

PEER-REVIEWED ARTICLE

## ICU dyspnoea assessment tools: Norwegian translation and inter-rater reliability

**Peder Sebastian Martinsen**

Clinical Nurse Specialist and Intensive Care Nurse\*  
Intensive Care, Rikshospitalet, Oslo University Hospital

**Runa Austad Haug**

Intensive Care Nurse\*  
Intensive Care, Rikshospitalet, Oslo University Hospital

**Silje Bådsvik**

Former Intensive Care Nurse  
Intensive Care 2, Rikshospitalet, Oslo University Hospital

**Hilde Vinje**

Associate Professor  
Faculty of Chemistry, Biotechnology and Food Science, Norwegian University of Life Sciences

**Kristin Hofsvø**

Professor, Senior Researcher and Intensive Care Nurse  
Lovisenberg Diaconal University College and Department of Postoperative and Critical Care Nursing, Oslo University Hospital

**Dyspnoea**

**Intensive care**

**Inter-rater reliability**

**IC-RDOS**

**MV-RDOS**

Sykepleien Forskning 2026;21(105043):e-105043

DOI: [10.4220/Sykepleienf.2026.105043en](https://doi.org/10.4220/Sykepleienf.2026.105043en)

**Background:** Dyspnoea is a common and stressful symptom in critically ill patients, and healthcare professionals tend to underestimate both the incidence and degree of this symptom. Currently, there is no Norwegian symptom assessment scale to aid in evaluating dyspnoea in intensive care unit (ICU) patients without the ability to self-report. The Intensive Care Observation Scale (IC-RDOS) and Mechanical Ventilation-Respiratory Distress

Observation Scale (MV-RDOS) are symptom assessment scales for identifying dyspnoea in non-mechanically ventilated and mechanically ventilated patients.

**Objective:** To translate the IC-RDOS and MV-RDOS into Norwegian and evaluate the inter-rater reliability of the two scales.

**Method:** We applied a prospective observational design to evaluate the inter-rater reliability and measurement errors of IC-RDOS and MV-RDOS. The translation process focused on linguistic and cultural adaption following international guidelines. Data for the inter-rater reliability analysis were collected in four ICUs in Norway. Two ICU nurses performed 110 assessments, and inter-rater reliability was evaluated using the intraclass correlation coefficient (ICC) for the total score and continuous items, while Gwet's agreement coefficient 1 was applied for the dichotomous items.

**Results:** IC-RDOS and MV-RDOS were translated into Norwegian. The total score of the scales, used in the evaluation of dyspnoea, showed a high ICC for both scales: ICC = 0.98 for IC-RDOS and ICC = 0.92 for MV-RDOS.

**Conclusion:** IC-RDOS and MV-RDOS had 'very good' inter-rater reliability. Based on the findings in this project, the scales may be ready for implementation in Norwegian ICUs.

## Introduction

In Norway, approximately 18,000 people need intensive care treatment annually (1). Critically ill patients are exposed to several sources of discomfort in the intensive care unit (ICU) (2). Pain, thirst, anxiety, dyspnoea and inadequate sleep are the five most common and stressful symptoms ICU patients can experience (3).

While pain has traditionally been the primary focus of symptom research, dyspnoea has received increasing attention in recent years (2, 4). Like pain, dyspnoea can be experienced by the patient despite analgosedation and should be evaluated daily (3). Dyspnoea is defined as 'a subjective experience of breathing discomfort that consists of qualitatively distinct sensations that vary in intensity' (5, p. 436).

Studies report that 34–55% of adult ICU patients reported significant dyspnoea (6–8). A recent multi-centre study that also included patients from our site found that one in three patients self-reported difficulty breathing during the first seven days of their ICU stay (9). Being a stressful symptom for the ICU patient, dyspnoea can also negatively impact treatment and result in failed or delayed weaning from mechanical ventilation (10, 11). Furthermore, ICU-related dyspnoea is associated with post-traumatic stress disorder (8).

The trend in recent decades has been for patients to remain more awake in the ICU, be mobilised earlier and receive lower doses of sedatives. The prevalence of dyspnoea may increase when patients receive less sedation or are ventilated with lower tidal volumes; however, many ICU patients are unable to communicate their symptoms (11–13).

Patients unable to communicate due to factors such as delirium, use of sedatives and endotracheal tubes make dyspnoea evaluation challenging (12). These factors, combined with the feeling of not being able to breathe as much as the body requires, may lead to anxiety for patients (4, 7, 14). Despite its impact, dyspnoea is often underestimated and underreported by healthcare professionals (6, 14, 15).

Haugdahl et al. enrolled 100 ICU patients and found that nurses and physicians underestimated dyspnoea in 56% and 48% of patients, respectively (15). These findings highlight the need for validated tools to identify dyspnoea in non-communicative ICU patients (16).

The Intensive Care-Respiratory Distress Observation Scale (IC-RDOS) and the Mechanical Ventilation-Respiratory Distress Observation Scale (MV-RDOS) are two hetero-evaluation symptom assessment scales intended to aid healthcare professionals in identifying dyspnoea in non-communicative critically ill patients (17, 18).

Both scales are meant to be used with ICU patients who are unable to self-report, IC-RDOS for patients not receiving MV and MV-RDOS for patients receiving MV. IC-RDOS and MV-RDOS are based on Campbell's work with the original RDOS to identify dyspnoea in palliative patients (17, 19).

Both scales consist of five items: heart rate, use of neck muscles during inspiration, abdominal paradox during inspiration, and facial expression of fear. Use of supplemental oxygen only applies for IC-RDOS, and respiratory rate only applies for MV-RDOS. Together, these items constitute a total score to assess dyspnoea. Patient self-report using a visual analogue scale (D-VAS), an alternative to a numeric rating scale (NRS) for alert and verbal patients, is the gold standard for assessing dyspnoea in the ICU (11).

Both IC-RDOS and MV-RDOS demonstrate acceptable to excellent predictive value for identifying dyspnoea on a visual analogue scale (D-VAS  $\geq 4$ ) with area under the curve values of 0.83 and 0.78, respectively (16, 17). IC-RDOS is recommended to identify dyspnoea in spontaneously breathing patients, and a score  $\geq 2.4$  predicts D-VAS  $\geq 4$  with a sensitivity and specificity of 72% (17).

MV-RDOS is designed for mechanically ventilated patients, and a score of  $\geq 2.6$  predicts dyspnoea with D-VAS  $\geq 4$  with a specificity of 57% and a sensitivity of 94% (16). Comparable metrics have been shown in a recent study (20). The scales have been used in recent studies, including to predict weaning outcomes and to investigate an association with mortality (10, 20).

Like other symptom assessment scales, both instruments are used as surrogates for self-reporting and should only be used when patients are unable to communicate (12). This is justified by dyspnoea being a subjective symptom, and self-reporting is the gold standard (5, 11).

Currently, there is no validated method to evaluate the dyspnoea of ICU patients without the ability to self-report in Norwegian ICUs. The need for a more systematic approach to identifying and managing dyspnoea in our ICUs has become evident. We believe that a Norwegian adaptation of IC-RDOS and MV-RDOS could support healthcare professionals in making an evidenced based evaluation of dyspnoea in ICU patients without the ability to self-report.

### ***Aim of the project***

The overall aim of the project was to translate both IC-RDOS and MV-RDOS into Norwegian and evaluate the inter-rater reliability (IRR) of the Norwegian versions.

### **Method**

To ensure the correct method and design, the study design checklist for patient-reported outcome measurement instruments from COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) was chosen as a guideline for the IRR testing (22). STrengthening the Reporting of OBservational studies in Epidemiology (STROBE) was applied to ensure appropriate reporting (23).

The project was conducted as a quality improvement project involving two phases. First, the translation process of IC-RDOS and MV-RDOS was performed (24). We then applied a prospective observational design to evaluate the IRR and measurement errors of the scales. A quality improvement project was chosen to provide a more systematic approach to evaluating dyspnoea in ICU patients who are unable to self-report in our ICUs (25).

To translate IC-RDOS and MV-RDOS, we used an international recognised guideline that emphasises the importance of linguistically and culturally adapted translation related to the setting in which the scale is supposed to be used (24).

### ***Setting and population***

Data for the reliability analysis were collected in four ICUs in a referral university hospital in Norway including both medical and surgical patients. Two of the units have ten beds, and the other two have six beds; both are categorised as level three units (1). Level three units have the capacity to manage advanced organ-supportive therapy for multiple organ systems including invasive mechanical ventilation (26).

The inclusion criteria for this project were patients over the age of 18 years, admitted to the ICU, and defined as ICU patients (ICU stay > 24 hours, use of vasoactive medication, or in need of mechanical ventilation) (1). Patients receiving muscle relaxants, declared brain dead, or able to self-report dyspnoea were excluded. We chose to include patients with various scores on the Richmond Agitation Sedation Scale (RASS) to ensure representation of all non-communicative ICU patients.

Prior to the data collection, we set a maximum threshold of 10% of mechanically ventilated patients (five patients) to be unresponsive, defined as RASS of  $-5$ , to ensure the instrument was tested on heavily sedated patients, despite it being likely that these patients have physical signs of dyspnoea. Some patients were assessed twice with a minimum of 48 hours between assessments to ensure independence.

## ***Data collection***

Patient ratings were performed using the Norwegian versions of IC-RDOS and MV-RDOS by two ICU nurses with six and eight years of ICU experience. This approach was chosen to minimise external factors influencing measurement errors. Dyspnoea assessment was performed simultaneously and independently, with no discussion or access to view each other's results until the project was completed.

Both raters participated in the translation of the scales. Descriptive clinical characteristics (i.e. Simplified Acute Physiology Score (SAPS) and RASS) of the patients were collected from the electronic health record. We included patients from November 2021 to March 2022.

## ***Sample size***

We estimated the sample size using power analysis, a priori, to be 50 assessments per scale with two raters. The sample size for Cohen's kappa was estimated using the Cicchetti and Fleiss (27) formula  $n = 2k^2$ , where  $n$  is the sample size and  $k$  is the number of items in the scale.

A parallel power analysis for intraclass correlation coefficient (ICC) indicated a smaller required sample. Consequently, we chose the power analysis of Cohen's kappa to increase the strength of the project. The intention was to use Cohen's kappa in the analysis, but after the data collection was completed, we changed this analysis to Gwet's agreement coefficient 1 ( $AC_1$ ) due to a better fit for the data collected. Therefore, the sample size was based on Cohen's kappa instead of Gwet's  $AC_1$ .

A sample size of 50 assessments per scale is, however, in accordance with the recommendations of COSMIN (22). During the project, we increased the sample size of MV-RDOS to 60 due to the low frequency of some of the dichotomous traits.

## ***Statistical analysis***

Descriptive statistics were used to summarise the patients' characteristics: continuous variables are reported as median with interquartile range (IQR) and categorical variables as frequencies (%) and counts. The IRR was analysed separately for each scale. For the total score and continuous items, we used a two-way random, average score, absolute agreement ICC (2,  $k$ ) (28, 29). Normality of the residuals was assessed with Q-Q plots. Gwet's  $AC_1$  was chosen for the dichotomous item due to robustness against trait prevalence (30).

We used Altman's scale to benchmark both Gwet's  $AC_1$  and the ICC. A Gwet's  $AC_1$ /ICC below 0.2 is considered 'poor', 0.21–0.4 as 'fair', 0.41–0.60 as 'moderate', 0.61–0.8 as 'good', and above 0.81 as 'very good' (31). The measurement errors are presented with 95% upper and lower limits of agreement (LoA) with confidence intervals for the total score and the continuous items (32).

In addition to total percentage agreement, we also calculated positive and negative agreement of the dichotomous items (33). Systematic differences between raters were tested using the paired t-test and McNemar's test (30).

The data were analysed using R version 4.1 (R Foundation for Statistical Computing, Austria, 2021), with a significance level at 0.05 and a 95% confidence interval (CI) where applicable.

### ***Ethical considerations***

This quality improvement project was first presented to the Regional Committees for Medical Research Ethics in Southeast Norway (reference number 2021/325676) but fell outside their mandate. The data protection office at the hospital approved the project, and the need for consent was waived (reference number 21/19413).

## **Results**

### ***Translation of IC-RDOS and MV-RDOS***

We describe each step of the translation process in Table 1, which constitutes the final report of the process. We want to emphasise the choice to use medical terminology over layman's terms to describe the items in the present project as a cultural adaptation. For example, 'Use of neck muscles during inspiration' was translated to 'Use of accessory muscles during inspiration', as the instrument will be used by health care professionals.

Furthermore, it is standard practice in Norway not to translate the name or abbreviation of scales used in the ICU. The translated versions of the IC-RDOS and MV-RDOS are presented in Figure 1. The back-translated version was accepted by the developer of the scales without disagreement.

**Table 1.** Description of the translation process

<b>The ten steps in the translation process (24)</b>	<b>Description of the translation of IC/MV-RDOS into Norwegian in our study</b>
<p><b>1 Preparation</b> The first step is to become familiar with the scale and obtain permission to translate the scale to respect copyright. The developer of the scale should be invited to be included in the process.</p> <p>Include translators who are familiar with the tool for the translation process. The translators should have both language skills and a professional understanding of the field in which the tool is used.</p>	<p>In this step the three translators focused on understanding IC-RDOS and MV-RDOS and read articles describing the development, validation, and reliability testing of the scales. We then got written consent to translate IC/MV-RDOS into Norwegian by one of the developers, Dr Decavèle, and invited him to be a part of the process. He accepted under certain terms, namely that the Norwegian version will not be used commercially.</p> <p>All the translators have Norwegian as their first language and are fluent in English. The translators had between 4-9 years of experience working in the ICU.</p>
<p><b>2 Forward translation</b> At least two independent forward translations should be made. The translations can be discussed when writing and must later be compared and discussed to detect errors or different interpretations.</p>	<p>The authors translated the scales independently. We discussed the differences and uncertainties and started the process individually again. The primary discussion was regarding medical versus layman's terminology.</p>
<p><b>3 Reconciliation</b> Reconciliation takes place through discussion in the group of individual translators. Both language and speech habits should be discussed and clarified</p>	<p>The authors reviewed each other's versions and agreed on a common reconciled version of the scales after several meetings to unite the translations. The discussion centred around the translation of the items 'Use of neck muscles during inspiration' and 'Abdominal paradox during inspiration' and the choice of medical versus lay terms. We decided on the use of precise medical terms and supplemented with a comprehensive explanation. We believe this is the most appropriate way to ensure user-friendliness for Norwegian ICU clinicians.</p>
<p><b>4 Back translation</b> The reconciled version of the tool is translated back to its original language to ensure that the tool has the same content and quality. It should be clarified whether the back translation should be done literally or conceptually</p>	<p>A professional medical translator translated the reconciled Norwegian version into English with the instruction to translate conceptually.</p>

<p><b>5 Back translation review</b> The back translation is reviewed against the original tool to identify any discrepancies in the meaning of the words.</p>	<p>We reviewed the back-translated version to ensure that the scales had the same linguistic and conceptual quality as the original. The back-translated version was then sent to Dr Decavèle, and he approved the Norwegian version with only a few comments on the new explanatory part of the scales.</p>
<p><b>6 Harmonisation</b> A review is performed to detect differences in the back-translation and the original version to ensure that the two versions are equal.</p>	<p>The back-translated version was compared to the original version by the three translators. We agreed that the re-translated version had the same content as the original scales. We also emphasised that Dr Decavèle accepted the translated version without commenting on the language or content of the scales.</p>
<p><b>7 Cognitive debriefing</b> The translated tool is tested on a group of 5-8 people to ensure ease of use and understanding of selected concepts.</p>	<p>Six ICU nurses with varying degrees of clinical experience (one to 19 years) reviewed the final Norwegian version of the scales. We emphasise that the nurses had not been involved in the translation process and had no prior knowledge of the scales. They had few or no comments on the understanding of use and relevance of the selected concepts. The few comments were related to the medical terms, which were understood when they read the explanation part of the scales. Comments were made regarding our choice to refrain from translating the name of the scales which is justified by the fact that it is not customary to translate the name of assessment scales used in Norwegian ICUs.</p>
<p><b>8 Review of cognitive debriefing results and finalisation</b> The results and comments from the cognitive debriefing are reviewed and the translation is finalised.</p>	<p>We changed one word after a discussion about hyphenation. We also changed the wording in the explanation relating to the Norwegian translation of 'non-communicative' to 'a patient who is not able to express dyspnoea to their healthcare professional'. No other changes were made.</p>
<p><b>9 Proofreading</b> The final version is proofread.</p>	<p>All the authors and an external person with excellent English skills proofread the Norwegian version multiple times.</p>
<p><b>10 Final report</b> A report is written to describe the translation process.</p>	<p>We consider this table as the final report for the process of translating IC-/MV-RDOS into Norwegian.</p>

**Figure 1.** The Norwegian IC-RDOS and MV-RDOS

Intensive Care-Respiratory Distress Observation Scale (IC-RDOS)			Mechanical Ventilation-Respiratory Distress Observation Scale (MV-RDOS)		
Delskår Item	Skår Score	Utregning Calculation	Delskår Item	Skår Score	Utregning Calculation
		3.3			3.3
1 Hjerterefreknens (slag/min) <i>Heart rate (beats/min)</i>	+ Hjerterefreknens/60 <i>+ Heart rate/60</i>		1 Hjerterefreknens (slag/min) <i>Heart rate (beats/min)</i>	+ Hjerterefreknens/60 <i>+ Heart rate/60</i>	
2 Bruk av aksessorisk hjelpemuskulatur under inspirasjon <i>Use of neck muscles during inspiration</i> Hvis til stede <i>If present</i> Hvis fraværende <i>If absent</i>	+1  -1		2 Bruk av aksessorisk hjelpemuskulatur under inspirasjon <i>Use of neck muscles during inspiration</i> Hvis til stede <i>If present</i> Hvis fraværende <i>If absent</i>	+1  -1	
3 Paradoksalt pustemønster, abdomen trekkes inn under inspirasjon <i>Abdominal paradox during inspiration</i> Hvis til stede <i>If present</i> Hvis fraværende <i>If absent</i>	+1  -1		3 Paradoksalt pustemønster, abdomen trekkes inn under inspirasjon <i>Abdominal paradox during inspiration</i> Hvis til stede <i>If present</i> Hvis fraværende <i>If absent</i>	+1  -1	
4 Ansiktsuttrykk som viser frykt <i>Facial expression of fear</i> Hvis til stede <i>If present</i> Hvis fraværende <i>If absent</i>	+1  -1		4 Ansiktsuttrykk som viser frykt <i>Facial expression of fear</i> Hvis til stede <i>If present</i> Hvis fraværende <i>If absent</i>	+1  -1	
5 Oksygentilskudd <i>Supplemental oxygen</i> Hvis til stede <i>If present</i> Hvis fraværende <i>If absent</i>	+0.7  -0.7		5 Respirasjonsfreknens (pust/min) <i>Respiratory rate (breaths/min)</i>	+ Respirasjonsfreknens/50 <i>+ Respiratory rate/50</i>	
<b>Totalskår Total score</b>			<b>Totalskår Total score</b>		

**Forklaring av de ulike delskårene i IC-RDOS og MV-RDOS**  
**Explanation of the different items in IC-RDOS and MV-RDOS**

Delskår Item	Beskrivelse Description
	Totalskåren starter med 3,3. <i>The total score begins at 3.3</i>
1 Hjerterefrekvens (slag/min) <i>Heart rate (beats/min)</i>	Pasientens hjerterefrekvens deles på 60 og legges til totalskåren. <i>The patient's heart rate is divided by 60 and added to the total score.</i>
2 Bruk av aksessorisk hjelpemuskel under i nspirasjon <i>Use of neck muscles during inspiration</i>	Dersom pasienten bruker hals-/nakkemuskel under inspirasjon, legges det til 1 i totalskåren, hvis ikke trekkes det fra 1. <i>If the patient uses neck muscles during inspiration, 1 is added to the total score. If not, 1 is subtracted.</i>
3 Paradoksalt pustemønster, abdomen trekkes inn under inspirasjon	Dersom pasientens abdomen går innover under inspirasjon, legges det til 1 i totalskåren, hvis ikke trekkes det fra 1. Normalt går abdomen utover under inspirasjon. Pasientens hodeende bør være elevvert maksimum 20°.

4 Ansiktsuttrykk som viser frykt	Dersom pasienten har et ansiktsuttrykk som viser frykt, legges det til 1 i totalskåren, hvis ikke trekkes det fra 1. Frykt kjennetegnes ved vidåpne øyne, øyenbryn hevet eller trukket sammen, og strukket eller åpen munn.
<i>Facial expression of fear</i>	If the patient has a facial expression that expresses fear, 1 is added to the total score. If not, 1 is subtracted. Fear is characterized by wide-open eyes, brows raised and furrowed, and a stretched or open mouth.
5 Oksygentilskudd*	Dersom pasienten har oksygentilskudd, uavhengig av hvor mye, legges det til 0,7 i totalskåren. Dersom pasienten får romluft, trekkes 0,7 fra. *Denne delskåren brukes kun på IC-RDOS
<i>Supplemental oxygen*</i>	If the patient receives supplemental oxygen, regardless of how much, 0.7 is added to the total score. If the patients receive room air, 0.7 is subtracted. *This item is only used in IC-RDOS.
5 Respirasjonsfrekvens (pust/min)**	Pasientens respirasjonsfrekvens deles på 50 og legges til totalskåren. Tell i ett minutt. **Denne delskåren brukes kun på MV-RDOS.
<i>Respiratory rate (breaths/min)**</i>	The patient's respiratory rate is divided by 50 and added to the score. Count one minute. **This item is only used on MV-RDOS.

IC-RDOS og MV-RDOS er symptomkartleggingsverktøy som kan hjelpe med identifisering av dyspné hos intensivpasienter over 18 år som ikke har mulighet til å angi dyspné eller grad av dyspné. Kartleggingen erstatter ikke helsepersonells selvstendige vurdering og observasjon.

*IC-RDOS and MV-RDOS are symptom assessment scales that can be used to identify dyspnoea in intensive care patients over the age of 18 who cannot self-report their dyspnoea. It does not replace the clinician's independent assessment and observation.*

Persichini R, Gay F, Schmidt M, Mayaux J, Demoule A, Morelot-Panzini C, et al. Diagnostic accuracy of respiratory distress observation scales as surrogates of dyspnea self-report in intensive care unit patients. *Anesthesiology* 2015; 123: 830-837. 2015/08/11. Doi: 10.1097/ALN.0000000000000805.

Decavéle M, Gay F, Persichini R, Mayaux J, Morelot-Panzini C, Similowski T, et al. The Mechanical Ventilation-Respiratory Distress Observation Scale as a surrogate of self-reported dyspnoea in intubated patients. *Eur Respir J* 2018; 52 2018/09/14. Doi: 10.1183/13993003.00598-2018.

## ***Patient characteristics***

This project enrolled 81 ICU patients. The median patient age was 60 (IQR 51, 69), and 64% were male. Most of the included patients were acute surgical cases (42%). The most frequent reason for admission was 'respiratory' (24%).

We performed 50 and 60 pairs of assessments for IC-RDOS and MV-RDOS, respectively. The median RASS at the time of the assessment was -1 (IQR -2, 0) for the non-mechanically ventilated patients and -3 (IQR -4, -2) for the mechanically ventilated patients, indicating drowsy and moderately sedated, respectively. The patient characteristics are summarised in Table 2.

**Table 2.** Patient characteristics

<b>Table 1: Patient characteristics</b>		
<b>Study sample (n = 81)</b>	<b>Norwegian version of IC-RDOS (50 assessments)</b>	<b>Norwegian version of MV-RDOS (60 assessments)</b>
<b>Variable</b>	<b>Median (IQR)/count (%)</b>	<b>Median (IQR)/count (%)</b>
Assessments		
Number of unique patients	42	46
Number of patients who were assessed two times	8	14
<b>Baseline characteristics of the patients</b>		
Age (years)	60 (52, 71)	60 (51, 69)
Sex (male)	22 (52%)	33 (72%)
SAPS II	51 (39, 59)	52 (39, 64)
Type of admission		
Acute non-operative	15 (36%)	23 (50%)
Acute surgery	20 (48%)	16 (35%)
Elective surgery	7 (17%)	7 (15%)
Cause of admission		
Respiratory	7 (17%)	15 (33%)
Circulatory and cardiovascular	2 (5%)	6 (13%)
Neurological	8 (19%)	6 (13%)
Trauma	8 (19%)	6 (13%)
Gastrointestinal	8 (19%)	8 (17%)
Sepsis	3 (7%)	1 (2%)
Other*	6 (14%)	4 (9%)

Characteristics of the patients at the time of assessment		
RASS	-1 (-2, 0)	-3 (-4, -2)
Day in the ICU	6 (3,14)	10 (3,18)
CAM-ICU		
Positive	16 (32%)	6 (10%)
Negative	10 (20%)	3 (5%)
Not applicable	12 (24%)	39 (65%)
Missing	12 (24%)	12 (20%)
Mechanical ventilation		
Supportive ventilation (IMV)	N/A	31 (52%)
Controlled ventilation (IMV)	N/A	25 (42%)
Non-invasive ventilation	N/A	4 (7%)
Mean IC-RDOS or MV-RDOS between rater 1 and rater 2	2.5 (2.1, 4.1)	2.5 (1.9, 4.2)

IC-RDOS = Intensive Care-Respiratory Distress Observation Scale  
MV-RDOS = Mechanical Ventilation-Respiratory Distress Observation Scale  
IQR = Interquartile range  
IMV = Invasive mechanical ventilation  
SAPS II = Simplified Acute Physiology Score 2  
RASS = Richmond Agitation and Sedation Scale  
CAM-ICU = Confusion Assessment Method for the ICU  
N/A = Not applicable  
\* Other includes haematological, renal, or post-operative

### ***Reliability of IC-RDOS***

The total score of IC-RDOS had an ICC of 0.98 (CI 95% 0.96 to 0.99), indicating 'very good' IRR. The mean difference between the raters for IC-RDOS was 0.08 ( $t = 0.21$ ,  $df = 49$ ,  $p = 0.98$ ), with LoA ranging from 1.19 to -1.03.

### ***Reliability of MV-RDOS***

The total score of MV-RDOS had an ICC of 0.92 (CI 95% 0.86 to 0.95), indicating 'very good' IRR. The mean difference between the raters for MV-RDOS was 0.003 ( $t = 0.01$ ,  $df = 59$ ,  $p = 0.99$ ), with LoA ranging from 1.77 to -1.77.

For both scales, the item with the lowest degree of IRR was 'use of neck muscles during inspiration' with an  $AC_1$  of 0.87 (CI 95% 0.74 to 1) for IC-RDOS and 0.77 for MV-RDOS (CI 95% 0.61 to 0.93). All ICC and Gwet's  $AC_1$ -values, LoAs, and percentage agreements are presented in Table 3. Measurement errors for the total scores are illustrated in Figure 2.

**Table 3.** Reliability analysis

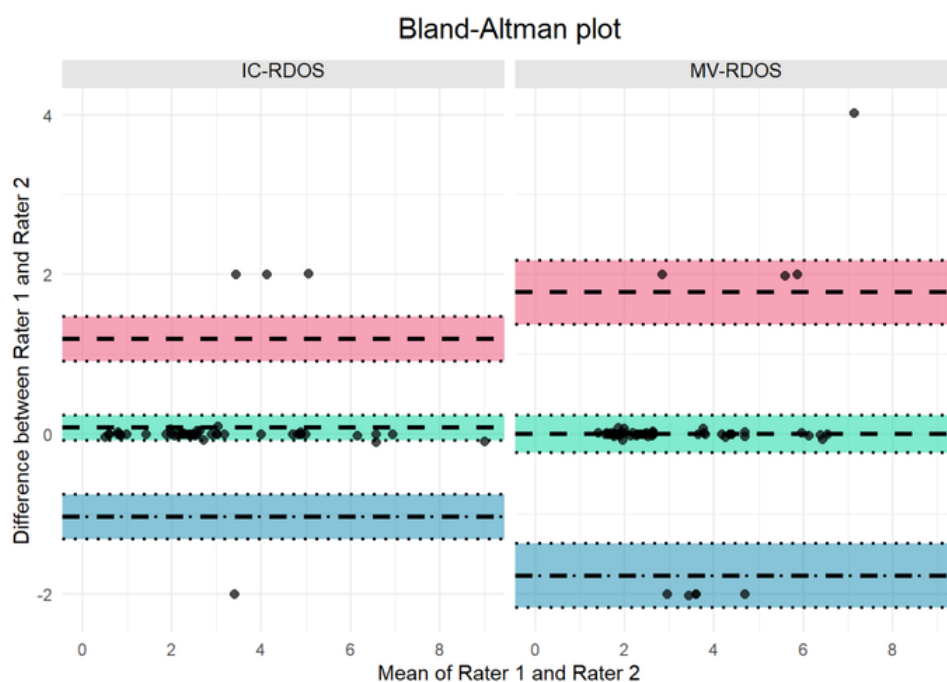
		Inter-rater reliability		Measurement error				
		Dichotomous items	Continuous items	Dichotomous items			Continuous items	
Item	Analysis	Gwet's AC1	ICC (2, k)	Percent agreement (total)	Percent agreement (sign present)	Percent agreement (sign absent)	Limit of agreement	
							Lower	Upper
		N/A	0.98 (0.96–0.99)	N/A	N/A	N/A	-1.03 (-1.31–[-0.76])	1.19 (0.92–1.47)
1. Heart rate (beats/min)		N/A	1 (1–1)	N/A	N/A	N/A	-3.61 (-4.51–[-2.71])	3.61 (2.71–4.51)
2. Use of neck muscle during inspiration		0.87 (0.74–1)	N/A	92%	85%	95%	N/A	
3. Abdominal paradox during inspiration		1 (1–1)	N/A	100%	100%	100%	N/A	
4. Facial expression of fear		1 (1–1)	N/A	100%	100%	100%	N/A	
5. Supplemental oxygen		1 (1–1)	N/A	100%	100%	100%	N/A	
		N/A	0.92 (0.86–0.95)	N/A	N/A	N/A	-1.77 (-2.17–[-1.37])	1.77 (1.37–2.18)
1. Heart rate (beats/min)		N/A	1 (1–1)	N/A	N/A	N/A	-1.66 (-2.04–[-1.27])	1.72 (1.34–2.11)
2. Use of neck muscle during inspiration		0.77 (0.61–0.93)	N/A	87%	78%	91%	N/A	
3. Abdominal paradox during inspiration		0.98 (0.93–1)	N/A	98%	94%	99%	N/A	
4. Facial expression of fear		0.94 (0.88–1)	N/A	95%	57%	97%	N/A	
5. Respiratory rate (breaths/min)		N/A	1 (0.99–1)	N/A	N/A	N/A	-1.69 (-2.1–[-1.29])	1.89 (1.49–2.3)

N/A = Not applicable

ICC = Intraclass correlation coefficient

95% CI in parentheses

**Figure 2.** Reliability analysis



Bland–Altman plot of the total score of the Norwegian version of Intensive Care-Respiratory Distress Observation Scale (IC-RDOS) and Mechanical Ventilation-Respiratory Distress Observation Scale (MV-RDOS). The red areas show 95% CI of the upper limit of agreement, the green areas show 95% CI of the bias, and the blue areas show 95% CI of the lower limit of agreement. The lines in the centre within each coloured area represent estimates of bias and both LoAs.

## Discussion

In this project, we translated IC-RDOS and MV-RDOS (Figure 1) and evaluated the Norwegian versions' inter-rater reliability. The main findings of this project are that the total scores of both instruments demonstrated 'very good' IRR. All the dichotomous items, except 'use of neck muscles during inspiration' had total agreement for the non-mechanically ventilated patients with the use of IC-RDOS. The total score of IC-RDOS had higher ICC values than MV-RDOS. All dichotomous items without complete agreement had greater negative than positive agreement.

To our knowledge, only one project has investigated IRR in IC-RDOS (17). Persichini et al. (17) found moderate consistency for the total score and variable IRR across the individual items. However, their study included multiple raters, which may have contributed to lower agreement than in our project, which had only two raters.

In our project, we found that the total score had a 'very good' IRR for both scales, but IC-RDOS had a slightly higher ICC than MV-RDOS. The differences in the ICC for the total scores of the two scales may be explained by an increase in subjectivity in the interpretation of the items. It is also possible that these behavioural signs are more subtle in mechanically ventilated patients than in non-mechanically ventilated patients due to a lower RASS or use of sedation.

While this project did not explore the underlying differences in ICC, it is a known challenge to identify behavioural signs in mechanically ventilated patients (34).

The item with the lowest degree of agreement on both scales was 'use of neck muscles during inspiration'. Nevertheless, the item still had 'good' and 'very good' agreement. The raters may have slightly different perceptions of what the item characterises. Alternatively, the item may be inherently more difficult to interpret than the other items on the scales. It is important to consider these findings in the further implementation of the scales as extra focus may be needed for this item.

Previous studies have found varying degrees of agreement for this item (17, 35). To address this, we designed a user guide to aid clinicians in using the scales correctly and thus minimise these events during future use. However, this item does not affect the total score in such a way as to make the scales less reliable.

A facial expression of fear is characterised by wide-open eyes with visible irises (36). We observed a lower IRR for this item for mechanically ventilated patients. Thus, these patients may not have had the ability or may have shown it more subtly because they were more sedated than non-mechanically ventilated patients (34). The subtle display of the sign may have led to a greater variation between the raters in this project.

Facial expressions can be a powerful predictor of dyspnoea (17) but they were uncommon in our sample. In mechanically ventilated patients, 'facial expression of fear' had the highest discrepancy between positive and negative agreement. This discrepancy may be influenced by the low prevalence of the sign in mechanically ventilated patients and was the reason we applied Gwet's  $AC_1$ , which is resilient to trait prevalence (30). Gwet's  $AC_1$  favours the overall agreement, but the low agreement in the actual presence of 'facial expression of fear' must, therefore, be taken into consideration in future implementation processes.

The measurement errors of IC-RDOS and MV-RDOS have not been examined in previous studies. To clarify, if both raters are in complete agreement except for on one dichotomous item, the total score will deviate by as much as two points. In this project, the LoAs of the total score of IC-RDOS and MV-RDOS are below two, which indicates most of the differences in the total score can be explained by the difference in only one dichotomous item or less.

In fact, the raters disagreed on two dichotomous items in only one assessment. We were not able to follow the recommendation to set a desired LoA prior to data analysis as our project is the first to describe the measurement error (37). Our findings offer a base for future studies to compare the measurement errors of the scales.

Following data collection, we observed clear differences between total percent agreement and Cohen's kappa, explained by the kappa paradox (33). This prompted a search for alternative methods. Cohen's kappa assumes that the raters evaluate all observations independently from the other evaluations, which was not the case in our population. The dichotomous items would

have been falsely deemed to have low IRR explained by low prevalence (38). In such cases, Gwet's AC<sub>1</sub> provides a more stable estimate (30).

We chose to use Gwet's AC<sub>1</sub> in the evaluation of the dichotomous items, which is an advantage of this project due to a more precise presentation of the results (38). If we had solely relied on Cohen's kappa, it might have led to delayed or failed implementation of a validated method for evaluating dyspnoea in ICU patients without the ability to self-report in Norway.

The patients included in this project do not fully represent the general ICU population. The first consideration relates to our inclusion of only patients from a referral hospital. This could have resulted in the inclusion of more severely ill patients or patients with rare conditions. A recent review demonstrated patients receiving non-invasive ventilation (NIV) self-reported both high prevalence and intensity of dyspnoea (39).

In addition, out of the 60 assessments performed on mechanically ventilated patients, only 4 (7%) used NIV. As a result of this we cannot say the Norwegian version of the scale has been fully tested on patients receiving NIV. A final consideration is that the assessments were done on median day 6 and 10 in the ICU for IC-RDOS and MV-RDOS respectively. This differs from our ICUs in general with a median length of stay of 3 days, indicating that the instruments have been tested in patients with a prolonged ICU length of stay (1).

The median SAPS II score of 51 indicates that the included patients were severely ill, reflecting the patient population at the hospital. Less severely ill ICU patients are more likely to be able to self-report their symptoms and consequently were excluded from the project. These factors combined point out that our sample of patients were severely ill ICU patients, but we argue that the scales still can be applicable to less severely ill ICU patients. Still, we have demonstrated the possible use of the translated version of IC-RDOS and MV-RDOS in Norwegian ICUs.

A recent statement from the European society of intensive medicine recommends the use of IC-RDOS and MV-RDOS in the evaluation of dyspnoea in ICU patients who cannot self-report (10).

Providing a Norwegian version of the scales will make it easier for clinicians to assess dyspnoea in ICU patients unable to self-report. Introducing these scales into clinical practise may promote a common language in the assessment of dyspnoea, hopefully resulting in improved patient care and reduced symptom burden. Further, being able to measure dyspnoea in a valid and reliable manner can facilitate future research on the associated factors related to dyspnoea in ICU patients.

## ***Limitations***

We chose not to perform dyspnoea evaluations on all non-communicative patients to ensure validity for a wider patient selection. Including all non-communicative patients may have biased the results in favour of deeply sedated patients, resulting in items such as 'facial

expressions of fear' having a lower prevalence. The raters gathered background information before every assessment.

In retrospect, knowing the background information before the assessment may have introduced confirmation bias, and this information therefore should have been collected after the assessments. The sample size was based on Cohen's kappa instead of Gwet's AC<sub>1</sub>. However, the sample sizes of 50 and 60 assessments are more than the recommended threshold (24).

Both raters were involved in the translation process. However, we believe the behavioural signs would have shown a higher prevalence or a systematic difference between the raters if confirmation bias was introduced. In future clinical use, a lower IRR must be expected due to a greater diversity of raters.

### ***Further research***

Regarding the validity and reliability of the Norwegian version of IC-RDOS and MV-RDOS in a larger population of ICU patients, further research is warranted after implementation with a greater diversity of raters. Power analysis should be based on Gwet's AC<sub>1</sub> due to the expected low prevalence in the dichotomous items.

### **Conclusion**

In this article, we presented a Norwegian version of the IC-RDOS and MV-RDOS and evaluated the scales' IRR. The primary result indicates the total score of the scales demonstrated 'very good' IRR. Further research regarding the Norwegian version of IC-RDOS and MV-RDOS should focus on a wider range of psychometric properties.

Dyspnoea is a subjective experience and is often underestimated by clinicians. We believe that IC-RDOS and MV-RDOS can contribute to both increased awareness and evidence-based assessment of dyspnoea experienced by ICU patients.

Based on the findings of this project, the scales may be suitable for implementation in Norwegian ICUs. Using the IC-RDOS and MV-RDOS in clinical practice is likely to increase awareness and improve symptom management.

*\*Peder Sebastian Martinsen and Runa Austad Haug share first authorship.*

*The authors declare no conflicts of interest.*

*Open access [CC BY 4.0](#)*



BREATHING DISCOMFORT: Dyspnoea is a subjective experience and is often underestimated by clinicians. *Illustration photo: Doug Olsson / Mostphotos*

1. Sjursæther EA, Vatnan A, Helland KF, Buanes EA. Årsrapport for 2023 med plan for forbedringstiltak (Annual report of 2023) [Internet]. Bergen: Norsk intensiv- og pandemiregister (Norwegian Intensive Care and Pandemic Registry); 2024 [cited 16 May 2025]. Available from: <https://www.helse-bergen.no/4a53e5/siteassets/seksjon/intensivregister/documents/arsrapporter/arsrapporter-nipar/nipar-arsrapport-2023.pdf>
2. Schmidt M, Banzett RB, Raux M, Morélot-Panzini C, Dangers L, Similowski T, et al. Unrecognized suffering in the ICU: addressing dyspnea in mechanically ventilated patients. *Intensive Care Med.* 2014;40(1): 1–10. DOI: [10.1007/s00134-013-3117-3](https://doi.org/10.1007/s00134-013-3117-3)
3. Chanques G, Nelson J, Puntillo K. Five patient symptoms that you should evaluate every day. *Intensive Care Med.* 2015;41(7):1347–50. DOI: [10.1007/s00134-015-3729-x](https://doi.org/10.1007/s00134-015-3729-x)
4. Demoule A, Similowski T. Respiratory suffering in the ICU: time for our next great cause. *Am J Respir Crit Care Med.* 2019;199(11):1302–4. DOI: [10.1164/rccm.201812-2248ED](https://doi.org/10.1164/rccm.201812-2248ED)
5. Parshall MB, Schwartzstein RM, Adams L, Banzett RB, Manning HL, Bourbeau J, et al. An official American Thoracic Society statement: update on the mechanisms, assessment, and management of dyspnea. *Am J Respir Crit Care Med.* 2012;185(4):435–52. DOI: [10.1164/rccm.201111-2042ST](https://doi.org/10.1164/rccm.201111-2042ST)

6. Gentzler ER, Derry H, Ouyang DJ, Lief L, Berlin DA, Xu CJ, et al. Underdetection and undertreatment of dyspnea in critically ill patients. *Am J Respir Crit Care Med*. 2019;199(11):1377–84. DOI: [10.1164/rccm.201805-0996OC](https://doi.org/10.1164/rccm.201805-0996OC)
7. Schmidt M, Demoule A, Polito A, Porchet R, Aboab J, Siami S, et al. Dyspnea in mechanically ventilated critically ill patients. *Crit Care Med*. 2011;39(9):2059–65. DOI: [10.1097/CCM.0b013e31821e8779](https://doi.org/10.1097/CCM.0b013e31821e8779)
8. Demoule A, Hajage D, Messika J, Jaber S, Diallo H, Coutrot M, et al. Prevalence, intensity, and clinical impact of dyspnea in critically ill patients receiving invasive ventilation. *Am J Respir Crit Care Med*. 2022;205(8):917–26. DOI: [10.1164/rccm.202108-1857OC](https://doi.org/10.1164/rccm.202108-1857OC)
9. Saltnes-Lillegård C, Rustøen T, Beitland S, Puntillo K, Hagen M, Lerdal A, et al. Self-reported symptoms experienced by intensive care unit patients: a prospective observational multicenter study. *Intensive Care Med*. 2023;49(11):1370–82. DOI: [10.1007/s00134-023-07219-0](https://doi.org/10.1007/s00134-023-07219-0)
10. Decavèle M, Rozenberg E, Niérat MC, Mayaux J, Morawiec E, Morélot-Panzini C, et al. Respiratory distress observation scales to predict weaning outcome. *Crit Care*. 2022;26(1):162. DOI: [10.1186/s13054-022-04028-7](https://doi.org/10.1186/s13054-022-04028-7)
11. Demoule A, Decavele M, Antonelli M, Camporota L, Abroug F, Adler D, et al. Dyspnoea in acutely ill mechanically ventilated adult patients: an ERS/ESICM statement. *Intensive Care Med*. 2024;50(2):159–80. DOI: [10.1007/s00134-023-07246-x](https://doi.org/10.1007/s00134-023-07246-x)
12. Decavèle M, Similowski T, Demoule A. Detection and management of dyspnea in mechanically ventilated patients. *Curr Opin Crit Care*. 2019;25(1):86–94. DOI: [10.1097/mcc.0000000000000574](https://doi.org/10.1097/mcc.0000000000000574)
13. Karlsen MW, Ølnes MA, Heyn LG. Communication with patients in intensive care units: a scoping review. *Nurs Crit Care*. 2019;24(3):115–31. DOI: [10.1111/nicc.12377](https://doi.org/10.1111/nicc.12377)
14. Binks AP, Desjardin S, Riker R. ICU clinicians underestimate breathing discomfort in ventilated subjects. *Respir Care*. 2017;62(2):150–5. DOI: [10.4187/respcare.04927](https://doi.org/10.4187/respcare.04927)
15. Haugdahl HS, Storli SL, Meland B, Dybwik K, Romild U, Klepstad P. Underestimation of patient breathlessness by nurses and physicians during a spontaneous breathing trial. *Am J Respir Crit Care Med*. 2015;192(12):1440–8. DOI: [10.1164/rccm.201503-0419OC](https://doi.org/10.1164/rccm.201503-0419OC)
16. Decavèle M, Campbell ML, Persichini R, Morélot-Panzini C, Similowski T, Demoule A, et al. Management of dyspnea in the noncommunicative patients: consider hetero-evaluation scales. *Chest*. 2018;154(4): 991–2. DOI: [10.1016/j.chest.2018.05.046](https://doi.org/10.1016/j.chest.2018.05.046)
17. Persichini R, Gay F, Schmidt M, Mayaux J, Demoule A, Morélot-Panzini C, et al. Diagnostic accuracy of respiratory distress observation scales as surrogates of dyspnea self-report in Intensive Care Unit patients. *Anesthesiology*. 2015;123(4):830–7. DOI: [10.1097/aIn.0000000000000805](https://doi.org/10.1097/aIn.0000000000000805)

18. Decavèle M, Gay F, Persichini R, Mayaux J, Morélot-Panzini C, Similowski T, et al. The mechanical ventilation-respiratory distress observation scale as a surrogate of self-reported dyspnoea in intubated patients. *Eur Respir J*. 2018;52(4):1800598. DOI: [10.1183/13993003.00598-2018](https://doi.org/10.1183/13993003.00598-2018)
19. Campbell ML, Templin T, Walch J. A respiratory distress observation scale for patients unable to self-report dyspnea. *J Palliat Med*. 2010;13(4):285–90. DOI: [10.1089/jpm.2009.0229](https://doi.org/10.1089/jpm.2009.0229)
20. Aikawa G, Imanaka R, Sakuramoto H, Hatozaki C, Unoki T, Okamoto S. Assessment of dyspnea in critically ill patients: a comparative analysis of evaluation scales. *Cureus*. 2024;16(1):e52751. DOI: [10.7759/cureus.52751](https://doi.org/10.7759/cureus.52751)
21. Decavèle M, Rivals I, Persichini R, Mayaux J, Serresse L, Morélot-Panzini C, et al. Prognostic value of the Intensive Care Respiratory Distress Observation Scale on ICU admission. *Respir Care*. 2022;67(7):823–32. DOI: [10.4187/respcare.09601](https://doi.org/10.4187/respcare.09601)
22. Mookink LB, Prinsen CA, Patrick DL, Alonso J, Bouter LM, Vet HCd, et al. COSMIN study design checklist for patient-reported outcome measurement instruments: Amsterdam University Medical Centers; 2019 [cited 10 May 2025]. Available from: [https://www.cosmin.nl/wp-content/uploads/COSMIN-study-designing-checklist\\_final.pdf](https://www.cosmin.nl/wp-content/uploads/COSMIN-study-designing-checklist_final.pdf)
23. Von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. Strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. *BMJ*. 2007;335(9596):806–8. DOI: [10.1136/bmj.39335.541782.AD](https://doi.org/10.1136/bmj.39335.541782.AD)
24. Wild D, Grove A, Martin M, Eremenco S, McElroy S, Verjee-Lorenz A, et al. Principles of good practice for the translation and cultural adaptation process for patient-reported outcomes (PRO) measures: report of the ISPOR task force for translation and cultural adaptation. *Value in Health*. 2005;8(2):94–104. DOI: <https://doi.org/10.1111/j.1524-4733.2005.04054.x>
25. Batalden PB, Davidoff F. What is 'quality improvement' and how can it transform healthcare? *Qual Saf Health Care*. 2007;16(1):2. DOI: [10.1136/qshc.2006.022046](https://doi.org/10.1136/qshc.2006.022046)
26. Ministry of Health and Care Services [Helse- og omsorgsdepartementet]. Report from the interregional working group on intensive care capacity [Rapport fra interregional arbeidsgruppe for intensivkapasitet]. Helse Midt-Norge, Helse Nord, Helse Sør-Øst, Helse Vest; 2022 [cited 13 November 2025]. Available from: <https://www.regjeringen.no/contentassets/859ef0b02c3248568fa820e4bcced1ab/vedlegg-1-rapport-interregional-arbeidsgruppe-for-intensivkapasitet-mai-2022.pdf>.
27. Cicchetti DV, Fleiss JL. Comparison of the null distributions of weighted kappa and the C ordinal statistic. *Appl Psychol Meas*. 1977;1(2):195–201. DOI: [10.1177/014662167700100206](https://doi.org/10.1177/014662167700100206)

28. Gwet KL. Handbook of inter-rater reliability: analysis of quantitative ratings. 5th ed. Gaithersburg, MD: AgreeStat Analytics; 2021.
29. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med. 2016;15(2):155–63. DOI: [10.1016/j.jcm.2016.02.012](https://doi.org/10.1016/j.jcm.2016.02.012)
30. Gwet KL. Handbook of inter-rater reliability: analysis of categorical ratings. 5th ed. Gaithersburg, MD: AgreeStat Analytics; 2021.
31. Altman DG. Practical statistics for medical research. London: Chapman & Hall/CRC; 1991.
32. Carkeet A. Exact parametric confidence intervals for Bland-Altman limits of agreement. Optom Vis Sci. 2015; 92(3):e71–80. DOI: [10.1097/OPX.0000000000000513](https://doi.org/10.1097/OPX.0000000000000513)
33. De Vet HC, Mokkink LB, Terwee CB, Hoekstra OS, Knol DL. Clinicians are right not to like Cohen's  $\kappa$ . BMJ. 2013;346:f2125. DOI: [10.1136/bmj.f2125](https://doi.org/10.1136/bmj.f2125)
34. Gélinas C, Arbour C. Behavioral and physiologic indicators during a nociceptive procedure in conscious and unconscious mechanically ventilated adults: similar or different? J Crit Care 2009;24(4):628:e7–17. DOI: [10.1016/j.jcrc.2009.01.013](https://doi.org/10.1016/j.jcrc.2009.01.013)
35. Tinti S, Destrebecq A, Terzoni S, De Maria B, Falcone G, Da Col D, et al. Respiratory distress observation scale Italian version: cultural-linguistic validation and psychometric properties. J Hosp Palliat Nurs. 2021;23(2):187–94. DOI: [10.1097/njh.0000000000000736](https://doi.org/10.1097/njh.0000000000000736)
36. Campbell ML. Fear and pulmonary stress behaviors to an asphyxial threat across cognitive states. Res Nurs Health. 2007;30(6):572–83. DOI: [10.1002/nur.20212](https://doi.org/10.1002/nur.20212)
37. Giavarina D. Understanding bland Altman analysis. Biochem Med (Zagreb). 2015;25(2):141–51. DOI: [10.11613/bm.2015.015](https://doi.org/10.11613/bm.2015.015)
38. Zec S, Soriani N, Comoretto R, Baldi I. High agreement and high prevalence: the paradox of Cohen's kappa. Open Nurs J. 2017;11(Suppl-1, M5):211–8. DOI: [10.2174/1874434601711010211](https://doi.org/10.2174/1874434601711010211)
39. Richardson BR, Decavèle M, Demoule A, Murtagh FEM, Johnson MJ. Breathlessness assessment, management and impact in the intensive care unit: a rapid review and narrative synthesis. Ann Intensive Care. 2024;14(1):107. DOI: [10.1186/s13613-024-01338-7](https://doi.org/10.1186/s13613-024-01338-7)