

Measurement instruments for breathlessness in palliative care

The patient's experience of breathlessness often do not correspond with the seriousness of the condition.

AUTHORS

Kjersti Solvåg

Spesial-/fagutviklingssykepleier

Sunniva senter for lindrende behandling Haraldsplass Diakonale Sykehus

Kristine Berg Titlestad

Høgskolelektor

Avdeling for helse- og sosialfag Høgskolen i Bergen

Kjetil Gundro Brurberg

Førstemanuensis

Senter for kunnskapsbasert praksis, HiB og forsker ved Nasjonalt kunnskapssenter for helsetjenesten

SUMMARY

Background: Breathlessness is a common and distressing symptom for many patients with advanced diseases. Due to the subjective nature of breathlessness, patient-reported outcome measures (PROMs) are required to measure the patient's own experience.

Objective: To identify PROMs used to measure breathlessness in palliative care and to synthesise their measurement properties. Instruments had to include dimensions for breathlessness and anxiety to be considered.

Method: A systematic literature search was performed in March 2014 and updated in December 2015. Two reviewers independently screened all references for relevance and quality assured these by means of the COSMIN-checklist. We performed a best evidence synthesis to summarise the measurement properties of each included PROM.

Results: We screened 1948 references for relevance, and included 15 studies evaluating the measurement properties of four different PROMs: CDS, DMQ, SRI and a respiratory symptom checklist. None of the included instruments were validated directly for use in a palliative setting, but they generally showed promising measurement properties in other relevant settings. We still lack data on important measurement properties for all the available instruments, and currently, only SRI seems to be available in a Norwegian validated version. Further research is therefore needed to translate and validate the PROMs for use in palliative care in Norway.

Conclusion: Several PROMs for breathlessness and anxiety show promising measurement properties, but further research is needed before we can draw firm conclusions and before the instruments are available for use in palliative care in Norway. Our review suggests that only SRI is available in a translated and validated Norwegian version.

INTRODUCTION

Dyspnea or breathlessness is a complex symptom in many patients with advanced disease (1,2). Research has shown that 94 per cent of patients with chronic lung disease and 78 per cent of patients with lung cancer suffer from breathlessness during the last year of life (3), and that breathlessness may be associated with poor quality of life, anxiety, reduced functioning, and reduced life expectancy (1,2).

The patient's experience of breathlessness frequently does not necessarily correspond with the disease's degree of severity and objective measures (4,5). In palliative care systematic use of instruments to measure the patient's own experience is therefore important in order to identify symptoms, gather information on disease progression, and evaluate the effect of interventions (6). Data based on such measurement instruments are commonly known as patient-reported outcome measures (PROMs).

Before we started working on this survey we performed a literature search to see if reviews answering our question had already been carried out. The search resulted in two relevant hits (7,8). Both reviews were based on ten-year-old literature searches and both revealed a need for a further evaluation as to which PROMs are suitable for measuring breathlessness in palliative care (7,8). The purpose of this systematic review has been to give an overview of the PROMs available for measuring breathlessness in palliative care patients, and the instruments' measurement properties. Measurement instruments must have dimensions for both breathlessness and anxiety to be considered for inclusion.

METHOD

We have used a methodological framework called CONsensus-based Standards for the selection of health Measurement INSTRUMENTS (COSMIN). The framework was developed through an international consensus process and gives specific recommendations on terminology, taxonomy, and use of method in studies dealing with PROMs and their measurement properties (9-11).

LITERATURE SEARCH

We searched systematically for literature in the databases MEDLINE (1946–), Embase (1974–), PsycINFO (1806–), AMED (1985–), CINAHL (1982–), Cochrane Library, and SveMed+. The search strategies entailed index terms and text words adapted to each database for the categories 1) breathlessness, 2) measurement properties, and 3) palliative care, cancer, or chronic obstructive pulmonary disease (COPD). We performed citation searches for the Bausewein et al. (7) and Dorman et al. (8) reviews in Science Citation Index and scanned the reference lists of relevant publications. The searches were performed without limitations as to language and time, and peer reviewed by a medical librarian. The literature searches were performed in March 2014 and updated in December 2015. Complete search strategies are available in master's thesis (12).

SELECTION

Titles and abstracts were considered for inclusion using predetermined criteria (table 1). Articles considered relevant were accessed in full text and evaluated for inclusion. All steps in the selection process were carried out by two reviewers (KS/KGB) independently, and disagreement was resolved through discussion.

Table 1: Inclusion and exclusion criteria

	Inclusion criteria	Exclusion criteria
Population	Adult (≥ 18 years) palliative patients suffering from breathlessness	Patients with acutely developed breathlessness, curable diseases
Setting	In hospital, day patients, outpatient clinic, or home based health services	
Measurement instrument	PROM for measuring breathlessness, including at least dimensions for breathlessness and anxiety	Generic PROMs. Measurement instruments developed for use in connection with exercise or in emergencies
Measurement properties	In the categories reliability, validity, responsiveness	

PROM (Patient Reported Outcome Measures)

MEASUREMENT PROPERTIES

Before using new measurement instruments it is important to ensure that the instruments have good measurement properties, i.e. that they measure what they are supposed to measure, and that the results are trustworthy (11). In the COSMIN taxonomy the measurement properties are divided into three main categories: reliability, validity, and responsiveness. Reliability refers to whether any sources of error are sufficiently small and if the results are sufficiently stable for the measurements to be trusted. In the COSMIN taxonomy reliability consists of three subcategories, called internal consistency, reliability (here test-retest reliability), and measurement errors. Validity describes whether an instrument measures the properties it is supposed to measure, such as whether persons who score differently on a depression scale, also score differently on other similar scales. The COSMIN taxonomy defines three subcategories of validity properties (content validity, criterion validity, and construct validity), where construct validity is further divided into the three categories structural validity, hypotheses testing, and cross-cultural validity. Responsiveness is a measure of whether the measurement instrument is able to capture important changes over time (10).

EVALUATION OF METHODOLOGICAL QUALITY

We evaluated methodological quality and risk of bias in accordance with the COSMIN checklist (13,14). The checklist consists of a variety of forms to be filled in to assess whether the measurement properties reported in a validation study are trustworthy. Each validation study generally measures a limited number of measurement properties, and only forms with relevance for the measurement properties that the study seeks to evaluate are completed. The studies' quality is assessed as excellent, good, fair, or poor. All quality assessments were carried out by two reviewers (KS/KBT), independently. Disagreement was resolved by discussion or by involving a third reviewer (KGB).

DATA EXTRACTION AND CATEGORISATION OF MEASUREMENT PROPERTIES

One reviewer (KS) went through all included studies and registered background information on the study, participants, measurement instrument, measurement properties, methodological quality, and outcome using a data extraction form. The data extraction was

then quality assured by another reviewer (KGB). For each study we evaluated the measurement properties of the relevant instrument as positive, indeterminate, or negative against a set of predetermined criteria (appendix I).

DOCUMENTATION ON THE INSTRUMENTS' PROPERTIES

We summarized the total documentation on the instruments' measurement properties in a best evidence synthesis taking into consideration methodological quality, measurement properties, and degree of consistency. Documentation quality is rated as strong, moderate, limited, conflicting, or unknown, based on the criteria listed in appendix II.

RESULTS

We identified 1948 references, and ended up including 15 studies (15-29) (figure 1). Table 2 gives an overview of the studies included.

Figure 1: Flow chart on identified literature

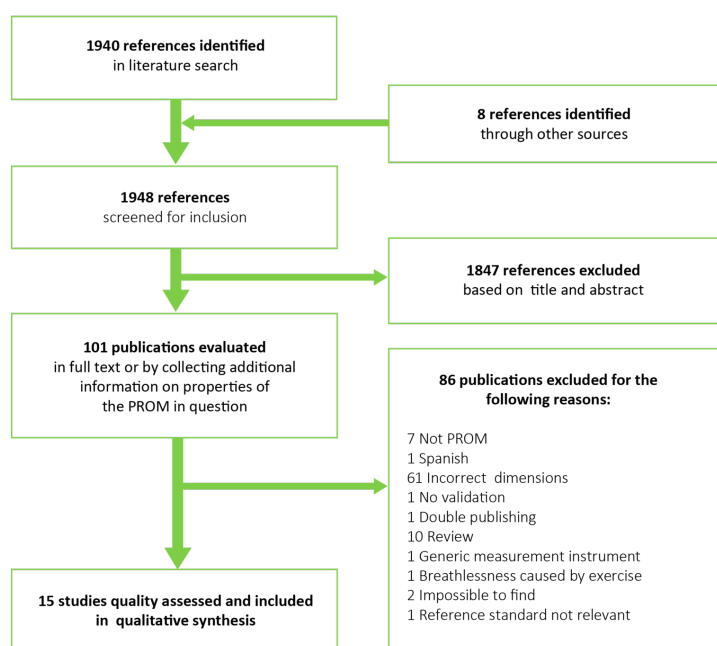


Table 2: Characteristics of included studies

PROM	Publications	Sample	N	Country	Setting	
CDS	Tanaka et al. (15)	Lung cancer	166	Japan	Outpatient clinic Hospital	
	Tanaka et al. (16)	Lung cancer	171	Japan	Outpatient clinic	
	Henoch et al. (17)	Lung cancer	99	Sweden	Outpatient clinic	
	Uronis et al. (18)	Lung cancer	109	USA, UK, AUS	Hospital*	
SRI	Windisch et al. (19)	Respiratory failure. LTMV. Various diagnoses	226	Germany	Hospital	
	Duierveman et al. (20)	Respiratory failure. COPD	72	The Netherlands	Not listed	
	Lopez-Campos et al. (21)	LTMV. Various diagnoses	115	Spain	Outpatient clinic	
	Windisch et al. (22)	Respiratory failure. LTMV. COPD	162	Germany	Hospital	
	Ghosh et al. (23)	Respiratory failure. LTMV. Various diagnoses	152	England	Outpatient clinic	
	Struik et al. (24)	Respiratory failure. COPD	170	The Netherlands	Hospital	
	Markussen et al. (25)	Respiratory failure. LTMV. Various diagnoses	127	Norway	In the home	
	DMQ-30	Norweg et al. (26)	COPD and asthma	85	USA	Not listed
	DMQ-56	Norweg et al. (27)	COPD	107	USA	Not listed
DMQ-CAT	Norweg et al. (28)	COPD	292	USA	Outpatient clinic	
RSC	Han et al. (29)	Heart/lung diseases	396	China	Hospital*	

PROM (Patient-Reported Outcome Measures); CDS (Cancer Dyspnea Scale); SRI (Severe Respiratory Insufficiency Questionnaire); DMQ (Dyspnea Management Questionnaire); CAT (Computer Adaptive Test); RSC (Respiratory Symptom Checklist); LTMV (Long-Term Mechanical Ventilation); COPD (Chronic Obstructive Pulmonary Disease); USA (United States of America); UK (United Kingdom); AUS (Australia); * (unclear whether hospitalised, outpatient clinic or both).

We identified four relevant measurement instruments: Cancer Dyspnea Scale (CDS), Severe Respiratory Insufficiency (SRI) Questionnaire, Dyspnea Management Questionnaire (DMQ), and Respiratory Symptom Checklist (RSC). The instruments' measurement properties are summarized in table 3.

Table 3: Characteristics of included PROMs

PROM	Target group	Number of questions	Dimensions	Measurement scale (per question)	Scoring system	Language	Expected completion time
CDS	Cancer	12	Sense of effort (5 qu) Sense of anxiety (4 qu) Sense of discomfort (3 qu) Total score	5-point Likert	Calculates total score per dimension: effort (0–20); anxiety (0–16); discomfort (0–12). Higher score indicates more severe breathlessness.	Japanese English Swedish	2 minutes
SRI	Chronic respiratory failure with LTMV treatment	49	Respiratory complaints (8 qu) Physical function (6 qu) Attendant symptoms and sleep (7 qu) Social relationships (6 qu) Anxiety (5 qu) Psychological well-being (9 qu) Social function (8 qu) Total score	5-point Likert	Calculates score per dimension as percentage of maximum (0–100). Higher score indicates better quality of life.	German Dutch Spanish English Norwegian	20 minutes
DMQ-30	COPD (rehabilitation)	30	Dyspnea intensity (7 qu) Dyspnea anxiety (7 qu) Fearful activity avoidance (7 qu) Self-efficacy for activity (5 qu) Satisfaction with strategy use (4 qu)	7-point Likert	Calculates average score per dimension (0–6). Total score calculated by summarising the dimensions intensity, anxiety and activity-avoidance. Higher score indicates better function.	English	Not given
DMQ-56	COPD (rehabilitation)	56	Dyspnea intensity (12 qu) Dyspnea-related anxiety (16 qu) Activity avoidance (15 qu) Self-efficacy for activity (6 qu) Satisfaction with strategy use (7 qu)	7-point Likert	Calculates average score per dimension (0–6). Higher score indicates better function.	English	Not given
DMQ-CAT	COPD (rehabilitation)	20	Dyspnea intensity Dyspnea-related anxiety Activity avoidance Self-efficacy for activity	6-point Likert	Not given	English	Not given
RSC	Breathlessness as result of heart/lung disease	46	Sensory qualities (23 qu) Affective aspect (12 qu) Behavioural impact (4 qu) Other symptoms (7 qu)	5-point Likert	Not given	Not given	10–15 minutes

PROM (Patient-Reported Outcome Measures); CDS (Cancer Dyspnea Scale); SRI (Severe Respiratory Insufficiency Questionnaire); DMQ (Dyspnea Management Questionnaire); CAT (Computer Adaptive Test); RSC (Respiratory Symptom Checklist); LTMV (LongTerm Mechanical Ventilation); COPD (Chronic Obstructive Lung Disease); qu (questions).

Table 4 shows our assessment of methodological quality and risk of bias, and the measurement properties of the included studies. A more detailed overview of various properties is available in master's thesis (12). The best evidence synthesis that summarize results across all included studies is presented in table 5. In the following we summarize the main results for each of the four measurement instruments.

Table 4: Overview of methodological quality and measurement properties of the studies included

PROM	Study	Internal consistency		Reliability		Content validity		Structural validity		Hypotheses testing		Criterion validity		Cross-cultural validity	
		M	R	M	R	M	R	M	R	M	R	M	R	M	R
CDS	Tanaka et al. (15)	Excellent	+	Fair	-	Excellent	+	Excellent	?	Good	+				
	Tanaka et al. (16)									Good	-				
	Henoch et al. (17)	Good	+					Excellent	?	Good	+				
	Uronis et al. (18)	Excellent	+					Good	?	Good/Fair	-				
SRI	Windisch et al. (19)					Excellent	+			Good	+				
	Duierveman et al. (20)	Poor	+	Poor	+					Good	+				
	Lopez-Campos et al. (21)	Poor	+	Fair	+			Poor	+	Fair	-				
	Windisch et al. (22)	Poor	+					Poor	+						
	Ghosh et al. (23)	Poor	+					Poor	+	Good	+				
	Struik et al. (24)	Poor	+							Good	+				
	Markussen et al. (25)	Poor	+			Good	+			Good	+			Poor	-
DMQ-30	Norweg et al. (26)	Good	+	Poor	+	Excellent	+			Good/Poor''	+				
DMQ-56	Norweg et al. (27)	Good	+	Fair	+			Poor	+						
DMQ-CAT	Norweg et al. (28)	Good	+			Excellent	+	Fair	+	Fair	+	Good	+		
RSC	Han et al. (29)							Excellent	+						

included PROM (Patient Reported Outcome Measures); M (methodological quality); R (results of measurement properties); + (positive result); ? (indeterminate result); - (negative result); CDS (Cancer Dyspnea Scale); SRI (Severe Respiratory Insufficiency Questionnaire); DMQ (Dyspnea Management Questionnaire); CAT (Computer Adaptive Test); RSC (Respiratory Symptom Checklist); ' (comparisons with VAS (Visual Analog Scale), BORG, ECOG (Eastern Cooperative Oncology Group) / HADS (Hospital Anxiety and Depression Scale), FACT (Functional Assessment of Cancer Therapy), SpO2 (peripheral oxygen saturation); '' (comparisons with SOLQ (Seattle Obstructive Lung Disease Questionnaire), SF-12 (Short Form Health Survey) / HADS-anxiety

Table 5: Best evidence synthesis

	Internal consistency	Reliability	Content validity	Structural validity	Hypotheses testing	Criterion validity	Cross-cultural validity
CDS	+++	-	+++	na	+/-	na	na
SRI	?	+	+++	?	+++	na	?
DMQ-30	++	?	+++	na	++	na	na
DMQ-56	++	+	na	?	na	na	na
DMQ-CAT	++	na	+++	+	+	++	na
RSC	na	na	na	+++	na	na	na

CDS (Cancer Dyspnea Scale); SRI (Severe Respiratory Insufficiency Questionnaire); DMQ (Dyspnea Management Questionnaire); CAT (Computer Adaptive Test); RSC (Respiratory Symptom Checklist); +++ or -- strong evidence positive/negative result; ++ or - moderate evidence positive/negative result; + or -, limited evidence positive/negative result; ± conflicting evidence; ? unknown due to poor methodological quality; na, no information available.

CANCER DYSPNEA SCALE (CDS)

CDS was developed in Japan for cancer patients and measures aspects of breathing difficulties (15). Development and validation are described in four publications (15-18) (table 2).

Representatives for the patient group and professional experts were involved in the development of CDS, and the measurement instrument was pilot tested in the target population (15). We assess the content validity of CDS as being probably very good (table 5).

Three studies (15,17,18) confirm that CDS has a three-factor structure (table 3), but we do not have sufficient data to assess the instruments' structural validity (table 5).

For hypotheses testing we have emphasised the comparisons between CDS and Hospital Anxiety and Depression Scale (HADS), Borg scale, and visual analogous scale (VAS-dyspnea).

The result for hypotheses testing was negative in two studies (16,18) and positive in two studies (15,17) (table 4). As such the results are contradictory, which makes it difficult to conclude univocally (table 5).

«Breathlessness results in poor quality of life, anxiety, lowered level of functioning and reduction of life expectancy.»

The internal consistency of CDS was reported as positive in three studies (15,17,18) (table 4), and we assess the internal consistency of CDS as probably very good (table 5). The result for test-retest reliability was negative (15) (table 4), but we have limited trust in the available documentation (table 5).

SEVERE RESPIRATORY INSUFFICIENCY (SRI) QUESTIONNAIRE

SRI was developed in Germany to measure health related quality of life in patients treated with long-term mechanical ventilation (LTMV) due to chronic respiratory failure. The respiratory failure is due to various underlying diseases (19). Seven publications (19-25) have described development and validation of the instrument (table 2). The samples of two publications (19,22) overlap in part, and data for structural validity and internal consistency is thus taken from the Windisch et al. study (22).

SRI is developed from social, psychological, and physical health domains, and both patients and professional experts were involved (19). The Norwegian version of SRI was pilot tested among users of LTMV (25), and the content validity is probably very good (table 5).

The factor structure is evaluated in three studies (21,23), which all found that the original seven dimensions consisted of several factors. Two studies (22,23) showed that the factors within one dimension corresponded, and the authors thus chose to keep the original structure of seven dimensions. One study (21) found 13 factors. All three studies showed positive properties for structural validity (table 4). There is nevertheless uncertainty attached to the structural validity (table 5), first and foremost due to methodological limitations in the studies where this is assessed.

For hypotheses testing we chose to emphasise the comparisons between SRI and Short Form Health Survey (SF-36), Chronic Respiratory Disease Questionnaire (CRQ), HADS, and Medical Research Council Dyspnea Scale (MRC) (19-21, 23-25). In general the correlation was strongest between SRI dimensions and dimensions of the other instruments that measure related aspects, while the correlation was weaker between dimensions that measure different aspects. Total documentation shows that SRI probably correlates very well with other PROMs that measure similar properties (table 5).

«Many patients with advanced serious disease experience breathlessness.»

SRI is the only instrument of the four included in our review that is available in a Norwegian version (25). We therefore wished to evaluate the cross-cultural validity of the Norwegian version. The absence of a comparison with the original version for factor structure and a low number of respondents compared to the number of questions in the measurement instrument made us conclude that there is a need for more research before we can say anything certain about the cross-cultural validity of the Norwegian version (table 5).

SRI (20-25) showed overall positive properties for internal consistency (table 4); however, due to methodological limitations, uncertainties attach to the total documentation (table 5). Test-retest reliability was assessed as positive in two studies (20,21) (table 4), but methodological weaknesses in the two studies leave us with limited trust in the total documentation (table 5).

DYSPNEA MANAGEMENT QUESTIONNAIRE (DMQ)

DMQ was developed in USA to measure the effect of lung rehabilitation and change over time in patients with COPD. The purpose was also to accommodate for more directed treatment for professions such as ergotherapy and psychology (26). Three studies describe development and validation of three different versions: DMQ-30 (26), DMQ-56 (27) and DMQ-CAT (28) (table 2).

Data from qualitative interviews, literature review on the areas of breathlessness, anxiety, avoidance behaviour, functional status, health related quality of life, user satisfaction, and lung rehabilitation, as well as a review of other measurement instruments, made up the basis for the development of DMQ-30. Both professional experts and patients were involved and a preliminary version was pilot tested among adults with COPD (26). DMQ-CAT is an electronic version and was developed based on an expanded collection of the questions from DMQ-56. Involvement of professional experts and patients along with an extensive literature review made up the basis for the expansion (28). The documentation shows that DMQ-30 and DMQ-CAT probably have excellent content validity (table 5).

The original five-factor model for DMQ-56 was confirmed (27). The assessment of the factor structure of DMQ-CAT (28) showed that a four-factor model was better suited. For both versions the factors explained more than 50 per cent of the variance, but even if the result for structural validity points in a positive direction for both DMQ-56 and DMQ-CAT, the quality of the documentation is limited (27,28) (table 4). It is therefore difficult to draw any certain over-all conclusions (table 5).

«No currently available measurement instrument is validated for use in measuring palliative care patients' experience of breathlessness.»

For hypotheses testing DMQ-30 is compared to Seattle Obstructive Lung Disease Questionnaire (SOLQ), Short Form Health Survey (SF-12), and HADS (26). DMQ-CAT was compared to University of California, San Diego Shortness of Breath Questionnaire (UCSD SOBQ), CRQ, COPD Self-efficacy Scale (CSES), and HADS (28). For both versions the strongest correlation was between the DMQ dimensions and dimensions for the other instruments that measure related aspects, and the weakest correlation was between dimensions that measure different aspects. The documentation shows that we may have moderate trust in that DMQ-36 correlates well with other PROMs that measure similar properties, while we have limited trust in that DMQ-CAT correlates well with other PROMs intended to measure similar properties (table 5).

Comparisons of DMQ-CAT with the total number of questions suggested high correlation for all dimensions, with Pearson's correlation coefficient 0.94-0.97 (28). In total we have moderate trust in the criterion validity of DMQ-CAT being good (table 5).

The internal consistency was positive for all three versions of DMQ (table 4). The documentation shows that we may have moderate trust in the internal consistency of DMQ-30, DMQ-56, and DMQ-CAT being good (table 5).

Test-retest reliability was positive for DMQ-30 (26) and DMQ-56 (27) (table 4). The documentation shows that we may have limited trust in the test-retest reliability being good for DMQ-56 (table 5). More uncertainty is linked to test-retest reliability for DMQ-30, due first and foremost to methodological limitations in the study where this was measured.

RESPIRATORY SYMPTOM CHECKLIST (RSC)

RSC was developed in China to measure multidimensional aspects of breathing difficulties in patients with heart and lung disease (29).

Assessment of the factor structure showed 12 consecutive and non-correlated factors. By adding additional criteria Han et al. (29) found seven factors that seem to measure three dimensions of breathlessness (table 3), while two remaining factors sorted under other symptoms. In sum the nine factors explained 64 per cent of the total variance (29). A total assessment indicates that the structural validity of RSC is probably excellent; we lack, however, knowledge on other important measurement properties (table 5).

DISCUSSION

We identified four PROMs for measuring breathlessness that satisfied our inclusion criteria: CDS, SRI, DMQ, and one respiratory symptom checklist, of which only SRI was available in a Norwegian validated version. The measurement properties of the four instruments were assessed in 15 studies. None of the included PROMs were validated directly for use in a palliative care setting, but on the whole they show promising properties for other relevant settings. The respiratory symptom checklist stand out as having been studied the least; there is, however, a need for more research to clarify important measurement properties for the other instruments as well.

STRENGTHS AND WEAKNESSES OF THE STUDY

We have worked out this systematic review in line with recommendations from the COSMIN initiative, which takes into account the special aspects of validating PROMs. The review builds on a wide and systematic literature search. All steps of the process were carried out by two reviewers independently, or carried out by one reviewer and quality assured by another person. Missing elements in the abstracts and the indexing of studies on measurement properties constitute a challenge to literature searches for reviews of PROMs (9). One weakness in our literature search is that search terms for respiratory failure were not included. Several of the studies on SRI were thus first identified in a review of relevant studies. This entails a risk that other relevant PROMs exist that we have not identified. We have not had capacity to search for “grey literature”, carry out supplemental searches on included PROMs, or contact professionals. All in all this entails a risk of there being relevant publications that we have not identified.

We evaluated the methodological quality of the included studies using the COSMIN checklist after first having carried out a thorough pilot test. The checklist enabled us to give a separate assessment of the various measurement properties that were assessed in one and the same study. A challenge has been that we have not always had the possibility of distinguishing between incomplete reporting and poor quality of a study. We have not had the capacity to contact authors for settling such issues.

THE METHODOLOGICAL QUALITY OF THE STUDIES

Two main challenges attaching to the included studies' methodological quality is sample size and incomplete reporting on hypotheses. The size of the samples was often too small, and individually the studies did not have enough respondents to assess test-retest reliability. When internal consistency and structural validity are to be assessed, recommended sample size is five to seven times the number of questions and minimum 100 (14,30). This entails

that instruments with numerous questions, such as SRI, score low on internal consistency and structural validity for all the included studies. Terwee et al. (31), who have summarized methodological quality in studies on measurement properties in a systematic review, also found that the sample size was a considerable challenge to the evaluation of many measurement properties.

«We found only one measurement instrument available in a Norwegian validated version.»

Several of the studies did not have pre-formulated hypotheses, or they were inadequately formulated with regard to expected size or direction of the correlations. It was not always clear whether this was due to shortcomings in the study or in the reporting. Terwee et al. (31) found in their study that less than 50 per cent of the studies had predefined hypotheses and only 48 per cent of these had described expected direction and size of the correlations. When hypotheses are missing there is a risk of coming up with alternative explanations of the results and the risk of bias increases (9).

IMPLICATIONS FOR PRACTICE AND FURTHER RESEARCH

Summary of the results and the best evidence synthesis showed that we have limited trust in the documentation and lack knowledge on important measurement properties for all instruments included. That we have limited trust in the documentation does not mean that the measurement properties were poor, but that they are still indeterminate (31). Even if some of the identified measurement instruments are more thoroughly studied than others, it is still difficult to give a clear answer to whether any of the available instruments have better measurement properties than others.

For PROMs that are to be used in palliative care patients in clinical practice, it is important that measurement instruments are not too extensive or demand too many resources for completion and administration (32). Here CDS stands out with 12 questions and an expected time of completion of two minutes (15). DMQ-CAT tailors the questions to the respondent based on one entry question (28). The demands on the patient will thus be reduced in that only the most informative and relevant questions have to be answered. Using DMQ-CAT requires access to and competence in using electronic instruments.

None of the instruments we identified are validated for use in palliative care patients independently of diagnosis, and only SRI is translated into Norwegian. In addition to the methodological weaknesses pointed out and measurement properties that have not yet been evaluated, this entails that the relevant PROMs must be validated further in high quality studies. The studies should have large enough samples, pre-formulated hypotheses, as well as good and sufficient reporting on the study's completion. It is also important to assess the measurement instruments' responsiveness, to gain knowledge on whether the instrument is capable of capturing important differences such as effect of treatment or deterioration of health condition. COSMIN may be used to advantage as guidance for planning and reporting of new validation studies.

CONCLUSION

Four measurement instruments for breathlessness satisfied our inclusion criteria. Of the three instruments most closely studied (CDS, SRI and DMQ), none stand out as having clearly better validity and measurement properties, but CDS and DMQ-CAT seem more user friendly when it comes to time of completion and demands on the patients. SRI is the only instrument available in a Norwegian validated version, which is an important condition for use in a Norwegian setting. Further research is needed to validate and make the

measurement instruments available for use in a palliative care setting in Norway.

We want to thank senior adviser and medical librarian Hilde Strømme for her peer review of our literature searches.

Appendix I: Quality criteria for measurement properties (33, 35)

Measurement property	Rating Quality	Criteria
Reliability		
Internal consistency	+	(Sub)scale unidimensional AND Cronbach's alpha(s) ≥ 0.70
	?	Dimensionality not known OR Cronbach's alpha not determined
	-	(Sub)scale not unidimensional OR Cronbach's alpha(s) < 0.70
Reliability	+	ICC/weighted Kappa ≥ 0.70 OR Pearson's $r \geq 0.80$
	?	Neither ICC/weighted Kappa, nor Pearson's r determined
	-	CC/weighted Kappa < 0.70 OR Pearson's $r < 0.80$
Validity		
Content validity	+	The target population considers all items in the questionnaire to be relevant
	?	No target population involvement
	-	The target population considers all items in the questionnaire to be irrelevant OR considers the questionnaire to be incompleteOriginal
Construct validity	+	Original factor structure confirmed OR no important DIF
	?	Original factor structure confirmed AND DIF not mentioned
	-	Original factor structure not confirmed OR important DIF
Structural validity	+	Factors should explain at least 50 % of the variance
	?	Explained variance not mentioned
	-	Factors explain < 50 % of the variance
Hypothesis testing	+	Correlation with an instrument measuring the same construct ≥ 0.50 OR at least 75 % of the results is in accordance with the hypotheses AND correlation with related constructs is higher than with unrelated constructs
	?	Solely correlations determined with unrelated constructs
	-	Correlation with an instrument measuring the same construct < 0.50 OR correlation with related constructs is lower than with unrelatedconstructs

ICC (intraclass correlation coefficient); DIF (differential item functioning)
 + (positive measurement properties); ? (indeterminate measurement properties);
 - (negative measurement properties)

Appendix II: Levels of evidence for the overall quality of the measurement properties adapted from Schellingerhout et al. (33, 34)

Level	Rating	Criteria
Strong	+++ OR	Consistent findings in several studies of good methodological quality
	---	OR in one study of excellent metodological quality
Moderate	++ OR	Consistent findings in several studies of fair methodological quality OR in one study of good methodological quality
	--	
Limited	+ OR -	One study of fair methodological qualityConflicting results
Conflicting	+ / -	Conflicting results
Unknown	?	Only studies of poor methodological quality

+(positive result); -(negative result)

REFERENCES

1. Bausewein C, Jolley C, Reilly C, Lobo P, Kelly J, Bellas H, et al. Development, effectiveness and cost-effectiveness of a new out-patient Breathlessness Support Service: study protocol of a phase III fast-track randomised controlled trial. *BMC Pulm Med.* 2012;12:58.

2. Helsedirektoratet. Nasjonalt handlingsprogram for palliasjon i kreftomsorgen. 2015.
Available from:
<https://helsedirektoratet.no/Lists/Publikasjoner/Attachments/918/Nasjonalt%20handlingsprogram%20for%20palliasjon%20i%20kreftomsorgen-IS-2285.pdf>. (Downloaded 01.12.15).
3. Edmonds P, Karlsen S, Khan S, Addington-Hall J. A comparison of the palliative care needs of patients dying from chronic respiratory diseases and lung cancer. *Palliat Med*. 2001;15(4):287-95.
4. Bausewein C, Booth S, Higginson IJ. Measurement of dyspnoea in the clinical rather than the research setting. *Curr Opin Support Palliat Care*. 2008;2(2):95-9.
5. American Thoracic Society. Dyspnea: mechanisms, assessment, and management. A consensus statement. *Am J Respir Crit Care Med*. 1999;159(1):321-40.
6. Antunes B, Harding R, Higginson IJ. Implementing patient-reported outcome measures in palliative care clinical practice: a systematic review of facilitators and barriers. *Palliat Med*. 2014;28(2):158-75.
7. Bausewein C, Farquhar M, Booth S, Gysels M, Higginson IJ. Measurement of breathlessness in advanced disease: a systematic review. *Respir Med*. 2007;101(3):399-410.
8. Dorman S, Byrne A, Edwards A. Which measurement scales should we use to measure breathlessness in palliative care? A systematic review. *Palliat Med*. 2007;21(3):177-91.
9. de Vet HCW, Terwee CB, Mokkink LB, Knol DL. *Measurement in medicine: a practical guide*. Cambridge University Press, Cambridge. 2011.
10. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol*. 2010;63(7):737-45.
11. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res*. 2010;19(4):539-49.
12. Solvåg K. Kartlegging av pustebesvær hos palliative pasienter – en systematisk kunnskapsoversikt over tilgjengelige kartleggingsverktøy og deres måleegenskaper (master's thesis). Bergen University College. 2015.
13. COSMIN. COSMIN checklist with 4-point scale. COSMIN; 2011. Available from: <http://www.cosmin.nl/>. (Downloaded 01.12.15).
14. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. COSMIN checklist manual. COSMIN; 2012. Available from: <http://www.cosmin.nl/images/upload/files/COSMIN%20checklist%20manual%20v9.pdf>. (Downloaded 01.12.15).
15. Tanaka K, Akechi T, Okuyama T, Nishiwaki Y, Uchitomi Y. Development and validation of the Cancer Dyspnoea Scale: a multidimensional, brief, self-rating scale. *Br J Cancer*. 2000;82(4):800-5.

16. Tanaka K, Akechi T, Okuyama T, Nishiwaki Y, Uchitomi Y. Factors correlated with dyspnea in advanced lung cancer patients: organic causes and what else? *J Pain Symptom Manage.* 2002;23(6):490–500.
17. Hensch I, Bergman B, Gaston-Johansson F. Validation of a Swedish version of the Cancer Dyspnea Scale. *J Pain Symptom Manage.* 2006;31(4):353–61.
18. Uronis HE, Shelby RA, Currow DC, Ahmedzai SH, Bosworth HB, Coan A, et al. Assessment of the psychometric properties of an English version of the cancer dyspnea scale in people with advanced lung cancer. *J Pain Symptom Manage.* 2012;44(5):741–9.
19. Windisch W, Freidel K, Schucher B, Baumann H, Wiebel M, Matthys H, et al. The Severe Respiratory Insufficiency (SRI) Questionnaire: a specific measure of health-related quality of life in patients receiving home mechanical ventilation. *J Clin Epidemiol.* 2003;56(8):752–9.
20. Duiverman ML, Wempe JB, Bladder G, Kerstjens HA, Wijkstra PJ. Health-related quality of life in COPD patients with chronic respiratory failure. *Eur Respir J.* 2008;32(2):379–86.
21. Lopez-Campos JL, Failde I, Masa JF, Benitez-Moya JM, Barrot E, Ayerbe R, et al. Transculturally adapted Spanish SRI questionnaire for home mechanically ventilated patients was viable, valid, and reliable. *J Clin Epidemiol.* 2008;61(10):1061–6.
22. Windisch W, Budweiser S, Heinemann F, Pfeifer M, Rzehak P. The Severe Respiratory Insufficiency Questionnaire was valid for COPD patients with severe chronic respiratory failure. *J Clin Epidemiol.* 2008;61(8):848–53.
23. Ghosh D, Rzehak P, Elliott MW, Windisch W. Validation of the English Severe Respiratory Insufficiency Questionnaire. *Eur Respir J.* 2012;40(2):408–15.
24. Struik FM, Kerstjens HA, Bladder G, Sprooten R, Zijnen M, Asin J, et al. The Severe Respiratory Insufficiency Questionnaire scored best in the assessment of health-related quality of life in chronic obstructive pulmonary disease. *J Clin Epidemiol.* 2013;66(10):1166–74.
25. Markussen H, Lehmann S, Nilsen RM, Natvig GK. The Norwegian version of the Severe Respiratory Insufficiency Questionnaire. *Int J Nurs Pract.* 2014.
26. Norweg A, Whiteson J, Demetis S, Rey M. A new functional status outcome measure of dyspnea and anxiety for adults with lung disease: the dyspnea management questionnaire. *J Cardiopulm Rehabil.* 2006;26(6):395–404.
27. Norweg A, Jette AM, Ni P, Whiteson J, Kim M. Outcome measurement for COPD: reliability and validity of the Dyspnea Management Questionnaire. *Respir Med.* 2011;105(3):442–53.
28. Norweg A, Ni P, Garshick E, O'Connor G, Wilke K, Jette AM. A multidimensional computer adaptive test approach to dyspnea assessment. *Arch Phys Med Rehabil.* 2011;92(10):1561–9.
29. Han JN, Xiong CM, Yao W, Fang QH, Zhu YJ, Cheng XS, et al. Multiple dimensions of cardiopulmonary dyspnea. *Chin Med J.* 2011;124(20):3220–6.
30. Terwee CB, Mokkink LB, Knol DL, Ostelo RW, Bouter LM, de Vet HC. Rating the

methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Qual Life Res.* 2012;21(4):651–7.

31. Terwee CB, Schellingerhout JM, Verhagen AP, Koes BW, de Vet HC. Methodological quality of studies on the measurement properties of neck pain and disability questionnaires: a systematic review. *J Manipulative Physiol Ther.* 2011;34(4):261–72.

32. Bausewein C, Daveson B, Benalia H, Simon ST, Higginson IJ. Outcome measurement in palliative care: the essentials. PRISMA, London. 2011.

33. Schellingerhout JM, Heymans MW, Verhagen AP, de Vet HC, Koes BW, Terwee CB. Measurement properties of translated versions of neck-specific questionnaires: a systematic review. *BMC Med Res Methodol.* 2011;11:87.

34. Schellingerhout JM, Verhagen AP, Heymans MW, Koes BW, de Vet HC, Terwee CB. Measurement properties of disease-specific questionnaires in patients with neck pain: a systematic review. *Qual Life Res.* 2012;21(4):659–70.

35. Terwee CB, Bot SDM, de Boer MR, van der Windt DAWM, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol.* 2007;60(1):34–42.