Development and Psychometric Evaluation of the Pediatric Anesthesia Emergence Delirium Scale

Nancy Sikich, M.Sc., R.N.,* Jerrold Lerman, B.A.Sc., M.D., F.R.C.P.C., F.A.N.Z.C.A.†

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 judgment 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."[1] It has been considered a common postanesthetic problem in children and adults since 1960.[2–4] The prevalence of ED in children ranges from 25 to 80%, depending on the definition of ED used to measure this phenomenon.[5,6] ED, which usually occurs within the first 30 min after ether anesthesia,[7–10] has been characterized as self-limiting but of variable duration.[5,10,11] 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.[12,13] 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.[5,14] These negative effects of ED have motivated clinicians to investigate possible etiologies and potential treatments for ED.[5–7,10,14–25] 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.[6,26]

Sixteen rating scales[3,5,6,8–10,15,16,18,20,21,26–30] and two visual analog scales that measure agitation have been used to measure ED in young children.[7,31] (table 1). 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. Shrout and Fleiss[32] 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,

* Research Nurse, Department of Anesthesia, St. Michael's Hospital, Toronto, Ontario, Canada. † Clinical Professor of Anesthesiology, Women and Children's Hospital of Buffalo, State University at New York, Buffalo, New York.

Address reprint requests to Ms. Sikich: St. Michael's Hospital, Department of Anesthesia, 30 Bond Street, Toronto, Ontario, Canada, M5B 1W8. Address electronic mail to: sikichn@smh.toronto.on.ca. Individual article reprints may be purchased through the Journal Web site, www.anesthesiology.org.

Table 1. Emergence Delirium Rating Scales

Author/Year	Scale	Population Assessed	Scale Type	Scale Development	Reliability	Validity
Cole et al.,[10] 2002	Emergence Delirium	10 months–6 yr	Ordinal	None described	None	None
Hallen et al.,[8] 2001	Excitation Scale	3–8 yr	Ordinal	None described	None	None
Cohen et al.,[18] 2001	Agitation Scale	2–9 yr	Ordinal	None described	None	None
Galinkin et al.,[5] 2000	Postoperative Behavior Scale	9 months–6 yr	Ordinal	None described	None	None
Cravero et al.,[6] 2000	Emergence Agitation Scale	6 months–10 yr	Ordinal	None described	None	None
Aono et al.,[15] 1999	Problematic Behavior Scale	3–6 yr, males	Ordinal	None described	None	None
Davis et al.,[21] 1999	Agitation Scale	1–5 yr	Ordinal	None described	None	None
Aono et al.,[16] 1997	Behavior Rating Scale	3–10 yr, males	Ordinal	None described	None	None
Johannesson et al.,[27] 1995	Behavior during Emergence Scale	1–7 yr	Ordinal	None described	None	None
Keegan et al.,[26] 1995	Excitement Score	1–15 yr	Ordinal	None described	Interrater reliability, ICC = 0.997	None
Davis et al.,[20] 1994	Quality of Anesthesia Recovery	1–6 yr	Ordinal	None described	None	None
Watcha et al.,[28] 1992	Agitation Scale	3 months–4 yr	Ordinal	None described	None	None
Heaman and Mattle,[29] 1982	Heaman-Mattle Excitement Scale	13–18 yr	Four-point adjectival scale	Literature, panel of experts, and experience of authors used to develop scale	"Reliability established through field testing"; no Reliability coefficient reported	Content validity assessed using a panel of experts; one construct validity test completed
Shields et al.,[30] 1980	Emergence Excitement Score Sheet	2–13 yr	Adjectival scale	None described	Interrater agreement, r = 0.82–0.84	None
Sheffer et al.,[9] 1973	Tranquility/Alertness Scale	3–11 yr	Semantic Differential Scale	None described	None	None
Smessaert et al.,[3] 1960	Mode of Recovery	10–65 yr	Ordinal	None described	None	None

ICC = intraclass correlation coefficient.

Canada), 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 determined the scale's reliability and validity.

Scale Development. First, ED was defined as a disturbance in a child's awareness of and attention to his or her environment with disorientation and perceptual alterations including hypersensitivity to stimuli and hyperactive 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*.[33–36] Second, the anesthesia, nursing, and psychiatric literature was reviewed, and interviews were conducted with pediatric anesthesiologists, PACU nurses, and a pediatric psychiatrist to collate behavioral descriptions of children thought to have ED or delirium.[37] From these behavioral descriptions, six categories of ED behaviors were derived: cognitive behavior, behavioral response to environmental 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 psychiatrist, to determine their content validity. These individuals 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 development process.[38]

The content validity evaluation was a two-step process for which specific instructions where given to each expert.[39] First, each expert was asked to rate the rele-

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 or 1), quite a bit (score or 2), very much (score of 3), and extremely (score of 4).[40] 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 scheduled to receive sevoflurane, isoflurane, or halothane for maintenance of anesthesia for an elective outpatient 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.[39] Items with a poor statistical profile were eliminated, and those with a good profile comprised the PAED scale.

Scale Evaluation. To determine interobserver reliability, the emergence behavior of 50 children was rated by a set of three observers using the PAED scale, 10 min after the child awakened and remained awake (did not fall back to sleep) postoperatively. Two of the three observers in each set were chosen at random. One of the authors (N. S.) was the third observer in all cases. All observers were blinded to the anesthetic agent administered during maintenance and were asked to refrain from discussing their evaluations with one another. A total of 37 observers participated, including 32 PACU nurses, 3 anesthesiologists, 1 paramedic, and the author (N. S.). To determine construct validity, five hypotheses were tested.[39]

Hypothesis 1. The PAED scale scores correlated negatively with the child's age.[4,6,7,41]

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.[4,9,11,18]

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.[42–47] Parents were telephoned on the second postoperative day to answer any questions regarding the questionnaires and to remind them to return the completed questionnaires. Questionnaires were returned to the investigator in self-addressed envelopes.

Hypothesis 5. The PAED scale scores in children who received sevoflurane were greater than in those who received halothane.[5–8,16,19,21,25,27,48] The choice of anesthetic administered was determined by the child's attending anesthesiologist.

ROC Curve Analysis. The sensitivity of the PAED scale was investigated using receiver operating characteristic (ROC) curve methodology.[49] A positive case of ED was defined as a child who received intravenous dimenhydrate postoperatively in the absence of vomiting to control an ED reaction. A negative case was defined as a child who did not receive dimenhydrate. Both morphine and dimenhydrate were used for their sedative effects to treat children with difficult emergence behavior. However, because it was unclear whether children who were given morphine were in pain, children who were treated with morphine were excluded from the ROC analysis.

Sample Size

Scale Development. It has been recommended that between 3 and 10 experts evaluate content validity.[50] The sample size for the item pretesting was based on enrolling five subjects for each item to be pretested.[51]

Scale Evaluation. The sample size for the interobserver reliability study was estimated using a Pearson product–moment correlation coefficient (r) of 0.75,[39] a half-width of the confidence interval (CI) of ± 0.1, and an α_2 of 0.05.[52] A sample size of 50 children was estimated.

The sample size for validity hypotheses 1–4 was based on the estimated maximum coefficient of $r = 0.86$ (or the $\sqrt{0.75}$).[39] A sample size of 47 children was estimated.[52]

The sample size required 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.[53]

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.[39,50] Statistical item analysis[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 α 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.[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 $H_A: \rho \neq 0$.[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 $P < 0.01$.[55] 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 (ρ) 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
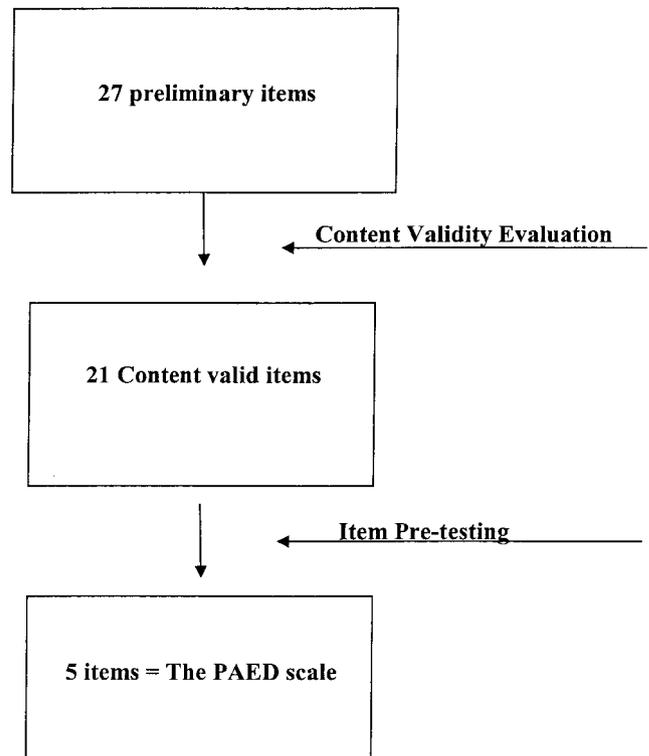

Fig. 1. Flow diagram outlining the steps taken to construct the Pediatric Anesthesia Emergence Delirium (PAED) scale, starting with a set of 27 preliminary scale items and ending with the final five items selected as the new scale.

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 $\pm$ 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 the caregiver's presence.
15. *The child is combative toward the caregiver who tries to comfort the child.*
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 ($r = -0.31$, $P < 0.08$) (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 admit-

Table 3. Surgical Procedures (n = 100)

Surgery	Proportion of Children, %	Duration of Surgery,* min
Ear, nose, throat	28	33.8 ± 24.2
Plastic surgery	9	66.6 ± 21.7
Dental procedures	52	83.5 ± 25.7
Ophthalmology	8	28.1 ± 11.9
Missing value	3	NA

* Data are presented as mean ± SD.
NA = not applicable.

Table 4. Anesthetic Agent (n = 100)

Anesthetic Agent	Proportion of Children, %	Duration of Surgery,* min
Sevoflurane	17	28.5 ± 10.4
Halothane	25	47.5 ± 34.4
Isoflurane	57	81.0 ± 25.6
Missing value	1	NA

* Data are presented as mean ± SD.

ted 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.[56]

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 dimenhydrate 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 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.
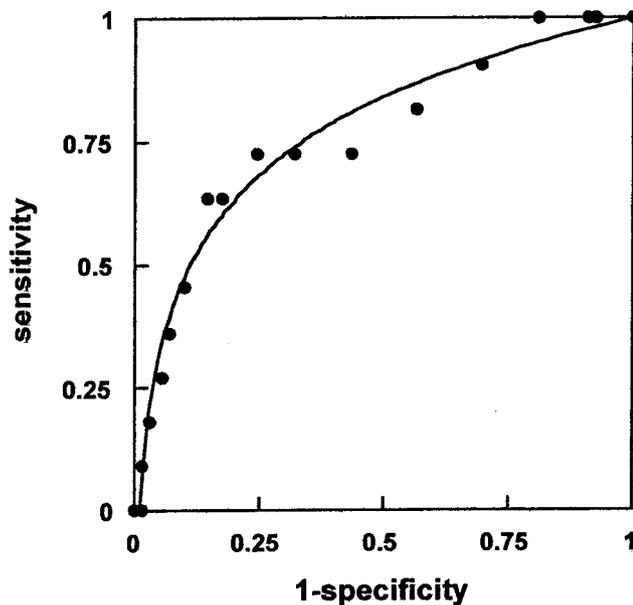
Fig. 2. Receiver operating characteristic (ROC) curve for the sensitivity (true-positive rate) and 1-specificity (false-positive rate) for scores on the Pediatric Anesthesia Emergence Delirium scale. A score of 10 or greater corresponds to a sensitivity of 0.64 and a 1-specificity of 0.14.

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.[36] A disturbance in consciousness includes a reduced awareness of the environment and impairment in the ability to focus, sustain, or shift attention.[36] 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.[39] 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.[39] 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.[39] 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.

References

1. Wilson TA, Graves SA: Pediatric considerations in a general postanesthesia care unit. J Post Anesth Nurs 1990; 5:16–24

2. Bastron RD, Moyers J: Emergence delirium. JAMA 1967; 200:883

3. Smessaert A, Schehr CA, Artusio JF: Observations in the immediate postanaesthesia period. Br J Anaesth 1960; 32:181–5

4. Eckenhoff JE, Kneale DH, Dripps RD: The incidence and etiology of postanesthetic excitement. ANESTHESIOLOGY 1961; 22:667–73

5. Galinkin JL, Fazi LM, Cuy RM, Chiavacci RM, Kurth CD, Shah UK, Jacobs IN, Watcha MF: Use of intranasal fentanyl in children undergoing myringotomy and tube placement during halothane and sevoflurane anesthesia. ANESTHESIOLOGY 2000; 93:1378–83

6. Cravero J, Surgenor S, Whalen K: Emergence agitation in paediatric patients after sevoflurane anaesthesia and no surgery: A comparison with halothane. Paediatr Anaesth 2000; 10:419–24

7. Beskow A, Westrin P: Sevoflurane causes more postoperative agitation in children than does halothane. Acta Anaesthesiol Scand 1999; 43:536–41

8. Hallen J, Rawal N, Gupta A: Postoperative recovery following outpatient pediatric myringotomy: A comparison between sevoflurane and halothane. J Clin Anesth 2001; 13:161–6

9. Sheffer L, Dean H, Steffenson J: Recovery room analgesia: A comparative study of drug effects. Anesth Analg 1973; 52:853–9

10. Cole J, Murray D, McAllister J, Hirshberg G: Emergence behavior in children: Defining the incidence of excitement and agitation following anaesthesia. Paediatr Anaesth 2002; 12:442–7

11. Voepel-Lewis T, Malviya S, Tait AR: A prospective cohort study of emergence agitation in the pediatric postanesthesia care unit. Anesth Analg 2003; 96:1625–30

12. Meyer-Pahoulis E, Williams SL, Davidson SI, McVey JR, Mazurek A: The pediatric patient in the post anesthesia care unit. Nurs Clin North Am 1993; 28:519–30

13. Haynes C: Emergence delirium: A literature review. Br J Theatre Nurs 1999; 9:502–3, 506–10

14. Uezono S, Goto T, Terui K, Ichinose F, Ishguro Y, Nakata Y, Morita S: Emergence agitation after sevoflurane versus propofol in pediatric patients. Anesth Analg 2000; 91:563–6

15. Aono J, Mamiya K, Manabe M: Preoperative anxiety is associated with a high incidence of problematic behavior on emergence after halothane anesthesia in boys. Acta Anaesthesiol Scand 1999; 43:542–4

16. Aono J, Ueda W, Mamiya K, Takimoto E, Manabe M: Greater incidence of delirium during recovery from sevoflurane anesthesia in preschool boys. ANESTHESIOLOGY 1997; 87:1298–300

17. Cohen IT, Finkel JC, Hannallah RS, Hummer KA, Patel KM: The effect of fentanyl on the emergence characteristics after desflurane or sevoflurane anesthesia in children. Anesth Analg 2002; 94:1178–81

18. Cohen IT, Hannallah RS, Hummer KA: The incidence of emergence agitation associated with desflurane anesthesia in children is reduced by fentanyl. Anesth Analg 2001; 93:88–91

19. Cravero J, Beach M, Dodge C, Whalen K: Emergence characteristics of sevoflurane compared to halothane in pediatric patients undergoing bilateral pressure equalization tube insertion. J Clin Anesth 2000; 12:397–401

20. Davis PJ, Cohen IT, McGowan FX, Latta K: Recovery characteristics of desflurane *versus* halothane for maintenance of anesthesia in pediatric ambulatory patients. ANESTHESIOLOGY 1994; 80:298–302

21. Davis PJ, Greenberg JA, Gendelman M, Fertal K: Recovery characteristics of sevoflurane and halothane in preschool-aged children undergoing bilateral myringotomy and pressure equalization tube insertion. Anesth Analg 1999; 88:34–8

22. Fan KT, Lee TH, Yu KL, Tang CS, Lu DV, Chen PY, Soo LY: Influences of tramadol on emergence characteristics from sevoflurane anesthesia in pediatric ambulatory surgery. Kaohsiung J Med Sci 2000; 16:255–60

23. Finkel JC, Cohen IT, Hannallah RS, Patel KM, Kim MS, Hummer KA, Choi SS, Pena M, Schreiber SB, Zalzal G: The effect of intranasal fentanyl on the emergence characteristics after sevoflurane anesthesia in children undergoing surgery for bilateral myringotomy tube placement. Anesth Analg 2001; 92:1164–8

24. Kulka PJ, Bressem M, Tryba M: Clonidine prevents sevoflurane-induced agitation in children. Anesth Analg 2001; 93:335–8

25. Viitanen H, Baer G, Annila P: Recovery characteristics of sevoflurane or halothane for day-case anaesthesia in children aged 1–3 years. Acta Anaesthesiol Scand 2000; 44:101–6

26. Keegan NJ, Yudkowitz FS, Bodian CA: Determination of the reliability of three scoring systems to evaluate children after general anesthesia. Anaesthesia 1995; 50:200–2

27. Johannesson GP, Floren M, Lindahl SG: Sevoflurane for ENT-surgery in children: A comparison with halothane. Acta Anaesthesiol Scand 1995; 39:546–50

28. Watcha MF, Ramirez-Ruiz M, White PF, Jones MB, Lagueruela RG, Terkonda RP: Perioperative effects of oral ketorolac and acetaminophen in children undergoing bilateral myringotomy. Can J Anaesth 1992; 39:649–54

29. Heaman DJ, Mattle LF: Adolescent emergence excitement. AORN J 1982; 35:230–42

30. Shields JR, Hovey JK, Fuller SS: A comparison of physostigmine and meperidine in treating emergence excitement. MCN Am J Matern Child Nurs 1980; 5:170–5

31. Sherwin T, Green S, Khan A, Chapman D, Dannenberg G: Does adjunctive midazolam reduce recovery agitation after ketamine sedation for pediatric procedures? A randomized, double-blind, placebo-controlled trial. Ann Emerg Med 2000; 35:229–38

32. Shrout PE, Fleiss JL: Intraclass correlations: Uses in assessing rater reliability. Psychol Bull 1979; 86:420–8

33. Przybylo HJ, Martini DR, Mazurek AJ, Bracey E, Johnsen L, Cote CJ: Assessing behaviour in children emerging from anaesthesia: Can we apply psychiatric diagnostic techniques? Paediatr Anaesth 2003; 13:609–16

34. Lipov EG: Emergence delirium in the PACU. Crit Care Nurs Clin North Am 1991; 3:145–9

35. Lindesay J: The concept of delirium. Dement Geriatr Cogn Disord 1999; 10:310–4

36. American Psychiatric Association: Diagnostic and Statistical Manual of Mental Disorders, 4th edition, text revision edition. Arlington, American Psychiatric Publishing, 2000

37. Streiner DL: A checklist for evaluating the usefulness of rating scales. Can J Psychiatry 1993; 38:140–8

38. Grant JS, Davis LL: Selection and use of content experts for instrument development. Res Nurs Health 1997; 20:269–74

39. Streiner DL, Norman GR: Health Measurement Scales: A Practical Guide to Their Development and Use, 2nd edition. New York, Oxford University Press, 1995, pp 15–27, 28–68, 104–27, 144–62

40. Conners CK: A teacher rating scale for use in drug studies with children. Am J Psychiatry 1969; 126:884–8

41. Meagher TF: The incidence of emergence excitement: A descriptive study. J Post Anesth Nurs 1988; 3:247–53

42. Thompson RH, Vernon DT: Research on children's behavior after hospitalization: A review and synthesis. J Dev Behav Pediatr 1993; 14:28–35

43. Kain Z: Postoperative maladaptive behavioral changes in children: Incidence, risk factors and interventions. Acta Anaesthesiol Scand 2000; 51:217–26

44. Holzki J: Halothane: the last word? (letter) Paediatr Anaesth 2001; 11: 249–52

45. Holzki J, Kretz FJ: Changing aspects of sevoflurane in paediatric anaesthesia: 1975–99 (editorial). Paediatr Anaesth 1999; 9:283–6

46. Foesel T, Reisch HJ: Postoperative behavioral changes in children: Comparison between halothane and sevoflurane. Paediatr Anaesth 2001; 11:719–23

47. Kain ZN, Mayes LC, Wang SM, Hofstadter MB: Postoperative behavioral outcomes in children: Effects of sedative premedication. ANESTHESIOLOGY 1999; 90:758–65

48. Lapin SL, Auden SM, Goldsmith LJ, Reynolds AM: Effects of sevoflurane anaesthesia on recovery in children: A comparison with halothane. Paediatr Anaesth 1999; 9:299–304

49. Vining DJ, Gladish GW: Receiver operating characteristic curves: A basic understanding. Radiographics 1992; 12:1147–54

50. Lynn MR: Determination and quantification of content validity. Nurs Res 1986; 35:382–5

51. Nunnally J: Introduction to Psychological Measurement. New York, McGraw-Hill, 1970, pp 212–9

52. Streiner DL: Sample-size formulae for parameter estimation. Percept Mot Skills 1994; 78:275–84

53. Cohen J: Statistical Power Analysis for the Behavioral Sciences, 2nd edition. Hillsdale, New Jersey, Lawrence Erlbaum, 1988, pp 19–61

54. Zar JH: Biostatistical Analysis, 2nd edition. New York, Prentice-Hall, 1984, pp 309–11

55. Nunnally J: Psychometric Theory. New York, McGraw-Hill, 1978, pp 138–9

56. Vernon DT, Schulman JL, Foley JM: Changes in children's behavior after hospitalization: Some dimensions of response and their correlates. Am J Dis Child 1966; 111:581–93

57. Hannallah RS, Broadman LM, Belman AB, Abramowitz MD, Epstein BS: Comparison of caudal and ilioinguinal/iliohypogastric nerve blocks for control of post-orchiopexy pain in pediatric ambulatory surgery. ANESTHESIOLOGY 1987; 66: 832–4

58. Merkel SI, Voepel-Lewis T, Shayevitz JR, Malviya S: The FLACC: A behavioral scale for scoring postoperative pain in young children. Pediatr Nurs 1997; 23:293–7

59. McGrath P, Johnson G, Goodman J, Schillinger J, Dunn J, Chapman J: CHEOPS: A behavioral scale for rating postoperative pain in children. Adv Pain Res Ther 1985; 9:395–402