absolute agreement between ratings and treats all misclassifications equally, KCC would consider the consequences of misclassifying a score of 10 as a score of 2 to be more serious than misclassifying a score of 10 as a score of 8.

To examine the relationship between FLACC scores preanalgesia administration and postanalgesia at 30 and at 60 minutes, repeated-measures analysis of variance was used. All analyses were conducted using the SAS statistical software (v9.2, SAS Institute Inc, Cary, NC).

RESULTS

Parents and caregivers of all the 101 patients from whom data were obtained received an information sheet and subsequently gave verbal consent to have their child's data included. Demographic and acute pain complaint data are listed in Table 1. The mean age of subjects in phase I was 26.1 months, and in phase II, it was 37.9 months. Kendall $W$, which was used to measure interrater reliability among the 6 dyads of evaluators, demonstrated a substantial to almost-perfect concordance for 90% (27/30) of the individual components of the scale (range, 0.63–1.00), and for 100% of the total scale scores (range, 0.85–0.96) (Table 2). Because patients in this phase were enrolled without regard to the perception of whether pain was considered to be present, Kendall $W$ was assessed in conditions presenting both with and in the absence of acute pain and was similar for all ranges of observed pain.

The distribution of chief complaints for patients in phase 2 is described in Table 1. Decreases in mean FLACC score were seen between the preanalgesia administration [5.54; 95% confidence interval (CI) 4.89–6.19] and at 30 and 60 minutes [4.87; 95% CI 4.22–5.52] postanalgesia.

![Table 1](image)

TABLE 1. Selected Demographic and Clinical Features of Patients Assessed by the FLACC Scale

<table>
<thead>
<tr>
<th>Category</th>
<th>Phase 1 (n = 67)</th>
<th>Phase 2 (n = 35)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, %</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6–18 mo</td>
<td>34</td>
<td>20</td>
</tr>
<tr>
<td>18–36 mo</td>
<td>42</td>
<td>43</td>
</tr>
<tr>
<td>3–5 y</td>
<td>24</td>
<td>37</td>
</tr>
<tr>
<td>Male, %</td>
<td>58</td>
<td>57</td>
</tr>
<tr>
<td>Acute pain complaint, %</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HEENT pain</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>Abdominal pain</td>
<td>14</td>
<td>14</td>
</tr>
<tr>
<td>Extremity pain</td>
<td>40</td>
<td>40</td>
</tr>
<tr>
<td>Other pain</td>
<td>26</td>
<td>26</td>
</tr>
</tbody>
</table>

HEENT indicates head, eyes, ears, nose, throat.
intervened on health care providers to ensure that all patients receive a formal assessment of their pain and, when appropriate, be treated for it. Children, especially those who are preverbal or nonverbal, are at high risk for having their pain overlooked and treatment subsequently disregarded.

The Pediatric Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials published a consensus statement based on the recommendations of its international panel detailing specific core outcome domains and measures for clinical trials of treatments for acute and chronic pain in children. The consensus meeting defined 4 distinct domains of acute pain: pain arising from medical procedures, postoperative pain, injury, and acute exacerbation of disease. After comprehensive literature review, the panel concluded that self-reported measures of pain intensity are more obvious limitations arising in the second phase of the study was the potential for expectation bias, as the FLACC measurements in the second phase of the study were performed without blinding to the administration of analgesia. It is possible that FLACC scores assigned after analgesia were influenced by the anticipation that analgesia should relieve pain and therefore manifest a lower score. This bias has been identified in several studies evaluating construct validity of the scale in the postoperative setting. To address this bias, we retrospectively reviewed clinical charts, which contained qualitative documentation of agreement with imputations that analgesia should relieve pain and therefore manifest a lower score.

The Pediatric Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials published a consensus statement based on the recommendations of its international panel detailing specific core outcome domains and measures for clinical trials of treatments for acute and chronic pain in children. The consensus meeting defined 4 distinct domains of acute pain: pain arising from medical procedures, postoperative pain, injury, and acute exacerbation of disease. After comprehensive literature review, the panel concluded that self-reported measures of pain intensity are not sufficiently valid for children up to age 4 and favored observational measures for this population.

In its evidence-based recommendations for observational measures of pain intensity, Pediatric Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials only supported their use in the setting of procedural and postoperative pain. None of the observational scales reviewed by the panel were approved for utilization in the other 2 domains of acute pain. It is precisely these domains—injury and disease-associated pain—that the population in the phase 2 of our study presented with. In 2003, Manworren and Hynan demonstrated clinical validation of the FLACC scale in a population of 147 children in various inpatient settings. Nearly 73% of these patients were either surgical, postoperative, or critically ill. Of the remaining 27%, 83% of these were neonates. None of these assessments occurred in an acute care setting or an emergency department.

In 2010, Voepel-Lewis et al validated the FLACC scale on a small population of critically ill patients, which included 29 adults and only 8 children. Of the 8 pediatric patients, 3 were on mechanical ventilation, and all but 1 were surgical patients.

More recently, Babl et al sought to validate the FLACC scale for the assessment of procedural pain in the pediatric emergency department. This prospective observational study utilized pediatric patients undergoing painful procedures in the emergency department and measured FLACC scores in 3 phases: preprocedural, during restraint, and during the procedure. To our knowledge, this is the only study to date that evaluates the FLACC scale in the emergency department.

Our study is unique in that it is the first to evaluate this observational pain scale both for interrater reliability and sensitivity after analgesia administration for the measurement of acute pain that was not in a postoperative or procedural setting. Our results confirm previous findings of strong interrater reliability for the FLACC scale.

One limitation of our study was some disparity in the frequency of measurements by each dyad. Ideally, there would have been an equal number of patients assessed by each dyad in the study’s initial phase; however, the practical limitation of requiring 2 study enrolers to be working clinically and available at the same time resulted in the disproportionate recruitments.

A more obvious limitation arising in the second phase of the study was the potential for expectation bias, as the FLACC measurements in the second phase of the study were performed without blinding to the administration of analgesia. It is possible that FLACC scores assigned after analgesia were influenced by the anticipation that analgesia should relieve pain and therefore manifest a lower score. This bias has been identified in several studies evaluating construct validity of the scale in the postoperative setting. To address this bias, we retrospectively reviewed clinical charts, which contained qualitative documentation of agreement with improvement in pain. The physicians caring for the patients were unaware of the ongoing study.

Malviya and colleagues validated a revised version of the FLACC observational tool in postoperative children with cognitive impairment. In their design, postanalgesia FLACC score assignments were correlated with those obtained by blinded nurses viewing randomized videotaped segments of patient behavior. Although this model would have lent additional support to the construct validity of the FLACC tool in our study in the pediatric emergency department, it was not practically feasible.

The determination of age criteria was based on several previous studies’ conclusions regarding the age at which children could reliably contribute to self-reported pain assessments. It has been suggested by some that children as young as 3 years possess this ability, whereas others argue that it is not possible before the age of 7 years. Several studies of self-report pain scales, including a systematic review of faces pain scales, demonstrate a paucity of evidence for the use of self-report before the age of 5 years. Therefore, this became the upper limit of our inclusion criteria for age.

We chose not to assess the predictive ability of the FLACC tool as it pertains to form of analgesia. More specifically, we did
not measure whether there was greater change in FLACC score assignment in patients who received an opioid versus those who received only acetaminophen or ibuprofen. Rather, subjects were eligible for enrollment if any analgesia was administered. It is possible in a subset of patients that acetaminophen or ibuprofen was administered not solely for acute pain but also for concurrent fever given the dual analgesic and antipyretic effects of these medications. However, patients in the second phase of the study were not included if they received these medications for fever in the absence of acute pain.

It has been suggested that the FLACC scale, like other observational tools, lacks discriminate validity and may measure both pain and distress in pediatric patients. This claim has also been directed against well-known self-report pain assessment tools. In our study, patients were not assessed in a preprocedure or preoperative state, perhaps reducing some degree of anxiety that children may experience in these settings. Until a more nuanced understanding of pain can be qualified, this will likely remain a limitation of all pain assessment tools used in this group.

CONCLUSIONS

The FLACC behavioral pain assessment scale is an observational tool used to assess pain in pediatric patients. Despite its widespread implementation in nearly all aspects of pediatric health care, it has only been validated in the postoperative environment. Our study demonstrates high reliability and sensitivity to acute pain assessment unrelated to surgery or procedure. Despite the limitations of observational measures of pain in comparison to self-reported measures, it is currently the best reliable method to quantify pain in young preverbal children.

REFERENCES