Development and Psychometric Evaluation of the Pediatric Anesthesia Emergence Delirium Scale


© 2004 American Society of Anesthesiologists, Inc. Lippincott Williams & Wilkins, Inc.

Anesthesiology 2004; 100:1138–45

Background: Emergence delirium has been investigated in several clinical trials. However, no reliable and valid rating scale exists to measure this phenomenon in children. Therefore, the authors developed and evaluated the Pediatric Anesthesia Emergence Delirium (PAED) scale to measure emergence delirium in children.

Methods: A list of scale items that were statements describing the emergence behavior of children was compiled, and the items were evaluated for content validity and statistical significance. Items that satisfied these evaluations comprised the PAED scale. Each item was scored from 1 to 4 (with reverse scoring where applicable), and the scores were summed to obtain a total scale score. The degree of emergence delirium varied directly with the total score. Fifty children were enrolled to determine the reliability and validity of the PAED scale. Scale validity was evaluated using five hypotheses: The PAED scale scores correlated negatively with age and time to awakening and positively with clinical emergence scores and Post Hospital Behavior Questionnaire scores, and were greater after sevoflurane than after halothane. The sensitivity of the scale was also determined.

Results: Five of 27 items that satisfied the content validity and statistical analysis became the PAED scale: (1) The child makes eye contact with the caregiver, (2) the child’s actions are purposeful, (3) the child is aware of his/her surroundings, (4) the child is restless, and (5) the child is inconsolable. The internal consistency of the PAED scale was 0.89, and the reliability was 0.84 (95% confidence interval, 0.76–0.90). Three hypotheses supported the validity of the scale: The scores correlated negatively with age (r = -0.31, P < 0.04) and time to awakening (r = -0.5, P < 0.001) and were greater after sevoflurane anesthesia than halothane (P < 0.008). The sensitivity was 0.64.

Conclusions: These results support the reliability and validity of the PAED scale.

EMERGENCE delirium (ED) has been described as “a mental disturbance during the recovery from general anesthesia consisting of hallucinations, delusions and confusion manifested by moaning, restlessness, involuntary physical activity, and thrashing about in bed.” It has been considered a common postanesthetic problem in children and adults since 1960.

The prevalence of ED in children ranges from 25 to 80%, depending on the anesthesia, from self-limiting but of variable duration. During an ED reaction, children risk injuring their surgical repair, themselves, and their caregivers. Their behavior is disruptive to the postanesthetic care unit and often requires constant nursing supervision, which strains nursing manpower resources. Moreover, when an ED reaction occurs, all members of the healthcare team as well as the parents express dissatisfaction with the quality of the child’s recovery.

These negative effects of ED have motivated clinicians to investigate possible etiologies and potential treatments for ED. However, none of the clinical investigations have used a reliable and valid tool to measure ED. Not only does this preclude comparisons among the clinical trials, but more importantly, it raises serious questions regarding measurement error and the reliability of the measurement and validity of the research results.

Sixteen rating scales and two visual analog scales that measure agitation have been used to measure ED in young children. These scales are deficient in two main respects: scale content and psychometric evaluation. Behaviors including crying, agitation, and lack of cooperation have been included as items in these ED rating scales. However, these behaviors are not specific to ED. They may also characterize children who are in pain or who are frightened or angry during emergence from general anesthesia. Of the rating scales listed in table 1, two scales report reliability estimates, and one, the Heaman-Mattle emergence excitement scale, has undergone both a reliability and a validity assessment. However, the Heaman-Mattle scale was developed for teenagers and is inappropriate for use with preschool and school-aged children. Because the content of the scales in table 1 was considered inadequate, further assessment of the psychometric properties of any one scale was not pursued by the authors.

To date, a reliable and valid rating scale to measure ED in children does not exist. Shroot and Fleiss state that “measurement error can seriously affect statistical analysis and interpretation [of data].” Therefore, to minimize measurement error in the clinical evaluation of ED in children, we sought to develop a reliable and valid rating scale to measure this phenomenon.

Materials and Methods

Methods

This study was approved by the Research Ethics Board at The Hospital for Sick Children (Toronto, Ontario, Ontario, Canada. † Clinical Professor of Anesthesiology, Women and Children’s Hospital of Buffalo, State University at New York, Buffalo, New York.
thought to have ED or delirium. From these behavioral
traits to collate behavioral descriptions of children
anesthesiologists, PACU nurses, and a pediatric psychia-
trist, to determine their content validity. These individ-
uals were considered experts because they had clinical
expertise with the emergence behavior of children,
and informed written consent was obtained
from the parents of all children who participated in this
study. The study methods consisted of two phases: scale
development and scale evaluation. Scale development
involved the construction of the Pediatric Anesthesia
Emergence Delirium (PAED) scale. Scale evaluation de-
tailed the scale’s reliability and validity.

**Scale Development.** First, ED was defined as a distur-
bance in a child’s awareness of and attention to his or
her environment with disorientation and perceptual al-
terations including hypersensitivity to stimuli and hyper-
active motor behavior in the immediate postanesthesia
period. This definition was predicated on the theoretical
framework of delirium found in the Diagnostic and
Statistical Manual of Mental Disorders. Second, the anesthesiology, nursing, and psychiatric literature was
reviewed, and interviews were conducted with pediatric
anesthesiologists, PACU nurses, and a pediatric psychia-
trist to collate behavioral descriptions of children
thought to have ED or delirium. From these behavioral
descriptions, six categories of ED behaviors were de-
vised: cognitive behavior, behavioral response to environ-
mental stimuli, behavior threatening patient safety, motor
behavior, affective behavior, and vocal behavior. Guided by
the definition of ED and the six behavioral categories, a list
of preliminary scale items or statements that described the
emergence behavior of children was compiled.

The preliminary scale items were evaluated by seven
experts, including four senior pediatric PACU nurses,
two pediatric anesthesiologists, and a pediatric psychia-
rist, to determine their content validity. These individ-
uals were considered experts because they had clinical
expertise with the emergence behavior of children,
knowledge of the conceptual framework of delirium
described in the Diagnostic and Statistical Manual of Mental Disorders, or knowledge of the scale develop-
ment process.

The content validity evaluation was a two-step process
for which specific instructions where given to each expert.
First, each expert was asked to rate the rele-

---

**Table 1. Emergence Delirium Rating Scales**

<table>
<thead>
<tr>
<th>Author/Year</th>
<th>Scale Description</th>
<th>Population Assessed</th>
<th>Scale Type</th>
<th>Scale Development</th>
<th>Reliability</th>
<th>Validity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cole et al., 2002</td>
<td>Emergence Delirium</td>
<td>10 months–6 yr</td>
<td>Ordinal</td>
<td>None described</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Hallen et al., 2001</td>
<td>Excitation Scale</td>
<td>3–8 yr</td>
<td>Ordinal</td>
<td>None described</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Cohen et al., 2001</td>
<td>Agitation Scale</td>
<td>2–9 yr</td>
<td>Ordinal</td>
<td>None described</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Galinkin et al., 2000</td>
<td>Postoperative Behavior Scale</td>
<td>9 months–6 yr</td>
<td>Ordinal</td>
<td>None described</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Cravero et al., 2000</td>
<td>Emergence Agitation Scale</td>
<td>6 months–10 yr</td>
<td>Ordinal</td>
<td>None described</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Aono et al., 1999</td>
<td>Problematic Behavior Scale</td>
<td>3–6 yr, males</td>
<td>Ordinal</td>
<td>None described</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Davis et al., 1999</td>
<td>Agitation Scale</td>
<td>1–5 yr</td>
<td>Ordinal</td>
<td>None described</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Aono et al., 1997</td>
<td>Behavior Rating Scale</td>
<td>3–10 yr, males</td>
<td>Ordinal</td>
<td>None described</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Johannesson et al., 1995</td>
<td>Behavior during Emergence</td>
<td>1–7 yr</td>
<td>Ordinal</td>
<td>None described</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Keegan et al., 1995</td>
<td>Excitement Score</td>
<td>1–15 yr</td>
<td>Ordinal</td>
<td>None described</td>
<td>Interrater reliability, ICC = 0.997</td>
<td>None</td>
</tr>
<tr>
<td>Davis et al., 1994</td>
<td>Quality of Anesthesia Recovery</td>
<td>1–6 yr</td>
<td>Ordinal</td>
<td>None described</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Watcha et al., 1992</td>
<td>Agitation Scale</td>
<td>3 months–4 yr</td>
<td>Ordinal</td>
<td>None described</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Heaman and Mattle, 1982</td>
<td>Heaman-Mattle Excitement Scale</td>
<td>13–18 yr</td>
<td>Ordinal</td>
<td>None described</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Shields et al., 1980</td>
<td>Emergence Excitement Score Sheet</td>
<td>2–13 yr</td>
<td>Adjectival scale</td>
<td>None described</td>
<td>Interrater agreement, ( r = 0.82–0.84 )</td>
<td>None</td>
</tr>
<tr>
<td>Sheffer et al., 1973</td>
<td>Tranquility/Alertness Scale</td>
<td>3–11 yr</td>
<td>Semantic Differential Scale</td>
<td>None described</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Smessaert et al., 1960</td>
<td>Mode of Recovery</td>
<td>10–65 yr</td>
<td>Ordinal</td>
<td>None described</td>
<td>None</td>
<td>None</td>
</tr>
</tbody>
</table>

**Note:** ICC = intraclass correlation coefficient.
vance of each scale item to the definition of ED using a seven-point scale ranging from not at all relevant (score of 1) to extremely relevant (score of 7). Second, the experts were asked to determine which of the six behavioral categories of ED each item best represented. The definition of ED and the behavioral categories were given to each expert. Items deemed content-valid were then pretested on a group of 100 children. For pretesting, items were scored as they would be for the final scale using the five response options: not at all (score of 0), just a little (score of 1), quite a bit (score of 2), very much (score of 3), and extremely (score of 4). Reverse scoring of items included the options not at all (4), just a little (3), quite a bit (2), very much (1), and extremely (0) and was used where applicable so that the greater the item score, the greater the degree of ED. During pretesting, each item was used by one of the authors (N.S.) to evaluate the emergence behavior of 100 children 10 min after the child awakened from anesthesia. Children were included if they were aged between 18 months and 6 yr; had an American Society of Anesthesiologists physical status class of I or II; had no known behavioral disorders; understood English; had no known contraindications to inhaled anesthetics; and were at risk for malignant hyperthermia. The evaluating author (N.S.) was blinded to the anesthetic administered was determined by the child’s surgical procedure. Children were excluded if they needed premedication, had cognitive impairment, or were at risk for malignant hyperthermia. The evaluating author (N.S.) was blinded to the type of anesthetic that the child received during surgery. The scores on each pretested scale item were analyzed (statistical item analysis) to obtain a statistical profile of each item. 

Hypothesis 1. The PAED scale scores correlated negatively with the child’s age. 

Hypothesis 2. The PAED scale scores correlated negatively with the child’s time to awakening, defined in minutes as the time from arrival in the PACU until consciousness is sustained.

Hypothesis 3. The PAED scale scores correlated positively with a clinical judgment score of ED measured on a seven-point scale from none (score of 1) to an extreme amount (score of 7). Each of the three observers in the reliability study completed the clinical judgment score after evaluating the child with the PAED scale.

Hypothesis 4. The PAED scale scores correlated positively with the child’s Post Hospital Behavior Questionnaire (PHBQ) scores as evaluated by a parent on postoperative days (PODs) 2 and 7. The choice of anesthetic administered was determined by the child’s attending anesthesiologist.

Hypothesis 5. The PAED scale scores in children who received sevoflurane were greater than in those who received halothane.

Sample Size

Scale Development. It has been recommended that between 3 and 10 experts evaluate content validity. The sample size for the item pretesting was based on enrolling five subjects for each item to be pretested.

Sample Evaluation. The sample size for the interobserver reliability study was estimated using a Pearson product–moment correlation coefficient ($r$) of 0.75, a half-width of the confidence interval (CI) of ± 0.1, and an $\alpha_2$ of 0.05. A sample size of 50 children was estimated.

Scale Evaluation. The sample size to test hypothesis 5 was based on an estimate of the expected effect size. Because the PAED scale is a new measure and no data exist to compute an effect size, the effect size was estimated. Assuming a medium effect size of 0.5 between the PAED scale scores of children who received sevoflurane and those who received halothane, the sample size for each group was estimated to be 63 children.
Statistical Analysis

Descriptive statistics were used to characterize the study sample. Age and duration of surgery were recorded as means and SDs. Type of surgery, type of inhalational anesthetic administered during surgery, and use of intraoperative narcotics were reported as proportions.

Scale Development. An item was deemed content relevant if it was rated at 4 or greater on the seven-point scale by six of the seven experts and if it represented only one of the six ED behavioral categories.\textsuperscript{39,50} Statistical item analysis\textsuperscript{39,51} included compiling the frequencies of the response options for each item (endorsement frequency) and the correlations between each item (item–item correlations) and between the item’s score and the scale’s total score (the item–total correlations). Items with response options that were selected with a frequency greater than 5% or less than 95% were retained. Of these, the item set with moderate item–item correlations, item–total correlations of 0.2 or greater, and an adequate internal consistency defined as an \( \alpha \) coefficient of greater than 0.7 but less than 0.9 was selected as the PAED scale.

Scale Evaluation. The interobserver reliability was determined using a one-way analysis of variance random-effects model and was reported as an intraclass correlation coefficient (for a single observer) with a 95% CI.\textsuperscript{32}

For validity hypotheses 1–4, the PAED scale scores of the three observers for each child were correlated with the age of the child, the time to awakening, the clinical judgment scores, and the PHBQ scores. An average correlation coefficient was determined and evaluated for statistical significance by testing the null hypothesis of \( H_0 : \rho = 0 \) against \( H_A : \rho \neq 0. \)\textsuperscript{54} Statistical significance was accepted at \( P < 0.05. \) Data were assessed for departure from normality. For those distributions that deviated from normality, the level for statistical significance was reduced to \( P < 0.01. \)\textsuperscript{35} For validity hypothesis 5, the PAED scale scores were compared using a two-sided unpaired \( t \) test or the comparable nonparametric Mann–Whitney test if the data deviated from normality. Statistical significance was accepted at \( P < 0.05. \) Data entry was double-checked and then analyzed using Statistical Package for the Social Sciences for Windows, version 11.0.0 (© 1989–2001; SPSS Inc., Chicago, IL).

To construct the ROC curve, the PAED scale scores were correlated using a Spearman (\( \rho \)) correlation coefficient with the dichotomous outcome of yes/no for treatment with dimenhydrinate. An ROC curve was generated using a nonparametric distribution assumption with the PAED scale score as the target variable and a response of yes for dimenhydrinate treatment as the positive state variable. The degree of ED increased directly with the PAED scale score. The PAED scale score that maximized the area under the curve of true positives (sensitivity) and minimized the area under the curve of false positives (1-specificity) was accepted as the cutoff point to define a case of ED that required treatment from one that did not.

Results

Scale Development (fig. 1)

Twenty-seven preliminary scale items were compiled (table 2). After evaluation, 21 items were deemed to be content-valid (table 2). These 21 items were pretested on 100 children, 56 males and 44 females, aged 3.7 ± 1.5 yr (tables 3 and 4), whose surgery lasted 63.2 ± 33.6 min (mean ± SD). Twenty percent of the children received an opioid intraoperatively. Five of the 21 items were deemed to have an adequate statistical profile. These items comprised the PAED scale (table 5). The internal consistency of the PAED scale was 0.89.

Scale Evaluation

The reliability of the PAED scale was evaluated in 46 of the 50 children. The interobserver reliability of the PAED scale was 0.84 (95% CI, 0.76–0.90). Results of the construct validity hypothesis testing are as follows.

Hypothesis 1. The PAED scale score correlated negatively with the child’s age (\( r = -0.31, P < 0.04 \) (n = 46).

Hypothesis 2. The PAED scale score correlated negatively with the child’s time to awakening (\( r = -0.50, \)
Table 2. Preliminary Scale Item List

1. The child can focus his attention on the caregiver.
2. The child pulls at the monitoring equipment or IV that is connected to him/her.
3. The child’s behavior threatens his/her safety.
4. The child’s movements are disruptive.
5. The child’s mood is irritable.
6. The child makes eye contact with the caregiver.
7. The child has hyperactive motor behavior.
8. The child’s actions are purposeful.
9. The child is agitated when touched by the caregiver.
10. The child’s behavior requires supervision.
11. The child is restless.
12. The child recognizes familiar objects (toys, blanket).
13. The child responds positively to comforting efforts by the caregiver.
14. The child is aware of his/her surroundings.  
15. The child is combative toward the caregiver who tries to comfort him/her.
16. The child’s behavior makes his/her postoperative nursing care difficult.
17. The child is vocally noisy.
18. The child is aware of his/her surroundings.
19. The child is distressed by the monitoring equipment connected to him/her.
20. The child is inconsolable.
21. The child’s behavior is uncontrollable.
22. The child seems panic stricken.
23. The child interacts purposefully with the caregiver.
24. The child responds purposefully to verbal stimuli.
25. The child is hypersensitive to tactile stimuli.
26. The child is uncooperative.
27. The child is attentive to his/her surroundings.

Items in italics are content-valid items.

P < 0.001). The times to awakening were not normally distributed (n = 46).

Hypothesis 3. The PAED scale score correlated positively with the clinical judgment scores (r = 0.86, P < 0.001). The clinical judgment scores were not normally distributed (n = 46).

Hypothesis 4. The PAED scale score correlated negatively with the PHBQ scores on PODs 2 and 7 (r = 0.31, P < 0.08) (n = 33) and 7 (r = 0.22, P = 0.20) (n = 34). The PHBQ scores on PODs 2 and 7 were not normally distributed.

Of the 50 parents who were given the PHBQ, 38 returned both questionnaires (POD 2 and 7 assessments). Of the 38 respondents, two were excluded because there was no corresponding PAED scale score, and two were excluded because their children were admitted to hospital postoperatively. These last two children were excluded from this evaluation because of concern for confounding effects of hospitalization on the child’s behavior. A fifth child was excluded because the assessment on POD 2 was incomplete.

Hypothesis 5. Seventeen children received sevoflurane for maintenance of anesthesia, and 25 children received halothane. The PAED scale scores were normally distributed in each treatment group. The average PAED scale scores of children who received sevoflurane was 7.2 ± 4.5 and of those who received halothane was 3.7 ± 2.6 (P < 0.008).

ROC Curve Analysis

Of the 100 children included in this analysis, 80 children did not receive morphine in the postoperative period. Of these, 11 received dimenhydrinate in the absence of vomiting. The ROC curve generated from these data accounted for 76.6% of the area under the curve. At a PAED scale score of 10 or greater, the true-positive rate (sensitivity) was 0.64, and the false-positive rate (1-specificity) was 0.14 (fig. 2).

Discussion

To minimize measurement error in the assessment of ED, clinicians require a reliable and valid measurement tool. Using a theoretical framework of delirium, we developed the PAED scale as a rating scale to measure ED in children. We conclude that the PAED scale is a reliable and valid tool based on the scale’s reliability, content, and initial construct validity profile determined in this study.

During the development of the PAED scale, ideas for scale items were collected from a variety of resources,

Table 3. Surgical Procedures (n = 100)

<table>
<thead>
<tr>
<th>Procedure</th>
<th>Proportion of Children, %</th>
<th>Duration of Surgery,* min</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ear, nose, throat</td>
<td>28</td>
<td>33.8 ± 24.2</td>
</tr>
<tr>
<td>Plastic surgery</td>
<td>9</td>
<td>66.6 ± 21.7</td>
</tr>
<tr>
<td>Dental procedures</td>
<td>52</td>
<td>83.5 ± 25.7</td>
</tr>
<tr>
<td>Ophthalmology</td>
<td>8</td>
<td>28.1 ± 11.9</td>
</tr>
<tr>
<td>Missing value</td>
<td>3</td>
<td>NA</td>
</tr>
</tbody>
</table>

* Data are presented as mean ± SD.

Table 4. Anesthetic Agent (n = 100)

<table>
<thead>
<tr>
<th>Anesthetic Agent</th>
<th>Proportion of Children, %</th>
<th>Duration of Surgery,* min</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sevoflurane</td>
<td>17</td>
<td>28.5 ± 10.4</td>
</tr>
<tr>
<td>Halothane</td>
<td>25</td>
<td>47.5 ± 34.4</td>
</tr>
<tr>
<td>Isoflurane</td>
<td>57</td>
<td>81.0 ± 25.6</td>
</tr>
<tr>
<td>Missing value</td>
<td>1</td>
<td>NA</td>
</tr>
</tbody>
</table>

* Data are presented as mean ± SD.

Table 5. The PAED Scale

1. The child makes eye contact with the caregiver.
2. The child’s actions are purposeful.
3. The child is aware of his/her surroundings.
4. The child is restless.
5. The child is inconsolable.

Items 1, 2, and 3 are reversed scored as follows: 4 = not at all, 3 = just a little, 2 = quite a bit, 1 = very much, 0 = extremely. Items 4 and 5 are scored as follows: 0 = not at all, 1 = just a little, 2 = quite a bit, 3 = very much, 4 = extremely. The scores of each item were summed to obtain a total Pediatric Anesthesia Emergence Delirium (PAED) scale score. The degree of emergence delirium increased directly with the total score.
including a review of the item content of three validated pediatric pain scales.57–59 Because of the known difficulty in differentiating pain from ED, it was important to preclude scale items that may also reflect pain.5,7,21 Of the three pain scales reviewed, only the Face, Legs, Activity, Cry, Consolability (FLACC) scale includes an item of consolability.58 All three scales use an aspect of restlessness to measure pain. Accordingly, it is possible that the PAED scale items “The child is inconsolable” and “The child is restless” may reflect pain as well as ED.

We included the salient features of delirium, i.e., a disturbance in consciousness and changes in cognition and the associated features, including a disturbance in psychomotor behavior and emotion, in the genesis of the PAED scale.56 A disturbance in consciousness includes a reduced awareness of the environment and impairment in the ability to focus, sustain, or shift attention.56 The PAED scale’s first item, “The child makes eye contact with the caregiver,” and third item, “The child is aware of his/her surroundings,” reflect disturbances in the child’s consciousness during an ED reaction. Cognitive changes may include impairment in perception and memory and disorganized thinking patterns. Purposeful movement may be altered in a child whose thinking is disorganized. The second item on the PAED scale, “The child’s actions are purposeful,” addressed changes in the child’s cognition during an ED reaction. The inclusion of items that reflect disturbances in consciousness and cognition may be pivotal to differentiating ED from pain.

The disturbance in psychomotor behavior and emotion, which are associated features of a delirium, have been captured in the fourth and fifth items on the PAED scale. “The child is restless” and “The child is inconsolable,” respectively. These are the features of ED that are most commonly incorporated in previous scales. Although these last two features may reflect pain as stated earlier, it is hoped that when they are grouped with indicators of consciousness and cognition such as items 1–3 (table 5), they better reflect ED than pain. Assessing children with the PAED scale and a valid and reliable pain scale may be required to test this assumption.

Reverse scoring was required for the first three items on the PAED scale. Reverse scoring can be easily applied by having all items scored in the conventional way (as per items 4 and 5 in table 5) and then subtracting the score of the item from a value of 4. This should make the scale easy to use even in a busy clinical setting. For example, if a conventional score of 4 (extremely) was chosen for item 1, then the actual reverse score for this item would be recorded as 0 (4 − 4), which is equal to the reverse-scored value of “extremely” in table 5.

The adjectives used for the response options were not operationally defined. This may be considered a limitation of the scale. However, large variability in the interpretation of the meaning of the response options for any item would have negatively affected the interobserver reliability coefficient. That the interobserver reliability of the PAED scale was 0.84, which exceeds the minimum acceptable reliability for a useful instrument of 0.75, suggests that the observers’ interpretations of the response options were similar enough so as to not compromise the scale’s reliability.

Whether the scores from rating scales can be considered interval data remains controversial. Unless the distribution of the scores from a rating scale is severely skewed, the data can be analyzed as if they were interval data, without introducing severe bias into the results.59 The scores from the PAED scale were all normally distributed in this analysis.

We tested five hypotheses to explore the construct validity of the PAED scale. This is consistent with the notion that construct validity is determined by a series of converging experiments.59 Of these five hypotheses, hypotheses 1 (age), 2 (awake time), and 5 (sevoflurane vs. halothane) supported the construct validity of the PAED scale. Hypothesis 3, which involved the clinical judgment scores, was rejected because of criterion contamination. Criterion contamination occurs when the results of one test bias the results of another.59 This bias artificially inflates the correlation between these two tests. In this study, the observers evaluated each child with the PAED scale first and with a seven-point scale of clinical judgment second. Because of this and the high correlation between the scores on these two scales, it is unknown to what extent the PAED scale scores biased the clinical judgment scores.

Our failure to find a statistically significant relation
between ED and any negative postoperative behavioral changes (validity hypothesis 4) may be attributed to the absence of a well-established theory associating these two constructs.39

The ROC analysis predicts a score above which an episode of ED requires treatment. The sensitivity of the scale is fair, although the false-positive rate is quite high. This may be a function of the positive state response variable used in this analysis. Further attempts to determine a cutoff point are needed, using other positive state response variables, to substantiate or improve on the ROC results determined in this study.

Our results showed that the PAED scale score in children who received sevoflurane was greater than that in those who received halothane. Although the estimated sample size for this comparison was not achieved, statistical significance was achieved because the effect size measured, 1.0, was double that used in the sample size estimation.

In conclusion, we detail the development and evaluation of a new rating scale to measure ED in children recovering from general anesthesia. Based on our results, the PAED scale is a reliable and valid measure of ED in children.

The authors thank the nurses in the Post Anesthetic Care Unit, The Hospital for Sick Children, Toronto, Ontario, Canada, for their participation in this study; David L. Streiner, Ph.D. (Professor Emeritus, Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, Canada, and Professor, Department of Psychiatry, University of Toronto, Toronto, Ontario, Canada), Geoffrey R. Norman Ph.D. (Professor, Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, Canada), and Peter Szatmari, M.D. (Professor, Department of Psychiatry and Behavioural Neurosciences, Offord Centre for Child Studies, McMaster University and Hamilton Health Sciences, Hamilton, Ontario, Canada), for their guidance; and Zeey N. Kain, M.D. (Professor, Anesthesiology, Pediatrics and Child Psychiatry, Yale University School of Medicine, and Anesthesiology in Chief, Yale New Haven Children’s Hospital, New Haven, Connecticut), and Arlette Lefebvre, M.D., D.C.P. (Psychiatrist, The Hospital for Sick Children, Toronto, and Associate Professor, Department of Psychiatry, University of Toronto), for their assistance during the scale development phase.

References


Anesthesiology, V 100, No 5, May 2004