# CHAPTER 7
## Palliative Sedation: reliability and validity of sedation scales

Jimmy J. Arevalo
Tijn Brinkkemper
Agnes van der Heide
Judith A. Rietjens
Miel Ribbe
Luc Deliens
Stephan A. Loer
Wouter W.A. Zuurmond
Roberto S.G.M. Perez

*Published in the Journal of Pain and Symptom Management 2012;44(5):704-14.*

## ABSTRACT

*Context* Observer-based sedation scales have been used to provide a measurable estimate of the comfort of non-alert patients in palliative sedation. However, their usefulness and appropriateness in this setting has not been demonstrated.

*Objectives* To study the reliability and validity of observer-based sedation scales in palliative sedation.

*Methods* A prospective evaluation of 54 patients under intermittent or continuous sedation with four sedation scales was performed by 52 nurses. Included scales were the Minnesota Sedation Assessment Tool (MSAT), the Richmond Agitation-Sedation Scale (RASS), the Vancouver Interaction and Calmness Scale (VICS) and a sedation score proposed in the Guideline for Palliative Sedation of the Royal Dutch Medical Association (KNMG). Inter-rater reliability was tested with the intraclass correlation coefficient (ICC) and Cohen's Kappa coefficient. Correlations between the scales using Spearmans' rho tested concurrent validity. We also examined construct, discriminative and evaluative validity. In addition, nurses completed a user-friendliness survey.

*Results* Overall moderate to high inter-rater reliability was found for the VICS interaction subscale (ICC=0.85), RASS (ICC=0.73) and KNMG (ICC=0.71). The largest correlation between scales was found for the RASS and KNMG (rho=0.836). All scales showed discriminative and evaluative validity, except for the MSAT motor subscale and VICS calmness subscale. Finally, the RASS was less time consuming, clearer, and easier to use than the MSAT and VICS.

*Conclusion* The RASS and KNMG scales stand as the most reliable and valid among the evaluated scales. In addition, the RASS was less time consuming, and clearer and easier to use than the MSAT and VICS. Further research is needed to evaluate the impact of the scales on better symptom control and patient comfort.

## 7.1 Introduction

Near the end of life, patients who remain alert may present with physical and psychological symptoms despite adequate palliative care. In some of these cases, relief from suffering may be achieved only if alertness is lowered by means of palliative sedation. However, the treatment of prolonged suffering with sedatives is only acceptable after all other reasonable options have been considered. Moreover, patients receiving sedatives may recover alertness and present with symptoms if insufficient medication is provided. Alternatively, excessive sedatives may cause respiratory depression and lead to an accelerated death.[1] For palliative sedation to be proportional, monitoring the effects of sedatives would seem desirable.

Nonetheless, the "how" of monitoring palliative sedation is currently an open question.[2] The purpose of palliative sedation, whether continuous or intermittent, is to provide comfort to patients with unbearable suffering.[3,4] As it is not possible to directly measure the comfort of non-alert patients, observer-based sedation scales have been used to provide a measurable estimate in palliative sedation.[5] Examples include the Glasgow Coma Scale,[2,6,7] the Ramsay Sedation Scale,[4,8] the Visual Analogue Scale[4,9] and the Richmond Agitation-Sedation Scale.[4] The clinimetric properties of several of these scales have been evaluated in the intensive care unit (ICU) setting. However, their usefulness and appropriateness in palliative sedation has not been demonstrated. One study described the use of a scale in the broader spectrum of palliative care patients (sedated and non-sedated).[6]

The usefulness of measurement scales in health care is dependent on their reliability, that is, how reproducible their results are under different conditions. In addition, scales must be valid, that is, measure what they are meant to measure.[10] To our knowledge, no studies have described how sedation scales perform in terms of reliability and validity in the setting of palliative sedation. Therefore, we studied the performance of four observer-based sedation scales as a first step to address this issue.

The observer-based sedation scales developed for the ICU setting include a group that use a single number to describe distinct behaviors and a group that comprises different subscales for the separate reporting of domains such as the level of consciousness, agitation, interaction and calmness. We included two scales from the latter group – the Minnesota Sedation Assessment Tool (MSAT) and the Vancouver Interaction and Calmness Scale (VICS) – and two single-number scales: the Richmond Agitation-Sedation Scale (RASS) and a sedation score proposed in the Guideline for Palliative Sedation of the Royal Dutch Medical Association (KNMG).[5,11-13] Although different scales have been proposed for assessment of sedation, we arbitrarily chose to evaluate scales that were already validated in the most controlled environment, in this case the ICU setting, and compare these to a scale proposed for palliative sedation. Because palliative sedation is performed in a variety of settings (including the ICU), a secondary reason for choosing the observer-based scales already in use in the ICU setting was to facilitate acceptance of these instruments in case reliability and validity of these scales could be confirmed in a palliative care setting.

## 7.2 Methods

*7.2.1 Scales*

Four observer-based scales were used to monitor palliative sedation. Three of the sedation scales we tested (MSAT, VICS, RASS) were originally designed to assess sedation of ICU patients, and one (KNMG) was proposed in the Dutch Guideline for Palliative Sedation.[5,11-13] No validated translations into the Dutch language were available for the MSAT, VICS and RASS. However, in one study the RASS scale was translated.[14] The English versions of the MSAT and VICS scales were translated into Dutch by a resident in anesthesiology and a member of our research group (R.S.G.M.P.). Then, a native English speaker retranslated the scales into English and the inconsistencies were corrected in the Dutch versions. The final translated versions were close to the original English texts.

Both the MSAT and the VICS are scales comprising two subscales to assess sedative effects; a total score cannot be computed. The MSAT has arousal ($MSAT_{(a)}$) and motor ($MSAT_{(m)}$) subscales, and the VICS has interaction ($VICS_{(i)}$) and calmness ($VICS_{(c)}$) subscales. Moreover, the MSAT has a third subscale about the quality of sedation ($MSAT_{(q)}$). For the purpose of statistical testing, these subscales were considered independent scales. In contrast, the RASS and the KNMG have no subscales. The terms used to describe the different theoretical backgrounds can be found in the original papers.[5,11-13] For the purpose of this study, we assumed that they are all assessing the level of consciousness. As described in the scales' guidelines, nurses performed physical or auditory stimulation to the patient during the assessments when required. Two investigators pilot-tested the feasibility and practical use of all scales in five patients.

*7.2.2 Patient Population*

Participating Institutions

This study was carried out in three palliative care institutions in the Amsterdam region of The Netherlands: Hospice Kuria, Hospice Alkmaar and Nursing Home Heemstede.

Sample

Eligible cases included all patients who received intermittent or continuous palliative sedation in a participating institution during the study period. The indication for palliative sedation was determined in all cases by the attending physician in accordance with local and national guidelines for sedation. In general, the indication for continuous sedation includes that the patient presents symptoms that do not respond to any other reasonable treatment and that death is expected in the near future (within one to two weeks). In cases where death is not expected in the near future but symptoms do not respond to other reasonable treatments, intermittent sedation is an option.

*7.2.3 Data Collection*

Patient Characteristics and Decision-making About Palliative Sedation

Data were collected between December 2007 and December 2010. General patient characteristics (age, gender, weight, diagnosis) were recorded by a health care provider. Information on the decision-making about palliative sedation, such as the initiative to start sedation, the type of sedation (intermittent or continuous), indication for and intended

depth of sedation (somnolent, soporific, sub-comatose, or comatose), were recorded by the attending physician.

Assessment of Sedation

The scales used to assess sedation were completed by nursing staff during their regular work schedule. Patients could be assessed at several time points (with a maximum of five times per day), for example, before the start of sedation, within one hour of the start of sedation, and subsequently during the course of sedation.

The protocol required an independent secondary assessment be performed by a nurse within 60 minutes at each new time point (paired assessments). The 60-minute limit for the secondary assessment was based on the hypothesis that minimal changes in the clinical sedation depth would occur within such a time frame. However, the time of each assessment was recorded to allow post-hoc analyses. In addition, nurses also stated their training level (qualified nurse, nurse in training) and completed a short survey about three symptom-specific scores (pain, delirium, respiratory distress) and a questionnaire about the user-friendliness of the scales (except for the KNMG).

### *7.2.4 Ethical Considerations*

The research protocol was approved by the Ethics Committee of VU University Medical Center. Written informed consent was obtained from either the patient or a family member. Codes were assigned to clinical records to protect the identity of patients.

### *7.2.5 Data Analyses*

The performance of the scales was analyzed in terms of reliability and validity. Descriptive statistics were calculated for the characteristics of patients and nurses. To test statistical differences, parametric and non-parametric tests were performed as appropriate and appear together with *P*-values. A $P < 0.05$ was considered statistically significant. For the analysis an SPSS PASW 17.0.2 for Windows (SPSS Inc., Chicago, IL) dataset was created.

Reliability

Paired assessments were used to evaluate the inter-rater reliability of the scales. All scales assess sedation in an ordinal order (except the MSAT quality subscale). Therefore, two forms of reliability coefficients were calculated. The intraclass correlation coefficient (ICC) was calculated for ordinal scales ($MSAT_{(a)}$, $MSAT_{(m)}$, $VICS_{(i)}$, $VICS_{(c)}$, RASS and KNMG) and Cohen's Kappa coefficient for the nominal scale, $MSAT_{(q)}$. The model for calculating the ICC accounted for the random effect of the different raters. The index type was set to address consistency rather than absolute agreement because the time within paired assessments varied between time points.

Validity

There is no gold standard to assess the level of consciousness. Therefore, a comparison between the different scales was done to test validity. For these analyses, only ordinal scales were considered and continuity of scores was assumed. To describe the different validity tests, the traditional terms are used (namely, content validity, internal consistency, criterion validity, construct validity).[10] The main hypothesis was that if all the scales assess the level of consciousness, their scores should be correlated.

Moreover, the purpose of the validity analyses was to compare the different scales rather than to address unique scale characteristics. Therefore, content validity and internal consistency of the scales were not evaluated.

Spearman's correlation coefficients between pairs of scales were calculated to assess one type of criterion validity (concurrent validity). We expected larger correlations between the pairs of scales where the common criterion measure (consciousness level) is readily assessed. For construct validity, we tested the hypothesis that a sedated patient should score high values on the KNMG and $VICS_{(c)}$ and low values on all other scales. No formal statistical test exists to determine construct validity. However, according to the design of the scales, they should correlate negatively. Moreover, the sample included two different populations of patients receiving palliative sedation (namely, intermittent and continuous). The dosages of sedatives vary (increase and/or decrease) in intermittent sedation whereas in continuous sedation, sedative dosages always increase. Therefore, patients under continuous sedation should have mean sedation scores describing a lower consciousness level when compared with those patients receiving intermittent sedation. Statistical differences in sedation scores between these groups would test if the scales have discriminative validity. In addition, differences were tested between the first assessment after the start of sedation and the last assessment recorded for the patient to determine evaluative validity. A significant difference would point out that the scale can be used to evaluate a patient during the course of sedation by recording changes in the level of consciousness.

## 7.3 Results

### *7.3.1 Characteristics of nurses and patients*

Fifty-two nurses performed the assessments of the level of consciousness at 254 time points in 54 patients under intermittent or continuous sedation. In addition, thirty-nine nurses performed 127 second assessments for inter-rater reliability tests in 43 patients. In general, nursing staff comprised qualified nurses (82%) and no statistical difference was found between the training level of the nurses that performed paired assessments (Sign Test, $P$=0.093).

Of the 54 patients in the study, 25 (46%) received intermittent sedation and 29 (54%) received continuous sedation (**table 1**). There was an overrepresentation of male patients in the group receiving continuous sedation, and of female patients in the group receiving intermittent sedation ($Chi^2$, $P$=0.012). Furthermore, most patients had been diagnosed with cancer (e.g., lung cancer), others had non-cancer diagnoses (e.g., heart failure, chronic obstructive pulmonary disease) or a combination of both cancer and non-cancer diagnoses.

### *7.3.2 Decision-making About Palliative Sedation*

According to the health care providers, the initiative to consider the option of palliative sedation was taken together by a combination of patient, physician, family and nurses in 25 (46%) cases (**table 1**). Patients themselves had considered the option of palliative sedation in 18 (33%) cases. The indication for sedation included symptoms like pain, delirium and fatigue and did not differ between intermittent and continuous sedation, except for delirium, which was more common in continuous sedation ($Chi^2$, $P$=0.01). The intended

depth of sedation for intermittent sedation was somnolent or soporific in 21 (76%) cases, whereas for continuous sedation, the intended depth was sub-comatose or comatose in 22 (76%) cases ($Chi^2$, $P<0.001$).

The median duration of sedation was two days, ranging from one to 13 days. The duration of sedation did not differ significantly between intermittent and continuous sedation (Mann-Whitney U, $P=0.179$). The sedative drug used during palliative sedation was midazolam for 52 (96%) cases, levopromazine for one (2%) case, and a combination of both for one (2%) case. Other medications used included an opioid (i.e., morphine, fentanyl) for 34 (63%) patients and dexamethasone for six (11%) patients. Thirteen (24%) patients received only the sedative drug.

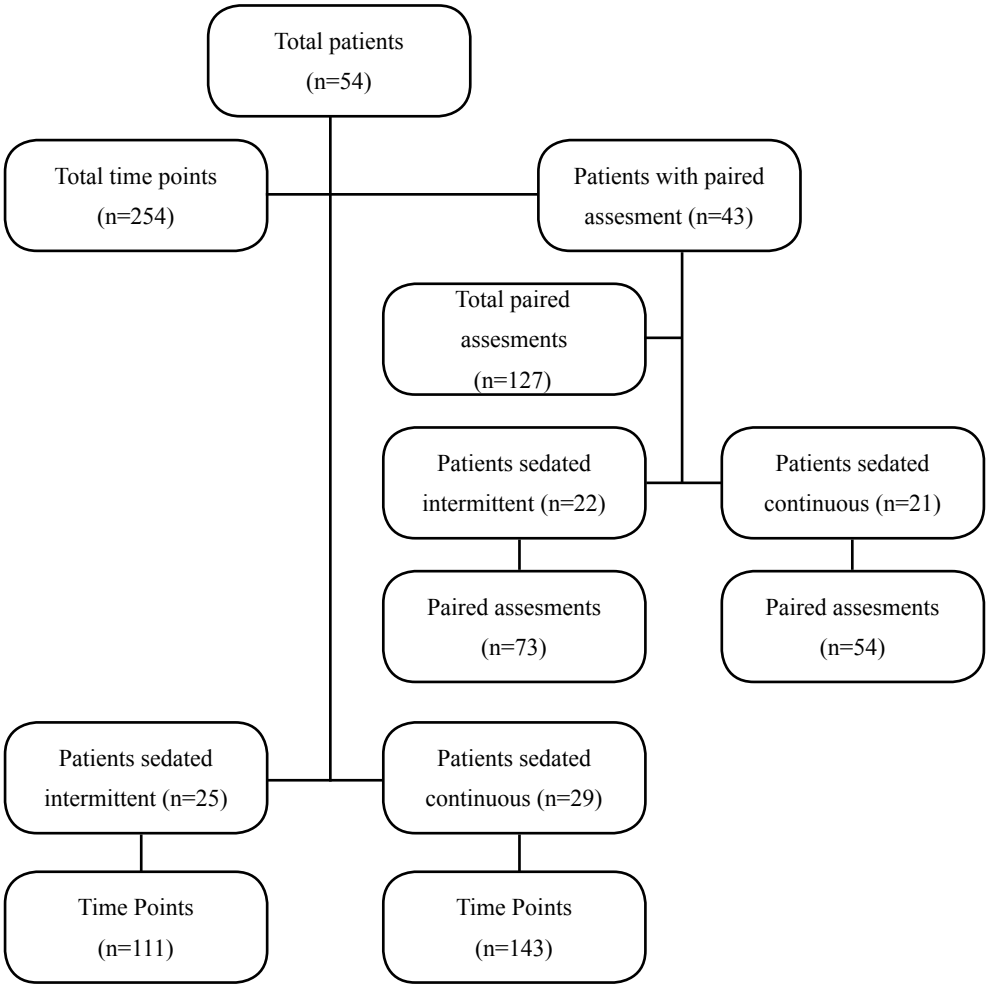
### *7.3.3 Time Points and Paired Assessments*

Figure 1 shows a flow diagram with the number of time points and paired assessments. The number of time points were 111 (44%) for intermittent sedation and 143 (56%) for continuous sedation. Moreover, an independent secondary assessment was performed in 43 (80%) patients, of whom 22 (51%) received intermittent sedation and 21 (49%) received continuous sedation.

In total, 127 paired assessments were performed. However, different numbers of paired assessments were available for the individual scales and subscales (**table 2**). In general, both raters filled in most of the variables. The $MSAT_{(a)}$, $MSAT_{(m)}$, KNMG and RASS had less than 10% missing values for paired assessments. Individual items of the VICS' subscales (i.e., patient needs encouragement to respond to questions, patient is pulling on lines and tubes) account for most missing values for this scale.

The median time within paired assessments was 15 minutes, ranging from 0 to 390 minutes. The number of valid paired assessments within 60 minutes was 111, and within 15 minutes, 65. Thus, 16 (13%) secondary assessments were performed outside the window of 60 minutes. The time within paired assessments did not differ significantly between intermittent and continuous sedation (Mann-Whitney U, $P= 0.384$).

Figure 1. Flow diagram regarding time points and paired




Intermittent = intermittent palliative sedation

Continuous = continuous palliative sedation

Time points=number of assessments

Paired assessments= number of available second assessments

Table 1. Characteristics of Patients and Decision-making About Palliative Sedation

| | *Purpose of Sedation* | | |
|---|---|---|---|
| | **Intermittent** ***n*=25 (46%)** | **Continuous** ***n*=29 (54%)** | **All N=54 (100%)** |
| Age (years) | | | |
| - Mean (SD) | 74 (12) | 72 (12) | 73 (12) |
| - Range (median) | 46-92 (78) | 50-87 (77) | 46-92 (77) |
| Gender * | | | |
| - Male | 4 (16) | 14 (48) | 18 (33) |
| - Female | 21 (84) | 15 (52) | 36 (67) |
| Diagnosis | | | |
| - Cancer | 22 (88) | 22 (76) | 44 (82) |
| - Non-cancer | 1 (5) | 4 (14) | 5 (9) |
| - Combination of cancer and non-cancer | 2 (8) | 3 (10) | 5 (9) |
| Initiative to consider palliative sedation [a] | | | |
| - Patient | 8 (32) | 10 (34) | 18 (33) |
| - Nursing staff | 3 (12) | 4 (14) | 7 (13) |
| - Family | 1 (0) | 0 | 1 (2) |
| - Combination of patient, family, physician and nursing staff | 11 (44) | 14 (48) | 25 (46) |
| Indication of sedation [b] | | | |
| - Pain | 10 (40) | 9 (31) | 19 (35) |
| - Delirium * | 4 (16) | 15 (52) | 19 (35) |
| - Fatigue | 9 (36) | 8 (28) | 17 (32) |
| - Dyspnea | 7 (26) | 6 (22) | 13 (24) |
| - Psychological distress | 5 (19) | 7 (26) | 12 (22) |
| - Nausea and/or vomiting | 4 (16) | 2 (7) | 6 (11) |
| - Not specified | 8 (30) | 8 (30) | 16 (29) |
| Intended depth of sedation [c] | | | |
| - Somnolent | 9 (36) | 3 (10) | 12 (22) |
| - Soporific | 10 (40) | 1 (3) | 11 (20) |
| - Sub-comatose | 0 | 7 (24) | 7 (13) |
| - Comatose | 0 | 15 (52) | 15 (28) |

* $Chi^2$= <0.05; [a] Six percent correspond to missing values

[b] Respondents could select more than one option for this item;

[c] Seventeen percent correspond to missing values

*7.3.4 Inter-Rater Reliability*

Table 2 shows the ICCs within paired assessments with a time difference less than 60 and 15 minutes. These coefficients varied for the different scales and subscales, ranging from 0 to 0.85. Overall moderate to high correlations were found for the $VICS_{(i)}$, RASS and KNMG. Low correlations involved the $MSAT_{(a)}$, whereas no correlation was found for the $MSAT_{(m)}$ and $VICS_{(c)}$. The Cohen's Kappa coefficient of the $MSAT_{(q)}$ was 0.436 for paired assessments within 60 minutes and 0.545 for paired assessments within 15 minutes.

When the coefficients were computed including the 16 paired assessments outside the 60-minute limit, they showed a drop of 0.04. Moreover, in the subsample where the time difference between the first and second assessments was less than 15 minutes, the coefficients were larger, from 0.01 to 0.1.

Table 2. Inter-Rater Reliability

| | **Time Between Paired Assesments** | | | | | |
|---|---|---|---|---|---|---|
| | **< 1 hour** | | | **< 15 minutes** | | |
| **Scales** | ***n*** | **ICC** | **(95% CI)** | ***n*** | **ICC** | **(95% CI)** |
| $VICS_{(i)}$ | 59 | 0.77 | (0.64 – 0.86) | 35 | 0.85 | (0.73 – 0.92) |
| RASS | 104 | 0.71 | (0.60 – 0.79) | 60 | 0.73 | (0.58 – 0.83) |
| KNMG | 102 | 0.66 | (0.54 – 0.76) | 59 | 0.71 | (0.55 – 0.82) |
| $MSAT_{(a)}$ | 106 | 0.59 | (0.45 – 0.70) | 61 | 0.64 | (0.46 – 0.77) |
| $VICS_{(c)}$ [a] | 77 | 0.34 | (0.13 – 0.52) | 44 | 0.12 | (-0.18 – 0.40) |
| $MSAT_{(m)}$ [a] | 103 | 0.11 | (-0.09 – 0.29) | 60 | $0.01^{-1}$ | (-0.25 – 0.25) |

ICC = intraclass correlation coefficient; CI = confidence interval; MSAT = Minnesota Sedation Assessment Tool, (a) = arousal subscale, (m) = motor activity subscale; VICS = Vancouver Interaction Calmness Scale, (i) = interaction subscale, (c) = calmness subscale; RASS = Richmond Agitation-Sedation Scale; KNMG = KNMG Sedation Score; [a] No statistical difference at $P<0.05$.

Table 3. Correlations Between Sedation Scales

| **Scales** | **Rho** | **Scales** | **Rho** | **Scales** | **Rho** |
|---|---|---|---|---|---|
| $MSAT_{(a)}$-KNMG | -0.81 | $MSAT_{(m)}$-$MSAT_{(a)}$ | 0.58 | $VICS_{(c)}$-KNMG | 0.44 |
| $MSAT_{(a)}$-RASS | 0.83 | $MSAT_{(m)}$-RASS | 0.61 | $VICS_{(c)}$-RASS | -0.57 |
| $MSAT_{(a)}$-$VICS_{(c)}$ | -0.48 | $MSAT_{(m)}$-$VICS_{(c)}$ | -0.44 | $VICS_{(i)}$-KNMG | -0.7 |
| $MSAT_{(a)}$-$VICS_{(i)}$ | 0.71 | $MSAT_{(m)}$-$VICS_{(i)}$ | 0.42 | $VICS_{(i)}$-RASS | 0.72 |
| $MSAT_{(m)}$-KNMG | -0.56 | RASS-KNMG | -0.84 | $VICS_{(i)}$-$VICS_{(c)}$ | -0.31 |

Rho = Spearman's correlation coefficient; RASS = Richmond Agitation-Sedation Scale; KNMG = KNMG Sedation Score; MSAT = Minnesota Sedation Assessment Tool, (a) = arousal subscale, (m) = motor activity subscale; VICS = Vancouver Interaction Calmness Scale, (i) = interaction subscale, (c) = calmness subscale.

*7.3.5 Validity*

Table 3 shows Spearman's correlation coefficients between scales. The largest correlation found was between the RASS and KNMG (rho =0.836). Other large correlations involved the $MSAT_{(a)}$ and RASS (rho =0.827), and the $MSAT_{(a)}$ and KNMG (rho =0.811). Moderate correlation coefficients were found for the $VICS_{(i)}$ and KNMG (rho =0.73), the $VICS_{(i)}$ and RASS (rho =0.717), and the $MSAT_{(a)}$ and $VICS_{(i)}$ (rho =0.708).

The hypothesis that the KNMG and $VICS_{(c)}$ would correlate negatively with the other scales was confirmed (**table 3**). Patients on continuous sedation had overall higher mean sedation scores when compared with patients on intermittent sedation. For all scales, a significant mean difference in sedation scores between intermittent and continuous sedation was found, except for the $MSAT_{(m)}$ and $VICS_{(c)}$ (**table 4**). A statistical difference was found between the score of the first and last assessment after the start of sedation for all scales, except for the $MSAT_{(m)}$ and $VICS_{(c)}$ (**table 4**).

*7.3.6 User-Friendliness of Scales*

Evaluations of time consumption, easiness and clarity of scales were performed at 197 time points (52% response rate). Fig. 2 shows that the RASS was reported as the least time-consuming scale when compared with the MSAT and VICS in the majority of assessments ($Chi^2$, $P$ <0.001), and the MSAT was less time consuming than the VICS ($Chi^2$, $P$=0.031). In addition, the VICS was both the least clear and the least easy to use scale when compared with the RASS and MSAT ($Chi^2$, $P$ <0.001). No difference was found between the RASS and MSAT scales regarding clarity ($Chi^2$, $P$=0.56) and ease of use ($Chi^2$, $P$=0.93).

Table 4. Discriminative and Evaluative Validity

| Scales | Discriminative Validity [a] | | | Evaluative Validity [b] | | |
|---|---|---|---|---|---|---|
| | Mean Difference | (95% CI) | *P*-value [c] | Sum of Ranks (+) | Sum of Ranks (-) | *P*-value [d] |
| KNMG | -0.86 | (-1.28 – -0.51) | **<0.001** | 397.0 | 68.00 | **0.001** |
| $MSAT_{(a)}$ | 1.04 | (0.58 – 1.50) | **<0.001** | 37.50 | 287.5 | **0.001** |
| $MSAT_{(m)}$ | 0.04 | (-0.24 – 0.31) | 0.639 | 47.50 | 72.50 | 0.462 |
| RASS | 1.07 | (0.54 – 1.59) | **<0.001** | 77.50 | 328.5 | **0.004** |
| $VICS_{(c)}$ | -0.01 | (-0.06 – 0.04) | 0.951 | 212.0 | 166.0 | 0.579 |
| $VICS_{(i)}$ | 0.18 | (0.12 – 0.24) | **<0.001** | 34.50 | 136.5 | **0.026** |

CI= confidence interval; KNMG = KNMG Sedation Score; MSAT = Minnesota Sedation Assessment Tool, a = arousal subscale, m = motor activity subscale; RASS = Richmond Agitation-Sedation Scale;

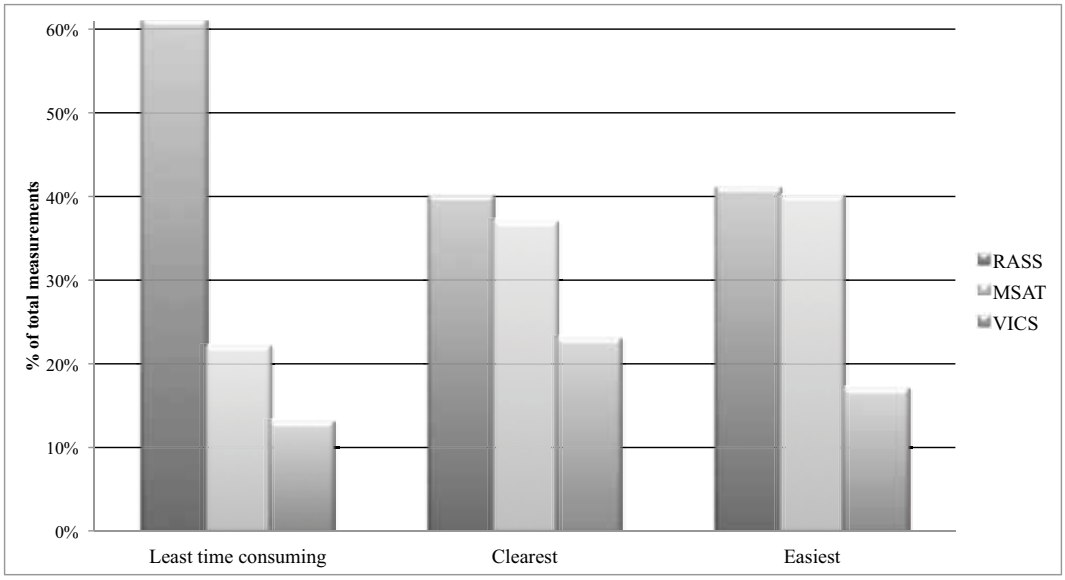
VICS = Vancouver Interaction Calmness Scale, i = interaction subscale,c = calmness subscale;

[a] Mean difference in scores between intermittent and continuous sedation.

[b] Within-patient difference between first and last recorded score for each scale after the start of sedation;

[c] Two-sample *t*-test; [d] Wilcoxon signed-rank test.

Figure 2. User-friendliness of scales




RASS = Richmond Agitation-Sedation Scale; MSAT = Minnesota Sedation Assessment Tool; VICS = Vancouver Interaction Calmness Scale.

## 7.4 Discussion

This is to our knowledge the first study reporting on the performance of different observer-based sedation scales in palliative sedation. Three of these scales were designed to assess sedation in ICU patients. However, the original papers describe sedation assessment from different perspectives. For example, the VICS uses two subscales with the aim of assessing sedation quality.[11] The MSAT measures the level of consciousness with three subscales that assess motor activity, arousal, and sedation quality.[13] In contrast, the RASS is a single-number scale used to perform a structured assessment of sedation and agitation.[12] Overall, the hypothesis that sedation can be assessed by all of these different scales in a particular patient was tested in terms of reliability and validity. In addition, we included in this analysis the sedation score proposed in the Dutch Guideline for Palliative Sedation (KNMG).[10]

Other observer-based sedation scales may be found in the literature like the Ramsay Sedation Scale,[15] the Motor Activity and Agitation Scale,[16] and the Sedation-Agitation Scale.[17] In addition, scales such as the Glasgow Coma Scale have been widely used to evaluate sedated patients.[18] For practical reasons, they all may not be evaluated in a single study on palliative sedation; therefore, we chose to include two single-number scales and two scales comprising multiple subscales. The only known consciousness level scale developed for the broader spectrum of palliative care patients (sedated and non-sedated) was published during the course of our study and, therefore, was not included.[6] We acknowledge that future research is needed to evaluate the clinimetric properties of other scales in palliative sedation for the selection of the most valid and reliable instrument.

Our findings are meant to provide the initial insight into the selection of a scale when monitoring the effects of palliative sedation. The results suggest that both the RASS and KNMG scales stand as the most reliable and valid among the evaluated scales. In addition, the RASS was less time consuming, clearer and easier to use than the MSAT and VICS. Whereas the $VICS_{(i)}$ subscale also had good reliability and validity when compared with the other subscales, it was found, together with the $VICS_{(c)}$, to be more time consuming, less clear and less easy to use. Although the arousal subscale of the MSAT had a moderate to high correlation with the other scales, it displayed moderate to low reliability, and the motor subscale was not reliable and did not strongly correlate with other scales. In addition, the $VICS_{(c)}$ and $MSAT_{(m)}$ did not differentiate between patients under intermittent or continuous sedation, or differences in the level of consciousness of patients during the course of palliative sedation.

### *7.4.1 Inter-Rater Reliability*

When interpreting the correlation coefficients of the reliability analysis, special attention should be given to the sometimes wide 95% confidence intervals (from 0.19 to 0.31). Most of these variations could have resulted from the intrinsic clinimetric properties of the scales. However, another explanation might be the effect of time within paired assessments, as a notable increase of the coefficients was observed when analyzing the subsample with less than 15 minutes between assessments (from 0.01 to 0.1). Although larger coefficients could result from doing concurrent paired assessments, the data suggest that only the $VICS_{(i)}$, RASS and KNMG would have high reliability.

### *7.4.2 Validity*

Although no gold standard for the evaluation of sedation exists in palliative sedation, it is reasonable to expect a strong correlation between two scales designed to measure sedation, meaning that they are measuring what they intend to measure. The results of the Spearman's correlation coefficient suggest that the KNMG and RASS are the pair of scales that largely measure the same criterion, hypothesized as sedation. However, the moderate to high correlation of the arousal subscale of the MSAT and the interaction subscale of the VICS also suggest that a similar criterion is being measured, and they cannot be excluded from being valid in palliative sedation.

Patients under intermittent sedation had a lower intended depth of sedation compared with patients under continuous sedation. Although the KNMG, $MSAT_{(a)}$, RASS and $VICS_{(i)}$ do identify statistical differences in the mean scores between patients under intermittent and continuous sedation, nothing can be said about the usefulness of this property in the clinical setting. Similarly, these scales showed statistically different scores between the first assessment after the start of sedation and the last recorded assessment, suggesting that they can be used to evaluate the level of consciousness of a patient throughout the course of palliative sedation.

The extent to which our results are applicable to other cases of palliative sedation relies on how representative our sample was from the general population receiving this procedure at the end of life. In general, patient characteristics were similar to other studies reporting on national and international samples. Comparable findings also include the types of sedatives used, duration, and indication for palliative sedation.[19,20]

When interpreting the findings of this study, some limitations should be considered. First, raters were able to perform more than one assessment per patient, which could increase the estimates in the analysis of concurrent validity. However, 52 different nurses were involved in this study, and they performed a maximum of five assessments per patient per day. Second, the time within paired assessments was larger than the predefined one-hour interval at 16 time points. As this could influence the responsiveness of the inter-rater reliability analyses, subsamples of paired assessments with a time difference of less than one hour were evaluated (60 and 15 minutes). In addition, it was not possible to test whether the scales have predictive validity (relate to future action, i.e., administration of sedatives). An omission in this study is the lack of data regarding the user-friendliness of the KNMG sedation scale. Some anecdotal data about the perceived time consumption required to fill in the individual scales was obtained from 11 nurses in one of the participating hospices. The reported mean (standard deviation) minutes for the KNMG, RASS, MSAT and VICS were 1.1 (0.6), 1.3 (0.8), 1.6 (1.2) and 2.9 (2.9), respectively. These data suggest a shorter time required for the KNMG as compared with the other scales (Kruskal-Wallis: $Chi^2$, $P$=0.01). Also, the KNMG scale is the most concise of the evaluated scales, which is often considered a prerequisite for user-friendliness. Nonetheless, we recognize that this is insufficient to make any assumptions with regard to the user-friendliness of this scale at this point. Therefore, for future studies, we suggest assessing the user-friendliness of the KNMG scale in a more standardized fashion, addressing time consumption, simplicity, ease of use and ease of recall.

With regard to the usefulness of the scales, we acknowledge that there is no strong evidence to support that monitoring palliative sedation with observer-based sedation scales results in better symptom control, patient comfort or more proportional administration of palliative sedation. For this to be the case, assessment of symptom burden should be performed in concert with assessment of consciousness level. Several authors have agreed that monitoring palliative sedation is important for both guiding the proportionality and for the appropriate documentation of the procedure in medical and nursing files.[21-24] Moreover, one of the few prospective studies on efficacy and safety of palliative sedation also supports this hypothesis.[25] Monitoring the level of consciousness in palliative sedation also might be useful when communicating decisions between health care providers and to the patients' families. Therefore, we acknowledge that further research is needed to evaluate the impact of sedation scales and (observational) instruments to assess symptom burden in achieving better symptom control and patient comfort and to guide the proportionality of palliative sedation.

## 7.5 Conclusion

The scales evaluated here were constructed under different hypothetical grounds; the goals of sedation in the ICU are probably different from those in palliative sedation (i.e., compliance with mechanical ventilation and tracheal stimulation). We suggest the use of the RASS and/or KNMG scales to provide a measurable estimate of the level of consciousness in palliative sedation given that these scales performed as the most reliable and valid among the evaluated scales.

## References


[1] Rady MY, Verheijde JL. Continuous deep sedation until death: palliation or physician-assisted death? *Am J Hosp Palliat Care* 2010 May;27(3):205-14.

[2] Claessens P, Menten J, Schotsmans P, Broeckaert B. Palliative Sedation, Not Slow Euthanasia: A Prospective, Longitudinal Study of Sedation in Flemish Palliative Care Units. *J Pain Symptom Manage* 2011 Jan;41(1):14-24.

[3] Alonso-Babarro A, Varela-Cerdeira M, Torres-Vigil I, Rodriguez-Barrientos R, Bruera E. At-home palliative sedation for end-of-life cancer patients. *Palliat Med* 2010 July;24(5):486-92.

[4] Aretha D, Panteli ES, Kiekkas P, Karanikolas M. Patient and/or family controlled palliative sedation with midazolam for intractable symptom control: a case series. *Cases J* 2009;2:136.

[5] Royal Dutch Medical Association (RDMA). Guidelines for palliative sedation. Utrecht, Netherlands; 2009.

[6] Goncalves F, Bento MJ, Alvarenga M, Costa I, Costa L. Validation of a consciousness level scale for palliative care. *Palliat Med* 2008 Sept.;22(6):724-9.

[7] Tendas A, Niscola P, Cupelli L et al. Palliative sedation therapy in a bone marrow transplant unit. *Support Care Cancer* 2009 Feb.;17(2):107-8.

[8] Porzio G, Aielli F, Verna L, Micolucci G, Aloisi P, Ficorella C. Efficacy and safety of deep, continuous palliative sedation at home: a retrospective, single-institution study. *Support Care Cancer* 2010 Jan.;18(1):77-81.

[9] Watanabe S, Pereira J, Tarumi Y, Hanson J, Bruera E. A randomized double-blind crossover comparison of continuous and intermittent subcutaneous administration of opioid for cancer pain. *J Palliat Med* 2008 May;11(4):570-4.

[10] *Health measurement scales: a practical guide to their development and use.* 3rd ed. Oxford, UK: Oxford Medical Publications; 2005.

[11] de Lemos J, Tweeddale M, Chittock D. Measuring quality of sedation in adult mechanically ventilated critically ill patients. the Vancouver Interaction and Calmness Scale. Sedation Focus Group. *J Clin Epidemiol* 2000 Sept.;53(9):908-19.

[12] Sessler CN, Gosnell MS, Grap MJ et al. The Richmond Agitation-Sedation Scale: validity and reliability in adult intensive care unit patients. *Am J Respir Crit Care Med* 2002 Nov. 15;166(10):1338-44.

[13] Weinert C, McFarland L. The state of intubated ICU patients: development of a two-dimensional sedation rating scale for critically ill adults. *Chest* 2004 Dec.;126(6):1883-90.

[14] Vreeswijk R, Toornvliet A, Honing MLH et al. Validation of the Dutch version of the confusion assessment method (CAM-ICU) for delirium screening in the intensive care unit. *Neth J Crit Care* 2009;13(2):73-8.

[15] Ramsay MA, Savege TM, Simpson BR, Goodwin R. Controlled sedation with alphaxalone-alphadolone. *Br Med J* 1974 June 22;2(5920):656-9.

[16] Devlin JW, Boleski G, Mlynarek M et al. Motor Activity Assessment Scale: a valid and reliable sedation scale for use with mechanically ventilated patients in an adult surgical

intensive care unit. *Crit Care Med* 1999 July;27(7):1271-5.
[17] Riker RR, Picard JT, Fraser GL. Prospective evaluation of the Sedation-Agitation Scale for adult critically ill patients. *Crit Care Med* 1999 July;27(7):1325-9.
[18] Teasdale G, Jennett B. Assessment of coma and impaired consciousness. A practical scale. *Lancet* 1974 July 13;2(7872):81-4.
[19] Kaasa S, Torvik K, Cherny N, Hanks G, De CF. Patient demographics and centre description in European palliative care units. *Palliat Med* 2007 Jan.;21(1):15-22.
[20] Swart SJ, Rietjens JA, Brinkkemper T et al. [Palliative sedation largely in accordance with Dutch national guideline.] *Ned Tijdschr Geneeskd* 2011;155(6):A2857.
[21] Cherny NI, Radbruch L. European Association for Palliative Care (EAPC) recommended framework for the use of sedation in palliative care. *Palliat Med* 2009 Oct.;23(7):581-93.
[22] Hasselaar JG, Verhagen SC, Vissers KC. When cancer symptoms cannot be controlled: the role of palliative sedation. *Curr Opin Support Palliat Care* 2009 Mar.;3(1):14-23.
[23] Hauser K, Walsh D. Palliative sedation: welcome guidance on a controversial issue. *Palliat Med* 2009 Oct.;23(7):577-9.
[24] Sulmasy DP, Curlin F, Brungardt GS, Cavanaugh T. Justifying different levels of palliative sedation. *Ann Intern Med* 2010 Mar. 2;152(5):332-3.
[25] Morita T, Chinone Y, Ikenaga M et al. Efficacy and safety of palliative sedation therapy: a multicenter, prospective, observational study conducted on specialized palliative care units in Japan. *J Pain Symptom Manage* 2005 Oct.;30(4):320-8.