Depth of sedation in children undergoing computed tomography: validity and reliability of the University of Michigan Sedation Scale (UMSS)

S. Malviya*, T. Voepel-Lewis, A. R. Tait, S. Merkel, K. Tremper and N. Naughton

Department of Anesthesiology, University of Michigan Health Systems, Ann Arbor, MI 48109, USA

**Corresponding author: Department of Anesthesiology, F3900/Box 0211, 1500 E. Medical Center Drive, Ann Arbor, MI 48109-0211, USA*

Background. Safe care of sedated children requires ongoing assessment of the depth of sedation to permit early recognition of progression to over-sedation. This study evaluated the validity and reliability of the University of Michigan Sedation Scale (UMSS) as a measure of sedation during procedures. The UMSS is a simple observational tool that assesses the level of alertness on a five-point scale ranging from 1 (wide awake) to 5 (unarousable with deep stimulation).

Methods. Thirty-two children aged 4 months to 5 yr (mean 1.5 yr), sedated for computed tomography (CT), were studied prospectively. The CT nurse assessed sedation using the UMSS before sedative administration and every 10 min thereafter. The child was videotaped during each assessment, and segments were edited and their order was randomized. Four nurses blinded to sedative administration viewed the segments and scored sedation using the UMSS. One of these nurses also scored sedation using a visual analogue scale (VAS) and another using the Observer's Assessment of Alertness/Sedation Scale (OAAS). To examine the test–retest reliability, 75 randomly selected video segments were viewed and scored on a second occasion.

Results. Changes in scores from baseline to discharge supported construct validity ($P<0.0001$). Criterion validity was demonstrated by significant correlations between the UMSS and the VAS and OAAS. There was good interobserver agreement between blinded observers' scores for each level of sedation and at discharge, and between blinded observers and the CT nurse for scores of 0 and 1 (lighter levels of sedation), but less agreement for scores 2 and 3 (deeper sedation) and discharge scores. Test–retest reliability was supported by agreement in the observers' UMSS scores.

Conclusion. The UMSS is a simple, valid and reliable tool that facilitates rapid and frequent assessment and documentation of depth of sedation in children.

Br J Anaesth 2002; **88**: 241–5

Keywords: monitoring; anaesthesia, paediatric; anaesthesia, depth

Accepted for publication: October 2, 2001

The increased use of sedation to facilitate diagnostic and therapeutic procedures in children has led to the provision of care by non-anaesthetists in many settings.[1–3] In response, a report from the Royal Colleges of Anaesthetists and Radiologists has emphasized careful titration of sedation depth to prevent excessively deep sedation and its associated risks by non-anaesthetists.[4] Furthermore, practice guidelines from the American Academy of Pediatrics (AAP) and the American Society of Anesthesiologists (ASA) have been developed to ensure consistent monitoring and care of sedated children regardless of who provides such care.[5 6] These guidelines stipulate that the level of consciousness and responsiveness should be assessed regularly and documented until the sedated patient satisfies specific discharge criteria. Frequent monitoring should permit early recognition of a child's progression into deep sedation and appropriate escalation of monitoring.

Accurate assessment of the depth of sedation requires a tool that is reliable and valid, yet is easy to use in the clinical setting. A variety of tools has been developed to assess the depth of sedation in both the research and the clinical setting.[7–15] Those scales that have been shown to be valid and reliable measures of sedation are somewhat lengthy[8–10] and, while useful as measures of sedation for research studies, they are cumbersome for frequent assessment of sedation in the clinical setting. The University of Michigan Sedation Scale (UMSS; see Appendix) was devised as a simple scale to facilitate the rapid assessment and documentation of the depth of sedation in all patients who receive a sedative agent for a diagnostic or therapeutic procedure. The purpose of this study was to evaluate the validity and reliability of the UMSS in a group of children undergoing sedation for a diagnostic procedure.

Methods

With approval from the Institutional Review Board at the University of Michigan and written consent from a parent or legal guardian, we studied healthy children (ASA I, II) aged 4 months to 5 yr who were scheduled to undergo sedation for an outpatient computed tomography (CT) procedure. Children who were severely medically compromised and those with moderate to severe cognitive impairment were excluded.

All children in this study received chloral hydrate 50–75 mg kg^{-1} orally, which is routine practice for outpatient CT at our institution. The paediatric CT nurse monitored all children throughout the procedure according to institutional sedation guidelines, which included pulse oximetry in all cases. The child's level of sedation was assessed and documented by the CT nurse before administration of the sedative agent and every 10 min after administration until the child was discharged, with the exception of the time during the CT scan, when the children were left undisturbed. The level of sedation was scored using the UMSS and proceeded with an assessment of the child's response to voice, light stimulation, such as stroking the face, and deeper physical stimulation, such as massaging the upper back or tickling the chest or axilla. A trained research assistant videotaped the child during each assessment. Children were discharged home when institutional discharge guidelines were met. Criteria for discharge included return to baseline vital signs and level of consciousness, ability to maintain a patent airway and the presence of protective airway reflexes (i.e. gag and swallow).

When all observations were complete, videotaped segments were coded and edited and their order was randomized. Four nurses, experienced in the care of sedated children and postanaesthesia recovery, independently viewed the videotapes and scored the level of sedation using the UMSS. To determine test–retest reliability, 75 randomly selected segments were viewed and scored on a second occasion, several weeks after the first session of observations. Additionally, one of the nurses observed all segments and assigned sedation scores using a 0–10 cm visual analogue scale (VAS; 0=awake and alert; 10=unarousable) that globally rates sedation, and another using the Observer's Assessment of Alertness/Sedation Scale (OAAS),[8] a sedation scale previously validated in children. Each of these nurse observers was blinded to the administration of sedative agents.

Statistics

All analyses were performed using SPSS software (SPSS, Chicago, IL, USA). Construct validity refers to the degree to which a tool actually measures the characteristic that it is intended to measure. In this case, scores were expected to increase after sedative administration and decrease over time before discharge. To evaluate the construct validity of the UMSS, the Wilcoxon signed rank test was used to evaluate the change in score over time after sedative administration. Criterion validity is the degree to which the measurement agrees with other approaches for measuring the same characteristic. To determine criterion validity, scores from the UMSS were compared with those from the VAS and OAAS. Because the metrics were different between the UMSS, VAS and OAAS, *z*-transformations of the scores were performed and comparisons were made using Spearman's rank order correlation coefficients (Spearman's ρ). Correlation coefficients and kappa (κ) statistics for levels of agreement were used to determine interobserver and test–retest reliability of the UMSS scores. The κ statistic is a measure of agreement that allows for observer variability for categorical data and corrects for chance levels of agreement.[16] Kappa values of 0.4 or greater were considered to provide acceptable agreement. Significance was accepted at the 5% level (i.e. $P<0.05$). Data are presented as mean (SD) where applicable.

Results

Thirty-two children, 4 months to 5 yr of age (mean 1.5 yr), were enrolled in this study. Validity and reliability analyses were based on 164 observations made throughout the care of these children during their sedation experiences. The mean time from sedative administration to the start of the procedure was 31 (SD 10) min (range 15–55). The mean duration of CT scan was 18 (10) min. The time to discharge varied: the mean time from administration of sedation to discharge was 71 (12) min (range 40–100) and from the end of the procedure to discharge it was 22 (7) min (range 10–35).

Validity

Construct validity was supported by the significant increase in sedation scores assigned by the blinded observers from baseline to 10, 20, 30 and 40 min ($P<0.0001$) and a significant decrease in scores from immediately after the

procedure to 10, 20 and 30 min thereafter ($P<0.0001$) (Table 1). Criterion validity was supported by the excellent correlations between the UMSS and VAS scores ($r=0.955$, $n=164$, $P<0.0001$) and between the UMSS and OAAS scores ($r=0.929$, $n=164$, $P<0.0001$).

Reliability

Interobserver agreement data for each level of sedation defined on the UMSS and for discharge scores are presented in Tables 2 and 3. Agreement between the blinded observers was good for scores at each level of sedation and at discharge. Additionally, there was excellent agreement between the CT nurse and blinded observers for sedation scores of 0 and 1, but less agreement for scores of 2 and 3, and poor agreement at discharge. There was a significant correlation between UMSS scores of the blinded observers and the CT nurse (Table 4). However, the CT nurses' scores tended to be lower than the blinded observers' average scores 30 and 40 min from baseline and immediately after the scan to discharge ($P=0.02$ immediately after scan; not significant for all other assessment times) (Table 1). Test–retest reliability of the UMSS scores at each level of sedation was supported by the excellent agreement in the observers' scores between the first and second observations (range 67–100% agreement; $\kappa=0.687$).

Table 1 Sedation scores at 10 min intervals before and after CT scanning. Data are presented as mean (SD) (median; interquartile range). *$P=0.02$ compared with CT nurse

	CT nurse	Blinded nurses
Baseline	0 (0;0)	0.15 (0.36) (0; 0.0)
10 min	0.9 (0.88) (1; 0.0)	1.01 (1.1) (1; 0.0)
20 min	2 (0.98) (2; 1.0)	1.97 (1.3) (2; 1.0)
30 min	2.27 (0.88) (3; 1.0)	2.6 (1.1) (3; 2.0)
40 min	2.6 (0.79) (3; 2.0)	3.1 (0.72) (3; 3.0)
Post Scan	1.97 (0.93) (2; 1.0)	2.6 (0.98) (3; 2.0)*
10 min	1.3 (0.72) (1; 1.0)	1.5 (0.98) (1; 1.0)
20 min	1.1 (0.76) (1; 0.5)	1.6 (1.03) (1; 1.0)
30 min	1 (1; 1.0)	1.75 (0.5) (2; 1.25)
Discharge	1.14 (0.76) (1; 1.0)	1.6 (1.07) (1; 1.0)

Discussion

One of the most important elements of monitoring children who are sedated for a diagnostic procedure is frequent assessment of the depth of sedation.[5 6 17] Children may move rapidly from lighter levels of sedation to deep sedation, so that escalation of monitoring and a greater degree of vigilance become necessary.[5] The practical aspects of repeated assessment of sedation require a tool that accurately detects changes in level of sedation and is reliable, yet is simple to administer and document without interfering with the procedure. The UMSS was devised as such a tool, and our study demonstrates that this scale has a high degree of validity in detecting changes in depth of sedation, as well as good interobserver reliability between blinded observers.

Several sedation scales have been devised and studied for use in the research setting and the intensive care unit, and for sedation during procedures.[8 18 12] The OAAS was designed to measure changes in level of alertness during procedures. However, its ability to capture the deep levels of sedation that are frequently encountered in the paediatric setting is somewhat limited.[8] The Neurobehavioral Assessment Scale and the Vancouver Sedative Recovery Scale (VSRS), on the other hand, provide more discriminating measures of the depth of sedation, with emphasis on the extreme ends of the scale.[9 18] Although validity has been established for each of these tools, their length and/or content preclude repeated use during procedures. Indeed, Macnab and colleagues have reported an average assessment time of 4 min for scoring the VSRS.[18] The UMSS was devised as a simple, efficient tool to assess the depth of sedation over the entire continuum from awake to

Table 2 Interobserver agreement for UMSS scores (total observations=164). *n* represents the range in number of observations scored at that level of sedation. Data are range/median

UMSS score	Agreement between each blinded observer and the CT nurse (four comparisons)	Agreement between the four blinded observers (six comparisons)
0=awake/alert ($n=38$–45)		
Agreement	68–82/80%	87–90/91%
κ statistic	0.642–0.797/0.719	0.78–0.872/0.842
1=sleepy/responds appropriately ($n=46$–58)		
Agreement	52–69/60%	72–87/78%
κ statistic	0.384–0.543/0.496	0.557–0.734/0.677
2=somnolent/arouses to light stimuli ($n=13$–34)		
Agreement	30–53/48%	38–87/59%
κ statistic	0.341–0.391/0.347	0.229–0.73/0.483
3=deep sleep/arouses to deeper physical stimuli ($n=23$–31)		
Agreement	16–68/54%	48–78/55%
κ statistic	0.029–0.587/0.435	0.283–0.778/0.510
4=unarousable to stimuli ($n=0$–30)		
Agreement	0	27–100/47%
κ statistic		0.373–0.588/0.465

Table 3 Interobserver agreement for UMSS scores at discharge (total observations=32). *n* represents the range in number of observations scored at that level of sedation. *One child was scored as UMSS 4 by all blinded observers. Data are range/median

UMSS score	Agreement between each observer and the CT nurse (four comparisons)	Agreement between the four blinded observers (six comparisons)
0=awake/alert (*n*=3–5)		
Agreement	60–80/60%	75–100/100%
κ statistic	0.604–0.868/0.658	0.708–1.0/0.837
1=sleepy/responds appropriately (*n*=10–15)		
Agreement	47–60/54%	69–90/79%
κ statistic	0.23–0.364/0.260	0.49–0.702/0.662
2=somnolent/arouses to light stimuli (*n*=2–8)		
Agreement	14–43/35%	25–86/50%
κ statistic	0.00–0.238/0.154	0.097–0.727/0.344
3 Deep sleep/arouses to deeper physical stimuli (*n*=1–8)		
Agreement	0	38–100/59%
κ statistic		0.507–0.868/0.589
4=unarousable to stimuli (*n**=1)		
Agreement	0	100%
κ statistic		1.0

Table 4 Spearman's correlation coefficients between the UMSS scores of the observers. All correlations were significant ($P<0.001$; two-tailed)

	CT nurse	Observer 1	Observer 2	Observer 3	Observer 4
CT nurse	1.00	0.823	0.862	0.779	0.841
Observer 1	0.823	1.00	0.894	0.898	0.869
Observer 2	0.862	0.894	1.00	0.880	0.928
Observer 3	0.779	0.898	0.880	1.00	0.858
Observer 4	0.841	0.869	0.928	0.858	1.00

unarousable, as discussed and defined in the AAP and ASA guidelines.[5] [6] This tool was intended to provide a standardized method for assessing sedation, primarily for areas where non-anaesthesia providers are caring for sedated patients.

There are several tests to determine the validity of an instrument, the most important of which, perhaps, is the measurement of construct validity. The construct validity of the UMSS as a measure of sedation was supported by the increase in sedation score after the administration of the sedative agent and the subsequent decrease in score as the sedative effects wore off. This measure of validity was strengthened by the fact that the independent observers were blinded to the timing of each videotaped segment relative to drug administration, thereby reducing the bias that may otherwise influence the sedation score. Additionally, the high degree of correlation between the UMSS scores and those of the VAS and OAAS demonstrates the criterion validity of the UMSS.

Our study found good agreement between the blinded observers' sedation scores as well as a high degree of test–retest reliability. Conversely, we found less agreement between the blinded observers' scores and those assigned by the CT nurse. The CT nurses' scores tended to be lower than those of the blinded observers from 30 min after drug administration to discharge. Although video recordings captured the nursing assessment and appropriate stimulation of the child, it is important to note that the blinded observers based their scores on these brief taped segments only, whereas the CT nurses may have incorporated other contextual indicators of the child's sedation course into their assessments. On the other hand, these findings may reflect a clinical bias towards underestimating the depth of sedation, perhaps to avoid escalation of monitoring or to expedite the discharge of children after the procedure. The emphasis on rapid patient turnover in busy radiology settings may, in some cases, influence patient assessment such that some children may be discharged prematurely. Indeed, Coté has previously alluded to, and cautioned against, a similar clinical bias towards liberalizing the definition of conscious sedation for the convenience of the practitioner.[19] These findings suggest that further education of non-anaesthesia caregivers may be necessary to ensure accurate and objective assessment of the depth of sedation. Additionally, in busy settings that require rapid patient turnover, the availability of a centralized recovery area, such as a postanaesthesia care area, may facilitate appropriate discharge criteria for all patients.

It has been recommended that sedative drugs administered by non-anaesthetists should have a wide margin of safety so that unintended loss of consciousness becomes unlikely.[4] All children in our sample received chloral hydrate in moderate doses because of its purported margin of safety. While the intent in each case was to produce moderate sedation, a significant number of children were observed to be deeply sedated, and even unarousable at one time-point or more during the sedation episode. All children were monitored with pulse oximetry according to institu-

tional standards, and none required escalation of care during the study period. As we did not follow these children beyond discharge, it is unknown whether any child experienced an adverse event at home. These data highlight the importance of careful monitoring of all children, regardless of the sedative agent used or its route of administration.

While our findings suggest good overall validity and reliability of the UMSS, there was some variability between observers at moderate to deep levels of sedation. Such variability may limit the usefulness of the UMSS in research settings that require a high degree of precision, such as in rigorous pharmacological studies that compare the potency of sedative regimes. However, our data support the ability of the UMSS to capture changes in the child's depth of sedation, and can therefore provide a useful tool in busy clinical settings that require frequent and rapid sedation assessment.

Ongoing assessment of the depth of sedation is important for early identification of the patient's progression into deep sedation and the potential loss of protective reflexes. The tool used to assess sedation must be able to capture this progression sufficiently. This study demonstrates that the UMSS is a valid measure of depth of sedation and has good interobserver and test–retest reliability. The tool has good clinical utility in that scores can be readily assigned when frequent assessment is indicated. Although the AAP and ASA guidelines require ongoing assessment throughout the procedure, it is impractical to do this just before and during the diagnostic procedure itself, as Hatch and Sury have suggested.[20] During this time, when sedation may be deepest, continuous pulse oximetry, assessment of vital signs and, in some cases, capnography may facilitate the safe monitoring of the child.

This study has evaluated the validity and reliability of the UMSS as a measure of sedation in young children undergoing non-painful procedures. Extrapolation of these data to other settings, including painful procedures and other populations, requires further investigation.

Appendix

University of Michigan Sedation Scale (UMSS)

0	Awake and alert
1	Minimally sedated: tired/sleepy, appropriate response to verbal conversation and/or sound
2	Moderately sedated: somnolent/sleeping, easily aroused with light tactile stimulation or a simple verbal command
3	Deeply sedated: deep sleep, arousable only with significant physical stimulation
4	Unarousable

Acknowledgements

The authors would like to thank Constance Burke, RN, Sue M. Hadden, RN, Sue Quinn, RN, Monica Siewert, BA, and Philip Rubin for their contributions to this project.

References

1 Cook BA, Bass JW, Nomizu S, Alexander ME. Sedation of children for technical procedures: current standard of practice. *Clin Pediatr (Philadelphia)* 1992; **31**: 137–42

2 Malviya S, Voepel-Lewis T, Tait AR. Adverse events and risk factors associated with the sedation of children by nonanesthesiologists. *Anesth Analg* 1997; **85**: 1207–13

3 Sury MR, Hatch DJ, Deeley T, Dicks-Mireaux C, Chong WK. Development of a nurse-led sedation service for paediatric magnetic resonance imaging. *Lancet* 1999; **353**: 1667–71

4 Royal College of Anaesthesists and Royal College of Radiologists. Sedation and anaesthesia in radiology: report of a joint working party. London: The Royal College of Anaesthetists, 1992

5 American Academy of Pediatrics Committee on Drugs. Guidelines for monitoring and management of pediatric patients during and after sedation for diagnostic and therapeutic procedures. *Pediatrics* 1992; **89**: 1110–5

6 Practice guidelines for sedation and analgesia by non-anesthesiologists. A report by the American Society of Anesthesiologists Task Force on Sedation and Analgesia by Non-Anesthesiologists. *Anesthesiology* 1996; **84**: 459–71

7 Ramsay MA, Savege TM, Simpson BR, Goodwin R. Controlled sedation with alphaxalone–alphadolone. *Br Med J* 1974; **2**: 656–9

8 Chernik DA, Gillings D, Laine H, *et al*. Validity and reliability of the Observer's Assessment of Alertness/Sedation Scale: study with intravenous midazolam. *J Clin Psychopharmacol* 1990; **10**: 244–51

9 Chernik DA, Tucker M, Gigli B, *et al*. Validity and reliability of the Neurobehavioral Assessment Scale. *J Clin Psychopharmacol* 1992; **12**: 43–8

10 Macnab AJ, Levine M, Glick N, Phillips N, Susak L, Elliott M. The Vancouver sedative recovery scale for children: validation and reliability of scoring based on videotaped instruction. *Can J Anaesth* 1994; **41**: 913–8

11 Malviya S, Voepel-Lewis T, Huntington J, Siewert M, Green W. Effects of anesthetic technique on side effects associated with fentanyl Oralet premedication. *J Clin Anesth* 1997; **9**: 374–8

12 Marx CM, Smith PG, Lowrie LH, *et al*. Optimal sedation of mechanically ventilated pediatric critical care patients. *Crit Care Med* 1994; **22**: 163–70

13 Riker RR, Fraser GL, Cox PM. Continuous infusion of haloperidol controls agitation in critically ill patients. *Crit Care Med* 1994; **22**: 433–40

14 Samra SK, Bradshaw EG, Pandit SK, Papanicolaou AC, Moore BD. The relation between lorazepam-induced auditory amnesia and auditory evoked potentials. *Anesth Analg* 1988; **67**: 526–33

15 Barker RA, Nisbet HI. The objective measurement of sedation in children: a modified scoring system. *Can Anaesth Soc J* 1973; **20**: 599–606

16 Fleiss JL. *Statistical Methods for Rates and Proportions*. New York: John Wiley and Sons, 1981

17 American College of Emergency Physicians. Clinical policy for procedural sedation and analgesia in the emergency department. *Ann Emerg Med* 1998; **31**: 663–77

18 Macnab AJ, Levine M, Glick N, Susak L, Baker-Brown G. A research tool for measurement of recovery from sedation: the Vancouver Sedative Recovery Scale. *J Pediatr Surg* 1991; **26**: 1263–7

19 Cote CJ. Monitoring guidelines: do they make a difference? *Am J Roentgenol* 1995; **165**: 910–2

20 Hatch DJ, Sury MR. Sedation of children by non-anaesthetists. *Br J Anaesth* 2000; **85**: 660–1