



Validity and reliability of outcome measures to assess dysfunctional breathing: a systematic review

Vikram Mohan ,¹ Chandrasekar Rathinam ,^{2,3} Derick Yates,³ Aatit Paungmali,⁴ Christopher Boos^{5,6}

To cite: Mohan V, Rathinam C, Yates D, *et al.* Validity and reliability of outcome measures to assess dysfunctional breathing: a systematic review. *BMJ Open Respir Res* 2024;**11**:e001884. doi:10.1136/bmjresp-2023-001884

► Additional supplemental material is published online only. To view, please visit the journal online (<https://doi.org/10.1136/bmjresp-2023-001884>).

Received 23 June 2023
Accepted 15 March 2024

ABSTRACT

Objective This study aimed to systematically review the psychometric properties of outcome measures that assess dysfunctional breathing (DB) in adults.

Methods Studies on developing and evaluating measurement properties to assess DB were included. The study investigated the empirical research published between 1990 and February 2022, with an updated search in May 2023 in the Cochrane Library database of systematic reviews and the Cochrane Central Register of Controlled Trials, the Ovid Medline (full), the Ovid Excerpta Medica Database, the Ovid allied and complementary medicines database, the Ebscohost Cumulative Index to Nursing and Allied Health Literature and the Physiotherapy Evidence Database. The included studies' methodological quality was assessed using the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) risk of bias checklist. Data analysis and synthesis followed the COSMIN methodology for reviews of outcome measurement instruments.

Results Sixteen studies met the inclusion criteria, and 10 outcome measures were identified. The psychometric properties of these outcome measures were evaluated using COSMIN. The Nijmegen Questionnaire (NQ) is the only outcome measure with 'sufficient' ratings for content validity, internal consistency, reliability and construct validity. All other outcome measures did not report characteristics of content validity in the patients' group.

Discussion The NQ showed high-quality evidence for validity and reliability in assessing DB. Our review suggests that using NQ to evaluate DB in people with bronchial asthma and hyperventilation syndrome is helpful. Further evaluation of the psychometric properties is needed for the remaining outcome measures before considering them for clinical use.

PROSPERO registration number CRD42021274960.

INTRODUCTION

The normal breathing pattern consists of thoracic and abdominal cavity expansion during inhalation and retraction during exhalation.¹ Dysfunctional breathing (DB) deviates from the typical biomechanical pattern.²⁻³ Barker and Everard (2015) proposed a definition for DB as 'an alteration in the normal biomechanical patterns

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Clinicians commonly use various outcome measures to examine dysfunctional breathing (DB). Currently, no review is available that examines these outcome measures psychometric properties.

WHAT THIS STUDY ADDS

⇒ The psychometric properties of the available DB outcome measures in adults are reviewed. Nijmegen Questionnaire (NQ) is the only available outcome measure graded as 'very high' quality and evaluated by the COnsensus-based Standards for the selection of health Measurement INstruments tool.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ The existing outcome measures need to establish content validity and other psychometric properties prior to consideration for clinical use. NQ can be used to assess DB in the adult population.

of breathing that results in intermittent or chronic symptoms that may be respiratory and/or non-respiratory'.³ The DB subtypes include thoracic and extrathoracic patterns.^{2,3} Thoracic DB is often observed in hyperventilation and extrathoracic DB in patients with paradoxical vocal cord dysfunction.³ A DB has historically been identified under a variety of nomenclature; a few examples include thora-coabdominal asynchrony, breathing pattern dysfunction, breathing pattern disorder, unexplained breathlessness, psychological breathlessness, panic breathing, apical breathing, periodic deep sighing, hyperventilation and paradoxical breathing.^{3,4} DB has an estimated prevalence of 29% and 8% in people with and without asthma, respectively.⁵ This signifies that the general adult population and those with lung disease may experience DB with symptoms that may improve with treatment, contributing to improved quality of life (QoL).⁶

Several respiratory disorders, such as bronchial asthma, sleep apnoea and chronic



© Author(s) (or their employer(s)) 2024. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

Correspondence to

Chandrasekar Rathinam;
c.rathinam@bham.ac.uk



obstructive pulmonary disease, are reported to be linked with DB.⁷⁻⁹ Breathlessness, chest tightness, anxiety, light-headedness and fatigue can occur in people with these illnesses and DB.⁶⁻⁹ QoL, anxiety, sense of coherence and asthma control are significantly reduced in patients with DB, and breathing retraining has been shown to improve DB and health-related QoL.^{10 11} Even though the DB is non-specific in some instances, it can lead directly to misdiagnosing respiratory disease in many situations.⁴ Despite the clinical importance of evaluating DB, a consensus on the assessment method still needs to be reached. The potential impacts of DB on constructs like bodily biochemistry, psychological functioning and social aspects must also be considered in a comprehensive evaluation.^{6 12-14}

Clinical judgement and outcome measures enhance symptom-specific DB evaluation. An outcome measure that examines DB is required to guide suitable treatments. A range of objective evaluation instruments are available, including respiratory movement measuring instruments and respiratory inductive plethysmography.^{15 16} These laboratory-based measurement methods offer identification of DB, and they have excellent reliability and validity.^{16 17} However, these outcome measures cannot be used in routine clinical practice, especially in the community, due to time consumption, expensive equipment and the need for specific clinical environments. Clinicians often use various outcome measures to assess DB.¹⁸⁻²⁰ These include Hi-Lo breathing,²¹ the Manual Assessment of Respiratory Motion (MARM),²¹ the Self-Evaluation of Breathing Questionnaire (SEBQ),²² the Breathing Pattern Assessment Tool (BPAT),²³ the Total Faulty Breathing Scale (TFBS)²⁴ and Nijmegen Questionnaire (NQ).²⁵

The available outcome measures use various methods to detect DB. For example, in MARM, the examiners use the palpation method to detect DB²¹; Hi-Lo and TFBS assess breathing motion through observation²⁶ and NQ through self-reported measures.^{15 21 25} Before any outcome measure is viable for routine clinical practice, validity and reliability must be established to ensure clinicians' confidence in the measurement. To determine best practices for the assessment of DB, a systematic review of the existing literature to explore the reliability and validity of outcome measures is imperative. The systematic retrieval and appraisal of all literature about DB with a quantitative synthesis will lead to best practice guidelines for clinicians and researchers. This systematic review aims to provide a synthesis of outcome measures used to evaluate DB and appraise the psychometric properties of these outcome measures.

METHODS

This study used the COnsensus-based Standards for the selection of health Measurement INSTRUMENTS (COSMIN) guidelines for systematic reviews of patient-reported outcome measures.²⁷⁻²⁹ The methods of this

systematic review follow the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) recommendations for systematic reviews and outcome measurement instrument selection, which are currently being piloted.³⁰ We registered this review protocol on PROSPERO (CRD42021274960) and updated the amendments regularly.

Search strategy

An experienced medical librarian (DY) carried out literature searches in the Cochrane Library database of systematic reviews and the Cochrane Central Register of Controlled Trials, the Ovid Medline (full), the Ovid Excerpta Medica Database (Embase), the Ovid Allied and complementary medicines database, the Ebscohost Cumulative Index to Nursing and Allied Health Literature and the Physiotherapy Evidence Database (PEDro).

To perform the literature searches, a construct (DB), instrument (assessment instruments) and outcome (validity and reliability) framework were employed. Following a scoping search, relevant synonyms were found and validated as suitable and informative by the review team's clinicians and academics. Searches were carried out to identify the relevant subject headings for those databases with a subject thesaurus (MeSH or Emtree) and text words in each database's title and abstract fields. Proximity operators were used to combine search words together in the title and abstract fields to increase search sensitivity. To increase the precision of the results returned by the searches, the review team decided to include a NOT operator in the search strategies to screen out papers related to sleep apnoea at the database search stage.

Searches were run in February 2022 and repeated in May 2023 before study completion to ensure the review considered the most recently published research. Due to the limited search functionality of the PEDro, this was searched using separate individual search phrases to identify relevant research on DB. On 22 February 2022, five of those phrases were identified as abstracts, and these were 'dysfunctional breathing', 'breathing disorder', 'thoracoabdominal synchrony', 'apical breathing' and 'respiratory dysfunction'. These phrases were searched again on 11 May 2023. Date limits were applied to screen out papers published before 1990. The rationale for this decision was that the term DB or breathing pattern dysfunction, only came into existence and began to be used commonly in the medical literature in 1990. A copy of the full search strategy run in Ovid Medline and other databases is available (online supplemental file S1). The resulting references identified by the database searches were uploaded into the Endnote reference management software package to allow for an initial screening.

Study selection

The following inclusion criteria were considered: (1) an outcome measure that investigated the validity and/

or reliability of DB in the adult population (18+ years) using clinician-reported and patient-reported outcome measures and (2) full articles and service evaluation reports published in a peer-reviewed journal in English. Exclusion criteria were studies that used laboratory-based outcome measures, systematic reviews, conference abstracts, research letters, commentaries and letters to the editor.

Data extraction (selection and coding)

Two independent reviewers (VM and CR) screened the titles and abstracts for relevancy using the inclusion and exclusion criteria. Reference lists of all included studies were also searched for relevant titles. The authors (VM and CR) retrieved full-text articles that met the study criteria. The first author (VM) article (TFBS) was included in this review; to mitigate conflict of interest and reduce bias, only CR investigated the articles related to TFBS. The PRISMA flow diagram of this procedure is depicted in [figure 1](#) using the PRISMA 2020 statement.³¹

Risk of bias and quality of results

The team used the COSMIN methodology for systematic reviews of patient-reported outcome measures (PROM) and clinician-reported outcome measures to evaluate the psychometric characteristics of outcome measures used in persons with DB.^{27–29} The COSMIN PROM recommends using an outcome measure with 'sufficient' content validity and internal consistency.^{27–29} The reviewers (VM and CR) individually extracted and evaluated the data for the first nine attributes listed in the COSMIN tool.

The COSMIN checklist was used to assess the methodological rigour of each outcome measure across the measurement attributes. These include reliability, validity and other psychometric properties. The methodologies provided for evaluating the measurement properties of all the outcome measures are included in this systematic review. Study quality was assessed separately for each measurement property using a four-point rating system (very good, adequate, doubtful, inadequate or not applicable).²⁹ The 'worst score counts' principle was used, where the overall rating for each measurement property is given by the lowest rating of any standard in the box.^{28,29} The results of individual studies on measurement characteristics were compared with COSMIN criteria for good measurement qualities. Each outcome was graded as sufficient (+), insufficient (–) or indeterminate (?). Relevance, comprehensiveness and comprehensibility criteria were used to grade the quality of the results in research reporting on content validity.

The result of each study on a measurement property is rated using the most recent standards for good measurement properties. The total ratings of the study outcomes for each measurement property per outcome measure were summarised as sufficient (+), insufficient (–), indeterminate (?) or inconsistent (±). An overall rating was calculated by summing the scoring of each study; if 75%

of the studies had the same scoring, that scoring became the overall rating (+ or –). However, if <75% of the studies had the same scoring, the overall rating would become inconsistent (±). If more than two articles were available, a summary of the overall evidence for measuring the properties of the outcome measures was determined. The lowest and highest results for each measurement property of an outcome measure are displayed to illustrate a set of findings that have been qualitatively aggregated.

The evidence's quality was rated using a modified Grading of Recommendations Assessment, Development and Evaluation (GRADE) system, with grades of 'high', 'moderate', 'low' or 'very low'.²⁷ The quality of the evidence was not rated for studies with an uncertain overall rating. For the quality assessment, two reviewers (VM and CR) independently worked on each stage while taking into account factors including the risk of bias, inconsistency, imprecision and indirectness. Starting with high-quality evidence, the quality of the evidence was reduced while considering all factors for the outcome measures. Disagreements were addressed by discussion and/or consultation with a third reviewer (AP).

Patient and public involvement

Patients were not involved in this review due to the complexity of evaluating the psychometric properties of the DB tools.

RESULTS

Our first search (22 February 2022) yielded 1735 references. After removing duplicates, 1246 references were included for title and abstract screening. In our second search (11 May 2023), we identified 144 references. After removing duplicates, 96 references were included for title and abstract screening. Sixteen papers met inclusion criteria, seven through database searching and nine through searching reference lists of included studies ([figure 1](#)).

Overview of outcome measures

Our search identified the following ten outcome measures that have examined reliability and/or validity components: Breathing Vigilance Questionnaires (Breathe-VQ),³² MARM,^{15 21} NQ,^{33–37} BPAT,^{38 39} Hi Lo test,²¹ clinical assessments of increased work of breathing,⁴⁰ Milstein Breathing Pattern Assessment Index (M-BPAI),⁴¹ SEBQ,^{14 22} TFBS^{24 26} and Dyspnoea-12 (D12) questionnaire.³⁸ The Hi-Lo and D-12 scales were not included in this review for evaluation because they are not the primary scales used to assess DB.^{21 38} Of the 16 studies, only nine included participants with DB, and the remaining seven included healthy participants. The COSMIN guidelines recommend testing the measurement properties on the target population.²⁷ However, the identified studies have used these outcome measures in patients and healthy people. Therefore, these groups'

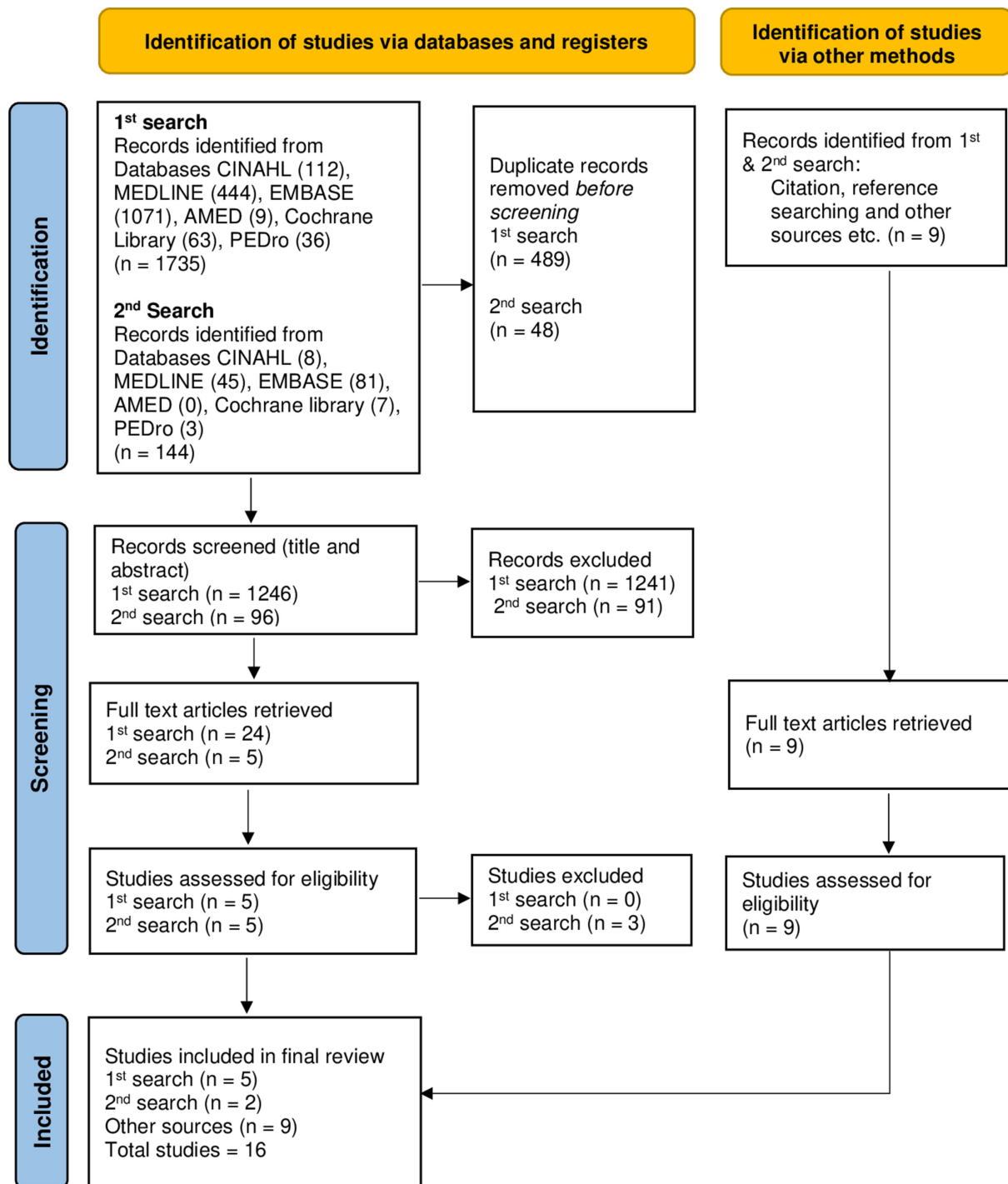


Figure 1 Preferred Reporting Items for Systematic reviews and Meta-Analyses flow chart. AMED, Ovid allied and complementary medicines database; CINAHL, Ebscohost Cumulative Index to Nursing and Allied Health Literature; EMBASE, Ovid Excerta Medica Database; MEDLINE, Medical Literature Analysis and Retrieval System Online; PEDro, Physiotherapy Evidence Database.

measurement properties were given separately (table 1 (online supplemental file S2)).

Developmental and content validity studies

Developmental studies

The evidence synthesis of the developmental and content validity of available outcome measures is summarised in

table 2. Of the eight outcome measures, only two were reported to have developmental and content validity properties.^{32 33 35 36} A representative patient sample and a cognitive interview are required to develop an outcome measure. A cognitive interview study offers information on the items' depth, especially their readability as an outcome measure. However, this was only followed in

Table 1 Characteristics of outcome measures in studies involving patients

Outcome measure (country, language, year)	No. of items/recall period	Response options (score range)	Sample size	Professionals involved	Patients condition	Measurement properties examined
Nijmegen Questionnaire* (Thailand, Thai, 2022) ³⁶	16/undefined	0–4 Likert scale (0–64)	100 (48 males and 52 females)	Two language experts and two physiotherapists	Hyperventilation syndrome	PROM development, content validity, internal consistency, cross-cultural validity,† test-retest reliability and hypothesis testing for construct validity†
Nijmegen Questionnaire*(New Zealand, English, 2019) ³⁷	15/undefined	0–4 Likert scale (0–60)	10 (six patients and four clinicians)	Four clinicians	Patients from respiratory clinics	PROM development, content validity and structural validity
Nijmegen Questionnaire*(Iran, Farsi, 2015) ³⁵	16/undefined	0–4 Likert scale (0–64)	100 (59 males and 41 females)	Lung specialists and physiotherapists	Mild to moderate asthma	PROM development, content validity, internal consistency, cross-cultural validity,† test-retest reliability and hypothesis testing for construct validity†
Nijmegen Questionnaire*(Greece, English, 2014) ³⁴	16/2 months	1–5 Likert scale (16–80)	162 (68 male and 94 female patients)	An expert chest physiotherapist and a pulmonologist	Mild to moderate asthma	Structural validity, internal consistency, cross-cultural validity,† test-retest reliability, hypothesis testing for construct validity† (discriminant validity†) and criterion validity†
Milstein Breathing Pattern Assessment Index‡ (USA, English, 2023) ⁴¹	6/5 hours	Yes/No; 0, 1 and 2 (0–15)	135 athletes (male/female)	Three speech-language pathologists and one sports medicine specialist	Athletes with concerns of exercise induced laryngeal obstruction	Inter-rater reliability, intrarater reliability and criterion validity
BPAT‡ (UK, English, 2022) ³⁸	7/undefined	0, 1 and 2 (0–14)	65 (30 males and 35 females)	One respiratory physician and one respiratory physiotherapist	Post-COVID patients	Criterion validity
BPAT‡ (UK, English, 2018) ³⁹	7/undefined	0, 1 and 2 (0–14)	150 (46 male and 104 female)	One respiratory physiotherapist	Refractory asthma	Criterion validity
Clinical assessment of increased work of breathing‡ (USA, English, 2016) ⁴⁰	1/3/undefined	Four-level scale (absent, slight, moderate and severe)	113 (72 males and 41 females)	Four critical care consultants and four critical care fellows	Patients with various major diagnosis	Inter-rater reliability and hypothesis testing for construct validity†

Continued



Table 1 Continued

Outcome measure (country, language, year)	No. of items/recall period	Response options (score range)	Sample size	Professionals involved	Patients condition	Measurement properties examined
Self-Evaluation of Breathing Questionnaire* (Australia, English, 2009) ¹⁴	17/undefined	0–3 Likert scale (0–51)	83 (29 males and 54 females)	Practitioners (specification unclear)	Mild asthma or healthy	Hypothesis testing for construct validity†

*Patient-reported outcome measure.
 †Reviewers inference of the measurement property.
 ‡Clinician reported outcome measure.
 BPAT, Breathing Pattern Assessment Tool.

some of the included studies. All four studies involved experts^{32 34–36} and three involved patients.^{34–36} Concept elucidation was deemed ‘inadequate’ for Breathe-VQ and Korean-NQ because only healthy participants engaged in the studies.^{32 33} Other NQ trials were rated ‘very good’ since the patients involved were typical of the target population.

Content validity studies

Three of the four articles on the content validity of NQ involved patients^{34–36} and all four involved experts.^{33–36} Of these three studies, patients’ relevance, comprehensiveness and comprehensibility were evaluated for one study only.³⁴ A cognitive interview was conducted for the Breathe-VQ, but the quality was ‘inadequate’ as it was not conducted in a patient population.³² No studies on the development and content validity of TFBS, MARM or SEBQ were found. Only the NQ has been considered for rating, and it was judged as ‘sufficient’.

Risk of bias assessment rating of other measurement properties

The evidence synthesis for all outcome measures and additional measurement properties is summarised in [table 3](#) (online supplemental file S3) and Supplementary file S4—<https://osf.io/49hju/>.

Internal structure

Among the included studies, only three reported structural validity.^{32 34 37} Two studies explored the structural validity of the NQ measure,^{34 37} and one study explored the Breathe-VQ measure.²⁶ NQ structural validity was examined using the Rasch model and exploratory factor analysis with ‘very good’ and ‘inadequate’ quality.^{34 37} For structural validity, the Breathe-VQ study employed exploratory/confirmatory factor analysis of ‘very good’ quality.³² The internal consistency of the NQ, as measured by Cronbach’s alpha, ranged from >0.70 to 0.92 with a ‘very good’ quality and ‘sufficient’ rating.^{34 35} However, Rasch and factor analysis do not apply to other outcome measures, especially clinician-reported outcome measures.

Reliability

In total, 10 studies reported reliability measures. M-BPAI⁴¹ and TFBS^{24 26} were rated to have ‘very good’ methodological quality and ‘sufficient’ rating.⁴¹ The methodological quality and rating were the same as the MARM and SEBQ.^{15 21 22} Only one study that reported the reliability of clinical assessment of the work of breathing exhibited ‘adequate’ methodological quality and was ‘indeterminate’ for the rating.⁴⁰ The test-retest reliability values for NQ were in the range of 0.90 and 0.98, corresponding to ‘very good’ to ‘adequate’ quality.^{34–36} It was also judged that the NQ’s overall rating was ‘sufficient’. The correlation value ranges from 0.81 to 0.82 when analysing NQ’s hypothesis testing for construct validity, indicating ‘very

Table 2 Evidence synthesis of developmental, content validity of Nijmegen Questionnaire and Breathing Vigilance Questionnaire using COnsensus-based Standards for the selection of health Measurement Instruments checklist

Outcome measure (country, language, year)	Quality of content validity				Reasons for quality rating and additional comments
	Quality of PROM development	Relevance – professionals	Comprehensiveness – professionals	Comprehensibility – patients	
Nijmegen Questionnaire (Thailand, Thai, 2022) ³⁶	D	V (+)	V (+)	D (+)	'D' – unclear whether group meeting or interview guide was used in quality of PROM development domain. 'D' – unclear if skilled group interviewers or moderators were used in comprehensibility domain.
Nijmegen Questionnaire (Korea, Korean, 2018) ³³	I	D (+)	D (+)	NA	'I' – target population not involved in the tool development stage under the quality of PROM development domain. 'D' – under the relevance and comprehensiveness domain, it was unclear if the individual items were tested with professionals; the appropriate approach was used for data analysis, and two researchers involved in the analysis. If at least two medical doctors were involved to examine the relevance and comprehensiveness.
Nijmegen Questionnaire (Iran, Farsi, 2015) ³⁵	D	A (+)	D (+)	D (+)	'D' – unclear whether qualitative data were collected until new knowledge was obtained from participants for the quality of PROM development domain. 'A' – it was assumed that the group meeting was recorded for the relevance of professional's domain. 'D' – unclear if the individual items were tested with professionals in comprehensiveness and with patient in comprehensibility domain.
Nijmegen Questionnaire (Greece, English, 2014) ³⁴	NA	(+) V	(+) V	V (+)	–
Breathing Vigilance Questionnaire (UK, English, 2022) ³²	I	(+) V	(+) V	NA	'I' – target population not involved in the development stage.

Quality rating: V=very good; A=adequate; D=doubtful and I=inadequate.
Quality score: sufficient (+); insufficient (-); indeterminate (?), and inconsistent (±).
NA, not applicable; PROM, patient-reported outcome measure.



Table 3 Methodological quality and rating of psychometric properties in studies involving patients

Outcome measures (country, language, year)	Structural validity	Internal consistency	Cross-cultural validity/ measurement invariance	Reliability	Measurement error	Criterion validity	Hypothesis testing for construct validity*	Reason for methodological quality/additional comments
NQ (Thailand, Thai, 2022) ³⁶	NA	V (+) Cronbach's alpha=0.79	V (?)	V (+) Test-retest ICC=0.90	NA	NA	V (+) p<0.0001	Cross-cultural validity: reviewers were unable to extract data for this item and therefore rated it as 'indeterminate'.
NQ (New Zealand, English, 2019) ³⁷	I (-)	NA	NA	NA	NA	NA	NA	Structural validity: Rasch model does not fit and therefore rated 'insufficient'.
NQ (Iran, Farsi, 2015)	NA	V (+) Cronbach's alpha>0.70	V (-) KMO index=0.65	V (+) Test-retest ICC=0.82	NA	NA	V (+) r=0.826	Cross-cultural validity: 55% data variance found and therefore rated 'insufficient'.
NQ (Greece, English, 2014) ³⁴	V (?)	V (+) Cronbach's alpha=0.92	A (+)	A (+) Test-retest ICC=0.98	NA	V (-)	V (+) 58.86% variance, convergent validity r=0.81, discriminant validity* ROC=0.95	Structural validity: reviewers were unable to extract data for this item and therefore rated it as 'indeterminate'. Cross-cultural validity*: data related to important difference variables are available for the following factors only – asthma control, asthma severity end tidal carbon dioxide and follow-up.
NQ – overall rating (pooled or summary result)		>0.70-0.92 (+)		0.82-0.98 (+)			0.81-0.826 (+)	

Continued

Table 3 Continued

Outcome measures (country, language, year)	Structural validity	Internal consistency	Cross-cultural validity/measurement invariance	Reliability	Measurement error	Criterion validity	Hypothesis testing for construct validity*	Reason for methodological quality/additional comments
Milstein Breathing Pattern Assessment Index (USA, English, 2023) ⁴¹	NA	NA	NA	V (+) ICC Inter-rater=0.99 Intrater=0.97	NA	V (+) AUC=0.87	NA	
BPAT (UK, English, 2022) ³⁸	NA	NA	NA	NA	NA	V (+) AUC=0.91	NA	
BPAT (UK, English, 2018) ³⁹	NA	NA	NA	NA	NA	V (+) AUC=0.88	NA	
BPAT – overall rating (pooled or summary result)						0.88–0.91 (+)		
Clinical assessment of increased work of breathing (USA, English, 2016) ⁴⁰	NA	NA	NA	V (-) K=(0.53–0.47); K(0.36–0.23)	NA	NA	V (+) Correlation 0.69	Reviewers rated reliability ‘insufficient’ as there was moderate to fair agreement between raters for the variables. Reviewers inferred this as a construct validity and intrater reliability study.
Self-Evaluation of Breathing Questionnaire (Australia, English, 2009) ¹⁴	NA	NA	NA	NA	NA	NA	V (+) Correlation 0.60	Reviewers inferred this as a construct validity.
Methodological quality rating: V, very good; A, adequate; D, doubtful and I, inadequate Result score: sufficient (+); insufficient (-) and indeterminate (?). *Reviewers inference of the measurement property. AUC, area under the curve; BPAT, Breathing Pattern Assessment Tool; ICC, intraclass correlation coefficient; k, kappa; K, Kruskal-Wallis test; KMO, Kaiser-Meyer-Olkin measure; MDC, minimum detectable change; NA, not applicable; NQ, Nijmegen Questionnaire; r, correlation coefficient.								



good' quality and 'sufficient' rating.^{34 35} A more comprehensive evidence synthesis for these and other outcome measures is available in Supplementary file S4—<https://osf.io/49hju/>.

GRADE quality

The reviewers used GRADE to assess the quality of studies that involved participants with respiratory disease since the clinical application would be acceptable in the actual patient population. As a result, only the NQ that included individuals with asthma and hyperventilation syndrome was included in the GRADE quality assessment. The evidence quality is 'high' for the NQ in reliability and hypothesis testing for construct validity domains but 'low' for cross-cultural and structural validity domains. The GRADE quality assessment cannot be applied to the remaining outcome measures.

DISCUSSION

This systematic review presented an overview of outcome measures used to assess DB and evaluated the psychometric properties of outcome measures used in healthy and DB populations. NQ is the only outcome measure with sufficient psychometric properties to be considered by clinicians for the DB assessment.

Nijmegen Questionnaire

NQ's measurement properties have gained much attention due to its long record and frequent use in DB assessment, notably in conditions including bronchial asthma and hyperventilation syndrome.^{39 42} The available evidence indicated that the NQ had been evaluated using rigorous methods, and its content validity, internal consistency and reliability were commonly reported. This outcome measure has been translated into other languages, but for one of the translated versions, the PROM development and content validity were not well documented.³⁴ However, other measurement properties were well established.³⁴

PROM development and content validity studies were not consistent across the included studies. This is due to variations in the methodological description, and it was the least reported psychometric property, followed by structural validity and hypothesis testing for construct validity. Despite this, the reviewers have used the COSMIN tool to infer the quality of PROM and content validity, and the NQ was found to have most of the measurement properties with 'sufficient' quality. This is an area that needs further exploration in future studies. In addition, the language and structure of the items used in the NQ need improvement. For instance, item NQ14 (cold hands or feet) does not fit the structural validity, and similarly, item NQ9 (bloated feeling in the stomach) also does not fit the Rasch model.³⁷ Since NQ looks at many DB dimensions, these factors could be considered for prospective use.

Breathe-VQ and BPAT

Breathe-VQ is the next potential outcome measure that can be used in the DB population because the measurement properties, such as structural validity, internal consistency, reliability, measurement error, criterion validity and hypothesis testing for construct validity, are well established.³² The Breathe-VQ is best suited to assess changes related to excessive conscious breathing rather than as an outcome measure for diagnosing the DB disorder. In contrast to the NQ, the Breathe-VQ has only been examined in one study; therefore, more research is required to determine its use in the DB population before considering it for clinical use.³² It may be helpful to use NQ with Breathe-VQ to identify excessive conscious breathing caused by anxiety. The same comments apply to the BPAT, which has proven criterion validity for patients with asthma, breathing pattern disorder and post-COVID breathless individuals.^{38 39} BPAT is more suitable for evaluating breathing irregularities in the DB population. However, BPAT is still in the trial phase, and its clinical utility has yet to be determined.

Other outcome measures

The remaining outcome measures, such as MARM, clinical evaluation of increased effort of breathing, TFBS, SEBQ and M-BPAI, had examined only a few psychometric properties.^{15 21 22 24 26 40 41} The reviewers could only comment on its clinical utility once the remaining properties had been thoroughly investigated.

Limitations

This review excluded grey literature, conference abstracts, poster abstracts and dissertations; therefore, potential studies could have been missed. The second ordered reference check was not done, which may lead to missing other relevant studies. Only English-language studies were considered for this review, which may have reduced the number of potentially acceptable studies in other languages in the DB population. Another limitation is the lack of primary data, which prevented the review team from conducting a meta-analysis. The reviewers had no specific training to use the COSMIN. Instead, they relied on their clinical and scholarly experience to reach an agreement. This might affect how studies are rated for quality. However, the review team mitigated this by sending the collected data to the corresponding authors of the included studies for verification, comments and triangulation.

Future consideration

Only five papers in our review briefly described the process of developing outcome measures and content validity. Determining whether the outcome measure development process had been rigorously carried out or was selectively reported is challenging. This might imply that the available outcome measures do not satisfy practitioners'

expectations or recognise the researchers' requirements. An outcome measure with 'inadequate' content validity or a lack of evidence of content validity has questionable use in clinical practice. Therefore, particular attention should be given to determining the content validity of those outcome measures that do not possess this property. Detailed information on the outcome measure development process and content validity should be provided in future research. The reviewers recommended addressing these aspects in future studies.

It should be noted that the COSMIN checklist is both comprehensive and rigorous in its quality. Any other outcome measures considered here are unlikely to fulfil the standards. As a result, some of the outcome measures are rated as 'inadequate' quality. However, the authors recommend considering these measurement properties when constructing an outcome measure that fulfils the stringent criteria.

CONCLUSIONS

This review found 10 outcome measures used to assess DB. The NQ is the only outcome measure that showed evidence quality to be 'high' for internal consistency and hypothesis testing for construct validity and reliability. The evidence quality is 'low' for NQ structural validity and cross-cultural validity. The measurement properties of NQ are sufficient to recommend its use as part of a clinical application of DB. Most outcome measures have examined only a few psychometric properties; a more comprehensive investigation of all psychometric properties is needed before considering their clinical use. Future research on the existing outcome measure or developing a new outcome measure may follow the COSMIN guidelines.

Author affiliations

¹Department of Rehabilitation and Sports Sciences, Faculty of Health and Social Sciences, Bournemouth University, Bournemouth, UK

²University of Birmingham, Birmingham, UK

³Birmingham Women's and Children's NHS Foundation Trust, Birmingham, UK

⁴Department of Physical Therapy, Faculty of Associated Medical Sciences, Chiang Mai University, Chiang Mai, Thailand

⁵Cardiology Department, University Hospitals Dorset NHS Foundation Trust, Poole, UK

⁶Faculty of Health and Social Sciences, Bournemouth University, Bournemouth, UK

X Vikram Mohan @VikramMohan10 and Derick Yates @DJY-LIB-EBP

Acknowledgements We thank Rachel Senior, Specialist Physiotherapist, Dorset Adult Integrated Respiratory Service, Dorset, for her expertise and comments about the outcome measures; Dr Kathryn Collins, Bournemouth University, for her contribution in editing and all the authors of the included studies for extracted data verification and corrections.

Contributors VM, CR, AP and CB were involved in study conceptualisation. VM and CR were responsible for screening, selecting articles and data entry, data interpreting, reporting and for preparing the final manuscript. DY was responsible for constructing search strategy and conducting searches. VM and CR are acting as guarantors for the work. All authors read, provided feedback and approved the final manuscript.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests VM and AP are authors of some of the included articles. They were not engaged in evaluating the methodological quality of these articles. They have no other competing interests. The first author (VM) article (TFBS) was included in this review; to mitigate conflict of interest and risk of bias, only CR investigated the articles related to TFBS.

Patient and public involvement statement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Ethics approval Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available in a public, open access repository. All the relevant data were available at <https://osf.io/49hju/>.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iDs

Vikram Mohan <http://orcid.org/0000-0002-9067-2817>

Chandrasekar Rathinam <http://orcid.org/0000-0002-0049-8430>

REFERENCES

- Ratnovsky A, Elad D, Halpern P. Mechanics of respiratory muscles. *Respir Physiol Neurobiol* 2008;163:82–9.
- Clifton-Smith T, Rowley J. Breathing pattern disorders and physiotherapy: inspiration for our profession. *Physical Therapy Reviews* 2011;16:75–86.
- Barker N, Everard ML. Getting to grips with 'dysfunctional breathing'. *Paediatr Respir Rev* 2015;16:53–61.
- Boulding R, Stacey R, Niven R, *et al*. Dysfunctional breathing: a review of the literature and proposal for classification. *Eur Respir Rev* 2016;25:287–94.
- Thomas M, McKinley RK, Freeman E, *et al*. The prevalence of dysfunctional breathing in adults in the community with and without asthma. *Prim Care Respir J* 2005;14:78–82.
- Vidotto LS, Carvalho CRF de, Harvey A, *et al*. Dysfunctional breathing: what do we know? *J Bras Pneumol* 2019;45:e20170347.
- Cluff RA. Chronic hyperventilation and its treatment by physiotherapy: discussion paper. *J R Soc Med* 1984;77:855–62.
- McKeown P, O'Connor-Reina C, Plaza G. Breathing re-education and phenotypes of sleep apnea: a review. *J Clin Med* 2021;10:471.
- Wells C, Cartwright M, Hirani S, *et al*. Identifying predictors for referral to a physiotherapy service for children with difficult asthma. ERS International Congress 2020 abstracts; 2020
- Hagman C, Janson C, Emtner M. Breathing retraining - a five-year follow-up of patients with dysfunctional breathing. *Respir Med* 2011;105:1153–9.
- Bruton A, Lee A, Yardley L, *et al*. Physiotherapy breathing retraining for asthma: a randomised controlled trial. *Lancet Respir Med* 2018;6:19–28.
- Connolly CK. Dysfunctional breathing in COPD. *Thorax* 2003;58:460–1.
- Kadam S, Prem V, Ughreja RA. Effect of breathing re-training on end-tidal carbon-Di-oxide in patients with asthma: a systematic review and meta-analysis. *J Asthma* 2023;60:1493–502.
- Courtney R, Greenwood KM. Preliminary investigation of a measure of dysfunctional breathing symptoms: the self evaluation of breathing questionnaire (SEBQ). *IJOM* 2009;12:121–7.
- Courtney R, van Dixhoorn J, Cohen M. Evaluation of breathing pattern: comparison of a manual assessment of respiratory motion (MARM) and respiratory induction plethysmography. *Appl Psychophysiol Biofeedback* 2008;33:91–100.



- 16 Gunnesson IL, Olsén MF. Validity in measuring breathing movements with the respiratory movement measuring instrument, RMMI. *Clin Physiol Funct Imaging* 2011;31:1–4.
- 17 Olsén MF, Romberg K. Reliability of the respiratory movement measuring instrument, RMMI. *Clin Physiol Funct Imaging* 2010;30:349–53.
- 18 Chapman EB, Hansen-Honeycutt J, Nasypany A, *et al.* A clinical guide to the assessment and treatment of breathing pattern disorders in the physically active: part 1. *Int J Sports Phys Ther* 2016;11:803–9.
- 19 Hansen-Honeycutt J, Chapman EB, Nasypany A, *et al.* A clinical guide to the assessment and treatment of breathing pattern disorders in the physically active: part 2, a case series. *Int J Sports Phys Ther* 2016;11:971–9.
- 20 Pryor JA, Prasad SA. *Physiotherapy for Respiratory and Cardiac Problems: Adults and Paediatrics*. 3rd edn. Edinburgh: Churchill Livingstone, 2002:643.
- 21 Courtney R, Cohen M, Reece J. Comparison of the manual assessment of respiratory motion (MARM) and the Hi lo breathing assessment in determining a simulated breathing pattern. *Int J Osteopath Med* 2009;12:86–91.
- 22 Mitchell AJ, Bacon CJ, Moran RW. Reliability and determinants of self-evaluation of breathing questionnaire (SEBQ) score: a symptoms-based measure of dysfunctional breathing. *Appl Psychophysiol Biofeedback* 2016;41:111–20.
- 23 Bondarenko J, Hew M, Button B, *et al.* Reliability of the breathing pattern assessment tool in asthma. *Respirology* 2021;26:38.
- 24 Mohan V, Perri M, Paungmal A, *et al.* Intra-rater and inter-rater reliability of total faulty breathing scale using visual observation and videogrammetry methods. *J Bodyw Mov Ther* 2017;21:694–8.
- 25 Dixhoorn JV, Folgering H. The Nijmegen questionnaire and dysfunctional breathing. *Eur Respir J* 2015;3–6.
- 26 Mohan V, Branney J, Clark C, *et al.* Reliability and preliminary reference values for the total faulty breathing scale (TFBS): a cross-sectional study. *Health Sci Rep* 2022;5:e661.
- 27 Prinsen CAC, Mokkink LB, Bouter LM, *et al.* COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res* 2018;27:1147–57.
- 28 Terwee CB, Prinsen CAC, Chiarotto A, *et al.* COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. *Qual Life Res* 2018;27:1159–70.
- 29 Mokkink LB, de Vet HCW, Prinsen CAC, *et al.* COSMIN risk of bias checklist for systematic reviews of patient-reported outcome measures. *Qual Life Res* 2018;27:1171–9.
- 30 Elsman EBM, Butcher NJ, Mokkink LB, *et al.* Study protocol for developing, piloting and disseminating the PRISMA-COSMIN guideline: a new reporting guideline for systematic reviews of outcome measurement instruments. *Syst Rev* 2022;11:121.
- 31 Page MJ, McKenzie JE, Bossuyt PM, *et al.* The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 2021;372:n71.
- 32 Steinmann J, Lewis A, Ellmers TJ, *et al.* Validating the breathing vigilance questionnaire for use in dysfunctional breathing. *Eur Respir J* 2023;61:2300031.
- 33 Ok JM, Park YB, Park YJ. Association of dysfunctional breathing with health-related quality of life: a cross-sectional study in a young population. *PLoS ONE* 2018;13:e0205634.
- 34 Grammatopoulou EP, Skordilis EK, Georgoudis G, *et al.* Hyperventilation in asthma: a validation study of the Nijmegen questionnaire--NQ. *J Asthma* 2014;51:839–46.
- 35 Ravanbakhsh M, Nargesi M, Raji H, *et al.* Reliability and validity of the Iranian version of Nijmegen questionnaire in Iranians with asthma. *Tanaffos* 2015;14:121–7.
- 36 Pakdee S, Piya-Amornphan N. The Thai version of the Nijmegen questionnaire. *Heliyon* 2022;8:e12296.
- 37 Li Ogilvie V, Kayes NM, Kersten P. The Nijmegen questionnaire: a valid measure for hyperventilation syndrome. *NZJP* 2019;47:160–71.
- 38 Hylton H, Long A, Francis C, *et al.* Real-world use of the breathing pattern assessment tool in assessment of breathlessness. *Clin Med* 2022;22:376–9.
- 39 Todd S, Walsted ES, Grillo L, *et al.* Novel assessment tool to detect breathing pattern disorder in patients with refractory asthma. *Respirology* 2018;23:284–90.
- 40 Tulaimat A, Patel A, Wisniewski M, *et al.* The validity and reliability of the clinical assessment of increased work of breathing in acutely ill patients. *J Crit Care* 2016;34:111–5.
- 41 Milstein CF, Patel RR, Laurash E, *et al.* Identification of breathing pattern disorder in athletes with exercise-induced laryngeal obstruction: a novel assessment tool. *J Voice* 2023.
- 42 Johnston R, Shaw F, Gow AM, *et al.* Does the Nijmegen correlate to the D12 when used as an outcome measure in patients with breathing pattern dysfunction. *Thorax* 2011;66:A151.