Manual Therapy xxx (2015) 1–13

Contents lists available at ScienceDirect

Manual Therapy

journal homepage: www.elsevier.com/math

Systematic review

Measurement methods to assess diastasis of the rectus abdominis muscle (DRAM): A systematic review of their measurement properties and meta-analytic reliability generalisation

A.T.M. van de Water [a], D.R. Benjamin [a, b, *]

[a] Department of Rehabilitation, Nutrition and Sport, School of Allied Health, La Trobe University, Bundoora, Victoria, Australia
[b] Physiotherapy Department, Angliss Hospital, Eastern Health, Upper Ferntree Gully, Victoria, Australia

ARTICLE INFO

Article history:
Received 24 June 2015
Received in revised form
19 September 2015
Accepted 23 September 2015

Keywords:
Validity
Reliability
Psychometrics
Women's health
DRAM

ABSTRACT

Study design: Systematic literature review.
Background: Diastasis of the rectus abdominis muscle (DRAM) has been linked with low back pain, abdominal and pelvic dysfunction. Measurement is used to either screen or to monitor DRAM width. Determining which methods are suitable for screening and monitoring DRAM is of clinical value.
Objectives: To identify the best methods to screen for DRAM presence and monitor DRAM width.
Methods: AMED, Embase, Medline, PubMed and CINAHL databases were searched for measurement property studies of DRAM measurement methods. Population characteristics, measurement methods/procedures and measurement information were extracted from included studies. Quality of all studies was evaluated using 'quality rating criteria'. When possible, reliability generalisation was conducted to provide combined reliability estimations.
Results: Thirteen studies evaluated measurement properties of the 'finger width'-method, tape measure, calipers, ultrasound, CT and MRI. Ultrasound was most evaluated. Methodological quality of these studies varied widely. Pearson's correlations of $r = 0.66–0.79$ were found between calipers and ultrasound measurements. Calipers and ultrasound had Intraclass Correlation Coefficients (ICC) of 0.78–0.97 for test–retest, inter- and intra-rater reliability. The 'finger width'-method had weighted Kappa's of 0.73–0.77 for test–retest reliability, but moderate agreement (63%; weighted Kappa = 0.53) between raters. Comparing calipers and ultrasound, low measurement error was found (above the umbilicus), and the methods had good agreement (83%; weighted Kappa = 0.66) for discriminative purposes.
Conclusions: The available information support ultrasound and calipers as adequate methods to assess DRAM. For other methods limited measurement information of low to moderate quality is available and further evaluation of their measurement properties is required.

1. Introduction

Diastasis of rectus abdominis muscle (DRAM) is relatively common and can have negative health consequences (Fast et al., 1990; Gilleard and Brown, 1996; Lee et al., 2008; Keeler et al., 2012). Therefore, screening for the presence of DRAM is often routinely practiced in inpatient and outpatient services, in particular in women during the pre- and post-natal periods (Keeler et al., 2012). If present, DRAM is monitored over time to assess natural recovery or to determine when interventions may be warranted. Appropriate and clinically useful measurement of DRAM is needed to help clinicians make such decisions.

DRAM or the increase of the inter-recti distance of the rectus abdominis muscle is due to stretching and thinning of the linea alba (Rath et al., 1996; Hsia and Jones, 2000). It is measured with one of several available methods, such as finger widths, calipers or ultrasound. Although consensus is still lacking, a widening of greater than 2.2–2.3 cm as identified by ultrasound measurements (Coldron et al., 2008; Liaw et al., 2011), has been considered a clinically important DRAM and has been connected to adverse musculoskeletal concerns in clinical practice.

* Corresponding author. Department of Rehabilitation, Nutrition and Sport, School of Allied Health, La Trobe University, Bundoora 3086, Australia. Tel.: +61 413599905.
E-mail address: deenikabenjamin@optusnet.com.au (D.R. Benjamin).

http://dx.doi.org/10.1016/j.math.2015.09.013
1356-689X/

Although the association is not definitive, presence, size and duration of DRAM have been linked to pelvic and low back pain (Taranto, 1990; Khushboo et al., 2014). It has been found to weaken abdominal muscles (Liaw et al., 2011) and disturb their functions in lumbo-pelvic stability (Khushboo et al., 2014). DRAM has also been associated with pelvic floor dysfunction (Spitznagle et al., 2007). Even though these findings are from small studies, it is considered important to identify DRAM presence and monitor DRAM width over time particularly in combination with such dysfunctions. For these measurement purposes, methods need to have sound measurement properties whilst being clinically feasible.

The purpose of clinical measurement can be to predict, to discriminate or to monitor (Kirshner and Guyatt, 1985). The choice of measurement method should be based on the purpose of measurement, its measurement properties and the clinical situation (Streiner and Norman, 2008). To screen for DRAM concerns discrimination between those with and without clinically important DRAM, which need reliable methods. For monitoring purposes of DRAM width over time to evaluate treatment, methods that are also responsive or sensitive to change in DRAM width are required.

Currently few studies exist that evaluate the different methods for measuring DRAM. Several methods, such as calipers, tape measures, ultrasound and the traditional 'finger width'-method (palpation), are used in clinical practice (Keeler et al., 2012) and clinical studies aiming to evaluate treatment effects on DRAM width (Mesquita and Machado, 1999; Benjamin et al., 2014). For example, the size of the diastasis can be measured by the number of finger widths that span the diastasis on palpation or by using calipers for which the distance between the tips of calipers (fitted across the width of the diastasis) is read. Ultrasound measurements, as well as measurements from Computed Tomography (CT) and Magnetic Resonance Imaging (MRI) are established from on-screen rulers within the software displaying the images. Claims have been made that the 'finger width'-method is 'unreliable' (Bursch, 1987; Mota et al., 2012), and that for example measurements on CT or MRI images can be considered 'gold standard' (Mendes et al., 2007; Mota et al., 2012; Barbosa et al., 2013). However, little supporting evidence on measurement properties or evidence based on incorrect statistics is provided with such claims, introducing potentially invalid statements.

When considering measurement purposes, it is important to evaluate the measurement properties of methods used to measure DRAM width. This could help guide clinicians and researchers towards appropriate and clinically meaningful measurement of DRAM. Therefore the aim of this systematic review was to summarise the literature and evaluate measurement properties of the available methods for measuring the presence or width of DRAM.

2. Methods

2.1. Data sources and searches

For the identification of studies evaluating measurement properties of methods for measurement of DRAM width, an electronic search strategy (Appendix A) consisting of two components was developed: population (with search combinations of, for example, "diastasis" AND "rectus abdominis muscles"), and -where possible- measurement properties (sensitive filter by Terwee et al., 2009). Since DRAM does not only occur in pre- and post-natal women (Moesbergen et al., 2009; De'Ath et al., 2010), and with the aim to collect all measurement information available on DRAM measurement methods, we did not limit the search of this review to a specific population, for example pre- and post-natal women.

Five databases were searched from their earliest available time until 21 December 2014, and included AMED (Allied and Complementary Medicine), CINAHL (Cumulative Index to Nursing and Allied Health Literature), Embase, Medline and PubMed. In addition to electronic searches, reference lists of included studies and other related articles were screened for potential relevant studies.

2.2. Study selection

Study selection on title/abstract and retrieved full text articles was performed independently by two reviewers. Specific selection criteria (Table 1) were set *a priori*, and discussed prior to the selection process. Discussion resolved any selection discrepancies.

2.3. Data extraction and quality assessment

A priori designed data extraction forms included details of study populations, measurement protocols, measurement information, and any notes on the analysed studies. Data were extracted by one reviewer (AvdW) and checked by a second reviewer (DB) for completeness and correctness.

Methodological quality of studies was assessed with a modified version of the "COnsensus-based Standards for the selection of health Measurement INstrument" (COSMIN checklist version 9) (Mokkink et al., 2010a, 2010b) and a component of the QUADAS-2 (Whiting et al., 2011). The COSMIN checklist was developed for quality assessment of development and psychometric studies of patient-reported health measurement instruments. Currently no established and validated tool for assessing quality of psychometric studies of clinician administered outcome measures is available, therefore a modified version of the COSMIN checklist was used. The COSMIN checklist consists of 9 sub-checklists for measurement properties, such as reliability or criterion validity. These sub-checklists are selected based on what properties a study evaluated. It also contains two additional checklists for 'Interpretability' and 'Generalisability' which are completed for all studies and relate to data presentation. Some items of the sub-checklists and two additional checklists were not relevant to clinician-administered measurement of DRAM; for example, reporting of the percentage of missing items. Hence, we used a modified version of the COSMIN where these items were marked Not Applicable (N/A). This has been done previously in performance tests for walking ability (van Bloemendaal et al., 2012).

Methodological quality scores are presented as 'number of positively rated items' out of 'total number of applicable items' (Table 3). As well as design requirements, the COSMIN also includes whether appropriate statistics have been used for the investigated measurement properties. Presentation of these results has been reported separately (Table 3) to highlight potential incorrect inferences made in the included studies. Two reviewers completed the modified COSMIN checklist for each study independently (Online Appendix). Any disagreements were resolved by discussion.

When studies presented data on the diagnostic accuracy of DRAM measurement methods, study quality was also evaluated using the Risk of Bias component of the QUADAS-2 (Whiting et al., 2011) by two independent reviewers. The QUADAS-2 is the successor of the original Quality Assessment of Diagnostic Accuracy Studies (QUADAS) tool. Only "Phase 3: Risk of Bias and applicability judgement" which consists of 4 domains (1. patient selection, 2. index test, 3. reference standard, and 4. flow and timing) was used. This was done as included studies were not primary diagnostic accuracy studies and reporting on these four domains provides the reader with information on potential risks of bias. All domains were judged on "low/high/unclear risk of bias" and domains 1 to 3 are

Please cite this article in press as: van de Water ATM, Benjamin DR, Measurement methods to assess diastasis of the rectus abdominis muscle (DRAM): A systematic review of their measurement properties and meta-analytic reliability generalisation, Manual Therapy (2015), http://dx.doi.org/10.1016/j.math.2015.09.013

Table 1
Study selection criteria.

Inclusion	Exclusion
Population	
• diastasis of rectus abdominis muscle (including, for example, pre- and post-natal women, male with abdominal aortic aneurysm, abdominoplasty patients)	• non-human • cadaver studies
Measurement properties	
• validity (correlations between methods) • sensitivity, specificity • reliability (e.g. intra-rater, inter-rater) • measurement error, minimal detectable change • longitudinal validity (responsiveness)	
Publication type	
• full-text study publications	• abstracts • conference proceedings • pilot studies

also judged on "low/high/unclear concerns regarding applicability". Any disagreements in judgement were resolved by discussion.

2.4. Data synthesis and analysis

Where publications of included studies reported raw data of their participants but no summarised sample data, descriptive statistics were used to describe characteristics of the sample of those studies.

Several guidelines were used to determine suitability of statistics for evaluation of measurement properties (Streiner and Norman, 2008; Mokkink et al., 2010a). A priori consensus was reached that studies had to use the proposed statistics of these guidelines for the results of these studies to be included in this review. These statistics included: Pearson's product–moment correlation coefficient (r) for concurrent validity, Standardised Error of Measurement (SEM), Minimal Detectable Change (MDC) or Bland Altman's Limits of Agreement for measurement error, and Intraclass Correlation Coefficient (ICC), Concordance Correlation Coefficient (CCC) or (weighted) Cohen's Kappa for reliability testing. Any results using statistics other than these were discarded as they would not evaluate measurement properties coherent with the review's primary interest.

Measurement properties of interest (for example, concurrent validity or measurement error) were estimated by the review authors under two conditions: measurement property values were not reported by the authors of included studies, and if presented data allowed their calculation. This means, included publications had to present raw data of each person in the sample, or, for evaluation of the diagnostic accuracy of DRAM methods, a 2×2 contingency table.

Where appropriate, the following statistics were used to evaluate measurement properties: for concurrent validity testing, Pearson's *r* was used for continuous data.

Measurement error between measurement methods of DRAM width was evaluated through Bland and Altman's Limits of Agreement. Sensitivity, specificity, positive predictive value and negative predictive value were calculated for dichotomous data (presence/absence) related to the diagnostic test accuracy of DRAM measurement methods.

When possible, pooling of reliability estimations, based on Charter (Charter, 2003; Streiner and Norman, 2008), was conducted on data from studies evaluating reliability. The measurement methods and locations from each study were considered prior to inclusion for pooled reliability calculations. Combining reliabilities was based on the mean score for the scale, its standard deviation, the sample size and ICC. The combined reliability coefficient is weighted by the sample sizes, mean and SD of DRAM width of the samples. For more detail on reliability generalisation including formulae, we refer to Charter (2003) and Streiner and Norman (2008) (pages 202–207). Forest plots were created following the methods described by Neyeloff et al. (2012).

3. Results

The electronic database searches resulted in 211 citations of which 18 were potentially eligible (Table 1) and retrieved to screen in full-text. Of these, 12 studies evaluated measurement properties of methods measuring DRAM width. Reference list screening resulted in one more study (Liaw et al., 2006) which was in Chinese language. This article was translated into English prior to data extraction. Fig. 1 shows the flow of identification, selection and inclusion of the 11 studies.

The included studies evaluated several clinical measurement methods for DRAM width (Table 2), including the 'finger width'-method (Bursch, 1987; Mota et al., 2013), tape measure (Emanuelsson et al., 2014), calipers (Boxer and Jones, 1997; Barbosa et al., 2013; Chiarello and McAuley, 2013), ultrasound (Liaw et al., 2006; Mendes et al., 2007; Liaw et al., 2011; Mota et al., 2012; Chiarello and McAuley, 2013; Iwan et al., 2014), Magnetic Resonance Imaging (MRI) (Elkhatib et al., 2011), Computed Tomography (CT) (Nahas et al., 2001; Emanuelsson et al., 2014) and intra-operative measurements with a ruler (Nahas et al., 2001; Elkhatib et al., 2011), surgical compass (Mendes et al., 2007).

3.1. Methodological quality

Methodological quality using the COSMIN checklists found variable quality between studies (Table 3; Online Appendix). The level of agreement between raters was Cohen's kappa = 0.91 (95%CI 0.87–0.96) (Percentage Agreement = 95.9%). Some of the included studies employed statistics (Bursch, 1987; Nahas et al., 2001; Mendes et al., 2007; Elkhatib et al., 2011; Mota et al., 2012) inconsistent with this review's criteria or had important design flaws such as lack of blinding for reliability studies (Liaw et al., 2011; Barbosa et al., 2013; Mota et al., 2013) or very small sample size (Elkhatib et al., 2011) and therefore had an increased risk of bias.

Of the seven studies that validated one measurement method against a reference standard, only one used statistics that were consistent with the recommended statistics (Table 3). Two studies (Nahas et al., 2001; Elkhatib et al., 2011) did not employ statistics fitting this review's criteria for validity testing, but reported raw data which were used to calculate correlations for concurrent validity testing.

Please cite this article in press as: van de Water ATM, Benjamin DR, Measurement methods to assess diastasis of the rectus abdominis muscle (DRAM): A systematic review of their measurement properties and meta-analytic reliability generalisation, Manual Therapy (2015), http://dx.doi.org/10.1016/j.math.2015.09.013

Table 2
Study characteristics.

First author, year	Sample characteristics	Delivery, parity, body composition	Measurement tools	Protocol
Post-partum participants				
Barbosa et al. (2013)	n = 106, all female, 27.1 years (SD 5.97), post-natal <72 h	**Delivery** caesarean 62.24% vaginal 37.76 **Parity** 2.2 (SD 1.4) **Body composition** BMI 29.95 (SD 6.80)	**Ultrasound** Medson SonoAce 8000, 5–7 MHz transducer **Caliper** UIU-STOOLS professional-Vernier CLA006	One assessor. Supine, legs flexed. Measurements at +3, +6, +9, +12 cm of umbilicus. Active measurements by flexing trunk until scapulae came off the bed.
Boxer and Jones (1997)	n = 30, all female, 31.3 years (SD 3.3), post-natal 11.1 weeks (SD 5.0)	**Delivery** vaginal 100% (caesarean excluded) **Parity** median 1,5 (1–3) primiparous n = 17 multiparous n = 13 **Body composition** Body fat (%) 20.8–39.1%	**Caliper** nylon dial, 0–150 mm, Baty International	One assessor with 15 years clinical physiotherapy experience, minimal experience with calipers. Supine, knees 90° flexed, hands resting on thighs. Measurement at midpoint umbilicus. 3 trials with 15 min rest; per trial measurements in resting, active (with scapulae just off the bed) status. Dial facing away for blinding, reporting by 'impartial recorder'
Bursch (1987)	n = 40, all female, 16–31 years, post-natal <4 days	**Delivery** vaginal 100% (caesarean excluded) **Parity** no info **Body composition** no info	**Finger width** no standardisation of finger width between raters	Four assessors; 2 experienced, 2 new. Supine, knees flexed. 3 measurements with 3 min rest, scores averaged Measurements at +4.5 cm of umbilicus Finger insertion depth to rater's proximal interphalangeal joints. After insertion of fingers, participant performed 3 partial sit-up with arms extended, scapulae just of the bed during which measurements were taken
Liaw et al. (2011)	n = 30 all female, 32.1 years (SD 3.0), 7wk and 6mo post-natal	**Delivery** vaginal 100% **Parity** primiparous n = 17 multiparous n = 13 **Body composition** BMI 21.5 (SD 2.8)	**Ultrasound** SSD-550, 7.5 MHz 38 mm linear transducer; in B mode	One assessor. Supine, two pillows under knees. Measurements at +2.5 cm, upper and lower margin umbilicus, −2.5 cm 3 images taken at end of normal expiration measurements taken using on-screen caliper; order of locations randomised
Mota et al. (2012)	n = 24, all female, 30.4 years (16–55) two subgroups: n = 12 at 10.9wk (9–13) post-partum n = 12 at 11.5 years (0–24) post-partum	**Delivery** no info **Parity** 0.75 (0–2) **Body composition** BMI 22.7 (18.9–28.5)	**Ultrasound** LOGIQ e, GE Healthcare; 4–12 MHz, 39 mm linear transducer; B-mode. Algorithm in MATLAB image-processing software (The MathWorks, Inc) helped finding the medial margins of rectus	One assessor. Supine, knees 90° flexed, arms alongside body on bed Measurements at +2 cm, −2 cm of umbilicus 3 situations: resting, 'draw in' and partial sit up; contractions held for 3 s, resting time 6–10 s. Images taken at end of exhalation (determined by visual inspection) 1 image per location per situation Retest on convenient day for participant
Healthy participants				
Chiarello and McAuley, 2013	n = 56, 11 male, 45 female, 34.8 years (SD 9.8; 19–64)	**Parity** (n = 45 female) 22 nulliparous 23 various parity **Body composition** BMI 24.3 (SD 4.3)	**Calipers** nylon digital calipers, Mitutoyo America Corporation, Aurora, IL, USA **Ultrasound** LOGIQ Book XP, GE Healthcare, Waukesha, WI, USA; 5-MHz curvilinear transducer	Two independent assessors. Measurements at +4.5 cm, −4.5 m of umbilicus 2 situations: resting (hook-lying, arms down by the side, with 1 pillow placed beneath the head) and active (crossed the arms over the chest and raised the head until the spine of the scapulae was off the table surface). Caliper measurements always preceded the Ultrasound measurements; caliper measurements included palpation with fingers to identify medial borders of recti
Iwan et al. (2014)	n = 30 14 male, 16 female 24.4 years (SD 7.4; 20–53)	**Parity** (n = 16 female) 13 nulliparous 3 various parity (1–4) **Body composition** BMI 23.9 (SD 2.8)	**Ultrasound** **Low-resolution** Chison 8300 Deluxe 7.5 MHz linear transducer (Chison Medical Imaging Co. Ltd., China) **High-resolution** Phillips iU22 12.5 MHz linear transducer (Royal Philips Electronics, The Netherlands)	Two independent assessors. Supine. Images taken at end of expiration. Measurements at +2 cm and −2 cm umbilicus with digital measurement caliper setting on the machine. 2 situations: resting and partial curl up as per Chiarello and McAuley (2013). 1 min rest between measurements within session (n = 30); 5 weeks between sessions (n = 10) Measurements were random for: rater 1 and 2, situations, measurement locations, Ultrasound machine.
Liaw et al. (2006)	n = 42, 16 male, 26 female, 21.6 years (SD 1.3)	**Parity** no info **Body composition** BMI 21.0	**Ultrasound** SSD-550, 7.5MHz linear array probe; in B-mode	One assessor for test–retest reliability. Supine. Images taken at maximum inhalation Measurements at +4.5 cm, upper and lower

Please cite this article in press as: van de Water ATM, Benjamin DR, Measurement methods to assess diastasis of the rectus abdominis muscle (DRAM): A systematic review of their measurement properties and meta-analytic reliability generalisation, Manual Therapy (2015), http://dx.doi.org/10.1016/j.math.2015.09.013

Table 2 (*continued*)

First author, year	Sample characteristics	Delivery, parity, body composition	Measurement tools	Protocol
				edge umbilicus, −4.5 cm Locations measured in random order; measurement taken directly on screen, read and reported by assistant for blinding Retest 2 days later
	n = 31 16 male, 15 female 22.1 years (SD 1.2)	**Parity** no info **Body composition** no info	**Ultrasound** SSD-550, 7.5 MHz linear array probe; in B-mode	Two assessors for inter-rater reliability. Same as above. Second assessor took images and measurements 15 min later
Mota et al. (2013)	n = 20, all female, 29.3 years (16–49)	**Parity** 0.7 (0–2) **Body composition** BMI 23.0 (18.9–28.5)	**'Finger width'-method** no standardisation procedures described **Ultrasound** GE Logic-e, 4–12 MHz, 39 mm linear transducer, B-mode	Two assessors, 7 and 31 years experience; One of them performed ultrasound imaging directly after their palpation assessment. Supine, knees 90° flexed, arms alongside body on bed. Measurements at +2 cm, −2 cm of umbilicus. 1 situation: abdominal crunch (shoulder blades just clearing the table) For inter-rater reliability, second assessor took measurements 2 min later; for test–retest reliability, retest period was 3.9 days (SD 3.9)
Abdominoplasty participants				
Elkhatib et al. (2011)	n = 20, all female, 33.6 years (SD 7.02) **Note**: Data on n = 10 only	**Delivery** no info **Parity** no info **Body composition** BMI 29.1 (22.9–35)	**Magnetic Resonance Imaging** Siemens scanner, Magnetom Avanto 1.5T **Ruler** Aspen, 15 cm	MRI images (T2-weighted axial scans) were taken in deep inspiration at levels of Lumbar 2 and Sacral 3. Intra-operative measurements were taken at supra and infra umbilical levels corresponding with L2 and S3 levels.
Emanuelsson et al. (2014)	n = 56 2 male, 54 female 39.8 years (25–60) **Note**: Data on n = 55	**Delivery** no info **Parity** no info **Body composition** BMI 23 (18–31)	**Computed Tomography** Siemens Definition AS machine **Tape measure** for clinical and intra-operative assessments	Measurement taken at mid-way xiphoid-umbilicus and mid-way umbilicus-pubic symphysis. Clinical and CT measurements (deep inspiration) were taken by the same radiologist (two radiologists collected data). Clinical tape measurements were performed three times by the same investigator and averaged.
Mendes et al. (2007)	n = 20, ~19/20 female, no further info	**Delivery** 19/20 had previous caesarean section; **Parity** no info **Body composition** no info	**Ultrasound** Medson SonoAce 8000 **Surgical compass** no info	One assessor (ultrasonographist) using ultrasound. Two assessors (surgeon, assistant) using surgical compass. 7 measurement levels: +12, +9, +6, +3, umbilicus, −2, −4 cm pre-operative ultrasound images taken at sustained maximal inspiration and expiration, values were averaged
Nahas et al. (2001)	n = 20, all female, 21–52 years	**Delivery** no info **Parity** no info **Body composition** moderate obese (9) average (6) thin (5) no further info	**Computed Tomography** scan ruler; no info **Ruler** no info	Supine Measurements at +3 cm and −2 cm of umbilicus no further info

Ten studies (Bursch, 1987; Boxer and Jones, 1997; Liaw et al., 2006; Mendes et al., 2007; Liaw et al., 2011; Mota et al., 2012; Barbosa et al., 2013; Mota et al., 2013; Emanuelsson et al., 2014; Iwan et al., 2014) evaluated a form of reliability or agreement (test–retest reliability, intra-rater reliability, inter-rater reliability or agreement between measurement methods (Table 5)), but two of these (Bursch, 1987; Mendes et al., 2007) did not report the statistics consistent with this review's criteria.

One study by Barbosa et al. (2013) reported on kappa agreement between ultrasound (as reference) and calipers for accurate diagnosis of DRAM, but did not report on other clinically useful values from their 2 × 2 contingency table such as sensitivity or positive predictive value. These were previously calculated by us (van de Water and Benjamin, 2014) using the data presented by the authors (Barbosa et al., 2013), and are also presented in this review (Table 4).

Barbosa et al. (2013) was the only study to provide information on diagnostic accuracy of calipers compared to ultrasound, and therefore was the only study assessed with the Phase 3-part of the QUADAS-2. Domain 1 "patient selection" was scored as 'unclear' risk of bias and 'high' concern regarding applicability. Domain 2 "index test" and Domain 3 "reference standard" were scored 'high' risk of bias and applicability concern was 'low'. Finally, domain 4 "flow and timing" was judged as 'low' risk of introducing bias.

3.2. Measurement information per measurement method

3.2.1. 'Finger width'-method

Two studies (Bursch, 1987; Mota et al., 2013) evaluated measurement properties of this method as an evaluative instrument, and not as tool to discriminate between presence and absence of DRAM. Mota et al. (2013) evaluated the reliability and agreement

Please cite this article in press as: van de Water ATM, Benjamin DR, Measurement methods to assess diastasis of the rectus abdominis muscle (DRAM): A systematic review of their measurement properties and meta-analytic reliability generalisation, Manual Therapy (2015), http://dx.doi.org/10.1016/j.math.2015.09.013

Table 3
Results of COSMIN methodological quality assessment.

Author, year	Measures	Measurement properties	Design requirements[a] (COSMIN v9)	Design flaws[b]	Statistics consistent with criteria	Statistics used
Barbosa et al. (2013)	Ultrasound Calipers	Reliability	6/7	Yes	Yes	Kappa
		Measurement error	6/7	Yes	Yes	LOA
		Criterion validity	2/2	Yes	Yes	Pearson's *r*
		Interpretability	2/4	Yes	–	
		Generalisability	5/6	–	–	
Boxer and Jones (1997)	Calipers	Test–retest reliability	6/7	No	Yes	$ICC_{3,1}$
		Measurement error	6/7	No	Yes	SEM
		Interpretability	2/4	No	–	
		Generalisability	4/6	–	–	
Bursch (1987)	Finger widths	Test–retest reliability	6/7	No	No	ANOVA
		Interpretability	2/4	No	–	
		Generalisability	3/6	–	–	
Chiarello and McAuley (2013)	Calipers Ultrasound	Criterion validity	2/2	No	Yes	$ICC_{3,2}$ (95%CI)
		Interpretability	3/4	No	–	
		Generalisability	6/6	–	–	
Elkhatib et al. (2011)	MRI Ruler	Measurement error	2/7	Yes	No	Not reported (raw data provided)
		Criterion validity	1/2	Yes	No	Spearman correlation (raw data provided
		Interpretability	1/4	Yes	–	
		Generalisability	5/6	–	–	
Emanuelsson et al. (2014)	CT	Measurement error	5/7	No	Yes	LOA
	Tape measure	Interpretability	3/4	No	–	
		Generalisability	6/6	–	–	
	- Clinical					
	- Intra-OP					
Iwan et al. (2014)	Ultrasound	Reliability	6/7	No	Yes	$ICC_{2,1(pooled)}$
		Measurement error	6/7	No	Yes	SEM/MDC_{95}/LOA
		Criterion validity	1/2	No	Yes	$ICC_{2,k}$
		Interpretability	2/4	No	–	
		Generalisability	6/6	–	–	
Liaw et al. (2006)	Ultrasound	Reliability	6/7	No	Yes	$ICC_{3,1}$ $ICC_{2,1}$ (95%CI)
		Measurement error	6/7	No	Yes	SEM/MDC_{95}/LOA
		Interpretability	2/4	No	–	
		Generalisability	4/6	–	–	
Liaw et al. (2011)	Ultrasound	Reliability	5/7	Yes	Yes	$ICC_{3,1(pooled)}$
		Measurement error	5/7	Yes	Yes	SEM/MDC_{95}
		Interpretability	2/4	Yes	–	
		Generalisability	6/6	–	–	
Mendes et al. (2007)	Ultrasound Surgical compass	Reliability	6/7	No	No	Wilcoxon's test
		Criterion validity	1/2	No	No	Wilcoxon's test
		Interpretability	0/4	No	–	
		Generalisability	3/6	–	–	
Mota et al. (2012)	Ultrasound	Reliability	6/7	No	Yes	$ICC_{1,1}$ (95%CI)
		Measurement error	6/7	No	Yes	SEM/MDC_{95}
		Interpretability	2/4	No	–	
		Generalisability	4/6	–	–	
Mota et al. (2013)	Finger width Ultrasound	Reliability	6/7	Yes	Yes/No[c]	Weighted kappa
		Criterion validity	1/2	No	No	ANOVA
		Interpretability	2/4	Yes	–	
		Generalisability	5/6	–	–	
Nahas et al. (2001)	CT Intra-operative ruler	Measurement error	1/7	Yes	No	Not reported (raw data provided)
		Criterion validity	1/2	Yes	No	Not reported (raw data provided)
		Interpretability	2/4	No	–	
		Generalisability	5/6	–	–	

[a] Design requirements presented as: Items scored 'yes'/total items.
[b] Design flaws including no blinding or very small sample size.
[c] Yes/No indicates that correct statistics (weighted kappa) were used, but the weighting scheme was not stated.

between finger width measurements (as number of finger widths with intervals of half finger widths) of two raters, finding a weighted kappa value of 0.53 and percentage agreement of 62.5%. They also evaluated the test–retest reliability and found weighted kappa values of 0.73 and 0.77 for the two women's health physiotherapists with 7 and 31 years of experience, respectively. In addition, they aimed to compare ultrasound and 'finger width' measurements, but did not report statistics that could be used. Another study by Bursch (1987) aimed to evaluate the inter-rater reliability of the 'finger width'-method but also employed statistics that were inconsistent with this review's criteria.

3.2.2. Tape measure

Clinical measurements with a tape measure were compared by Emanuelsson et al. (2014) with intra-operative tape measurements and measurements on pre-operative CT images. Concordance

Please cite this article in press as: van de Water ATM, Benjamin DR, Measurement methods to assess diastasis of the rectus abdominis muscle (DRAM): A systematic review of their measurement properties and meta-analytic reliability generalisation, Manual Therapy (2015), http://dx.doi.org/10.1016/j.math.2015.09.013

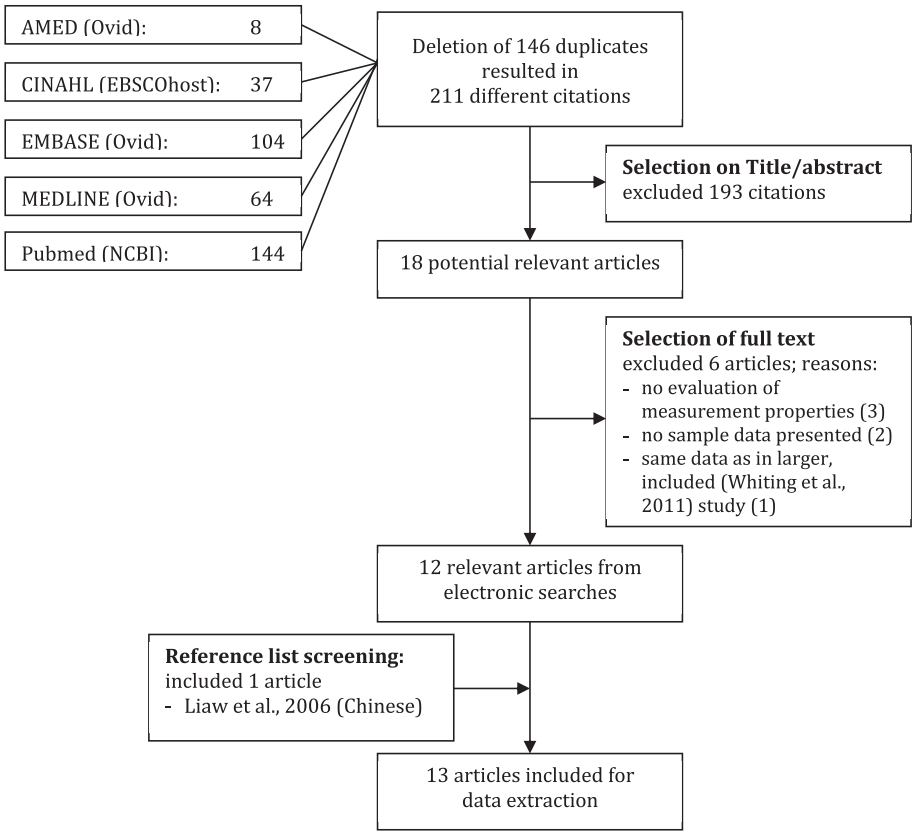

Fig. 1. Flowchart detailing identification and selection of studies.

Table 4
Validity of DRAM measurement methods.

First author, year	Outcome measures	Correlations[a]	Other	Notes
Barbosa et al. (2013)	Calipers Ultrasound	+3 cm $r = 0.66$ +6 cm $r = 0.71$ +9 cm $r = 0.69$ +12 cm $r = 0.79$	Sensitivity 89.7% Specificity 75.0% PPV 82.5% NPV 84.6%	"other" statistics were calculated (van de Water and Benjamin, 2014) using raw data reported by Barbosa et al.
Chiarello and McAuley (2013)	Calipers Ultrasound			statistics used for validity testing non-compliant with review criteria (paired *t*-test)
Elkhatib et al. (2011)	MRI ruler (intra-operative)	L2 $r = 1.00$ S3 $r = 0.99$		correlations were calculated using raw data (n = 10) presented
Iwan et al. (2014)	Ultrasound High vs Low resolution			statistics used for validity testing non-compliant with review criteria (*t*-tests)
Mendes et al. (2007)	Ultrasound surgical compass			statistics used for validity testing non-compliant with review criteria (Wilcoxon's test)
Mota et al. (2013)	'Finger width'-method Ultrasound			statistics used for validity testing non-compliant with review criteria (ANOVA)
Nahas et al. (2001)	CT ruler (intra-operative)	+3 cm $r = 0.89$ −2 cm $r = 0.78$		correlations were calculated using raw data (n = 20) presented

PPV, Positive Predictive Value; NPV, Negative Predictive Value; ICC, Intra-class correlation; MRI, Magnetic Resonance Imaging; L2, level of Lumbar 2; S3, level of Sacral 3; CT, Computed Tomography.
[a] Correlations are Pearson's *r*.

Correlation Coefficients (CCC; <0.90 indicating poor agreement (Emanuelsson et al., 2014)) of 0.37–0.48 were found between clinical and intra-operative measurements. Tape measurements and CT also had lower agreement (CCCs of 0.00–0.22) with wide Limits of Agreement of up to 2.5 cm (Table 5).

3.2.3. Calipers

Concurrent validity testing was conducted by Barbosa et al. (2013) who compared measurements on ultrasound images with caliper measurements and found, depending on the measurement location above the umbilicus, correlations of $r = 0.66$ to 0.79 between the two methods. Their mean differences of DRAM width estimates (above the umbilicus) between methods were less than 1 mm.

Also Chiarello and McAuley (2013) found that, for measurements above the umbilicus, calipers had similar estimations of DRAM width compared to ultrasound (SEM 0.01–0.17 cm). However, for measurements 4.5 cm under the umbilicus systematic differences of 0.74–1.43 cm were found between methods.

Please cite this article in press as: van de Water ATM, Benjamin DR, Measurement methods to assess diastasis of the rectus abdominis muscle (DRAM): A systematic review of their measurement properties and meta-analytic reliability generalisation, Manual Therapy (2015), http://dx.doi.org/10.1016/j.math.2015.09.013

Table 5
Reliability, agreement and measurement error of DRAM measurement methods.

First author, year	Reliability/agreement		Measurement error (in cm)	Notes
'Finger width'-method				
Mota et al. (2013)	Test–retest reliability	**$kappa_{weighted}$** 0.73–0.77 **Percentage agreement** 73–80%		
Mota et al. (2013)	Inter-rater reliability	**$kappa_{weighted}$** 0.53 **Percentage agreement** 62.5%		
Bursch (1987)	Inter-rater reliability			statistics used for validity testing non-compliant with review criteria (ANOVA); no raw data available
Calipers				
Boxer and Jones (1997)	Test–retest reliability	**$ICC_{3,1}$** rest: 0.93 crunch: 0.95	**SEM** 0.31 0.16	
Ultrasound				
Iwan et al. (2014)	Intra-rater reliability	**$ICC_{2,1}$ – Pooled** rest +2 cm: 0.96 rest −2 cm: 0.96 crunch +2 cm: 0.95 crunch −2 cm: 0.92		Pooled reliability based on data from 2 raters (experienced and novice), measuring 4 times each on 2 machines (high and low resolution ultrasound).
Liaw et al. (2011)	Intra-rater reliability	**$ICC_{3,1}$ – Pooled** +2.5 cm: 0.96 upper edge: 0.92 lower edge: 0.92 −2.5 cm: 0.96		Pooled reliability based on data at two time points (7 weeks and 6 months post-partum)
Mota et al. (2012)	Intra-rater reliability	**$ICC_{1,1}$ (95%CI)** rest +2 cm: 0.98 (0.95–1.00) rest −2 cm: 0.96 (0.90–0.98) crunch +2 cm: 0.94 (0.88–0.98) crunch −2 cm: 0.97 (0.93–1.00) draw in +2 cm: 0.93 (0.85–0.97) draw in −2 cm: 0.99 (0.97–1.00)	**SEM/MDC_{95}** 0.10/0.29 0.10/0.27 0.16/0.43 0.12/0.32 0.20/0.55 0.07/0.18 **Limits of Agreement** 1.1 (−0.20 to 0.21)	
Iwan et al. (2014)	Test–retest reliability	**$ICC_{2,1}$ – Pooled** rest +2 cm: 0.92 rest −2 cm: <0.65* crunch +2 cm: 0.82 crunch −2 cm: 0.74		Pooled reliability based on data from 2 raters (experienced and novice), measuring 4 times each on 2 machines (high and low resolution ultrasound). * One negative ICC (ICC = −0.51) was set as 0.00 for pooling.
Liaw et al. (2006)	Test–retest reliability	**$ICC_{3,1}$ (95%CI)** +4.5 cm: 0.95 (0.90–0.97) upper edge: 0.91 (0.82–0.95) lower edge: 0.85 (0.72–0.92) −4.5 cm: 0.90 (0.82–0.95)	**SEM/MDC_{95}** 0.07/0.20 0.08/0.23 0.10/0.29 0.08/0.23	
Mota et al. (2012)	Test–retest reliability	**$ICC_{1,1}$(95%CI)** rest +2 cm: 0.87 (0.73–0.94) rest −2 cm: 0.78 (0.56–0.90) crunch +2 cm: 0.83 (0.65–0.92) crunch −2 cm: 0.50 (0.14–0.75) draw in +2 cm: 0.90 (0.79–0.96) draw in −2 cm: 0.74 (0.48–0.88)	**SEM/MDC_{95}** 0.28/0.76 0.23/0.63 0.25/0.69 0.44/1.21 0.24/0.66 0.32/0.87 **Limits of Agreement** −0.03 (−8.67 to 8.34)	
Iwan et al. (2014)	Inter-rater reliability	**$ICC_{2,1}$ – Pooled** rest +2 cm: 0.97 rest −2 cm: 0.65 crunch +2 cm: 0.95 crunch −2 cm: 0.83		Pooled reliability based on data from 2 raters (experienced and novice), measuring 4 times each on 2 machines (high and low resolution ultrasound).
Liaw et al. (2006)	Inter-rater reliability	**$ICC_{2,1}$ (95%CI)** +4.5 cm: 0.86 (0.72–0.93) upper edge: 0.89 (0.75–0.95) lower edge: 0.78 (0.55–0.90) −4.5 cm: 0.83 (0.65–0.92)		
Between-methods/Parallel-forms reliability				
Barbosa et al. (2013)	Calipers Ultrasound	**kappa (presence/absence)** 0.66	**Limits of Agreement** +3 cm: −0.02 (<−2 to >2) +6 cm: 0.04 (−1.8 to 1.8) +9 cm: 0.08 (−1.9 to 2) +12 cm: −0.07 (−1.3 to 1.2)	based on n = 102; Limits of Agreement were approximated; blinding issues
Chiarello and McAuley (2013)	Calipers Ultrasound	**$ICC_{3,2}$ (95%CI)** rest +4.5 cm: 0.79 (0.64–0.88)	**SEM/MDC_{95}/Limits of Agreement** 0.01/0.04/0.03 (−2 to 2)	

Please cite this article in press as: van de Water ATM, Benjamin DR, Measurement methods to assess diastasis of the rectus abdominis muscle (DRAM): A systematic review of their measurement properties and meta-analytic reliability generalisation, Manual Therapy (2015), http://dx.doi.org/10.1016/j.math.2015.09.013

Table 5 (*continued*)

First author, year	Reliability/agreement		Measurement error (in cm)	Notes
		rest −4.5 cm: 0.40 (−0.02–0.65)	0.52/1.45/1.43 (0.5–2.5)	
		crunch +4.5 cm: 0.71 (0.51–0.83)	0.17/0.48/−0.03 (−1.5 to 1.5)	
		crunch −4.5 cm: 0.43 (0.03–0.67)	0.50/1.38/0.74 (−1 to 2.5)	
Emanuelsson et al. (2014)	CT	**CCC (95% CI)**	**Limits of Agreement**	X–U, mid-way xiphoid-umbilicus
	Tape measure:	CT vs. Intra-OP	X–U: −1.47 (−3.95–1.13)	U–S, mid-way umbilicus-pubic symphysis
	Intra-operative	X–U: 0.16 (0.02–0.30)	U–S: −2.37 (−5.29–0.55)	CCC was interpreted (Emanuelsson et al., 2014) as >0.99, almost perfect
	Clinical	U–S: 0.00 (−0.08–0.07)	no info	0.95–0.99, substantial
		CT vs. Clinical	no info	0.90–0.95, moderate
		X–U: 0.22 (0.06–0.37)	X–U: −0.04 (−2.46–2.37)	<0.90, poor
		U–S: 0.00 (−0.06–0.06)	U–S: 0.47 (−1.64–2.58)	
		Clinical vs. Intra-OP		
		X–U: 0.37 (0.12–0.57)		
		U–S: 0.48 (0.27–0.65)		
Mota et al. (2013)	Finger width			statistics used for validity testing non-compliant with review criteria (ANOVA)
	Ultrasound			

CCC, Concordance Correlation Coefficient; ICC, intraclass correlation coefficient; SEM, standard error of measurement; MDC, minimal detectable change.

Boxer and Jones (1997) evaluated test–retest reliability of calipers in post-partum women (with 15 min retest interval) and reported ICCs of 0.93–0.95 with small SEMs of 1.5–3 mm for measurements at the umbilicus.

Barbosa et al. (2013) also dichotomised their DRAM width measurement data into presence/absence of DRAM as measured by calipers and ultrasound. They found that the agreement between the methods in discriminating between absence and presence of DRAM was good (Landis and Koch, 1977) with a kappa of 0.66. When taking ultrasound as their reference standard and calipers as index test, from the data presented by Barbosa et al. (2013) diagnostic accuracy values could be calculated (van de Water and Benjamin, 2014). Sensitivity, or the chance to detect DRAM with calipers when it is indeed present, was 89.7% with a specificity of 75%. The positive predictive value, or the chance that DRAM is indeed present when positively tested with calipers, was 82.5% (Table 4).

3.2.4. Ultrasound

Measurements on ultrasound images have been compared with caliper measurements by two studies for concurrent validity testing and measurement error (Barbosa et al., 2013; Chiarello and McAuley, 2013), and have been used as reference standard for determining DRAM presence (see above, Calipers and Table 4). Mendes et al. (2007) compared measurement on ultrasound images with surgical compass measurements but reported statistics for validity testing inconsistent with this review's criteria.

Three studies (Liaw et al., 2011; Mota et al., 2012; Iwan et al., 2014) evaluated intra-rater reliability of ultrasound. Pooled ICCs were between 0.95 and 0.97 (ICC range 0.89–0.99) for different locations, and for rested and active measurement situations (Table 5; Figs. 2 and 3).

Test–retest reliability of ultrasound was evaluated by three studies (Liaw et al., 2006; Mota et al., 2012; Iwan et al., 2014) (Table 5). Pooling of reliability estimates resulted in pooled ICCs of 0.81–0.94 in resting situation and pooled ICCs of 0.68–0.86 for active (partial sit-up) measurement situation (Figs. 4 and 5). Reliability was found lower for measurements below the umbilicus (Mota et al., 2012; Iwan et al., 2014), in active situations (Mota et al., 2012; Iwan et al., 2014) and for novice sonographers (Iwan et al., 2014).

Liaw et al. (2006) and Iwan et al. (2014) evaluated inter-rater reliability and found ICCs of 0.65–0.97 between measurements of two raters (Table 5), Lower ICCs were found when comparing measurements below the umbilicus between an experienced and novice sonographer (ICC 0.65–0.83).

3.2.5. MRI and CT versus intra-operative measurements

Emanuelsson et al. (2014) compared intra-operative tape measurements with CT image measurements, and found a very low agreement (CCC of 0.00–0.16) with large average differences of 1.5–2.4 cm between methods (Table 5).

In two studies (Nahas et al., 2001; Elkhatib et al., 2011), MRI and CT were used as evaluative tools of DRAM width. Both studies presented raw data from MRI (Elkhatib et al., 2011) or CT (Nahas et al., 2001) and intra-operative measurements, which allowed calculations for criterion/concurrent validity testing. Raw data from a very small sample size ($n = 10$) (Elkhatib et al., 2011) indicated that MRI and intra-operative findings were strongly associated (Table 4). Bland and Altman-plots (Online Fig. 1) showed a significant systematic difference between measurement methods of 0.58 cm. Measurements on CT scan images (sample size $n = 20$) (Nahas et al., 2001) correlated well ($r = 0.78–0.88$) with intra-operative findings and there were no significant differences (Online Fig. 2).

4. Discussion

4.1. Findings

This systematic review presents measurement information of clinical methods used to assess DRAM, and can help clinicians and researchers determine the most appropriate method for their specific measurement purpose and situation. Ultrasound has been most widely researched with regards to its reliability, and showed a reliable method when images were taken by experienced sonographers. Calipers also seem a reliable method to measure DRAM width (measurements above the umbilicus). The clinically widely used 'finger width'-method has been under-evaluated. Some reliability data might indicate that the method is sufficient for retesting (one rater) and potentially for screening of DRAM presence. There are no measurement data available to support MRI or CT scan measurements. Available measurement information showed that different DRAM measurement methods correlated well with each other, but no data are available for any method on longitudinal validity or responsiveness.

4.2. Implications for clinical practice

4.2.1. Screening and diagnosis of clinically important DRAM

In clinical practice, measurement of DRAM width is often performed to screen for presence of clinically important DRAM. Determining whether a widening is over or under an accepted

Please cite this article in press as: van de Water ATM, Benjamin DR, Measurement methods to assess diastasis of the rectus abdominis muscle (DRAM): A systematic review of their measurement properties and meta-analytic reliability generalisation, Manual Therapy (2015), http://dx.doi.org/10.1016/j.math.2015.09.013

cut-off value (Rath et al., 1996; Coldron et al., 2008; Liaw et al., 2011), with consistency between raters in ranking and categorising the patients, is important. Ultrasound and calipers are satisfactory for this purpose. This has been supported by the results of Barbosa et al. (2013) and further calculations of diagnostic accuracy values based on their results (van de Water and Benjamin, 2014). Although supporting information on measurement properties and diagnostic accuracy is required, the 'finger width'-method might also have potential as a screening method for presence of clinically important DRAM. Results from Mota et al. (2013) supported that the 'finger width'-method can be reliable (retesting by one rater). Claims that the 'finger width'-method would be unreliable (Mota et al., 2012) have been based on the findings by Bursch (1987) only. However, they used the 'finger width'-method inappropriately: individual 'number of finger widths' between raters was compared and their intended measurement purpose was to evaluate DRAM width rather than to screen for DRAM presence. In addition, they employed statistics inconsistent with the review's criteria to evaluate inter-rater reliability. This highlights the importance of considering the measurement purpose before selecting measurement methods.

4.2.2. Monitoring DRAM width

Reliable and responsive methods are required to monitor DRAM width over time. Ultrasound was found to be reliable in healthy and post-natal women, regardless of the measurement location or active or resting situation. However, data on longitudinal validity or responsiveness is still lacking. Similarly, calipers have evidence of being a reliable method to measure DRAM in post-natal women in active or resting situations and at different measurement locations, in particular at and above the umbilicus. This finding is strengthened by two pilot-test reliability studies (Hsia and Jones, 2000; Chiarello et al., 2005) that reported on reliability using calipers. Chiarello et al. (2005) used a nylon digital caliper in a blinded manner in pregnant women, and reported test–retest reliability ICCs of 0.995 and 0.997, and an inter-rater reliability ICC of 0.87. Hsia and Jones (2000) evaluated the test–retest reliability of dial calipers in 9 ante- and post-natal women reporting ICCs of 0.99 for active and resting situations.

4.2.3. Clinical feasibility of methods

Clinical feasibility of a tool is important when choosing a measurement method. For clinical out- and inpatient services and many research purposes, CT and MRI scans are not feasible methods to measure DRAM width. Also, we found insufficient evidence that they can be considered 'gold standard' as often claimed (Mendes et al., 2007; Mota et al., 2012; Barbosa et al., 2013). Moreover, Emanuelsson et al. (2014), Nahas et al. (2001) and Elkhatib et al. (2011) compared measurements on CT and MRI images to intra-operative measurements, of which the latter they considered their 'reference standard'.

Although recommended for use in clinical studies (Mota et al., 2012) and based on reliability evidence in this review, the use of ultrasound as a measurement tool for DRAM width might be limited in daily practice. Costs and availability of the device, required training to obtain clear and accurate images, reading of images and on-screen measurement of DRAM are some factors influencing its feasibility.

Palpatory methods (calipers, tape measures and 'finger width') are clinically feasible methods to determine DRAM (Keeler et al., 2012). These are inexpensive, easy to use and results are simple and quick to record. However, standardisation of these methods might be a clinical challenge, and therefore require a standardisation protocol. For example, knowing how broad one's grouped fingers are in cm/mm for the 'finger with'-method, can help with screening for DRAM presence and increase consistency when ranking or categorising patients with multiple raters.

4.3. Strengths

All the current available evidence examining measurement of DRAM and its measurement properties were included in this systematic review. This review followed the PRISMA guidelines on high quality reporting of systematic reviews (Moher et al., 2009) and reliability generalisation studies (Streiner and Norman, 2008). Its comprehensive search strategies without limits, methodological quality assessment of included studies with a modified COSMIN checklist (Mokkink et al., 2010a, 2010b) recommended for systematic reviews of psychometric property studies (Mokkink et al., 2010b), specific data analysis (Streiner and Norman, 2008) and reporting (Mokkink et al., 2010a, 2010b) highlight several strengths.

4.4. Limitations

Although the findings of this review are clinically relevant, caution is required when interpreting and applying the measurement methods due the design and methodological quality of

First author, year - study data	ICC	95% CI
Iwan, 2014 - LRUS, novice	0.89	(0.78-0.95)
Iwan, 2014 - LRUS, experienced	0.96	(0.92-0.98)
Iwan, 2014 - HRUS, novice	0.98	(0.96-0.99)
Iwan, 2014 - HRUS, experienced	0.98	(0.96-0.99)
Mota, 2012	0.96	(0.90-0.98)
Liaw, 2011 - 6 months	0.95	(0.91-0.97)
Liaw, 2011 - 7 weeks	0.97	(0.95-0.99)
Pooled	0.97	

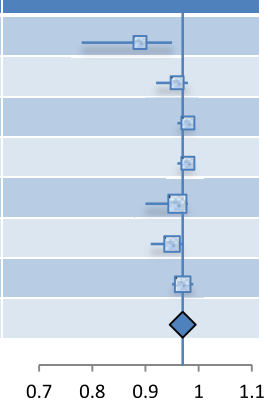

Fig. 2. Forest plot ultrasound, intra-rater reliability below umbilicus, rested situation LRUS/HRUS, low and high resolution ultrasound. The vertical line represents the value of the pooled estimate.

Please cite this article in press as: van de Water ATM, Benjamin DR, Measurement methods to assess diastasis of the rectus abdominis muscle (DRAM): A systematic review of their measurement properties and meta-analytic reliability generalisation, Manual Therapy (2015), http://dx.doi.org/10.1016/j.math.2015.09.013

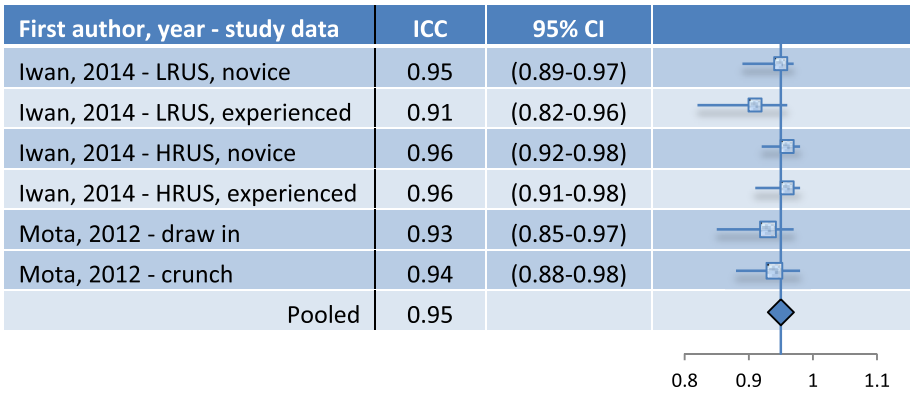

Fig. 3. Forest plot ultrasound, intra-rater reliability above umbilicus, active situation LRUS/HRUS, low and high resolution ultrasound. The vertical line represents the value of the pooled estimate.

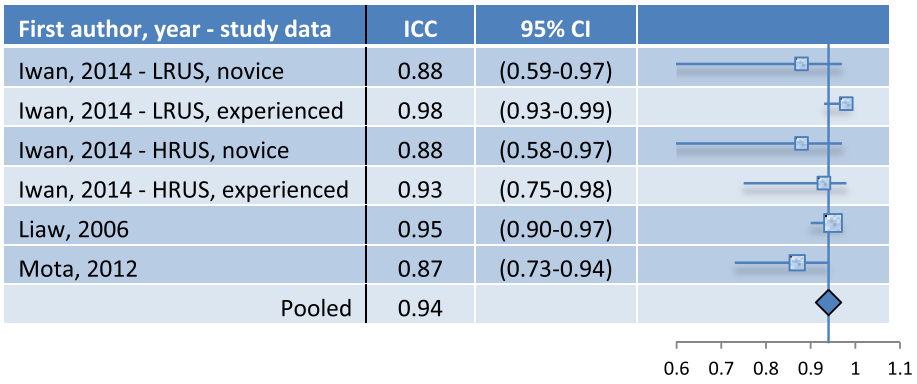

Fig. 4. Forest plot ultrasound, test–retest reliability above umbilicus, rested situation LRUS/HRUS, low and high resolution ultrasound. The vertical line represents the value of the pooled estimate.

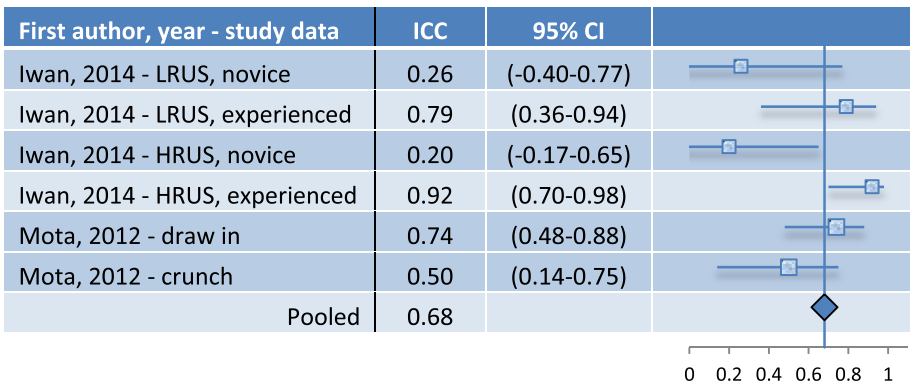

Fig. 5. Forest plot ultrasound, test–retest reliability below umbilicus, active situation LRUS/HRUS, low and high resolution ultrasound. The vertical line represents the value of the pooled estimate.

the included studies. These studies had a prospective clinical observational design with primarily small to moderate sample sizes and varying methodological quality. Few of these studies were sparse in data (Bursch, 1987; Nahas et al., 2001; Elkhatib et al., 2011) and some had important design flaws such as lack of blinding for reliability studies (Liaw et al., 2011; Barbosa et al., 2013; Mota et al., 2013), and employed statistics inconsistent with this review's criteria (Bursch, 1987; Nahas et al., 2001; Mendes et al., 2007; Elkhatib et al., 2011; Mota et al., 2012). Additionally, none of the studies presented data on longitudinal validity or responsiveness.

4.5. Directions for future research

Given the limitations of the current evidence on the psychometric properties of DRAM measurement methods, future research needs to ensure adequately powered high quality prospective studies are undertaken which are evaluated using appropriate statistical methods. These studies should be focussed on examining forms of reliability (test–retest, inter- and intra-rater reliability) and responsiveness of clinically feasible 'standardised palpatory measurement methods' such as calipers, tape-measures and standardised finger width. Also criterion validity of these palpatory

Please cite this article in press as: van de Water ATM, Benjamin DR, Measurement methods to assess diastasis of the rectus abdominis muscle (DRAM): A systematic review of their measurement properties and meta-analytic reliability generalisation, Manual Therapy (2015), http://dx.doi.org/10.1016/j.math.2015.09.013

methods compared to ultrasound as the reference standard should be investigated. This future psychometric research could help determine the suitability of these methods related to their measurement purpose (screening or measurement).

5. Conclusion

Ultrasound has been found to be a reliable method across a range of measurement situations, but might have clinical feasibility issues. Calipers are clinically feasible and available measurement information supports its use for clinical practice. Although supportive evidence on longitudinal validity is advised, it may also have potential as a measurement method to monitor DRAM width in evaluative research. Despite its wide use in daily practice, the 'finger width'-method has been under-evaluated with regards to measurement properties, but may be a valuable method for screening women for DRAM presence. Additional high quality measurement studies are warranted to evaluate longitudinal responsiveness of these methods, substantiate further the potential of calipers to adequately screen for presence of DRAM and monitor DRAM width (compared to ultrasound), and confirm the potential of the 'finger width'-method to screen for presence of clinically important DRAM.

Conflict of interest

No author received or will receive any financial benefits from this work. The authors have no conflict of interest to declare.

Acknowledgments

The authors would like to thank Scarlett Wu, PT for her help in translating the 2006 Chinese language article by Liaw et al. (2006) into the English language, and Associate Professor Megan Davidson, PT, PhD for her critical comments on an earlier version of the manuscript.

Appendix A. Example search strategy for PubMed

#1	(diastasis OR separation OR gapping OR widening OR divarication)
#2	(recti OR rectus) AND (abdominis OR abdominus)
#3	("abdominal muscles" [mesh] OR abdominals OR (abdominal AND muscles))
#4	(#2 OR #3)
#5	(#1 AND #4)
#6	((inter-recti OR between-recti) AND distance)
#7	(#5 OR #6)
#8	Sensitive search filter Part 1 by Terwee et al. (2009)
#9	Sensitive search filter Part 2 by Terwee et al. (2009)
#10	#7 AND #8
#11	#10 NOT #9
	Sensitive filter by Terwee et al. (2009) (partially displayed)
AND	(instrumentation[sh] OR methods[sh] OR Validation Studies[pt] OR Comparative Study[pt] OR "psychometrics"[MeSH] OR psychometr*[tiab] OR clinimetr*[tw] OR clinometr*[tw] OR "outcome assessment (health care)"[MeSH] OR outcome assessment[tiab] OR outcome measure*[tw] OR "observer variation"[MeSH] OR observer variation[tiab] OR "Health Status Indicators"[Mesh] OR "reproducibility of results"[MeSH] OR reproducib*[tiab] OR "discriminant analysis"[MeSH] OR reliab*[tiab] OR unreliab*[tiab] OR valid*[tiab] OR coefficient[tiab] OR homogeneity[tiab] OR homogeneous[tiab] OR "internal consistency"[tiab] OR (cronbach*[tiab] AND (alpha[tiab] OR alphas[tiab])) OR (item[tiab] AND (correlation*[tiab] OR selection*[tiab] OR reduction*[tiab])) OR agreement[tiab] OR precision[tiab] OR imprecision[tiab] OR "precise values"[tiab] OR test-retest[tiab] OR (test[tiab] AND retest[tiab]) OR (reliab*[tiab] AND (test[tiab] OR retest[tiab])) OR stability[tiab] OR interrater[tiab] OR inter-rater[tiab] (…)
NOT	("addresses"[Publication Type] OR "biography"[Publication Type] OR "case reports"[Publication Type] OR "comment"[Publication Type] OR "directory"[Publication Type] OR "editorial"[Publication Type] OR "festschrift"[Publication Type] OR "interview"[Publication Type] OR "lectures"[Publication Type] OR "legal cases"[Publication Type] OR "legislation"[Publication Type] OR "letter"[Publication Type] (…)

Appendix B. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.math.2015.09.013.

Ethical approval statement

No ethics approval was required for this study.

References

Barbosa S, de Sa RA, Coca Velarde LG. Diastasis of rectus abdominis in the immediate puerperium: correlation between imaging diagnosis and clinical examination. Arch Gynecol Obst 2013;288(2):299–303.
Benjamin DR, van de Water AT, Peiris CL. Effects of exercise on diastasis of the rectus abdominis muscle in the antenatal and postnatal periods: a systematic review. Physiotherapy 2014;100(1):1–8.
Boxer S, Jones S. Intra-rater reliability of rectus abdominis diastasis measurement using dial calipers. Aust J Physiother 1997;43:109–14.
Bursch G. Interrater reliability of diastasis recti abdominis measurement. Phys Ther 1987;67:1077–9.
Charter RA. Combining reliability coefficients: possible application to meta-analysis and reliability generalization. Psychol Rep 2003;93:643–7.
Chiarello CM, Falzone LA, McCaslin KE, Patel MN, Ulery KR. The effects of an exercise program on diastasis recti abdominis in pregnant women [corrected]. J Womens Health Phys Ther 2005;29(1):11–6.
Chiarello CM, McAuley JA. Concurrent validity of calipers and ultrasound imaging to measure interrecti distance. J Orthop Sports Phys Ther 2013;43:495–503.
Coldron Y, Stokes M, Newham D, Cook K. Postpartum characteristics of rectus abdominis on ultrasound imaging. Man Ther 2008;13:112–21.
De'Ath HD, Lovegrove RE, Javid M, Peter N, Magee TR, Galland RB. An assessment of between-recti distance and divarication in patients with and without abdominal aortic aneurysm. Ann R Coll Surg Engl 2010;92:591–4.
Elkhatib H, Buddhavarapu SR, Henna H, Kassem W. Abdominal musculoaponeuretic system: magnetic resonance imaging evaluation before and after vertical plication of rectus muscle diastasis in conjunction with lipoabdominoplasty. Plast Reconstr Surg 2011;128:733e–40e.
Emanuelsson P, Dahlstrand U, Stromsten U, Gunnarsson U, Strigard K, Stark B. Analysis of the abdominal musculo-aponeurotic anatomy in rectus diastasis: comparison of CT scanning and preoperative clinical assessment with direct measurement intraoperatively. Hernia 2014;18:465–71.
Fast A, Weiss L, Ducommun EJ, Medina E, Butler JG. Low-back pain in pregnancy. Abdominal muscles, sit-up performance, and back pain. Spine (Phila Pa 1976) 1990;15:28–30.
Gilleard W, Brown J. Structure and function of the abdominal muscles in primigravid subjects during pregnancy and the immediate postbirth period. Phys Ther 1996;76:750–62.
Hsia M, Jones S. Natural resolution of rectus abdominis diastasis. Two single case studies. Aust J Physiother 2000;46:301–7.

Please cite this article in press as: van de Water ATM, Benjamin DR, Measurement methods to assess diastasis of the rectus abdominis muscle (DRAM): A systematic review of their measurement properties and meta-analytic reliability generalisation, Manual Therapy (2015), http://dx.doi.org/10.1016/j.math.2015.09.013

Iwan T, Garton B, Ellis R. The reliability of measuring the inter-recti distance using high resolution and low-resolution ultrasound imaging comparing a novice to an experienced sonographer. N. Z J Physiother 2014;42:154–62.
Keeler J, Albrecht M, Eberhardt L, Horn L, Donnelly C, Lowe D. Diastasis recti abdominis: a survey of women's health specialists for current physical therapy clinical practice for postpartum women. J Womens Health Phys Ther 2012;36:131–42.
Khushboo D, Amrit K, Mahesh M. Correlation between diastasis rectus abdominis and lumbopelvic pain and dysfunction. Indian J Physiother Occup Ther 2014;8:210–4.
Kirshner B, Guyatt G. A methodological framework for assessing health indices. J Chronic Dis 1985;38:27–36.
Landis JR, Koch GG. The measurement of agreement for categorical data. Biometrics 1977;33:159–74.
Lee DG, Lee LJ, McLaughlin L. Stability, continence and breathing: the role of fascia following pregnancy and delivery. J Bodyw Mov Ther 2008;12:333–48.
Liaw LJ, Hsu MJ, Liao CF, Liu MF, Hsu AT. The relationships between inter-recti distance measured by ultrasound imaging and abdominal muscle function in postpartum women: a 6-month follow-up study. J Orthop Sports Phys Ther 2011;41:435–43.
Liaw LJ, Liu SM, Hsiao SF. The reliability of measuring of inter-recti distance using real-time ultrasonography [chinese]. Formos J Phys Ther 2006;31:213–8.
Mendes DDA, Nahas FX, Veiga DF, Mendes FV, Figueiras RG, Gomes HC, et al. Ultrasonography for measuring rectus abdominis muscles diastasis. Acta Cir Bras 2007;22(3):182–6.
Mesquita LA, Machado AV. Physiotherapy for reduction of diastasis of the recti adbominis muscles in the postpartum period [Portuguese]. Rev Bras Ginecol Obstet 1999;21:267–72.
Moesbergen T, Law A, Roake J, Lewis DR. Diastasis recti and abdominal aortic aneurysm. Vascular 2009;17:325–9.
Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009;6:e1000097.
Mokkink LB, Terwee CB, Knol DL, Stratford PW, Alonso J, Patrick DL, et al. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. BMC Med Res Methodol 2010a;10:22.
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 2010b;19:539–49.
Mota P, Pascoal AG, Sancho F, Bo K. Test-retest and intrarater reliability of 2-dimensional ultrasound measurements of distance between rectus abdominis in women. J Orthop Sports Phys Ther 2012;42:940–6.
Mota P, Pascoal AG, Sancho F, Carita AI, Bo K. Reliability of the inter-rectus distance measured by palpation. Comparison of palpation and ultrasound measurements. Man Ther 2013;18:294–8.
Nahas FX, Augusto SM, Ghelfond C. Nylon versus polydioxanone in the correction of rectus diastasis. Plast Reconstr Surg 2001;107:700–6.
Neyeloff JL, Fuchs SC, Moreira LB. Meta-analyses and forest plots using a microsoft excel spreadsheet: step-by-step guide focusing on descriptive data analysis. BMC Res Notes 2012;5:52.
Rath AM, Attali P, Dumas JL, Goldlust D, Zhang J, Chevrel JP. The abdominal linea alba: an anatomo-radiologic and biomechanical study. Surg Radiol Anat 1996;18:281–8.
Spitznagle TM, Leon FC, Dillen LR. Prevelence of daistasis recti abdominals in urogynecological population. Int Urogynecol J Pelvic Floor Dysfunct 2007;18:321–8.
Streiner DL, Norman GR. Health measurement scales – a practical guide to their development and use. 4th ed. Oxford: Oxford University Press; 2008.
Taranto I. The relief of low back pain with WARP adominoplasty: a preliminary report. Plast Reconstr Surg 1990;85:545–55.
Terwee CB, Jansma EP, Riphagen II , de Vet HC. Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. Qual Life Res 2009;18:1115–23.
van Bloemendaal M, van de Water ATM, van de Port IG. Walking tests for stroke survivors: a systematic review of their measurement properties. Disabil Rehabil 2012;34:2207–21.
van de Water ATM, Benjamin DR. Measure DRAM with a purpose: diagnose or evaluate. Arch Gynecol Obst 2014;289(1):3–4.
Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 2011;155:529–36.

Please cite this article in press as: van de Water ATM, Benjamin DR, Measurement methods to assess diastasis of the rectus abdominis muscle (DRAM): A systematic review of their measurement properties and meta-analytic reliability generalisation, Manual Therapy (2015), http://dx.doi.org/10.1016/j.math.2015.09.013