



Review article

Reliability of diaphragmatic mobility assessment: A systematic review

**Vikram Mohan^{1,2}, Aatit Paungmali¹, Patraporn Sitilerpisan¹, Sulaiman Md Dom³,
Ummi Farhana Hashim³, Siti Nor Binti Daud², Manikumar Muthiah⁴**

¹Neuro-Musculoskeletal and Pain Research Unit, Department of Physical Therapy, Faculty of Associated Medical Sciences, Chiang Mai University, Chiang Mai, Thailand

²Faculty of Health Sciences, Centre of Physiotherapy, Universiti Teknologi MARA Selangor, Bandar Puncak Alam, Puncak Alam, Malaysia

³Faculty of Health Sciences, Centre of Medical Imaging, Universiti Teknologi MARA Selangor, Bandar Puncak Alam, Puncak Alam, Malaysia

⁴Saveetha College of Physiotherapy, Saveetha University, Chennai, India

ARTICLE INFO

Article history

Received 17 March 2017

Accepted 22 May 2017

Available online 4 June 2018

Keywords

Diaphragm

Test–retest reliability

Ultrasonography

Radiography

Doi

10.29089/2017.17.00044

User license

This work is licensed under a
Creative Commons Attribution –
NonCommercial – NoDerivatives
4.0 International License.



ABSTRACT

Introduction: Diaphragm Mobility (DM) assessment is gaining interest in the field of medicine and in the healthcare sector. Despite its clinical usage, the measure of reliability in assessing DM is not clearly known.

Aim: To critically appraise the evidence describing the reliability measures of DM assessment using any of the diagnostic modalities.

Material and methods: A systematic search across five databases was carried out from January 1990 to September 2016. Quality Appraisal of Reliability Studies (QUAREL) and the Grading of Recommendations Assessment, Development and Evaluation (GRADE) system were used to assess the risk of bias and for rating the quality of the evidence. In addition, levels of evidence grading which synthesize all the included articles for grading were also used.

Results and discussion: Four papers were included for assessing both intra-rater and inter-rater reliability using ultrasound and radiography. Three papers reported ICC measures of reliability, with one paper reporting CV% of reliability. The results demonstrate that, overall, lower levels of evidence exist among the selected articles between moderate and good for intra-rater reliability and good for inter-rater reliability measures. The synthesis of all the included articles demonstrated that, overall, moderate evidence exists.

Conclusions: There were moderate-to-good reliability measures with a low risk of bias in both the forms of reliability for assessing diaphragmatic mobility.

Corresponding author: Aatit Paungmali, Neuro-Musculoskeletal and Pain Research Unit, Department of Physical Therapy, Faculty of Associated Medical Sciences, Chiang Mai University, Chiang Mai 50200, Thailand. Tel.: +66 53949246,

Fax: +66 53936042.

E-mail address: aatit.p@cmu.ac.th.

1. INTRODUCTION

Breathing is a natural physiological process which is required to withstand life. In this context, the diaphragm is an important muscle of respiration. Thus, it is understood that dysfunction or impairment of diaphragm disturbs the breathing cycle. Such undue changes which occur due to the phenomena of abnormal physiology and biomechanical abnormality could eventually lead to altered work of breathing and deterioration in exercise tolerance, which would affect the quality of life.¹⁻³

Diaphragmatic mobility (DM) is one of the important parameters to be assessed in recent years to identify dysfunction of the diaphragm. In general, DM assessment was developed to improve screening of individuals who suffer from respiratory concerns. It has been theorized that individuals whose breathing is compromised because of respiratory illness and musculoskeletal disorders may exhibit poor movement patterns of the diaphragm, thus predisposing an individual to respiratory dysfunction.^{1,2,4} In order to know the extent of DM, a few imagery assessment methods are used to evaluate the function and position of the diaphragm. The imagery assessment methods that are in practice are X-ray, fluoroscopy, magnetic resonance imaging (MRI), and ultrasonography.⁵⁻⁷ Out of all these techniques proposed to assess DM, ultrasound has the advantage of being a safe operating procedure, as described in the earlier literature.⁸ Even though various techniques of assessment are available to assess DM, the commonly used methods of estimation to ensure reliability of results are not understood clearly.

In general, reliability takes two forms, one is relative reliability and the other is absolute reliability.⁹ In relation to DM, relative reliability is the degree to which DM values differ on two occasions, and this can be expressed by means of Pearson's correlation coefficient and intraclass correlation coefficient (ICC). On the other hand, absolute reliability is the degree to which repeated measurements of DM vary for individuals, which can be expressed by means of the standard error of measurements (SEMs), the coefficient of variation (CV), and Bland and Altman's 95% limits of agreement. The current consensus in the literature is that these two forms of reliability estimates need to be used together while performing reliability statistics in order to test repeatability and reproducibility.⁹

On considering these two reliability forms, the term reliability of DM has been reported in a few studies.^{10,11} These studies have utilized differing forms of reliability and different methods of evaluating DM using different diagnostic modalities. In addition, the authors have utilized raters with different levels of clinical expertise and different backgrounds in the field of medicine. Therefore, there is no consensus on which forms of reliability measures are commonly used to test reliability measures and which mode of assessment technique is reliable for evaluating DM.

To date, there has been no synthesis of the evidence regarding the reliability of DM to make a definite statement regarding the clinical applicability or use of this method in

practice using a particular modality for assessment. If this assessment method of using any of the modalities is reliable within and between the raters, clinicians can be confident in their assessments and begin to utilize interventions to improve the patient's status. Furthermore, it will enable clinicians and researchers to assess the effectiveness of their treatments through reevaluation to improve the patient's condition following respiratory rehabilitation.

2. AIM

The specific objective of this systematic review was to critically appraise published evidence describing the reliability measures of DM using any of the diagnostic modalities described in earlier literature.

3. MATERIAL AND METHODS

3.1. Search strategy and selection criteria

The details of the search are presented in Figure 1. Two reviewers (VM and AP) independently selected the eligible studies based on the inclusion and the exclusion criteria as stated in the study protocol. All studies that investigated the reliability of DM assessment using any of the measurement devices were included. The criteria for the inclusion comprised the following: the articles had to assess human subjects with no restriction regarding methods of the assessment instrument. Furthermore, only articles that are published in English were considered in the present study. Studies, which are not published in relation to reliability and DM were excluded. In the initial review, the extracted data included the type of study design, purpose of study, and examination of DM with statistical analysis and conclusion. In case of disagreement in the article selection, a consensus was reached by consulting with a third reviewer (MM).

3.2. Quality assessment

The methodological quality of the included studies in this present systematic review was assessed using a Quality Appraisal of Reliability Studies (QUAREL) scale.¹² A score of 60% or more indicates high-quality studies.^{13,14} The tool was found to be a reliable tool for assessing diagnostic reliability studies.¹⁵ The present study adapted the QUAREL scale to identify whether the equipment used was able to detect DM. The quality assessment was carried out by two reviewers (PS and UFH) independently for all the four included studies. In case of discrepancy in rating the articles, a consensus was reached by consulting with a third reviewer (SD).

3.3. Level of evidence

The Grading of Recommendations Assessment, Development, and Evaluation (GRADE) approach was used to rate the quality of evidence and the grading strength of the recommendations.¹⁶ The GRADE method of approach describes the evidence as high, moderate, low, and very low. In

addition, the level of evidence was analyzed using updated method guidelines for systematic reviews in the Cochrane group, as proposed by van Tulder et al.¹⁷

4. RESULTS

The computer software used for this review included Microsoft Office 2008, SPSS v. 21 (IBM Corporation; Armonk, New York), Microsoft Office Excel 2008, and Mendeley v. 1.16.3 for reference formatting. The SPSS data sheet imported all, the averaged data of QUAREL between the observers for inter-rater reliability of Kappa statistic measures of assessment from a Microsoft Excel spreadsheet.

4.1. Literature search

The PRISMA flow chart depicting the systematic search and review process for selection can be found in Figure 1. Of the initial 70 articles retrieved through the electronic search engines, 6 articles met the criteria set by the study protocol. In addition, 2 other articles were retrieved through other resources, thus totaling the number of articles to be 8. After screening and removing the duplicates, the total number of

articles in the qualitative synthesis was 4. Three articles determined the intra-observer and the inter-observer reproducibility in relation to DM assessment using ultrasound. Only 1 article determined both reproducibility and repeatability in relation to DM assessment using radiograph. The characteristics of all the included studies are presented in Table 1.

4.2. Assessment of risk of bias within studies

The two reviewers initially agreed on 39 out of 44 (88.63%) items on the QUAREL checklist with a Kappa score (0.42). Differences in the QUAREL scores were resolved through discussion among the reviewers. The quality scores ranged from 45.45% to 81.81% with one high-quality study (>60%) and three low-quality studies (<60%). The internal component ranged from 14.28% to 71.42%, while the external validity component was 100%. All the studies rated at 100% on the statistical portion of the QUAREL scale.

4.3. Study characteristics

Three modalities including B-mode, M-mode ultrasonography, and radiographic equipment were used for assessing DM in four of the included studies.^{10,11,18,19} Three studies

Figure 1. Flowchart of study selection process according to the preferred reporting items for systematic review and meta-analysis (PRISMA).

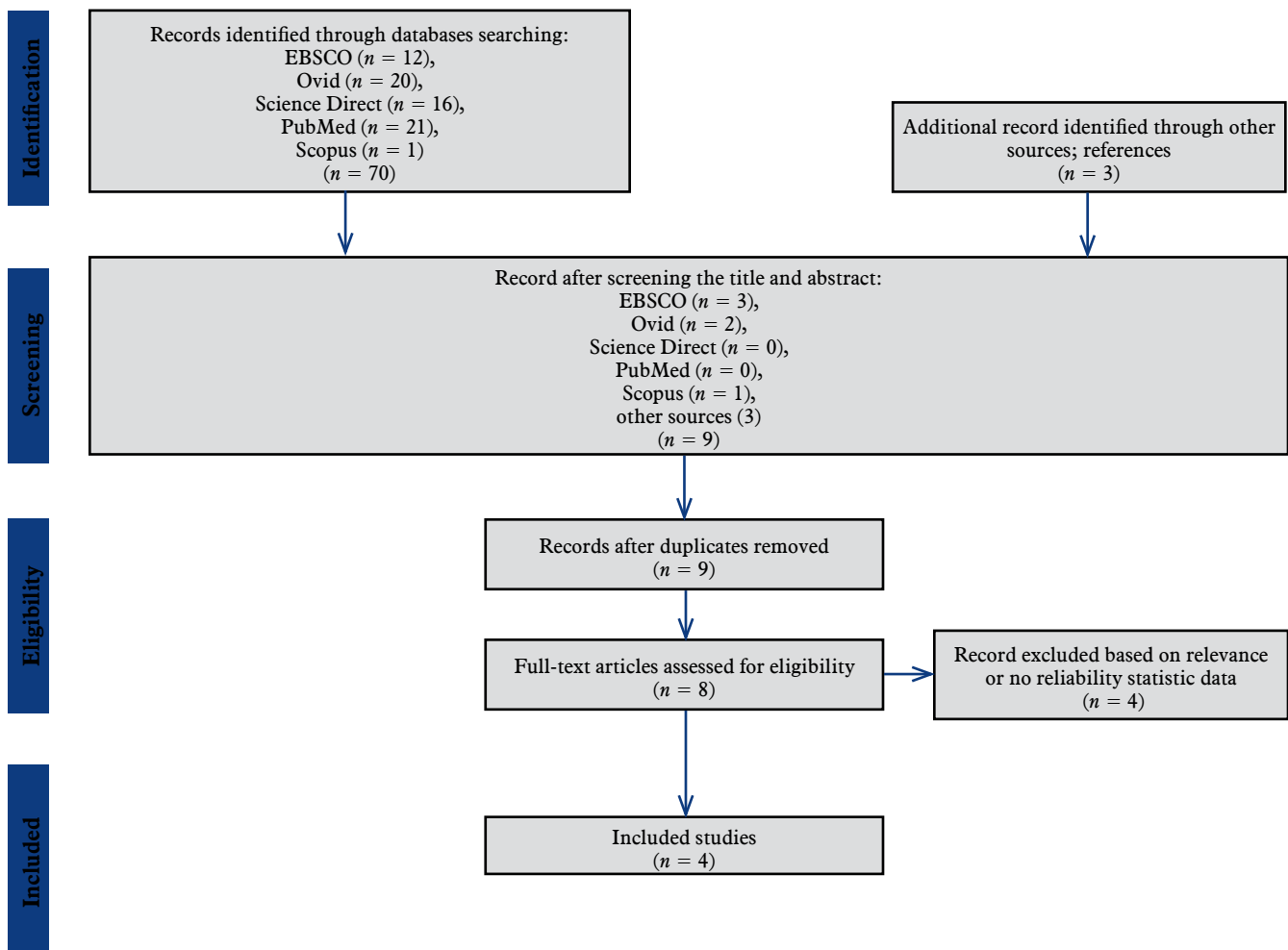


Table 1. Characteristics of included studies.

Author	Reliability design	Sample size (raters)	Rater (experience)	Modality	Side of diaphragm mobility	Position of subjects / Probe position	Sample size (participants)
Grams et al. (2014)	Intra-observer						
Inter observer RP & repeatability	2	1 year	B-Mode US	RT HD	Dorsal decubitus/ right subcostal region in the sagittal plane	40 HS	
Boussuges et al. (2009)	Intra and inter observer RP	2	NC	M-mode US	RT & LT HD	Standing/ subcostal or low intercostal probe placed between the anterior & mid axillary lines	180 HS
Houston et al. (1992)	Intra and inter observer RP	2	NS	Real time US &			
M-mode US	RT and LT HD	Supine/intercostal probe between mid-axillary & mid-clavicular line	55 general medical ward patients				
Saltiel et al. (2013)	Intra and inter observer RP	2	NS	Chest radiography: Antero-posterior view	RT & LT HD	Supine/NS	44 patients who waited for cholecystectomy surgery

Comments: HD – hemi-diaphragm, HS – healthy subjects, LT – left, NS – not specified, NC – not clear, RT – right, RP – reproducibility, US – ultrasound.

used an ultrasound device in which two of the studies used M-mode ultrasonography and one study used B-mode ultrasound.^{11,18,19} One of the included studies used the radiography method to assess the reliability measures of DM.¹⁰ Among the four included studies, only one study reported that the rater had one year of clinical experience, whereas the remaining four studies did not specify anything about experience, or it was not clearly mentioned. Three of the included studies evaluated both right and left hemi-DM. One of the included studies evaluated only right hemi-DM.

The interpretation of the ICC values for all the included studies was based on evidence as poor (0.00–0.25), fair (0.26–0.50), moderate (0.51–0.75), and good (0.76–1.00).²⁰ The inter-rater reliability according to the ICC interpretation for three of the included studies are good. The intra-rater reliability according to the ICC interpretation for three of the included studies ranged from moderate to good. Only one of the included studies utilized coefficient of variation percentage (CV %) to report reliability measures for DM using real time and M-mode ultrasonography. Both the measures of reproducibility of quiet and deep breathing using the real-time method of assessment were acceptable, with a CV of 13%, and the deep real time of DM was good, with 6.5%.

Statistical pooling of the data was not performed as the number of studies that used B-Mode, M-mode ultrasound, and radiography for assessing DM was limited. Second, methodological variation in terms of the position of the patients by identifying a landmark varied between the studies. Third, the inclusion criteria between the studies were heterogeneous in terms of samples. A few of the studies decided not to perform funnel plot analysis to rule out publication bias.

4.4. Level of evidence

The results of the inter-observer and the intra-observer reliability statistics demonstrate that, overall, low levels of evidence existed in the selected articles, which were between moderate and good for intra-rater reliability. For inter-rater reliability measures, it was good, on using the GRADE method of assessment.¹⁶ Equally, the van Tulder et al. approach indicates that, overall, moderate evidence exists in both inter-observer and intra-observer reliability of DM.¹⁷

5. DISCUSSION

The present study is, to our knowledge, the first systematic review assessing the risk of bias and summarizing the results of reliability measures such as relative and absolute for DM using various modalities. The included studies had different measuring equipment and were conducted with different methodological measures. Hence, a universal methodology for assessing DM either using ultrasound or through radiograph is necessary. All the included studies were conducted with low risk of bias and had a moderate level of evidence for the measures of reliability in measuring DM.

All the included studies examined intra-observer reliability and inter-observer reliability. However, the statistical measures that were utilized differed between the studies. Three of the studies included in the review reported relative reliability (ICC) and Pearson correlation coefficient analysis.^{10,11,19} One of the included studies reported absolute reliability (CV %).¹⁸ Only one study used both absolute and relative forms of reliability measure, which were Pearson and

Bland-Altman methods of analysis.¹⁹ The studies reported for inter-rater reliability were, overall, shown to have good reliability, indicating that clinicians could replicate the measurement of DM using an ultrasound device. Measurement of the other type of reliability, which is intra-observer reliability measurement, had a distinction between moderate and good reliability. This signifies that neither of the forms of reliability measures, which are relative and absolute reliability measures, was reported in any of the included studies for the present systematic review. A report which was carried on diaphragmatic displacement using ultrasound was excluded when it was identified through other resources.²¹ The reason behind the exclusion is that the majority of the characteristics which were to be extracted for the present study were not clear. Nevertheless, the results of the short report showed that the measurement of the diaphragmatic displacement at tidal breathing was reliable. This further supports that the studies which were carried out in this area did not report clearly the measures of reliability.

5.1. Methodological considerations

Three of the included studies demonstrated low internal validity (1/7) and one of the included studies showed high internal validity (5/7) as assessed by the QUAREL checklist. Hence, it can be inferred that the included studies ranged between low and high quality in terms of internal validity measures. Most of the parameters of internal validity as assessed through the QUAREL checklist scored 'unclear' or 'no' in three of the included studies, indicating that the quality of the studies was low. The intimidation to internal validity as evaluated by the QUAREL checklist concerns the parameters of blinding raters, blinding clinical information, additional cues, and diseases.

In order to avoid bias between the reviewers who assessed the QUAREL checklist, the review team opted for one clinical content expert and a non-expert as recommended by earlier guidelines for performing a systematic review.¹⁷ In addition, the review team carried out a pilot test of methodological quality assessment on the various articles that were not included in the review. This could be the probable reason why the quality scores were graded as low, as operationalization and interpretation of each of the parameters are discussed earlier for methodological quality assessment using the QUAREL scoring method. Furthermore, the review group opted to have an international group of authors to reduce bias in the inclusion of articles.

5.2. Practical implications

The results of this systematic review imply that inter-observer reliability is typically good while intra-observer reliability is typically between moderate and good. Therefore, it can be interpreted that the results are encouraging enough to conclude that clinicians can replicate the DM values. The range of interpretation which is between moderate and good for intra-rater reliability is concerning the credibility of the assessment. The lack of consistency within the rater's reading challenges the credibility of the measurement. This can be overwhelmed by identifying factors such as knowledge of the anatomi-

cal landmark of the abdomen, experience of the investigator, the position of the patient, and positioning of the transducer. Therefore, it can be conceded that these factors need to be deliberated on to improve the consistency within the investigator's reading. In addition, currently, there are no reference standards for DM values, meaning there are no acceptable ranges of healthy and diseased subjects. Hence, there is a need to generate the reference values with regard to both healthy and clinical populations. Overall, the intra-observer and the inter-observer measures of reliability were found to be good. However, the results of the systematic review need to be interpreted with caution as there is limited literature in this particular field.

5.3. Limitations and recommendations of review

The systematic review, which assessed intra-observer and inter-observer reliability of DM using ultrasound and radiography was accomplished based on the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines. In addition, the measures of QUAREL, GRADE, and level of evidence grading were employed based on earlier guidelines.^{12,15–17}

Even though the measures of systemic review guidelines were followed, the study has certain limitations. Firstly, only articles that were published in English for assessing intra-observer and inter-observer reliability were included. Even though no articles were retrieved during the search other than articles in English, this could be considered as one of the limitations. This could be because of the selection of the databases for the present systematic review: it could be that these databases were not able to retrieve articles other than those in English.

Out of the four articles included in the systematic review, two articles were from Brazil, one was from France, and the fourth was from the United Kingdom. The impression is that the study was including two articles from South America and two from Europe. This indicates that articles relevant to the topic may have been published in languages other than English which the team may have presumably missed out. Secondly, most of the articles which were included in this review performed DM assessment on healthy subjects, which means that there is a possibility that the results of the assessment may be different in various clinical conditions. Hence, the measure of reliability needs to be tested on clinical populations to ascertain the reliability measures of DM. Studies with similar methods need to be carried out on various clinical populations, such as those suffering from musculoskeletal, neurological, and cardio-respiratory conditions for the study findings of the reliability measures to be generalized.

6. CONCLUSIONS

The results of this systematic review indicate that diaphragmatic mobility assessment is presented in studies with moderate-to-good reliability with low risk of bias. The clinical implications of the tests may be suggested with caution.

Therefore, utilization of this technique may be recommended. However, future research is important to evaluate diaphragmatic mobility in both healthy and diseased subjects.

Conflict of interest

None declared.

Acknowledgements

The authors would like to thank the library assistance of Universiti Teknologi MARA, Puncak Alam, Malaysia.

References

- 1 Kolar P, Sulc J, Kyncl M, et al. Postural function of the diaphragm in persons with and without chronic low back pain. *J Orthop Sports Phys Ther.* 2012;42(4):352–362. <https://doi.org/10.2519/jospt.2012.3830>.
- 2 Paulin E, Yamaguti WPS, Chammass MC, et al. Influence of diaphragmatic mobility on exercise tolerance and dyspnea in patients with COPD. *Respir Med.* 2007;101(10):2113–2118. <https://doi.org/10.1016/j.rmed.2007.05.024>.
- 3 Rocha T, Souza H, Brandão DC, et al. The Manual Diaphragm Release Technique improves diaphragmatic mobility, inspiratory capacity and exercise capacity in people with chronic obstructive pulmonary disease: A randomised trial. *J Physiother.* 2015;61(4):182–189. <https://doi.org/10.1016/j.jphys.2015.08.009>.
- 4 Santana PV, Prina E, Albuquerque AL, Carvalho CR, Caruso P. Identifying decreased diaphragmatic mobility and diaphragm thickening in interstitial lung disease: the utility of ultrasound imaging. *J Bras Pneumol.* 2016;42(2):88–94. <https://doi.org/10.1590/S1806-37562015000000266>.
- 5 Haji K, Royse A, Green C, Botha J, Canty D, Royse C. Interpreting diaphragmatic movement with bedside imaging, review article. *J Crit Care.* 2016;34:56–65. <https://doi.org/10.1016/j.jcrc.2016.03.006>.
- 6 Kolář P, Neuwirth J, Šanda J, et al. Analysis of diaphragm movement during tidal breathing and during its activation while breath holding using MRI synchronized with spirometry. *Physiol Res.* 2009;58(3):383–392. <https://doi.org/10.1093/ehjci/jeu087>.
- 7 Rocha FR, Brüggemann AKV, Francisco D de S, Medeiros CS de, Rosal D, Paulin E. Diaphragmatic mobility: relationship with lung function, respiratory muscle strength, dyspnea, and physical activity in daily life in patients with COPD. *J Bras Pneumol.* 2017;43(1):32–37. <https://doi.org/10.1590/s1806-37562016000000097>.
- 8 Toledo NSG, Kodaira SK, Massarollo PCB, Pereira OI, Mies S. Right hemidiaphragmatic mobility: assessment with US measurement of craniocaudal displacement of left branches of portal vein. *Radiology.* 2003;228(2):389–394. <https://doi.org/10.1148/radiol.2282011554>.
- 9 Bruton A, Conway JH, Holgate ST. Reliability: What is it, and how is it measured? *Physiotherapy.* 2000;86(2):94–99. [https://doi.org/10.1016/S0031-9406\(05\)61211-4](https://doi.org/10.1016/S0031-9406(05)61211-4).
- 10 Saltiel R V, Grams ST, Pedrini A, Paulin E. High reliability of measure of diaphragmatic mobility by radiographic method in healthy individuals. *Brazilian J Phys Ther.* 2013;17(2):128–136. <https://doi.org/10.1590/S1413-35552012005000076>.
- 11 Grams ST, von Saltiel R, Mayer AF, et al. Assessment of the reproducibility of the indirect ultrasound method of measuring diaphragm mobility. *Clin Physiol Funct Imaging.* 2014;34(1):18–25. <https://doi.org/10.1111/cpf.12058>.
- 12 Lucas NP, Macaskill P, Irwig L, Bogduk N. The development of a quality appraisal tool for studies of diagnostic reliability (QAREL). *J Clin Epidemiol.* 2010;63(8):854–861. <https://doi.org/10.1016/j.jclinepi.2009.10.002>.
- 13 May S, Chance-Larsen K, Littlewood C, Lomas D, Saad M. Reliability of physical examination tests used in the assessment of patients with shoulder problems: A systematic review. *Physiotherapy.* 2010;96(3):179–190. <https://doi.org/10.1016/j.physio.2009.12.002>.
- 14 May S, Littlewood C, Bishop A. Reliability of procedures used in the physical examination of non-specific low back pain: A systematic review. *Aust J Physiother.* 2006;52(2):91–102. [https://doi.org/10.1016/S0004-9514\(06\)70044-7](https://doi.org/10.1016/S0004-9514(06)70044-7).
- 15 Lucas N, Macaskill P, Irwig L, et al. The reliability of a quality appraisal tool for studies of diagnostic reliability (QAREL). *BMC Med Res Methodol.* 2013;13:111. <https://doi.org/10.1016/j.jclinepi.2009.10.002>.
- 16 Guyatt G, Oxman AD, Akl EA, et al. GRADE guidelines: 1. Introduction – GRADE evidence profiles and summary of findings tables. *J Clin Epidemiol.* 2011;64(4):383–394. <https://doi.org/10.1016/j.jclinepi.2010.04.026>.
- 17 Van Tulder M, Furlan A, Bombardier C, Bouter L. Updated Method Guidelines for Systematic Reviews in the Cochrane Collaboration Back Review Group. *Spine (Phila Pa 1976).* 2003;28(12):1290–1299. <https://doi.org/10.1097/BRS.0b013e3181b1c99f>.
- 18 Houston JG, Morris AD, Howie CA, Reid JL, McMillan N. Technical report: Quantitative assessment of diaphragmatic movement – A reproducible method using ultrasound. *Clin Radiol.* 1992;46(6):405–407. [https://doi.org/10.1016/S0009-9260\(05\)80688-9](https://doi.org/10.1016/S0009-9260(05)80688-9).
- 19 Boussuges A, Gole Y, Blanc P. Diaphragmatic motion studied by M-mode ultrasonography. *Chest.* 2009;135(2):391–400. <https://doi.org/10.1378/chest.08-1541>.
- 20 Portney LG, Watkins MP. *Foundations of Clinical Research Applications to Practice.* 3rd Ed. New Jersey: Prentice Hall Health; 2009.
- 21 Blaney F, English CS, Sawyer T. Sonographic measurement of diaphragmatic displacement during tidal breathing manoeuvres – a reliability study. *Aust J Physiother.* 1999;45:41–43.