Review Article

Reliability of trunk strength measurements with an isokinetic dynamometer in non-specific low back pain patients: A systematic review

Waleska Reyes-Ferrada[a,b], Luis Chirosa-Rios[a], Darío Martinez-Garcia[b], Ángela Rodríguez-Perea[a] and Daniel Jerez-Mayorga[b,*]

[a]*Department Physical Education and Sports, Faculty of Sport Sciences, University of Granada, Granada, Spain*
[b]*Exercise and Rehabilitation Sciences Laboratory, School of Physical Therapy, Faculty of Rehabilitation Sciences, Universidad Andres Bello, Santiago, Chile*

Abstract.
BACKGROUND: Imbalance or decreased trunk strength has been associated with non-specific low back pain (NSLBP).
OBJECTIVE: This systematic review aimed (I) to evaluate the quality of evidence of studies evaluating the reliability of trunk strength assessment with an isokinetic dynamometer in NSLBP patients, (II) to examine the reliability of trunk strength assessment using an isokinetic dynamometer in NSLBP patients and (III) to determine the most reliable protocol for trunk strength assessment in NSLBP patients.
METHOD: PRISMA guidelines were followed. Three databases were used: PubMed, Scopus, and Web of Science with the following keywords: Isokinetic, Dynamometer, Trunk strength testing, Muscle testing, Isokinetic measurement, CORE, Abdominal muscles, Abdominal wall, Torso, Trunk, Spine, Reliability and, Reproducibility. We included only test-retest studies, focused on the reliability of isometric and isokinetic strength assessed with an isokinetic dynamometer in NSLBP adults' patients, published in English and from inception to March 30, 2021. The methodological quality was evaluated with the CAT scale and QAREL checklist.
RESULTS: Five hundred and seventy-seven articles were retrieved, of which five are included in this review. Three articles provide good quality of evidence, the reliability of trunk strength assessment in NSLBP patients is excellent, and the most reliable protocol for isometric assessment is in a seated position (ICC = 0.94–0.98) and for isokinetic strength in standing position, at 60°/s and 120°/s (ICC = 0.98).
CONCLUSION: There is good quality evidence regarding the trunk strength assessment's reliability. Reliability is excellent in NSLBP patients; however, a familiarization process should be considered to obtain clinically reliable data. The most reliable protocol is in a seated position for isometric strength and a standing position for isokinetic strength.

Keywords: Reproducibility, test re-test, torso, core muscles, trunk strength testing, back pain

1. Introduction

Low back pain (LBP) is the leading cause of decreased productivity worldwide and is one of the lead-

---

*Corresponding author: Daniel Jerez-Mayorga, PhD in Biomedicine, Fernández Concha 700, Las Condes, Santiago, Chile. E-mail: daniel.jerez@unab.cl.

ing causes of years lived with disability [1]. In addition; it has been associated with other musculoskeletal injuries, such a fragility fractures [2,3] and poor quality of life [4]. In 2017, LBP affected 577 million people worldwide [5]. LBP is defined as pain, muscle tension, or stiffness between the lower costal edge and the lower limit of the gluteal fold, with or without irradiation [4]. In addition, it can be characterized in terms of temporality as acute pain, less than six weeks, subacute, and chronic, when the pain extends beyond 12 weeks [6,7]. It has been estimated that LBP will affect 90% of the population at least once in their lives [8,9]. Of these acute episodes, most will recover within two weeks. However, about 70% will have recurrences, of which 40% will need to use health services [10], and it is expected that at least 5% of these patients with low back pain will develop chronic low back pain (cLBP) [11].

LBP is understood as multifactorial and involves several risk factors [12]. Thus, LBP is classified as specific when the anatomical structure can be identified, as in the presence of fractures, metastases, infections, etc. [13]. However, in 90% of the cases, it is impossible to find an anatomical cause, so it is called nonspecific low back pain (NSLBP) [13]. However, several risk factors can be attributed to the development of NSLBP, such as the altered neuromuscular response of the trunk [14,15], deconditioning of the lumbar musculature [16,17], the reduced muscle mass [18], imbalance, and reduced trunk flexors and extensors muscle strength [19,20].

Concerning trunk strength, there are records of its assessment since the 1940s [21]. Multiple evaluation systems have been developed to assess trunk strength [22–24], with isokinetic evaluation being the gold standard [25]. The measurement of trunk strength with an isokinetic dynamometer can be performed isometrically, at different angular positions, and isokinetically, i.e., at different angular velocities [26]. This type of assessment has proven valid for measuring trunk strength [27]. However, the assessments need to be reliable given the importance of trunk strength in health and performance. Reliability is defined as the consistency of measurements or the absence of measurement errors [28]. Reliability can be relative (intraclass correlation coefficient (ICC)) or absolute (standard error of measurement (SEM) or the coefficient of variation (CV)). Relative reliability indicates how similar the rank orders of the participants in the test are to the retest [29], whereas absolute reliability is related to the consistency of individual scores [30,31]. For this, reliable measurements are relevant in sports medicine and

research [31,32] to objectively reflect the increase or decrease in strength rather than the product of procedural or equipment error.

Recently, Estrázulas et al. [33] reviewed the protocols for isokinetic and isometric measurements using a dynamometer in healthy subjects, recommending a protocol in seated and standing positions to increase the reliability of these measurements. Unfortunately, the results are contradictory in subjects with LBP since Gruther et al. [34], when comparing the isometric and isokinetic trunk assessment in healthy subjects and those with LBP, reported low reliability and therefore did not recommend this type of assessment in LBP patients. However, Verbrugghe et al. [35] reported substantial reliability (ICC = 0.93–0.98; SEM 5.5%–9.3%) in isometric trunk assessment using an isokinetic dynamometer when comparing healthy subjects with LBP patients.

Thus, the reliability of isokinetic trunk strength assessment in healthy subjects is well established; however, given the characteristics of pain and muscle function in LBP patients, to the best of our knowledge, the reliability of the trunk strength assessment using an isokinetic dynamometer in this type of patient has not been proven. Nevertheless, it is important from a clinical and researchers' point of view since reliable measurements allow a better evaluation and monitoring of objective parameters, such as trunk strength, in these patients. Therefore, the aims of the present systematic review were: (I) to evaluate the quality of evidence of studies evaluating the reliability of trunk strength assessment with an isokinetic dynamometer in NSLBP patients, (II) to examine the reliability of trunk strength assessment using an isokinetic dynamometer in NSLBP patients and (III) to determine the most reliable protocol for trunk strength assessment using an isokinetic dynamometer in NSLBP patients.

2. Method

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines were used [36]. PRISMA was designed to help researchers transparently report why the review was done, what the authors did, and what they found (Supplementary Table S1). The protocol for this review was registered in PROSPERO (CRD42021247943).

2.1. Study search

The search was performed by two authors (WR-F and DJ-M). The databases used were PubMed, Scopus, and

Web of Science. The search was performed on March 30th, 2021, with no restriction on publication dates, i.e., from inception until March 2021. The following keywords were included: "Isokinetic", "Dynamometer", "Trunk strength testing", "Muscle testing", "Isokinetic measurement", "CORE", "abdominal muscles", "abdominal wall", "torso", "trunk", "Spine", "Reliability", "Reproducibility". We also manually searched the references of selective articles to identify additional potentially relevant studies. The search strategy is presented in Supplementary Table S2.

2.2. Eligibility criteria

Articles that met the following criteria were included in this review: (I) subjects > 18 years old, (II) subjects with NSLBP, (III) studies with repeated measures design assessing isokinetic trunk flexors and extensors strength, (IV) studies reporting measures of reliability: coefficient of variation (CV), intraclass correlation index (ICC), standard error (TE), standard error of measurement (SEM), minimum detectable change (MDC) or Pearson correlation (r), (V) full text available, and (VI) articles in English. In addition, we excluded all those articles that (I) only considered healthy subjects or subjects with specific LBP, (II) conference presentations, theses, books, editorials, review articles and expert opinions, (III) duplicate articles, and (IV) articles in which the principal or secondary authors did not respond to e-mail requests.

2.3. Study selection

The articles retrieved from the search were entered into the Rayyan QCRI application [37], an app that assists the article selection process, optimizing review time and allowing collaborative work among researchers (available for free at http://rayyan.qcri.org, accessed on 27 May 2021). Rayyan QCRI is very easy to learn how to use, with an intuitive and user-friendly interface [38], and has been previously used in systematic reviews [39].

Duplicate articles were eliminated, and two investigators (WR-F and DJ-M) independently reviewed titles and abstracts to identify articles that met the eligibility criteria. In case of discrepancies, a third investigator (LC-R) was consulted and resolved by consensus. Finally, the selected articles were read in total, and the reference list was reviewed for relevant articles that could be included.

2.4. Quality of evidence assessment

Two authors (WR-F and AR-P) independently assessed the quality of evidence of the articles included in this review; in case of discrepancies, a third assessor (LC-R) was consulted and resolved by consensus. The Critical Appraisal Tool (CAT) scale was used to assess the quality of the evidence of the studies included in this review [40] and the Quality Appraisal for Reliability Studies (QAREL) checklist [41]. The agreement rate be-tween the reviewers was calculated using kappa statistics.

The CAT is a scale developed to evaluate the methodological quality of studies that verify the validity and reliability of objective clinical tools [40] and contains 13 items categorized as "yes" if the information is described in sufficient detail, "no" when the information is not clear enough, or "not applicable." In addition, five items are related to validity and reliability, four to validity, and four to reliability only. For this reason, only nine items were considered in this review. Finally, the percentage of the evaluation was calculated ((Items "yes" $\times$ 100)/9), assuming a maximum of 100% (nine items), which was the highest methodological quality. Studies that scored over 45% were considered to be of high quality [42].

The quality appraisal tool for studies of diagnostic reliability (QAREL) checklist is an assessment tool for evaluating the quality of diagnostic reliability studies [41]. QAREL contains 11 items encompassing seven key domains (subjects, examiners, examiner blinding, order of assessment, time interval between repeated measurements, test application and interpretation, and statistical analysis). Each item is labeled as "yes", "no", or "unclear". In addition, some items include the option "not applicable." Quality was calculated ((Items "yes" $\times$ 100)/11), and the maximum value was 100%. Based on previous studies [42–44], a score higher or equal to 60% was considered high quality.

3. Results

No systematic reviews with a similar objective to the present study were found. From the initial search, 577 articles were found (Fig. 1), of which 201 articles were eliminated because they were duplicates. After evaluating titles and abstracts, 366 articles did not meet the inclusion criteria, leaving ten articles for full-text reading. Of the ten articles, one was not available because the authors could not be contacted. In addition, three arti-
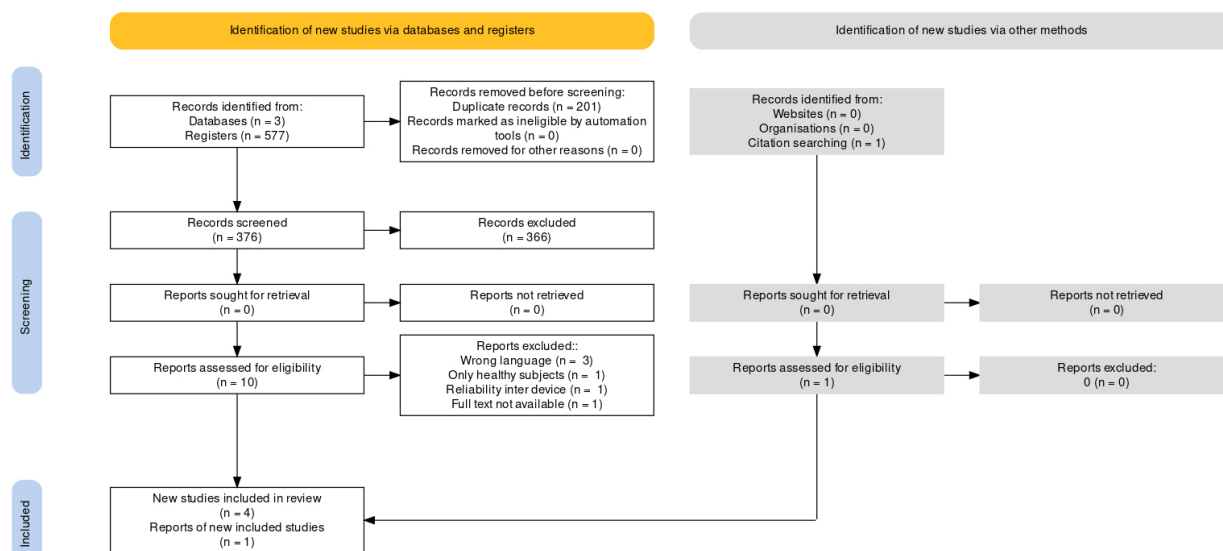
Fig. 1. PRISMA flowchart [63].

cles were in languages other than English (German and Turkish), one did not evaluate subjects with NSLBP, and another compared inter-rater reliability. One additional article was identified from other sources. Finally, five studies were included in this systematic review.

3.1. Characteristics of the articles included

The sample size among the studies ranged from 39 [35] to 66 [34] subjects, with ages ranging from 32 [45] to 45.1 [46] years and with a total of 141 patients with NSLBP. The initial test and retest were performed between two days [47] and three weeks [34].

Two of the five included articles were evaluated in a seated position [34,35], while three were assessed in a standing position [45–47]. Regarding the type of contraction, two evaluated the isometric flexors and extensors strength (seated and at 20°, 60°, and 100°) [34,35], four evaluated isokinetic strength at velocities of 90°/s [34], 60°/s, 90°/s y 120°/s [46,47]. Meanwhile, one study evaluated only the isokinetic extensors strength at 60°/s, 120°/s, and 150°/s [45]. All studies analyzed peak torque except Keller et al. [45], who considered total work (Nm) (Table 1).

3.2. Quality of evidence

In this review, 85 items (85%) were evaluated in an agreement between two investigators. 82.2% for the CAT scale and 87.2% for the QAREL checklist. The remaining 15% was decided by consensus. Considering

the total number of items evaluated, the kappa agreement rate between reviewers was 0.82.

The quality of evidence of the articles using the CAT scale varied between 44% and 67%, with a maximum of 100%. Three articles were classified as high quality (Table 2).

Concerning the QAREL checklist, the quality of the articles varied between 36% and 55%, with a maximum of 100%. None of the articles were classified as high quality (Table 3).

For the sample used, all the studies retrieved in this review describe it correctly and represent the population to be studied. As for the evaluators, four studies describe their qualifications, while Gruther et al. [34] only explain that it was a study assistant. Regarding the evaluation blinding, only Newton et al. [47] specify that the evaluator was blinded from the results of the clinical and psychometric assessments. However, it is not clear whether they were blinded from the results of their assessments, baseline values, extra clinical information, or other characteristics of the subjects under study. The remaining studies do not provide sufficient details about blinding. None of the studies varied the order of the assessments; however, all studies respected the theoretical stability of the evaluation to perform the retest. Newton et al. [47] and Gruther et al. [34] did not clearly specify the position, familiarization, and rest times between assessments regarding the protocol. However, all applied and interpreted the evaluation correctly. Regarding withdrawals during the test, all, except for Keller et al. [45] and Hupli et al. [46], explained

Table 1
Characteristics of the articles included

Author	Participants	Dynamometer type	Procedure	Motion/velocity	Position	Contraction	Measurement unit
Verbrugghe et al. [35]	HC: 19 (8M; 11F); 40.2 ± 11.9 yrs. LBP: 20 (10M; 10F); age: 44.0 ± 11.2 yrs.	Biodex system 3 pro, Enraf-Nonius, USA	Test-retest: 5–10 days	Flexion/extension Isometric	Seated: Semi flex and Isolate lumbar	Isometric: 3 reps/ 5 sec	Peak torque
Gruther et al. [34]	HC: 19 cLBP: 32 cHA: 15	Biodex 2000, NY, USA	Test-retest: 2–3 weeks (2.38 ± 0.9)	Flexion/extension Isometric at 20°, 60° and 100°. Isokinetic 90°/s ROM: 20–100°.	Seated	Isometric: 3 reps; Isokinetic: 4 reps	Isometric: Peak torque (Nm) Isokinetic: Peak power (W), work (J) and peak torque (Nm)
Keller et al. [45]	HC: 31 (7M; 24F); 32yrs. CLBP: 31 (7M; 24F); 36 yrs.	Cybex 6000 TEF Modular Component (Ronkonkoma, NY)	3 measurements, interval: 5–10 days	Isokinetic flexion/extension, 60°/s, 120°/s, and 150°/s; ROM: upright position to 80° forward flexion.	Standing	Isokinetic	Total work (Nm)
Hupli et al. [46]	HC: 22; 43.5 ± 61.1 yrs. Severe LBP: 18; 45.1 ± 8.4.yrs. Mild LBP: 20; 44.3 ± 7.1 yrs.	Lidoback® (Loredan Biomedical Inc., Davis,TX)	3 measurements, Interval: 1 week	Isokinetic flexion/extension, 60°/s, 90°/s, and 120°/s. ROM: 80° flexion and 5° extension.	Standing	Isokinetic, 5 reps at each velocity	Peak torque, average peak torque, coefficient of variation, total work, and peak torque to body weight ratio.
Newton et al. [47]	HC: 70 (35M; 35F); 37.9 ± 10.4 yrs. LBP REF1r: 94 (47M; 47F); 35.5 ± 9.9 yrs. REF3r: 26 (12M; 14F); 34.5 ± 8.7 yrs. Test-retest: HC: 21 LBP: 20	Cybex II Back Testing System	4 measurements: interval: 2–3 days	Isokinetic flexion/extension, 60°/s, 90°/s, and 120°/s. ROM: 0–60°.	Standing	Isokinetic, 4 reps at each velocity	Peak torque (Ft-lb), Ratio flexion/ extension, average point variance.

HC: healthy control; LBP: Low back pain; cLBP: chronic Low Back Pain; cHA: chronic headache; REF1r: primary referrals; REF3r: tertiary referrals M: males; F: females; ROM: Range of Motion; Rep: repetitions; Nm: Newton-meters; W: Watts; J: Joules.

Table 2
Evaluation of the quality of the studies with clinical evaluation tool (CAT)

Study	1	2	3	4	5	6	7	8	9	%
Verbrugghe et al. [35]	Yes	Yes	No	No	No	Yes	Yes	Yes	Yes	67
Gruther et al. [34]	Yes	No	NA	No	No	Yes	No	Yes	Yes	44
Keller et al. [45]	Yes	Yes	NA	No	No	Yes	Yes	No	Yes	56
Hupli et al. [46]	Yes	Yes	NA	No	No	Yes	Yes	No	No	44
Newton et al. [47]	Yes	Yes	Yes	No	No	Yes	No	Yes	Yes	67

%: (Items "yes" x 100)/9; 1. If human subjects were used, did the authors give a detailed description of the sample of subjects used to perform the isokinetic test on? 2. Did the author clarify the qualification, or competence of the rater(s) who performed the isokinetic test? 3. If interrater reliability was tested, were raters blinded to the finding of the other raters? 4. If intrarater reliability was tested, were raters blinded to their own prior findings of the test under evaluation? 5. Was the order of examination varied? 6. Was the stability (or theoretical stability) of the variable being measured taken into account when determining the suitability of the time interval between repeated measures? 7. Was the execution of the test described in sufficient detail to permit replication of the test? 8. Were withdrawals from the study explained? 9. Were the statistical methods appropriate for the purpose of the study? %: final percentage of reliability. NA: not applicable.

Table 3
Evaluation of the quality of the studies with Quality Appraisal of Reliability Studies (QAREL)

Study	1	2	3	4	5	6	7	8	9	10	11	%
Verbrugghe et al. [35]	Yes	Yes	UC	UC	UC	UC	UC	No	Yes	Yes	Yes	45
Gruther et al. [34]	Yes	UC	No	UC	UC	UC	UC	No	Yes	Yes	Yes	36
Keller et al. [45]	Yes	Yes	NA	UC	UC	UC	UC	No	Yes	Yes	Yes	45
Hupli et al. [46]	Yes	Yes	NA	UC	UC	UC	UC	No	Yes	Yes	No	36
Newton et al. [47]	Yes	Yes	Yes	UC	UC	UC	UC	No	Yes	Yes	Yes	55

%: (Items "yes" x 100)/11; Was the test evaluated in a sample of subjects who were representative of those to whom the authors intended the results to be applied? 2. Was the test performed by raters who were representative of those to whom the authors intended the results to be applied? 3. Were raters blinded to the findings of other raters during the study? 4. Were raters blinded to their own prior findings of the test under evaluation? 5. Were raters blinded to the results of the reference standard for the target disorder (or variable) being evaluated? 6. Were raters blinded to clinical information that was not intended to be provided as part of the testing procedure or study design? 7. Were raters blinded to additional cues that were not part of the test? 8. Was the order of examination varied? 9. Was the time interval between repeated measurements compatible with the stability (or theoretical stability) of the variable being measured? 10. Was the test applied correctly and interpreted appropriately? 11. Were appropriate statistical measures of agreement used? Yes; No; UC: unclear; NA: not applicable.

the dropouts. All studies used relative reliability (ICC), except Hupli et al. [46], which only used the *t*-test and Pearson's correlation.

3.3. Reliability

The evaluation's reliability was estimated in all the articles with the ICC, except in the study by Hupli et al. [46] in which the percentage of change was used. In this review, to classify relative reliability, we used the criteria proposed by Koo et al. [48] for the ICC: < 0.50, poor; between 0.50 and 0.75, moderate, between 0.75 and 0.90 good; above 0.90, excellent. Overall, the ICC values ranged between 0.81 and 0.98 for patients with NSLBP and 0.91 to 0.98 for healthy subjects. For isometric strength assessment, the ICC values ranged between 0.81–0.98 for patients with NSLBP and 0.94–0.97 for healthy subjects. In the isokinetic assessment,

the ICC varied between 0.95–0.98 and 0.91–0.98, respectively (Table 4).

Only Verbrugghe et al. [35] provide absolute reliability values through the standard error of measurement (SEM) and Keller et al. [45] through the coefficient of variation (CV).

Regarding the most reliable protocol for evaluating LBP patients, for the isometric testing, the highest reliability was reported by Verbrugghe et al. [35] evaluating in a seated functional position (semi-flex), three series of five seconds, with excellent reliability values for both flexion (ICC = 0.98; SEM = 6.0%) and extension (ICC = 0.94; SEM = 8.9%). For the isokinetic evaluation, considering peak torque, the most reliable protocol for trunk flexors was that reported by Newton et al. [47] in standing position, knees 15° of semi-flexion and the axis of rotation adjusted at L5-S1, range of motion (ROM) of 60°, concentric mode and velocities of 60°/s, 90°/s, and 120°/s with an ICC of 0.98. For trunk

Table 4
Reliability of trunk flexion and extension strength of the studies

Author	Evaluation parameters	Mean (Nm)/(SD) 1 test	Mean (Nm)/(SD) 2 test	ICC (CI 95%)	Other measures
Verbrugghe et al. [35]	Semiflex – isometric ext (H)	238.3 (89.3)	247.8 (92.8)	0.94 (0.85–0.98)	SEM 8.9% MDC 61.3
	Isolated lumbar isometric ext (H)	232.2 (86.7)	239 (87.7)	0.97 (0.93–0.99)	SEM 6.4% MDC 40.9
	Semiflex – isometric flex (H)	150.6 (55.5)	152.9 (55.9)	0.97 (0.91–0.99)	SEM 5.8% MDC 29.8
	Isolated lumbar isometric flex (H)	104.2 (35.4)	106.3 (34.1)	0.94 (0.84–0.98)	SEM 8.2% MDC 23.4
	Semiflex – isometric ext (LBP)	269.5 (95.6)	268.3 (93.1)	0.94 (0.86–0.98)	SEM 8.9% MDC 66.9
	Isolated lumbar isometric ext (LBP)	244 (83.3)	249.2 (87.3)	0.93 (0.83–0.97)	SEM 9.3% MDC 62.7
	Semiflex – isometric flex (LBP)	155.2 (58.3)	155.4 (55.4)	0.98 (0.95–0.99)	SEM 6.0% MDC 25.0
	Isolated lumbar isometric flex (LBP)	108.2 (32.1)	108.4 (34.5)	0.97 (0.92–0.99)	SEM 5.5% MDC 16.4
Gruther et al. [34]	Isometric ext 20° (LBP)	160.2 (58.0)	161.6 (68.0)	0.85	$p = 0.856$
	Isometric ext 60° (LBP)	171.7 (53.6)	128.2 (63.2)	0.85	$p = 0.136$
	Isometric ext 100° (LBP)	163.9 (62.8)	179.1 (73.7)	0.81	$p = 0.098$
	Isometric flex 20° (LBP)	28.1 (31.2)	35.2 (33.0)	–	$p = 0.036*$
	Isometric flex 60° (LBP)	69.6 (37.9)	81.8 (44.2)	–	$p = 0.019*$
	Isometric flex 100° (LBP)	77.5 (35.5)	95.2 (35.7)	–	$p = 0.001*$
	Concentric ext 90°/s (LBP)	87.7 (74.3)	117.9 (87.9)	–	$p = 0.006*$
	Concentric flex 90°/s (LBP)	50.5 (35.5)	63.8 (38.1)	–	$p = 0.008*$
Keller et al. [45]	Concentric ext 60°/s (H)	140 (118–158)[#]	137 (111–165)[#]	0.96	CV 10% CD 27%
	Concentric ext 120°/s (H)	1847 (1525–2421)[#]	1911 (1405–2284)[#]	0.98	CV 8% CD 21%
	Concentric ext 150°/s (H)	105 (76–143)[#]	112 (81–142)[#]	0.96	CV 14% CD 39%
	Concentric ext 60°/s (LBP)	162 (124–197)[#]	151 (126–210)[#]	0.98	CV 10% CD 28%
	Concentric ext 120°/s (LBP)	2061 (1421–2510)[#]	1971 (1444–2637)[#]	0.97	CV 14% CD 38%
	Concentric ext 150°/s (LBP)	100 (59–121)[#]	90 (49–137)[#]	0.95	CV 23% CD 63%
Hupli et al. [46]	Concentric ext 60°/s- 90°/s & 120°/s (H)	–	–	–	–
	Concentric flex 60°/s- 90°/s & 120°/s (H)	–	–	–	–
	Concentric ext 60°/s- 90°/s & 120°/s (LBP)	–	–	–	–
	Concentric flex 60°/s- 90°/s & 120°/s (LBP)	–	–	–	–
Newton et al. [47]	Concentric ext 60°/s (H)	142.9 (–)[‡]	148.5 (–[‡])	0.98	–
	Concentric ext 90°/s (H)	132.4 (–)[‡]	136.0 (–)[‡]	0.94	–
	Concentric ext 120°/s (H)	113.4 (–)[‡]	119.6 (–)[‡]	0.91	–
	Concentric flex 60°/s (H)	115.8 (–)[‡]	115.8 (–)[‡]	0.94	–
	Concentric flex 90°/s (H)	111.9 (–)[‡]	113.3 (–)[‡]	0.96	–
	Concentric flex 120°/s (H)	103.4 (–)[‡]	107.3 (–)[‡]	0.95	–
	Concentric ext 60°/s (LBP)	122.0 (–)[‡]	123.8 (–)[‡]	0.98	–
	Concentric ext 90°/s (LBP)	100.9 (–)[‡]	106.5 (–)[‡]	0.97	–
	Concentric ext 120°/s (LBP)	81.7 (–)[‡]	84.7 (–)[‡]	0.98	–
	Concentric flex 60°/s (LBP)	114.1 (–)[‡]	113.6 (–)[‡]	0.98	–
	Concentric flex 90°/s (LBP)	104.0 (–)[‡]	106.0 (–)[‡]	0.98	–
	Concentric flex 120°/s (LBP)	88.7 (–)[‡]	100.5 (–)[‡]	0.98	–

Nm: Newton meter peak torque; Ext: extension; Flex: flexion; H: healthy subjects; LBP: low back pain patients; SEM: standard error of measurement; MDC: minimal detectable change; CV: Coefficient of variation; CD: Critical Difference; *paired t-test $p < 0.05$; [‡]: ft-lb; [#]Median and (Quartiles) Nm.

extensors, the most reliable protocol considering peak torque was reported by Newton et al. [47] in concentric mode but at velocities of 60°s and 120°/s (ICC = 0.98).

However, considering the total work, the most reliable protocol was that reported by Keller et al. [45] in standing position, with the pelvis fixed by an adhesive belt

below the iliac crest, from an upright position to 80° forward flexion and back to the upright position (ROM 80°), concentric mode at 60°/s with an ICC of 0.98 and a CV = 10%.

3.4. Adverse outcome from trunk isokinetic assessment

From the reviewed studies, none reported adverse effects during or after isokinetic strength assessment in LBP patients. In addition, the assessment did not increase pain even in the group of patients with severe LBP [46]. Only one healthy subject had to drop out of the evaluation for an episode of acute LBP at the initial isometric evaluation [35].

4. Discussion

The present review aimed to (I) assess the quality of evidence from studies evaluating the reliability of trunk strength assessment using an isokinetic dynamometer in NSLBP patients, (II) examine the reliability of trunk strength assessment using an isokinetic dynamometer in NSLBP patients, and (III) determine the most reliable protocol in trunk strength assessment in NSLBP patients. The main findings of this review indicate that (I) there is good quality evidence from studies regarding the reliability of trunk strength assessment in patients with NSLBP, (II) the reliability of isometric and isokinetic assessment of trunk flexor and extensor strength in patients with NSLBP using an isokinetic dynamometer is excellent and (III) the most reliable protocol for isometric assessment is in functional seated (semi-flex) position, while for isokinetic assessment of flexors and extensors is in standing position with velocities of 60°/s and 120°/s and ROM of 60°.

Concerning the quality of the evidence, three of the five articles retrieved presented good quality evidence when the CAT scale was used; however, when the QAREL checklist was used, none of the articles included were classified as high quality. This difference could be explained by the fact that, although both scales complement each other in the reliability assessment for objective evaluations [40,41], the QAREL checklist has 36% of its items (four) corresponding to the process of blinding. In contrast, the CAT scale only considers one item according to whether intra- or inter-rater reliability was tested. In the case of this review, all the studies, except for Newton et al. [47], did not report information regarding whether or not a blinding process was per-

formed. Hence, they were classified as "unclear," and the QAREL checklist assessment score decreased.

Regarding the isometric assessment reliability using an isokinetic dynamometer in NSLBP patients, the evidence shows that this type of measurement has excellent reliability for flexors (ICC = 0.98 (0.95–0.99)) and good to excellent for extensors (ICC = 0.94 (0.86–0.98)) using the ICC 3.1 and the 95% confidence interval (95% CI) [35]. In addition, when the agreement was evaluated, SEM values of less than 10% were obtained. The protocol used by the authors can explain this high reliability [35], which consisted of a comprehensive trunk-specific warm-up to familiarize the subjects with the procedure, followed by an education period on the correct execution of the test. Grabiner et al. [49] reported variations between 17% and 26.5% in the retest of subjects with a history of LBP compared to healthy subjects in the strength evaluation, suggesting that clinicians and researchers should provide a substantial familiarization session when evaluating LBP patients to obtain clinically reliable data. Another reason for excellent reliability could be because when the assessment is conducted at zero velocity and with no change in ROM, there is less possibility of misalignment of the axis of motion or changes in the position and fixation of the subjects; this allows for minor variation between the test and retest. Having reliable protocols for measuring trunk isometric strength is essential for monitoring interventions in LBP patients but also for detecting individuals at risk for LBP since the incidence and severity of LBP is related to isometric and isokinetic weakness of trunk muscles [19].

Considering high methodological quality studies, the reliability of the isokinetic assessment of trunk flexors and extensors was also excellent considering the ICC. However, both Newton et al. [47] and Keller et al. [45] do not specified the 95% CI. If we thought the data reported by Keller et al. [45], who only measured extensor strength as total work (Nm), the most reliable condition was the concentric mode at 60°/s with an ICC 0.98 and a CV = 10%. However, it is crucial to consider that Keller et al. [45] report the reliability of measurements two and three; since they found statistically significant differences in the strength between the first and second measurements. Thus, they did not evaluate the reliability and considered it as a learning effect. Something similar occurs with Newton et al. [47], who assessed the reliability in a subsample of 20 patients, reporting a learning effect between evaluation one and two, so reliability was evaluated between measurement two and four, reporting excellent reliability for trunk

flexors and extensors at velocities of 60, 90 and 120°/s with ICC between 0.97 and 0.98. Gruther et al. [34], only performed two evaluations separated by two weeks and reports "a few trials for familiarization," finding a significant increase, between 45 and 160%, compared to baseline in flexors ($p = 0.008$) and extensors ($p = 0.006$) concentric isokinetic strength at 90°/s. Hupli et al. [46], whose does not specify a familiarization process in their protocol, reported a variation in strength between measurements closer to 15% in the mild LBP group and 43–50% in the severe LBP group. This highlights the importance of familiarization when assessing muscle strength to optimize strength production while decreasing the learning effect. When familiarization is not performed, we could underestimate the results. On the other hand, excessive familiarization could produce training effects or fatigue without considering the loss of time in the evaluation [50]. For trunk isokinetic strength, Roth et al. [51], evaluating young, healthy subjects, reported good reliability (ICC = 0.85–0.96) with acceptable CVs between day one to day four measurement. However, reliability was lower when comparing familiarization to day one testing, reaffirming familiarization's importance in decreasing basal variability, especially at high velocities. Urzica et al. [52] compared isokinetic strength measurements on days one, two, and 21 days of admission to a trunk strengthening treatment in LBP patients. They found significant differences between day one and day two measurements, possibly attributable to the learning effect, reaffirming the need for a familiarization process, especially when the isokinetic evaluation situates the patients in a condition unrelated to their natural movements.

Regarding the measurement position, when healthy subjects are testing, the most reliable protocol is in the standing position, at velocities of 60°/s and 90°/s with a ROM of 80° in concentric mode [33]. Therefore, according to this review, the most reliable protocol for LBP patients is standing at 60°/s and 120°/s and ROM of 60°. Concerning the variable analyzed, peak torque has been widely used in healthy subjects [27] and LBP patients [39]. In this review, all the articles except for Keller et al. [45] used peak torque to determine the measurement reliability. Regarding the type of contraction, it is essential to consider that none of the articles retrieved in this review evaluated the reliability of the trunk flexors and extensors eccentric strength evaluation. Eccentric contraction occurs when the external strength is greater than the muscular strength, thus plays a vital role in daily life activities and sports, decelerating the body during movement [53]. Specif-

ically, in the trunk, the spinal erectors are responsible for initiating the extension movement from the standing position, while the flexors must eccentrically control this movement [54,55]. In addition, the extensor group has a clear eccentric antigravitational function [56]. In healthy men, eccentric trunk strength reliability is good to excellent (ICC = 0.78–0.91) [57]. In patients with a giant ventral hernia, the reliability of eccentric flexor strength was excellent (ICC = 0.92–0.96) [58].

To our knowledge, the reliability of eccentric trunk strength in LBP patients has not been probed. Therefore, we can suggest that determining the reliability of these measurements is necessary to understand trunk dynamics in these patients. Finally, from a clinical point of view, it is essential to note that measuring trunk strength using an isokinetic dynamometer does not generate adverse effects or aggravation of pain in LBP patients. It should encourage clinicians and researchers to evaluate and monitor these patients. In addition, after reviewing the evidence, it is clear that the familiarization process is essential in LBP patients. For this reason, it would be interesting to determine the best familiarization program in terms of series and repetitions and to determine whether it should be performed on the same day or on different days. Given the criticisms regarding unnatural movements during isokinetic assessment with classical dynamometers [59], it is necessary to know the reliability of the new generations of isokinetic dynamometers [60,61], which have a more functional approach and could be a new assessment option in LBP patients.

This review is not exempt from limitations; we only use three databases and include only articles in English, which may have affected the number of articles retrieved. In addition, this review considered articles from two to 28 years old, which did not allow to characterize each study's sample correctly due to heterogeneity in the presentation of the data in each study. It could be explained by the fact that the standards of scientific publication have changed, and new guidelines have been developed [62]. Notwithstanding this, we can consider as a strength the fact that we reviewed all the available evidence, with no publication deadline until March 2021.

5. Conclusion

There is good quality evidence regarding the trunk strength assessment's reliability. Reliability is excellent in NSLBP patients; however, a familiarization process

should be considered to obtain clinically reliable data. The most reliable protocol is in a seated position for isometric strength and a standing position for isokinetic strength.

Clinical message:

– The reliability of trunk strength assessment using an isokinetic dynamometer is excellent in patients with low back pain.
– A familiarization process is necessary to obtain clinically reliable data.
– Trunk strength assessment using an isokinetic dynamometer does not produce adverse effects or aggravation of symptoms in patients with low back pain.
– Isometric strength should be measured in seated position, while isokinetic strength should be measured in standing position, at velocities of 60°/s and 120°/s.

Acknowledgments

Author contributions

All authors contributed to the development of the research design, concept, data acquisition, data analysis, and interpretation, prepared the manuscript, revised it critically and approved the final version.

Conflict of interest

The authors declare that there is no conflict of interest.

Funding

Supplementary data

The supplementary files are available to download from http://dx.doi.org/10.3233/BMR-210261.

References

[1]   GBD 2017 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet* 2018; 392: 1789-1858.

[2]   Scaturro D, Lauricella L, Tumminelli LG, et al. Is there a relationship between mild-moderate back pain and fragility fractures? Original investigation. *Acta Medica Mediterr* 2020; 36: 2149-2153.

[3]   Scaturro D, Asaro C, Lauricella L, et al. Combination of rehabilitative therapy with ultramicronized palmitoylethanolamide for chronic low back pain: An observational study. *Pain Ther* 2020; 9: 319-326.

[4]   Vlaeyen JWS, Maher CG, Wiech K, et al. Low back pain. *Nat Rev Dis Prim* 2018; 4: 52.

[5]   Wu A, March L, Zheng X, et al. Global low back pain prevalence and years lived with disability from 1990 to 2017: Estimates from the Global Burden of Disease Study 2017. *Ann Transl Med* 2020; 8: 299-299.

[6]   Heuch I, Foss IS. Acute low back usually resolves quickly but persistent low back pain often persists. *J Physiother* 2013; 59: 127.

[7]   Menezes Costa LDC, Maher CG, Hancock MJ, et al. The prognosis of acute and persistent low-back pain: A meta-analysis. *Can Med Assoc J* 2012; 184: E613-E624.

[8]   Cassidy JD, Carroll LJ, Côté P. The saskatchewan health and back pain survey. *Spine* 1998; 23: 1860-1866.

[9]   Freburger JK, Holmes GM, Agans RP, et al. The rising prevalence of chronic low back pain. *Arch Intern Med* 2009; 169: 251-258.

[10]   da Silva T, Mills K, Brown BT, et al. Recurrence of low back pain is common: A prospective inception cohort study. *J Physiother* 2019; 65: 159-165.

[11]   Downie AS, Hancock MJ, Rzewuska M, et al. Trajectories of acute low back pain: A latent class growth analysis. *Pain* 2016; 157: 225-234.

[12]   Knezevic NN, Candido KD, Vlaeyen JWS, et al. Low back pain. *Lancet*; 6736. Epub ahead of print June 2021. doi: 10.1016/S0140-6736(21)00733-9.

[13]   Maher C, Underwood M, Buchbinder R. Non-specific low back pain. *Lancet* 2017; 389: 736-747.

[14]   Cholewicki J, Greene H, Polzhofer G, et al. Neuromuscular function in athletes. *J Orthop Sport Phys Ther* 2002; 32: 568-75.

[15]   Radebold A, Cholewicki J, Panjabi MM, et al. Muscle response pattern to sudden trunk loading in healthy individuals and in patients with chronic low back pain. *Spine* 2000; 25: 947-954.

[16]   Catalá MM, Schroll A, Laube G, et al. Muscle strength and neuromuscular control in low-back pain: Elite athletes versus general population. *Front Neurosci* 2018; 12: 436.

[17]   Steele J, Bruce-Low S, Smith D. A reappraisal of the deconditioning hypothesis in low back pain: Review of evidence from a triumvirate of research methods on specific lumbar extensor deconditioning. *Curr Med Res Opin* 2014; 30: 865-911.

[18]   Hori Y, Hoshino M, Inage K, et al. ISSLS PRIZE IN CLINICAL SCIENCE 2019: Clinical importance of trunk muscle mass for low back pain, spinal balance, and quality of life – a multicenter cross-sectional study. *Eur Spine J* 2019; 28: 914-921.

[19]   Cho KH, Beom JW, Lee TS, et al. Trunk muscles strength as a risk factor for nonspecific low back pain: A pilot study. *Ann*

Rehabil Med 2014; 38: 234-240.

[20] Lee JH, Hoshino Y, Nakamura K, et al. Trunk muscle weakness as a risk factor for low back pain. A 5-year prospective study. *Spine* 1999; 24: 54-57.

[21] Mayer L, Greenberg BB. Measurements of the strength of trunk muscles. *JBJS* 1942; 24: 842-856.

[22] De Blaiser C, De Ridder R, Willems T, et al. Reliability and validity of trunk flexor and trunk extensor strength measurements using handheld dynamometry in a healthy athletic population. *Phys Ther Sport* 2018; 34: 180-186.

[23] Dvir Z, Keating J. Reproducibility and validity of a new test protocol for measuring isokinetic trunk extension strength. *Clin Biomech* 2001; 16: 627-630.

[24] Glenn JM, Galey M, Edwards A, et al. Validity and reliability of the abdominal test and evaluation systems tool (ABTEST) to accurately measure abdominal force. *J Sci Med Sport* 2015; 18: 457-462.

[25] Stark T, Walker B, Phillips JK, et al. Hand-held dynamometry correlation with the gold standard isokinetic dynamometry: A systematic review. *PM R* 2011; 3: 472-479.

[26] Kannus P. Isokinetic evaluation of muscular performance. *Int J Sports Med* 1994; 15: S11-S18.

[27] Mueller S, Stoll J, Mueller J, et al. Validity of isokinetic trunk measurements with respect to healthy adults, athletes and low back pain patients. *Isokinet Exerc Sci* 2012; 20: 255-266.

[28] Safrit MJ, Wood TM. *Measurement concepts in physical education and exercise science*. Human Kinetics Books Champaign, 1989.

[29] Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *J Strength Cond Res* 2005; 19: 231-240.

[30] Hopkins WG, Marshall SW, Batterham AM, et al. Progressive statistics for studies in sports medicine and exercise science. *Med Sci Sports Exerc* 2009; 41: 3-12.

[31] Hopkins WG. Mesaures of reliability in sports medicine and science. *Sport Med* 2000; 30: 1-15.

[32] Atkinson G, Nevill AM. Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sport Med* 1998; 26: 217-238.

[33] Estrázulas JA, Estrázulas JA, de Jesus K, et al. Evaluation isometric and isokinetic of trunk flexor and extensor muscles with isokinetic dynamometer: A systematic review. *Phys Ther Sport* 2020; 45: 93-102.

[34] Gruther W, Wick F, Paul B, et al. Diagnostic accuracy and reliability of muscle strength and endurance measurements in patients with chronic low back pain. *J Rehabil Med* 2009; 41: 613-619.

[35] Verbrugghe J, Agten A, Eijnde BO, et al. Reliability and agreement of isometric functional trunk and isolated lumbar strength assessment in healthy persons and persons with chronic non-specific low back pain. *Phys Ther Sport* 2019; 38: 1-7.

[36] Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ* 2021; 372: n71.

[37] Ouzzani M, Hammady H, Fedorowicz Z, et al. Rayyan – a web and mobile app for systematic reviews. *Syst Rev* 2016; 5: 210.

[38] Cleo G, Scott AM, Islam F, et al. Usability and acceptability of four systematic review automation software packages: A mixed method design. *Syst Rev* 2019; 8: 145.

[39] Reyes-Ferrada W, Chirosa-Rios L, Rodriguez-Perea A, et al. Isokinetic trunk strength in acute low back pain patients compared to healthy subjects: A systematic review. *Int J Environ Res Public Health* 2021; 18: 2576.

[40] Brink Y, Louw QA. Clinical instruments: Reliability and validity critical appraisal. *J Eval Clin Pract* 2012; 18: 1126-1132.

[41] Lucas NP, Macaskill P, Irwig L, et al. The development of a quality appraisal tool for studies of diagnostic reliability (QAREL). *J Clin Epidemiol* 2010; 63: 854-861.

[42] Muñoz-Bermejo L, Pérez-Gómez J, Manzano F, et al. Reliability of isokinetic knee strength measurements in children: A systematic review and meta-analysis. *PLoS One* 2019; 14: e0226274.

[43] Powden CJ, Hoch JM, Hoch MC. Reliability and minimal detectable change of the weight-bearing lunge test: A systematic review. *Man Ther* 2015; 20: 524-532.

[44] Adhia DB, Bussey MD, Ribeiro DC, et al. Validity and reliability of palpation-digitization for non-invasive kinematic measurement – A systematic review. *Man Ther* 2013; 18: 26-34.

[45] Keller A, Hellesnes J, Jens I. Brox A. Reliability of the Isokinetic Trunk Extensor Test, Biering-Sørensen Test, and Åstrand Bicycle Test. *Spine (Phila Pa 1976)* 2001; 26: 771-777.

[46] Hupli M, Hurri H, Luoto S, et al. Isokinetic performance capacity of trunk muscles. Part I: The effect of repetition on measurement of isokinetic performance capacity of trunk muscles among healthy controls and two different groups of low-back pain patients. *Scand J Rehabil Med* 1996; 28: 201-206.

[47] Newton M, Thow M, Somerville D, et al. Trunk strength testing with iso-machines. Part 2: Experimental evaluation of the Cybex II Back Testing System in normal subjects and patients with chronic low back pain. *Spine (Phila Pa 1976)* 1993; 18: 812-24.

[48] Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med* 2016; 15: 155-163.

[49] Grabiner MD, Jeziorowski JJ, Divekar AD. Isokinetic measurements of trunk extension and flexion performance collected with the biodex clinical data station. *J Orthop Sports Phys Ther* 1990; 11: 590-598.

[50] Chan JPY, Krisnan L, Yusof A, et al. Maximum isokinetic familiarization of the knee: Implication on bilateral assessment. *Hum Mov Sci* 2020; 71: 102629.

[51] Roth R, Donath L, Kurz E, et al. Absolute and relative reliability of isokinetic and isometric trunk strength testing using the IsoMed-2000 dynamometer. *Phys Ther Sport* 2017; 24: 26-31.

[52] Urzica I, Tiffreau V, Popielarz S, et al. Isokinetic trunk strength testing in chronic low back pain. The role of habituation and training to improve measures. *Ann Readapt Med Phys Rev Sci la Soc Fr Reeduc Fonct Readapt Med Phys* 2007; 50: 271-274.

[53] Shirado O, Ito T, Kaneda K, et al. Concentric and eccentric strength of trunk muscles: Influence of test postures on strength and characteristics of patients with chronic low-back pain. *Arch Phys Med Rehabil* 1995; 76: 604-611.

[54] Kalimo H, Rantanen J, Viljanen T, et al. Lumbar muscles: Structure and function. *Ann Med* 1989; 21: 353-359.

[55] Thorstensson A, Oddsson L, Carlson H. Motor control of voluntary trunk movements in standing. *Acta Physiol Scand* 1985; 125: 309-321.

[56] Floyd WF, Silver PHS. The function of the erectores spinae muscles in certain movements and postures in man*. *J Physiol* 1955; 129: 184-203.

[57] Dervisevic E, Hadzic V, Burger H. Reproducibility of trunk isokinetic strength findings in healthy individuals. *Isokinet Exerc Sci* 2007; 15: 99-109.

[58] Gunnarsson U, Johansson M, Strigård K. Assessment of abdominal muscle function using the Biodex System-4. Validity

and reliability in healthy volunteers and patients with giant ventral hernia. *Hernia* 2011; 15: 417-421.

[59]   Bouilland S, Loslever P, Lepoutre FX. Biomechanical comparison of isokinetic lifting and free lifting when applied to chronic low back pain rehabilitation. *Med Biol Eng Comput* 2002; 40: 183-192.

[60]   Dvir Z, Müller S. Multiple-joint isokinetic dynamometry: A critical review. *J Strength Cond Res* 2019; 00: 1-15.

[61]   Rodriguez-Perea Á Jerez-Mayorga D, García-Ramos A, et al. Reliability and concurrent validity of a functional electrome-

chanical dynamometer device for the assessment of movement velocity. *Proc Inst Mech Eng Part P J Sport Eng Technol* 2021; 1-6.

[62]   Schulz KF, Altman DG, Moher D. CONSORT 2010 Statement: Updated guidelines for reporting parallel group randomised trials. *BMJ* 2010; 340: 698-702.

[63]   Haddaway NR, LA M. PRISMA2020: R package and ShinyApp for producing PRISMA 2020 compliant flow diagrams (Version 0.0.1). *Zenodo*.