Reliability and validity of clinically accessible smartphone applications to measure joint range of motion: A systematic review

Keogh, Justin W L; Cox, Alistair; Anderson, Sarah; Liew, Bernard; Olsen, Alicia; Schram, Ben; Furness, James

Published in:
PLoS One

DOI:
10.1371/journal.pone.0215806
10.1371/journal.pone.0215806
10.1371/journal.pone.0215806
Published: 08/05/2019

Document Version:
Publisher's PDF, also known as Version of record

Licence:
CC BY

Link to publication in Bond University research repository.

Recommended citation(APA):
Keogh, J. W. L., Cox, A., Anderson, S., Liew, B., Olsen, A., Schram, B., & Furness, J. (2019). Reliability and validity of clinically accessible smartphone applications to measure joint range of motion: A systematic review. PLoS One, 14(5), [e0215806]. https://doi.org/10.1371/journal.pone.0215806, https://doi.org/10.1371/journal.pone.0215806

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

For more information, or if you believe that this document breaches copyright, please contact the Bond University research repository coordinator.
RESEARCH ARTICLE

Reliability and validity of clinically accessible smartphone applications to measure joint range of motion: A systematic review

Justin W. L. Keogh¹,²,³∗, Alistair Cox⁴, Sarah Anderson⁴, Bernard Liew⁵, Alicia Olsen¹, Ben Schram⁴,⁶ James Furness⁴,⁶

¹ Faculty of Health Sciences and Medicine, Bond University, Gold Coast, Queensland, Australia, ² Sports Performance Research Institute New Zealand (SPRINZ), AUT University, Auckland, New Zealand, ³ Cluster for Health improvement, Faculty of Science, Health, Education and Engineering, University of the Sunshine Coast, Sunshine Coast, Queensland, Australia, ⁴ Department of Physiotherapy, Faculty of Health Sciences and Medicine, Bond University, Gold Coast, QLD, Australia, ⁵ Centre of Precision Rehabilitation for Spinal Pain (CPR Spine), School of Sport, Exercise and Rehabilitation Sciences, University of Birmingham, Edgbaston, Birmingham, United Kingdom, ⁶ Water Based Research Unit, Bond Institute of Health and Sport, Bond University, Gold Coast, QLD, Australia

* jkeogh@bond.edu.au

Abstract

Measuring joint range of motion is an important skill for many allied health professionals. While the Universal Goniometer is the most commonly utilised clinical tool for measuring joint range of motion, the evolution of smartphone technology and applications (apps) provides the clinician with more measurement options. However, the reliability and validity of these smartphones and apps is still somewhat uncertain. The aim of this study was to systematically review the literature regarding the intra- and inter-rater reliability and validity of smartphones and apps to measure joint range of motion. Eligible studies were published in English peer-reviewed journals with full text available, involving the assessment of reliability and/or validity of a non-videographic smartphone app to measure joint range of motion. Eligible studies were published in participants >18 years old. An electronic search using PubMed, Medline via Ovid, EMBASE, CINAHL, and SPORTSDiscus was performed. The risk of bias was assessed using a standardised appraisal tool. Twenty-three of the eligible 25 studies exceeded the minimum 60% score to be classified as a low risk of bias, although 3 of the 13 criteria were not achieved in >50% of the studies. Most of the studies demonstrated adequate intra-rater or inter-rater reliability and/or validity for >50% of the range of motion tests across all joints assessed. However, this level of evidence appeared weaker for absolute (e.g. mean difference ± limit of agreement, minimal detectable change) than relative (e.g. intraclass correlation, correlation) measures; and for spinal rotation than spinal extension, flexion and lateral flexion. Our results provide clinicians with sufficient evidence to support the use of smartphones and apps in place of goniometers to measure joint motion. Future research should address some methodological limitations of the literature, especially including the inclusion of absolute and not just relative reliability and validity statistics.
Introduction

The measurement of joint range of motion (ROM) in static and dynamic, passive and active, human movements is an essential skill in the musculoskeletal assessments commonly performed by physiotherapists, as well as some strength and conditioning coaches, to examine joint function, detect joint asymmetry and evaluate treatment efficacy as an objective outcome measure [1]. In the present study, static ROM is defined as the range of a joint held motionless at either of its limit of movement. Dynamic ROM is the range a joint moved to and from the limits of movement. When a joint is moved passively by an assessor or external device, passive ROM is assessed. When a joint moves as a result of muscular contraction, active ROM is assessed. The universal goniometer has long been the preferred method of clinical ROM measurement (especially static ROM) due to its ease of use, low cost, and demonstrated reasonable levels of reliability and validity in numerous studies [2–4].

However, the universal goniometer is not without its drawbacks, even when assessing static joint ROM. When assessing static ROM such as the angle of hinge joints like the knee and elbow in adults, there may always be some degree of error due to the universal goniometer not typically being long enough to be aligned directly with the appropriate landmarks on both proximal and distal adjacent joints. Spinal rotation may also be difficult to measure with a universal goniometer due to the difficulty in palpating anatomical landmarks to use as a reference point [5–7]. It is perhaps no surprise then that reliability is reduced when measuring the spinal compared to upper and lower limb motion with a universal [6–9]. These potential issues highlighted for the use of the universal goniometer in assessing static joint ROM may be further exacerbated in inexperienced clinicians who have a relative inability to correctly locate anatomical landmarks; as well as the assessment of dynamic rather than static ROM [10, 11].

The development of smartphone technology and software applications (apps), coupled with the ubiquity of smartphone ownership, now allows smartphones to measure joint ROM. Like the universal goniometer, smartphones are similarly easy to use, relatively inexpensive, and highly accessible [12]. Their inbuilt sensors such as an accelerometer, gyroscope, and magnetometer provide the necessary equipment to allow the smartphone to measure angles and displacements [12]. With the use of apps that can be downloaded onto the smartphone, these measurements can be transformed into meaningful assessment data such as joint ROM. One possible advantage of smartphone apps is that their use may circumvent some of the difficulties of using the universal goniometer regarding landmark identification and alignment. Where smartphone apps can altogether overcome the aforementioned drawbacks of the universal goniometer may depend upon the technology used and the experience of the clinician with this alternative approach. The emergence of smartphone apps therefore presents the clinical practitioners with a new set of tools to incorporate into clinical practice, especially for some of the more difficult joint ROMs to quantify.

In order for clinicians to be willing to replace the universal goniometer (at least in some contexts) with smartphone apps as a tool to clinically assess ROM, the validity and reliability of smartphone apps must be comparable or better than the universal goniometer. In psychometric terminology, reliability deals with the consistency in angle and displacement measures produced by smartphone apps, when used by multiple assessors (inter-rater), and when the same assessor performs multiple measurements (intra-rater) [13]. Validity deals with the extent that the measurement obtained from one device, such as smartphone apps, correlates or matches the criterion laboratory devices such as 3-D motion capture or criterion clinical tools such as the universal goniometer [13].

On the topic of synthesizing the psychometric properties of smartphone apps, a number of systematic reviews have been conducted [14, 15]. However, the review of Milani et al. [14] is
considered to be outdated due to the relative explosion of research into human movement analysis apps and as such, only included 12 studies assessing joint angle measurements. Further, while the review of Rehan Youssef and Gumaa [15] was more recent and well conducted in most aspects, there were several methodological limitations. First, their literature search was completed in August 2016 (including 15 studies and one case study assessing joint ROM) [15]. Second, they utilised a non-validated risk of bias assessment tool that they personally developed [15]. Third, there was a relative lack of reporting of specific reliability and validity data for each of the multiple actions that can occur at some joints such as the spine (trunk) and shoulder joints [15]. The relative lack of reporting specific data for each joint action is a major issue for clinicians, as it is quite possible that a particular smartphone and app may have sufficient reliability and/or validity or measuring some actions at a particular joint in certain planes of motion (e.g. flexion and extension in the sagittal plane) but that more complicated actions such as rotation in the transverse plane may be less reliable and/or valid.

The purpose of this systematic review was to address some of the limitations of the previous review in this area so as to better assist the clinician identify which smartphone apps may show adequate inter-rater and intra-rater reliability as well as validity for the measurement of ROM at particular joints and actions in clinical practice. This state-of-the-art review will assist clinical practitioners in deciding the appropriateness and choice of smartphone apps for clinical ROM assessment.

Search methodology

Search strategy

The protocol for this systematic review has not been registered. A database search of PubMed, Medline via Ovid, EMBASE, CINAHL, and SPORTSDiscus was initially performed on 20th October 2017 by two independent reviewers. This search was repeated on 20th December 2018 to maximise the currency of the findings of this review. The search strategy is described in Appendix 1.

Inclusion and exclusion criteria. Studies retrieved from the search process were determined by two independent reviewers, with a third reviewer used to assist with consensus, were any discrepancies being initially reported by the first two independent reviewers. The eligibility of the studies to be included in this review was determined by the following criteria: published in peer-reviewed journals; measure human participants aged over 18 years old; used a smartphone app to measure joint ROM and assessed validity and/or reliability of these apps; published from 2007 as this was the year the iPhone was launched; published in English and have full text available. Case studies, abstracts only or grey literature were not included. Smartphone apps which required either image/video recordings and/or post data collection analyses to generate joint angles were excluded, as such an approach is unlikely to be used in clinical practice due to privacy concerns with the storage of video footage and the additional analysis time that would be required.

Quality assessment

The Critical Appraisal Tool (CAT), developed by Brink and Louw [16] was used to appraise the methodological quality of studies reporting a reliability and/or validity component. The included studies to be appraised were rated on a set of specific criteria involving 13 items that assessed a variety of methodological factors including subject and rater characteristics, rater blinding, testing order, criterion measure characteristics and statistical analyses performed [16]. Consistent with a recent study that has used the CAT [17], in order to satisfy Criteria 13 (Statistical Methods) the study had to report absolute reliability and/or validity statistics (e.g. SEM, MDC or MD±LOA) in addition to the more commonly reported relative reliability and/
or validity statistics (e.g. r or ICC). As not all included studies assessed validity, not all the CAT criteria were relevant to each study. In this case, each validity item was scored as not applicable (NA) and that criteria not included in the overall assessment of the particular study’s risk of bias. Consistent with the use of the CAT in previous studies, a threshold of $\geq 60\%$ was considered as high quality, and a quality of $< 60\%$ was rated as poor quality, consistent with previous systematic reviews [4, 17, 18].

The methodological quality of the studies identified by the search was assessed by two independent reviewers. Across all the 13 items of the CAT, there was an overall agreement of 86.2% between the raters when reviewing the methodological quality of the 37 articles included in this review, resulting in a Cohen’s unweighted Kappa statistic of 0.64, indicating good agreement between the two raters [19].

Data extraction

Data was obtained from studies that met the inclusion and exclusion criteria, which included: the CAT assessment, participants, application and smartphone device, joint movement assessed and position that the participant was in whilst being assessed. Where applicable, data was extracted for intra-rater and inter-rater reliability as well as validity. Both relative and absolute reliability and validity statistics were reported where available to provide an index of the correlation or rank order (relative measure) and change/difference in the mean (absolute measure) [20, 21]. Common measures of relative reliability and validity include the intra-class correlation coefficient (ICC), concordance correlation coefficient (CCC) and Pearson’s product moment correlation (r). Alternatively, common measures of absolute reliability and validity include the standard error of measurement (SEM), minimal detectable change (MDC), mean difference (MD) and limits of agreement (LoA).

Data analysis

A critical narrative approach was applied to synthesize and analyse the data. For each measure, the following criteria were used to judge the level of intra-rater and inter-rater reliability and validity. For relative measures, the following criteria were used: ICC: Poor = ICC $< 0.40$, Fair = ICC $0.40–0.59$, Good = ICC $0.60–0.74$, Excellent $\geq 0.75$ [22]; for r: negligible r $= 0–0.29$, low r $= 0.30–0.49$, moderate r $= 0.50–0.69$, high r $= 0.70–0.89$, very high r $= 0.90–1$ [23]; and for CCC: Poor CCC $< 0.90$, Moderate CCC $= 0.90–0.94$, Substantial CCC $= 0.95–0.99$, Almost perfect CCC $> 0.99$ [24]. For absolute measures of reliability and validity, the following criteria were used: SEM: Poor SEM $> 5˚$ and Good SEM $\leq 5˚$ [25]; for MDC: Poor MDC $> 5˚$ and Good MDC $\leq 5˚$ [25, 26] for LOA, a standard deviation threshold of $5˚$ [25–28] multiplied by 1.96 to derive the 95% LOA bandwidth: poor $> \pm 9.8˚$ and Good $< \pm 9.8˚$.

Results

Selection of studies

Fig 1 represents the article review process based on the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [29]. Our initial literature search identified 1066 studies, with the second literature search identified an additional 170 studies, leading to combined total of 1236 identified studies. Within the 1236 identified studies, 268 duplicates were removed prior to the title and abstract screening, with an additional five duplicates subsequently identified when screening the second literature search. The search strategy yielded 36 eligible studies, with one additional study identified through other sources for a total of 37 studies.

Study characteristics and methodology

A description of the broad methodology of each included study was depicted in Table 1. In order of the most common to least common joints assessed, were spine (trunk), knee,
shoulder, wrist, elbow, ankle and hip. As the trunk and shoulder allow movement in more directions than other joints, the studies assessing the trunk and shoulder typically looked at a greater number of joint movements across the multiple planes of movement. For example, the studies assessing trunk motion typically looked at trunk flexion/extension, lateral flexion and
| Reference | Participants | Movement assessed | Position | Application | Device | Criterion instrument |
|-----------|--------------|------------------|----------|-------------|--------|----------------------|
| Spine (trunk) | | | | | | |
| Bedekar et al. [38] | 30 healthy students (5 males and 25 females, age $21.5 \pm 1.5$ years) | Active, static lumbar flexion | Standing | Goniometer | iPod (model not stated) | Dual inclinometer |
| Furness et al. [39] | 30 healthy students (20 females, 10 males; age: $29.8 \pm 8.9$ years) | Active, static thoracic rotation | Seated | Compass app | IPhone 6S | Universal goniometer |
| Grondin et al. [40] | 22 healthy adults (15 male, 7 female, age: $29.9 \pm 5.4$ years) | Active, static cervical flexion-extension ROM | Sitting and supine | Clinometer | IPhone 5 |
| Jung et al. [41] | 17 male adults (age: $22.2 \pm 1.6$ years) | Active, static, pelvic rotation | Supine | Clinometer level and slope finder (Plaincode Software Solutions, Stephanskirchen, Germany) | Not reported | Vicon 3D Motion Analysis |
| Kolber & Hanney [42] | 30 healthy adults (12 male, 18 female, age: $25.6 \pm 2.1$ years) | Active, static thoracolumbo-pelvic flexion, isolated lumbar flexion, thoracolumbo-pelvic extension, right lateral flexion, left lateral flexion | Standing | iHandy | IPhone 4 | Bubble inclinometer |
| Pourahmadi et al. [43] | 30 healthy adults (15 male, 15 female, age: $27.9 \pm 6.3$ years) | Active, static lumbar flexion & extension | Standing | The TiltMeter | IPhone 5 | Gravity based inclinometer |
| Pourahmadi et al. [30] | 40 adults with non-specific neck pain (20 male, 20 female, age: $31.1 \pm 6.4$ years) | Active, static cervical flexion, extension, lateral flexion, rotation | Sitting | G-pro | IPhone 7 | Universal goniometer |
| Quek et al. [44] | 21 healthy adults (11 male, 10 female, age $31.0 \pm 9.1$ years) | Active, static cervical ROM (flexion, extension, right lateral flexion, left lateral flexion, right rotation, left rotation) | Sitting | Custom-built app designed by a co-author (RC) of this study using MIT App Inventor. | Samsung Galaxy S3 | Vicon 3D Motion Analysis |
| Stenneberg et al. [31] | Validity study 30 patients with neck pain (19 female, 11 male, age: $53.4 \pm 9.1$ years) Reliability study 26 patients with neck pain (19 female, 7 male, age: $45.2 \pm 15.3$ years) | Active, static cervical ROM (flexion-extension, lateral flexion, rotation) | Sitting | 3D range of motion | IPhone 4s | Polhemus 3D Motion Analysis |
| Tousignant-Laflamme et al. [45] | 28 healthy adults (9 male, 19 female, age $23 \pm 5$ years) | Active, static cervical ROM (flexion, extension, right lateral flexion, left lateral flexion, right rotation, left rotation.) | Sitting | Clinometer (frontal and sagittal planes) Compass (transverse plane) | IPhone (model not stated) | Cervical range of Motion Device (CROM) |
| Ullucci et al. [46] | 38 healthy adults (19 female, 19 male, age: $28 \pm 1.2$ years) | Passive cervical flexion and right and left rotation | Sitting | Clinometer (plaincode, Stephanskirchen, Germany) | IPhone (model not stated) & Android (model not stated) |
| Shoulder | | | | | | |
| Lim et al. [47] | 47 healthy adults (28 male, 19 female, age $24.9 \pm 3.5$ years) | Passive, static shoulder horizontal adduction | Supine vs. sidelying position | Goniometer Pro | IPhone 5 |

(Continued)
| Reference                          | Participants                                                                 | Movement assessed                                      | Position                      | Application                  | Device                        | Criterion instrument          |
|-----------------------------------|-------------------------------------------------------------------------------|--------------------------------------------------------|-------------------------------|-----------------------------|-------------------------------|--------------------------------|
| Mejia-Hernandez et al. [33]       | 75 patients (21 female, 54 male, age: 46 years (range, 24–94 years) with shoulder disorders) | Forward flexion, Total active abduction, Active glenohumeral Abduction, Passive glenohumeral Abduction, Active internal rotation, Passive internal rotation, Active external rotation, Passive external rotation | Seated and supine             | GetMyROM                    | IPhone 5s                     | Universal goniometer           |
| Mitchell et al. [48]              | 94 healthy adults (37 male, 57 female, age: 26.4 ± 7.6 years)                 | Active, static shoulder external rotation               | Supine                        | GetMyROM                    | IPhone 4                      | Standard goniometer            |
| Ramkumar et al. [49]              | 10 healthy adults (5 male, 5 female, age 27 years)                             | Active, static, flexion, abduction, internal and external rotation | Not reported                  | Built-in iPhone sensors     | IPhone (model not reported)   | Standard goniometer            |
| Shin et al. [27]                  | 41 patients with unilateral symptomatic shoulders (20 males, 21 females, age: 52.7 ± 17.5 years) | Active and passive, static shoulder ROM: forward flexion, abduction, external rotation, external rotation at 90° abduction, and internal rotation. | Standing                      | Clinometer Application, Clinometer-Level and Slope Finder (Plaincode Software Solutions) | Samsung Galaxy S | Standard goniometer |
| Werner et al. [28]                | 24 healthy adults (9 male, 15 female) 15 symptomatic adults (all undergone total shoulder replacement 6–12 weeks earlier) | Static abduction, forward flexion, external rotation with arm at side, external rotation with shoulder abducted to 90°, internal rotation with arm abducted at 90°, Passive/active ROM not reported | Standing and supine            | Smartphone clinometer (Plaincode Software Solutions) | IPhone (model not stated) | Standard goniometer |
| Behnoush et al. [50]              | 60 healthy adults (47 male, 13 female, age: 42.3 ± 11.4 years).                | Active, static elbow flexion, supination, and pronation. | Sitting                        | Bubble inclinometer          | HTC (model not stated)        | Universal goniometer           |
| Cruz & Morais [51]                | 41 healthy adults (21 male, 20 female, age: 31.3 ± 13.2 years)                 | Passive, static ULNT1 sequence                         | Supine                        | Compass                     | IPhone 4                      |                                |
| Vauclair et al. [52]              | 30 healthy adults (11 females, 9 males, age: 52 years (range 21–74))           | Active, static flexion extension, pronation, supination | Sitting                        | Clinometer                   | Not reported                  | Standard goniometer            |
| Lendner et al. [53]               | 306 wrists from 171 healthy participants (50% male, 50% female wrists, age: 45.9 ± 20.2 years) | Static, wrist flexion-extension & ulnar-radial deviation. Passive/active ROM not reported | Sitting                        | Gyroscope                    | IPhone 4                      | Goniometer                     |
| Modest et al. [35]                | 30 wrist-injured subjects (age: 47 ± 19 years) 30 wrist-healthy subjects (age: 38 ± 15 years) | Active, static wrist flexion, extension, pronation, supination | Sitting and standing           | In built goniometer          | IPhone 5                      | Universal goniometer           |
| Pourahmadi et al. [54]            | 120 wrists from 70 healthy adults (38 male, 32 female, age: 27.5 years)        | Active, static wrist flexion, extension radial deviation and ulnar deviation | Sitting with the forearm placed on a supporting surface | G-pro                        | IPhone 5                      | Universal goniometer           |
| Santos et al. [34]                | 20 healthy adults (10 male, 10 female, age: 52.5 ± 15.5 years) 20 participants with upper limb injuries (10 male, 10 female, age: 41.5 ± 15.7 years) | Active, static forearm pronation and supination | Sitting                        | Gyroscope, with and without selfie-stick | IPhone 4                      | Goniometer                     |
| Reference                  | Participants                                                                 | Movement assessed                                                                 | Position                                         | Application                                      | Device                        | Criterion instrument               |
|----------------------------|------------------------------------------------------------------------------|------------------------------------------------------------------------------------|--------------------------------------------------|-------------------------------------------------|---------------------------------|----------------------------------|
| Charlton et al. [55]       | 20 healthy male adults (age: 23.8 ± 4.6 years)                               | Passive, static flexion, abduction, adduction, supine internal and external rotation.  
|                            |                                                                              | Supine: hip flexion, internal and external rotation.  
|                            |                                                                              | Sidelying: abduction, adduction. Seated: seated hip internal and external rotation.  
|                            |                                                                              | Hip ROM Tester, designed by a co-author (RC) of this study using MIT App Inventor.  
|                            |                                                                              | Samsung Galaxy S2                                                                | Vicon 3D motion capture                          |
| Derhon et al. [56]         | 34 healthy females (age: 21 ± 2 years)                                       | Passive, static knee extension                                                    | Supine                                           | free ROM                                        | Samsung Galaxy S5 smartphone    |                                  |
| Dos Santos et al. [57]     | 34 healthy females (age: 21 ± 2 years)                                       | Passive, static knee extension                                                    | Supine                                           | free ROM                                        | Samsung Galaxy S5 smartphone    | Universal goniometer             |
| Hambly et al. [58]         | 96 healthy adults (79 male, 17 female, age: 31 ±11 years)                    | Active, static maximum knee joint angle                                          | Supine                                           | iGoniometer                                     | IPhone 3GS                      |                                  |
| Hancock et al. [59]        | 3 healthy adults (absent demographics)                                       | Passive, static knee flexion                                                      | Supine                                           | Goniometer Pro                                  | IPhone 7 Plus                   |                                  |
| Jones et al. [60]          | 36 healthy adults (8 male, 28 female, age: 60.6 ± 6.2 years)                 | Active, static knee joint angle during a lunge                                    | Standing lunge                                   | Simple Goniometer                               | IPhone 3GS                      | Universal goniometer             |
| Mehta et al. [36]          | 60 Orthopedic clinic patients (22 male, 38 female, age: 62.9 ± 8.9 years)    | Active, static knee flexion-extension                                             | Supine                                           | i-Goni                                         | IPhone (model not stated)       | Universal goniometer             |
| Milanese et al. [1]        | 6 healthy adults (3 male, 3 female)                                          | Passive, static knee flexion                                                      | Supine                                           | Knee Goniometer                                 | IPhone 4                        | Universal goniometer             |
| Ockendon & Gilbert [61]    | 5 healthy males, age: 30–40 years                                            | Passive, static knee flexion                                                       | Supine with simulated fixed-flexion deformity.   | Knee Goniometer                                 | IPhone 3GS                      | Conventional goniometer          |
| Pereira et al. [37]        | 20 healthy adults 2 groups of 20 adults post-operative knee surgery Group 1: 8 male, 12 female, age: 72.3 ± 8.8 years, hospitalized 8.5 ± 7.4 days; Group 2: 6 male, 14 female, 72.9 ± 8.9 years, hospitalized 6.9 ± 5.1 days | Active & passive, static knee flexion & extension                                 | Supine                                           | Knee goniometer                                | IPhone 4S                       | Standard goniometer              |
| Ankle                      |                                                                              |                                                                                    |                                                  |                                                 |                                 |                                  |
| Morales et al. [62]        | 33 healthy older adults (age: 71 ± 3.6 years)                                 | Active, static ankle dorsiflexion                                                  | Weight-bearing lunge test                        | Inclinometer                                    | IPhone 5S                       |                                  |
| Vohralik et al. [63]       | 20 healthy adults (7 male, 13 female, age: 22.4 ± 2.0 years)                  | Active, static ankle dorsiflexion                                                  | Weight-bearing lunge test                        | iHandy Level                                    | IPhone (model not stated)       | Digital inclinometer and Fastrak 3D motion capture |

(Continued)
axial rotation; with the studies assessing shoulder motion typically examining flexion, abduction, horizontal adduction as well as external/internal rotation.

The majority of studies involved healthy participants, although some studies involved patients with neck pain [30, 31], shoulder pathology [27, 32, 33], various upper limb injuries [34, 35] or knee pain [36, 37]. A relatively wide variety of smartphones, applications and criterion devices (for the assessment of validity) were utilised in the studies. The most common smartphones were iPhones which were used in 28 studies, with the most common model being the iPhone 4 which was used in nine studies. Samsung phones were used in another six studies, with one study also using an iPod. A wide variety of apps were utilised, with only the most frequently used being the Clinometer (n = 5) and Knee Goniometer (n = 3). All other apps were used in either one or two studies. For the 30 studies that looked at some aspects of validity, the validity of the app was most commonly compared to goniometers (n = 19), 3D motion capture (n = 5) or inclinometers (n = 4).

Critical appraisal
A critical appraisal of the included articles is summarised in Table 2. The percentage of CAT score ranged from 55% [65] to 100% [32]. Papers with ‘NA’ in their appraisal were not assessed against that particular criteria. Two studies were considered to be of low quality with a score < 60% [53, 55], with another one study close to this criteria with an overall quality score of 62% [38]. Only two of the CAT criteria were achieved in less than 50% of the studies (Criteria Six: Order of Examination and Criteria 13: Statistical Methods). This contrasted with one other criteria being achieved in all studies Criteria #10: Execution of the Index Test).

Reliability and validity
The reliability and validity of the assessments are summarised in Table 3. For the sake of simplicity, the following three text sections will summarise the key results for intra-rater reliability, inter-rater reliability and validity, respectively.

**Intra-rater reliability.** Twenty-six studies assessed aspects of intra-rater reliability, with 10 studies reporting relative metrics only, one study reporting absolute metrics only, and the remaining 15 studies reporting both relative and absolute metrics. Twenty-five of 26 studies reported excellent intra-rater relative reliability as defined by an ICC > 0.75 for more than 50% of the joint movements they examined, the only exception being Tousignant-Lafllamme et al. [45]. However, this classification of poor relative intra-rater reliability for Tousignant-Lafllamme et al. [45] was primarily due to the results of one examiner using an iPhone 3, compared to the other examiner who used an iPhone 4. If we were to consider all the studies that assessed relative intra-rater reliability with an iPhone 4, all six studies demonstrated that smartphone apps had adequate relative intra-rater reliability [1, 34, 37, 42, 48, 64]. Thirteen of 17 studies reported good absolute intra-rater reliability as defined by a SEM or MDC < 5° or LOA < ± 9.8° for more than 50% of the joint movements they examined, with only four
Table 2. Critical appraisal of the eligible studies.

| Critical Appraisal Tool (CAT) for Validity and Reliability | 1 V+R | 2 V+R | 3 V | 4 R | 5 R | 6 R | 7 V | 8 R | 9 V | 10 V+R | 11 V | 12 V+R | 13 V+R | % |
|-----------------------------------------------------------|-------|-------|-----|-----|-----|-----|-----|-----|-----|--------|-----|--------|--------|
| **Spine (trunk)**                                         |       |       |     |     |     |     |     |     |     |        |     |        |        |
| Bedekar et al. [38]                                       | Y     | Y     | N   | Y   | N   | N   | Y   | Y   | Y   | N      | Y   | N      | 62%    |
| Furness et al. [39]                                       | Y     | Y     | Y   | Y   | Y   | N   | Y   | Y   | Y   | Y      | Y   | Y      | 92%    |
| Grondin et al. [40]                                       | Y     | Y     | NA  | NA  | Y   | NA  | N   | NA  | Y   | NA     | Y   | Y      | 88%    |
| Jung et al. [41]                                          | Y     | Y     | Y   | NA  | N   | N   | Y   | Y   | Y   | Y      | Y   | Y      | 75%    |
| Kolber & Hanney [42]                                      | Y     | Y     | Y   | Y   | Y   | N   | Y   | Y   | Y   | Y      | Y   | Y      | 85%    |
| Pourahmadi et al. [43]                                    | Y     | Y     | Y   | Y   | Y   | N   | Y   | Y   | Y   | Y      | Y   | Y      | 92%    |
| Pourahmadi et al. [30]                                    | Y     | Y     | Y   | Y   | Y   | N   | Y   | Y   | Y   | Y      | Y   | Y      | 85%    |
| Quek et al. [44]                                          | Y     | Y     | Y   | NA  | Y   | Y   | Y   | Y   | N   | Y      | Y   | N      | 83%    |
| Stenneberg et al. [31]                                    | Y     | Y     | Y   | NA  | Y   | N   | N   | Y   | Y   | N      | Y   | N      | 75%    |
| Tousignant-Laflamme et al. [45]                           | Y     | Y     | Y   | N   | N   | N   | Y   | Y   | Y   | Y      | Y   | N      | 69%    |
| Ullucci et al. [46]                                       | Y     | Y     | NA  | Y   | Y   | Y   | NA  | Y   | NA  | Y      | NA  | Y      | 89%    |
| **Shoulder**                                              |       |       |     |     |     |     |     |     |     |        |     |        |        |
| Lim et al. [47]                                           | Y     | Y     | NA  | Y   | Y   | Y   | NA  | Y   | NA  | Y      | NA  | Y      | 89%    |
| Mejia-Hernandez et al. [33]                               | Y     | Y     | N   | Y   | NA  | N   | Y   | Y   | Y   | N      | Y   | N      | 67%    |
| Mitchell et al. [48]                                      | Y     | Y     | Y   | Y   | Y   | Y   | Y   | Y   | Y   | Y      | Y   | N      | 92%    |
| Ramkumar et al. [49]                                      | Y     | N     | Y   | NA  | NA  | N   | Y   | Y   | Y   | Y      | Y   | N      | 73%    |
| Shin et al. [27]                                          | Y     | Y     | NA  | Y   | Y   | N   | NA  | Y   | NA  | Y      | NA  | Y      | 89%    |
| Werner et al. [28]                                        | Y     | Y     | Y   | NA  | Y   | Y   | Y   | Y   | Y   | Y      | Y   | Y      | 100%   |
| **Elbow**                                                 |       |       |     |     |     |     |     |     |     |        |     |        |        |
| Behnoush et al. [50]                                      | Y     | Y     | Y   | Y   | Y   | Y   | Y   | Y   | Y   | Y      | Y   | Y      | 92%    |
| Cruz & Morais [51]                                        | Y     | Y     | NA  | NA  | N   | N   | Y   | Y   | NA  | Y      | NA  | Y      | 67%    |
| Vauclair et al. [52]                                      | Y     | N     | Y   | NA  | NA  | N   | Y   | Y   | Y   | Y      | Y   | N      | 73%    |
| **Wrist**                                                 |       |       |     |     |     |     |     |     |     |        |     |        |        |
| Lendner et al. [53]                                       | Y     | N     | Y   | NA  | NA  | N   | Y   | N   | Y   | Y      | Y   | N      | 55%    |
| Modest et al. [35]                                        | N     | N     | Y   | NA  | NA  | N   | Y   | Y   | Y   | Y      | Y   | Y      | 73%    |
| Pourahmadi et al. [54]                                    | Y     | Y     | Y   | Y   | Y   | N   | Y   | Y   | Y   | Y      | Y   | Y      | 92%    |
| Santos et al. [34]                                        | Y     | Y     | NA  | Y   | Y   | NA  | N   | NA  | Y   | NA     | Y   | N      | 78%    |
| **Hip**                                                   |       |       |     |     |     |     |     |     |     |        |     |        |        |
| Charlton et al. [55]                                      | Y     | Y     | N   | NA  | N   | N   | Y   | Y   | N   | Y      | Y   | N      | 58%    |
| **Knee**                                                  |       |       |     |     |     |     |     |     |     |        |     |        |        |
| Derhon et al. [56]                                        | Y     | Y     | NA  | NA  | N   | N   | Y   | N   | Y   | NA     | Y   | Y      | 89%    |
| Dos Santos et al. [57]                                    | Y     | Y     | Y   | Y   | N   | Y   | Y   | Y   | Y   | Y      | Y   | Y      | 92%    |
| Hambly et al. [58]                                        | Y     | Y     | Y   | NA  | NA  | N   | Y   | Y   | Y   | Y      | Y   | Y      | 91%    |
| Hancock et al. [59]                                       | Y     | Y     | Y   | NA  | NA  | N   | Y   | Y   | Y   | Y      | Y   | Y      | 92%    |
| Jones et al. [60]                                         | Y     | Y     | Y   | NA  | NA  | Y   | Y   | Y   | Y   | Y      | Y   | N      | 91%    |
| Mehta et al. [36]                                         | Y     | Y     | Y   | Y   | N   | Y   | Y   | Y   | Y   | Y      | Y   | N      | 85%    |
| Milanese et al. [1]                                       | N     | Y     | Y   | Y   | N   | Y   | Y   | Y   | Y   | Y      | Y   | N      | 77%    |
| Ockendon & Gilbert [61]                                   | Y     | N     | Y   | Y   | Y   | N   | Y   | Y   | Y   | Y      | Y   | N      | 77%    |
| Pereira et al. [37]                                       | Y     | Y     | Y   | Y   | Y   | Y   | Y   | Y   | Y   | Y      | N   | Y      | 85%    |
| **Ankle**                                                 |       |       |     |     |     |     |     |     |     |        |     |        |        |
| Morales et al. [62]                                       | Y     | Y     | Y   | Y   | NA  | N   | Y   | Y   | Y   | Y      | Y   | Y      | 92%    |
| Vohralik et al. [63]                                      | Y     | Y     | Y   | Y   | N   | Y   | Y   | Y   | Y   | Y      | Y   | Y      | 92%    |
| Williams et al. [64]                                      | Y     | Y     | NA  | NA  | Y   | Y   | Y   | Y   | Y   | Y      | Y   | Y      | 91%    |

(Continued)
studies not satisfying this threshold \[40, 44, 51, 59\]. It should however be noted that the study by Quek et al. \[44\] satisfied the criteria for more than 50% of the movements when quantified by the SEM (the three of the four movements) but failing this when reliability was assessed by the MDC for all four movements.

Inter-rater reliability. Twenty-five studies assessed aspects of inter-rater reliability, in which 13 studies reported relative metrics only, and 12 studies reported both relative and absolute metrics. Twenty-three of 25 studies demonstrated excellent inter-rater reliability as defined by an ICC \(>0.75\) for more than 50% of the joint movements they examined, with only two studies not satisfying this threshold for relative inter-rater reliability \[37, 45\]. Six out of 11 studies reported good absolute inter-rater reliability as defined by a SEM or MDC < 5° or LOA < ± 9.8° for more than 50% of the joint movements they examined, with five studies not satisfying this criteria \[27, 30, 31, 39, 42\]. While Pourahmadi et al. \[30\] was deemed to not meet this threshold of absolute inter-rater reliability, this was based on all four MDC values > 5°, although the SEM values for the same movements were all < 5°.

Validity. Thirty studies measured some aspects of validity, of which seven studies reported relative metrics only, five studies reported absolute metrics only, and 18 studies reported both relative and absolute metrics. Twenty of 25 studies observed excellent/substantial relative validity as defined by ICC > 0.75, \(r > 0.9\) or CCC > 0.95 for more than 50% of the joint movements examined, with five studies not meeting this threshold for this criteria \[28, 30, 36, 45, 54\]. Seventeen of 23 studies observed excellent/substantial absolute validity as defined by SEM or MDC < 5° or LOA < ± 9.8° for more than 50% of the joint movements they examined, with six studies not meeting this threshold of absolute validity \[27, 34, 36, 42, 53, 57\].

Discussion

This study systematically reviewed the literature for studies which examined the reliability and/or validity of smartphones and apps to quantify joint ROM. Thirty-seven studies were found to be eligible, with the studies assessing joint ROM across most of the body’s major joints. Specifically, the most common joints assessed were the spine/trunk (n = 11), knee (n = 9) and shoulder (n = 6), with a smaller number of studies examining the wrist (n = 4), elbow (n = 3), ankle (n = 3) and hip (n = 1) joints. The primary result of the systematic review was that the apps generally demonstrated adequate intra-rater and inter-rater reliability as well
### Table 3. Reliability and validity of the selected studies.

| Reference                     | Intra-rater reliability | Inter-rater reliability | Validity |
|-------------------------------|-------------------------|-------------------------|----------|
|                               | Relative                | Absolute                | Relative | Absolute | Relative | Absolute |
| Spine (trunk)                 |                         |                         |          |
| Bedekar et al. [38]           | Trunk Flexion ICC = 0.92| Trunk Flexion ICC = 0.812| Trunk Flexion r = 0.95 |
| Furness et al. [39]           | Thoracic rotation ICC = 0.94–0.98 | Thoracic rotation SEM = 2.23–4.36 | Thoracic rotation ICC = 0.72–0.89 | Thoracic rotation r² = 0.697 | Thoracic rotation LoA = 2.8˚ (-9.5˚ to 15.3˚) |
| Bedekar et al. [38]           | Spine (trunk)            |                         |          |
| Grondin et al. [40]           | Cervical Flexion-Rotation ICC = 0.95 | Sagittal plane cervical ROM: ICC = 0.90 | Cervical Flexion-Rotation Test: ICC = 0.81 | Cervical Lateral Flexion ICC = 0.92 |
| Kolber & Hanney [41]          | Pelvic rotation ICC = 0.77–0.83 | Pelvic rotation SEM = 0.64˚ - 0.73˚ | Pelvic rotation ICC = 0.99 | Pelvic rotation LoA = -1.16˚ [-2.2˚ to -0.12˚] |
| Pourahmadi et al. [43]        | Trunk Flexion ICC = 0.87–0.92 | Trunk Extension: SEM = 2.1–3.0˚ trunk Extension: SEM = 2.3–2.74 | Trunk Flexion: r = 0.85 | Trunk Extension: r = 0.91 | Trunk Flexion: LoA from 6.9˚ to 6.3˚ | Trunk Extension: LoA from -5.4˚ to 4.9˚ |
| Pourahmadi et al. [43]        | Cervical Flexion ICC = 0.76 | Cervical Extension ICC = 0.87–0.92 | Cervical Lateral Flexion ICC = 0.90–0.94 | Cervical Rotation ICC = 0.95 |
| Quek et al. [44]              | Cervical Flexion-Rotation ICC = 0.86 | Sagittal plane cervical ROM: ICC = 0.90 | Cervical Flexion-Rotation ICC = 0.87–0.92 | Cervical Lateral Flexion ICC = 0.76 |
| Stenneberg et al. [31]        | Flexion-extension ICC = 0.90 | Rotation ICC = 0.96 | Flexion-extension ICC = 0.95 | Flexion-extension ICC = 0.95 | Flexion-extension ICC = 0.95 |

(Continued)
| Reference | Intra-rater reliability | Inter-rater reliability | Validity |
|-----------|-------------------------|-------------------------|----------|
|           | Relative                | Absolute                | Relative | Absolute |           |
| Tousignant-Laflamme et al. [45] | Examiner 1 with iPhone 4 | Trunk Flexion ICC = 0.48 | \(\text{Trunk Extension ICC} = 0.76\) | | Examiner 1 with iPhone 4 |
|           | Trunk Extension ICC = 0.49 | \(\text{Lateral Flexion ICC} = 0.54\) | | \(\text{Trunk Rotation ICC} = 0.07–0.07\) | |
|           | Trunk Rotation ICC = 0.07–0.07 | | \(\text{Trunk Flexion ICC} = 0.76\) | | |
|           | Trunk Extension ICC = 0.84 | \(\text{Lateral Flexion ICC} = 0.77–0.78\) | | \(\text{Trunk Extension ICC} = 0.58\) | |
|           | Trunk Rotation ICC = 0.66–0.74 | \(\text{Lateral Flexion ICC} = 0.42\) | | \(\text{Trunk Rotation ICC} = 0.70–0.85\) | |
|           | Examiner 2 with iPhone 3 GS | \(\text{Trunk Flexion ICC} = 0.68\) | | \(\text{Trunk Rotation ICC} = 0.43–0.55\) | |
|           | Trunk Flexion ICC = 0.68 | \(\text{Lateral Flexion ICC} = 0.68\) | | | |
|           | Trunk Extension ICC = 0.42 | \(\text{Trunk Rotation ICC} = 0.17–0.80\) | | | |
| Ullucci et al. [46] | Cervical Right rotation | IPhone: ICC = 0.98 | \(\text{Peak ROM ICC} = 0.87\) | Total ROM ICC = 0.82 |
|           | Android: ICC = 0.91 | IPhone: ICC = 0.951 | | | |
|           | Cervical Left rotation | Android: ICC = 0.962 | | | |
| Shoulder | Lim et al. [47] | Supine Shoulder Horizontal Adduction ICC = 0.72–0.89 | Supine Shoulder Horizontal Adduction ICC = 0.94 | | |
|           | Side Lying Shoulder Horizontal Adduction ICC = 0.95–0.97 | | | | |
| Mejia-Hernandez et al. [33] | Forward flexion | ICC = 0.99 | Total active abduction ICC = 0.99 |
|           | Total active abduction ICC = 0.99 | Active glenohumeral Abduction ICC = 0.98 | Passive glenohumeral Abduction ICC = 0.97 |
|           | Active internal rotation ICC = 0.98 | Passive internal rotation ICC = 0.98 | Active external rotation ICC = 0.99 |
|           | Passive internal rotation ICC = 0.99 | Active external rotation ICC = 0.99 | Passive external rotation ICC = 0.99 |
| Mitchell et al. [48] | Shoulder External Rotation ICC = 0.79 | | Shoulder External Rotation ICC = 0.94 |
|           | | | Shoulder External Rotation ICC = 0.94 |
| Ramkumar et al. [49] | Shoulder Flexion ICC = 0.96–0.99 | Flexion Mean difference = 4˚±2˚ |
|           | Shoulder Abduction SEM = 0.95–0.98 | Abduction Mean difference = 3˚±3˚ |
| Shin et al. [27] | Second session, Observer A Shoulder Flexion | Flexion Mean difference = 4˚±2˚ |
|           | Shoulder Flexion SEM = 2.3–2.7 | Abduction Mean difference = 3˚±3˚ |
|           | Shoulder Abduction SEM = 4.5–6.3 | Internal rotation Mean difference = 2˚±4˚ |
|           | Shoulder External Rotation SEM = 2.8–3.3 | External rotation Mean difference = 3˚±3˚ |
|           | Shoulder Internal Rotation SEM = 1.9–3.2 | | | | |

(Continued)
Table 3. (Continued)

| Reference | Intra-rater reliability | Inter-rater reliability | Validity |
|-----------|-------------------------|-------------------------|----------|
|           | Relative | Absolute | Relative | Absolute | Relative | Absolute |
| Werner et al. [28] | | | | | | |
| Healthy: | | | | | | |
| Shoulder Abduction | ICC = 0.72 | | | | | |
| Shoulder Flexion | ICC = 0.75 | | | | | |
| Shoulder External Rotation | ICC = 0.86 | | | | | |
| Shoulder Internal Rotation | ICC = 0.81 | | | | | |
| Symptomatic: | | | | | | |
| Shoulder Abduction | ICC = 0.91 | | | | | |
| Shoulder Flexion | ICC = 0.97 | | | | | |
| Shoulder External Rotation | ICC = 0.85–0.88 | | | | | |
| Shoulder Internal Rotation | ICC = 0.86 | | | | | |
| Relative Absolute | | | | | | |
| Shoulder Abduction | SEM = 3.0˚ | | | | | |
| Shoulder Flexion | SEM = 3.1˚ | | | | | |
| Shoulder External Rotation | SEM = 3.7–4.0˚ | | | | | |
| Shoulder Internal Rotation | SEM = 3.1˚ | | | | | |
| Healthy (as assessed by Orthopaedic Sports Medicine fellow) | | | | | | |
| Shoulder Abduction | MD ± LOA = 4.0 ± 9.5˚ | | | | | |
| Shoulder Flexion MD ± LOA = 6.2 ± 10.8˚ | | | | | | |
| Shoulder External Rotation | MD ± LOA = 6.9 ± 12.1˚ and 9.7 ± 14.9˚ | | | | | |
| Shoulder Internal Rotation | MD ± LOA = 6.9 ± 10.9˚ | | | | | |
| Symptomatic (as assessed by Orthopaedic Sports Medicine fellow) | | | | | | |
| Shoulder Abduction | MD ± LOA = 2.6 ± 4.1˚ | | | | | |
| Shoulder Flexion MD ± LOA = 2.4 ± 4.3˚ | | | | | | |
| Shoulder External Rotation | MD ± LOA = 2.0 ± 3.4˚ and 2.5 ± 5.8˚ | | | | | |
| Shoulder Internal Rotation | MD ± LOA = 1.7 ± 4.0˚ | | | | | |
| Relative Absolute | | | | | | |
| Elbow | Behnoush et al. [50] | | | | | |
| Elbow flexion: ICC = 0.95 | | | | | | |
| Pronation: ICC = 0.98 | | | | | | |
| Supination: ICC = 0.98 | | | | | | |
| Elbow flexion: ICC = 0.84 | | | | | | |
| Pronation: ICC = 0.90 | | | | | | |
| Supination: ICC = 0.96 | | | | | | |
| Elbow flexion MD ± LoA = -0.4˚ (-3.9˚ to 3.0˚) | | | | | | |
| Pronation MD ± LoA = 0.4˚ (-5.8˚ to 5.0˚) | | | | | | |
| Supination MD ± LoA = -0.4˚ (-6.0˚ to 5.2˚) | | | | | | |
| Cruz & Morais [51] | | | | | | |
| Elbow flexion at onset of pain dominant side: SEM = 6.6˚ | | | | | | |
| Elbow flexion at onset of pain non-dominant side: SEM = 6.8˚ | | | | | | |
| Elbow flexion at onset of pain dominant side: MDC95 = 18.4˚ | | | | | | |
| Elbow flexion at onset of pain non-dominant side: MDC95 = 18.8˚ | | | | | | |
| Elbow flexion at max tolerable pain dominant side: SEM = 4.8˚ | | | | | | |
| Elbow flexion at max tolerable pain non-dominant side: SEM = 4.2˚ | | | | | | |
| Elbow flexion at max tolerable pain dominant side: MDC95 = 13.2˚ | | | | | | |
| Elbow flexion at max tolerable pain non-dominant side: MDC95 = 11.7˚ | | | | | | |
| Vasclair et al. [52] | | | | | | |
| Flexion SEM = 1˚ | | | | | | |
| Extension SEM = 0.8˚ | | | | | | |
| Flexion SEM = 1.9˚ | | | | | | |
| Flexion SEM = 1.2˚ | | | | | | |
| Wrist | Lendner et al. [53] | | | | | |
| Wrist Range of Motion MD (LoA) = 0.5˚. (-16.7˚ to 17.3˚) | | | | | | |
| Modest et al. [35] | | | | | | |
| Healthy: | | | | | | |
| Wrist flexion ICC = 0.97 | | | | | | |
| Wrist extension ICC = 0.96 | | | | | | |
| Wrist supination ICC = 0.95 | | | | | | |
| Injured: | | | | | | |
| Wrist flexion: ICC = 0.99 | | | | | | |
| Wrist extension ICC = 0.99 | | | | | | |
| Wrist supination ICC = 0.99 | | | | | | |
| LoA = average absolute deviation < 2˚ | | | | | | |

(Continued)
Table 3. (Continued)

| Reference         | Intra-rater reliability | Inter-rater reliability | Validity          |
|-------------------|--------------------------|--------------------------|-------------------|
|                    | Relative                 | Absolute                 | Relative          | Absolute                 |
| Pourahmadi et al. | Within day              | Wrist Flexion ICC = 0.89 | Wrist Flexion ICC = 0.79 | Wrist Flexion SEM = 2.2 |
|                   |                          | Wrist Extension ICC = 0.90 | Wrist Extension ICC = 0.81 | Wrist Extension SEM = 1.7 |
|                   |                          | Wrist Radial Deviation ICC = 0.81 | Wrist Radial Deviation ICC = 0.80 | Wrist Radial Deviation SEM = 1.1 |
|                   |                          | Wrist Ulnar Deviation ICC = 0.80 | Wrist Ulnar Deviation ICC = 0.82 | Wrist Ulnar Deviation SEM = 1.6 |
|                   |                          | Wrist Extension MDC = 4.3 | Wrist Extension MDC = 2.9 | Wrist Extension MDC = 4.6 |
|                   |                          | Wrist Ulnar Deviation MDC = 2.6 | Wrist Ulnar Deviation MDC = 2.9 | Wrist Ulnar Deviation MDC = 4.5 |
|                   |                          | Wrist Extension MD (LOA) = 0.99 | Wrist Extension SEM (LOA) = 0.63 | Wrist Extension MD (LOA) = 0.96 |
|                   |                          | Wrist Extension SEM (LOA) = 0.63 | Wrist Extension SEM (LOA) = 0.73 | Wrist Extension SEM (LOA) = 0.84 |
| Santos et al.     | Injured i-Phone 5 with self-stick Wrist Pronation and Supination: ICC = 0.94 | Injured i-Phone 5 with self-stick Wrist Pronation and Supination: ICC = 0.94 | Injured i-Phone 5 with self-stick Wrist Pronation and Supination: ICC = 0.94 | Injured i-Phone 5 with self-stick Wrist Pronation and Supination: ICC = 0.94 |
|                   |                          | Non-injured i-Phone 5 with self-stick Wrist Pronation and Supination: ICC = 0.89 | Non-injured i-Phone 5 with self-stick Wrist Pronation and Supination: ICC = 0.89 | Non-injured i-Phone 5 with self-stick Wrist Pronation and Supination: ICC = 0.89 |
|                   |                          | i-Phone 5 handheld Wrist Pronation and Supination: ICC = 0.77 | i-Phone 5 handheld Wrist Pronation and Supination: ICC = 0.77 | i-Phone 5 handheld Wrist Pronation and Supination: ICC = 0.77 |
|                   |                          | Hip Flexion ICC = 0.86 | Hip Flexion ICC = 2.3 | Hip Flexion ICC = 0.92 |
|                   |                          | Hip Abduction ICC = 0.68 | Hip Abduction SEM = 4.6 | Hip Abduction ICC = 0.68 |
|                   |                          | Hip Adduction ICC = 0.68 | Hip Adduction SEM = 2.5 | Hip Adduction ICC = 0.68 |
|                   |                          | Hip Supine IR ICC = 0.94 | Hip Supine IR SEM = 3.2 | Hip Supine IR ICC = 0.94 |
|                   |                          | Hip Supine ER ICC = 0.87 | Hip Supine ER SEM = 2.6 | Hip Supine ER ICC = 0.87 |
|                   |                          | Hip Sitting IR ICC = 0.94 | Hip Sitting IR SEM = 3.4 | Hip Sitting IR ICC = 0.94 |
|                   |                          | Hip Sitting ER ICC = 0.63 | Hip Sitting ER SEM = 2.8 | Hip Sitting ER ICC = 0.63 |
|                   |                          | Hip Flexion Measurement 1 SDMD = 2.6 and SEMD = 0.4 | Hip Flexion Measurement 1 SDMD = 2.6 and SEMD = 0.4 | Hip Flexion Measurement 1 SDMD = 2.6 and SEMD = 0.4 |
|                   |                          | Hip Flexion Measurement 2 SDMD = 3.3 and SEMD = 0.6 | Hip Flexion Measurement 2 SDMD = 3.3 and SEMD = 0.6 | Hip Flexion Measurement 2 SDMD = 3.3 and SEMD = 0.6 |
|                   |                          | Hip Flexion Measurement 3 SDMD = 2.3 and SEMD = 0.4 | Hip Flexion Measurement 3 SDMD = 2.3 and SEMD = 0.4 | Hip Flexion Measurement 3 SDMD = 2.3 and SEMD = 0.4 |
|                   |                          | Knee Flexion Measurement 1 SDMD = 2.6 and SEMD = 0.4 | Knee Flexion Measurement 1 SDMD = 2.6 and SEMD = 0.4 | Knee Flexion Measurement 1 SDMD = 2.6 and SEMD = 0.4 |
|                   |                          | Knee Flexion Measurement 2 SDMD = 3.3 and SEMD = 0.6 | Knee Flexion Measurement 2 SDMD = 3.3 and SEMD = 0.6 | Knee Flexion Measurement 2 SDMD = 3.3 and SEMD = 0.6 |
|                   |                          | Knee Flexion Measurement 3 SDMD = 2.3 and SEMD = 0.4 | Knee Flexion Measurement 3 SDMD = 2.3 and SEMD = 0.4 | Knee Flexion Measurement 3 SDMD = 2.3 and SEMD = 0.4 |
|                   |                          | Knee Flexion r = 0.96–0.98 | Knee Flexion r = 0.96–0.98 | Knee Flexion r = 0.96–0.98 |
|                   |                          | Knee Flexion i-Phone 5 handheld | Knee Flexion i-Phone 5 handheld | Knee Flexion i-Phone 5 handheld |
|                   |                          | Knee Flexion r = 0.96–0.98 | Knee Flexion r = 0.96–0.98 | Knee Flexion r = 0.96–0.98 |
|                   |                          | Knee Flexion Measurement 1 MD = 0.5° (-4.6° to 6.1°) | Knee Flexion Measurement 1 MD = 0.5° (-4.6° to 6.1°) | Knee Flexion Measurement 1 MD = 0.5° (-4.6° to 6.1°) |
|                   |                          | Knee Flexion Measurement 2 MD = 0.5° (-6.0° to 7.1°) | Knee Flexion Measurement 2 MD = 0.5° (-6.0° to 7.1°) | Knee Flexion Measurement 2 MD = 0.5° (-6.0° to 7.1°) |
|                   |                          | Knee Flexion Measurement 3 MD = 0.5° (-4.0° to 5.3°) | Knee Flexion Measurement 3 MD = 0.5° (-4.0° to 5.3°) | Knee Flexion Measurement 3 MD = 0.5° (-4.0° to 5.3°) |
| Mehta et al. [36] | Knee Flexion ICC = 0.97 | Knee Flexion ICC = 0.99 | Knee Flexion r = 0.92 | Knee Flexion r = 0.92 |
|                   | Knee Extension ICC = 0.94 | Knee Extension r = 0.92 | Knee Extension r = 0.92 | Knee Extension r = 0.92 |
|                   | Knee Flexion Measurement 1 SDMD = 2.7 | Knee Extension MDC = 6.3 | Knee Extension MDC = 6.3 | Knee Extension MDC = 6.3 |
|                   | Knee Flexion Measurement 2 SDMD = 3.2 | Knee Extension SEM = 1.18 | Knee Extension SEM = 1.18 | Knee Extension SEM = 1.18 |
|                   | Knee Flexion Measurement 3 SDMD = 2.3 | Knee Extension MDC = 4.3 | Knee Extension MDC = 4.3 | Knee Extension MDC = 4.3 |
| Milanese et al. [1] | Knee flexion 3 Clinicians: CCC = 0.99 | Knee flexion 3 Clinicians: ICC = 0.99 | Knee flexion r = 0.95 | Knee flexion r = 0.95 |
|                   | Knee flexion 3 Students | Knee flexion 3 Students | Knee flexion 3 Students | Knee flexion 3 Students |
|                   | SEM = 1.4 | SEM = 1.4 | SEM = 1.4 | SEM = 1.4 |
|                   | SEM = 1.5 | SEM = 1.5 | SEM = 1.5 | SEM = 1.5 |
| Ockendon & Gilbert [61] | Knee flexion r = 0.99 | Knee flexion r = 0.98 | Knee flexion r = 0.95 | Knee flexion r = 0.95 |
|                   | Knee flexion MD ± LoA = -0.4° ± 3.9° | Knee flexion MD ± LoA = -0.4° ± 3.9° | Knee flexion MD ± LoA = -0.4° ± 3.9° | Knee flexion MD ± LoA = -0.4° ± 3.9° |

(Continued)
as validity when compared to criterion devices such as goniometers, inclinometers and 3D motion capture. However, there was a trend for the reliability outcomes that these results were somewhat stronger for relative (e.g. ICC, r) than absolute measures (e.g. SEM, MDC).

The tendency for the relative measures to be stronger than absolute measures is something that needs to be clearly understood by the clinician. Historically, many reliability and/or validity studies have only reported relative statistics such as the ICC and Pearson’s product moment correlation [3, 7, 8]. Relative statistical measures are typically used to describe the resemblance of two or more units within a group (e.g. the similarity of measurements undertaken by two clinicians) as a function of the resemblance between different groups. ICC is thus operationalized as a ratio between two variance measures [66]. To illustrate, the inter-rater reliability ICC measure of Pourahmadi et al. [30] is derived by the ratio of variance between (1) the variance between two measurements from the same participant, repetition, and session, against (2) the variance between two measurements from the same participant, repetition, session, from different raters [66]. While these relative statistics provide important information regarding the correlation or rank order of two or more measurements, they provide no detail regarding the magnitude of change/difference in the measurement across these time points [20, 21]. In contrast, absolute statistical measures of reliability/validity simply report the resemblance of two or more units within a group—in other words, it simply represents the individual variance components [66]. Clinically, an ICC is useful for a manager wanting to train a team of clinicians in the use of a mobile app, where the aim is to achieve a value as close to one. However, the individual variance components of between repetition and between sessions are more useful for the day-to-day practice of individual clinicians. Knowing the inherent variation in

Table 3. (Continued)

| Reference | Intra-rater reliability | Inter-rater reliability | Validity |
|-----------|-------------------------|-------------------------|----------|
|           | Relative | Absolute | Relative | Absolute | Relative | Absolute |
| Pereira et al. [37] | Postoperative | Knee flexion active motion | ICC = 0.99 | Knee flexion passive motion: ICC = 0.92 | Healthy | Knee flexion active motion | ICC = 0.86 | Knee flexion passive motion: ICC = 0.90 |
|           | Postoperative | Knee flexion active motion | ICC = 0.43 | Knee flexion passive motion: ICC = 0.27 | Healthy | Knee flexion active motion | ICC = 0.12 | Knee flexion passive motion: ICC = 0.13 |
|           | Postoperative | Knee Active extension | ICC = 0.80 | Knee Active flexion | ICC = 0.97 | Knee Passive extension | ICC = 0.72 | Knee Passive flexion | ICC = 0.99 | Healthy | Knee Active extension | ICC = 0.88 | Knee Active flexion | ICC = 0.60 | Knee Passive extension | ICC = 0.90 | Knee Passive flexion | ICC = 0.50 |

Ankle

Morales et al. [62] Ankle dorsiflexion ICC = 0.97–0.98 Ankle dorsiflexion SEM = 0.29 cm–0.43 cm MDC = 0.79 cm–1.19 cm

Vohralik et al. [63] Ankle Dorsiflexion ICC = 0.97 Ankle Dorsiflexion SEM = 1.4° Ankle Dorsiflexion ICC = 0.76 Ankle Dorsiflexion SEM = 3.4° Ankle Dorsiflexion ICC = 0.76 Ankle Dorsiflexion SEM = 3.4° Ankle Dorsiflexion SEM = 3.4° Ankle Dorsiflexion ICC = 0.76 Ankle Dorsiflexion ICC = 0.80 Ankle Dorsiflexion with bent knee ICC = 0.85 Ankle Dorsiflexion with straight knee ICC = 0.81 Ankle Dorsiflexion with bent knee ICC = 0.85

Williams et al. [64] Ankle dorsiflexion with straight knee ICC = 0.81 Ankle dorsiflexion with bent knee ICC = 0.85

CCC = Concordance correlation coefficient, Coefficient of Variation = CoV, ER = external rotation, ICC = intraclass correlation, IR = internal rotation, LoA = limits of agreement, MD = mean difference, MDC = minimal detectable change, r = Pearson’s product moment correlation, SDMD = Standard deviation of the mean difference, SEM = standard error of the measurement, SEMD = standard error of the mean difference, SRD = standard real difference, TCP = thoracolumbar-pelvic.

https://doi.org/10.1371/journal.pone.0215806.t003

Reliability and validity of apps for measuring joint range of motion

PLOS ONE | https://doi.org/10.1371/journal.pone.0215806 May 8, 2019 16 / 24
outcomes between each measurement repetition and between clinical visits, allows a clinician to judge the clinical importance of any kinematic change using a mobile app.

With respect to the validity of smartphones and apps to quantify ROM, it was apparent that the majority of studies included in this review assessed the validity of the smartphone app against a universal goniometer as the criterion test. However, it could be argued that the most appropriate criterion measure to determine joint ROM would be radiographic images such as x-ray or 3D motion capture. Only five studies utilised 3D motion capture as the criterion method [31, 41, 44, 55, 63]. All five of these studies demonstrated that the apps had adequate levels of relative validity with respect to 3D motion capture [31, 41, 44, 55, 63], with a similar result observed for all of the three studies assessing absolute metrics also reporting adequate validity [31, 41, 44]. It should also be noted that Charlton et al. [55] compared the relative validity of their smartphone app and inclinometer to the criterion method of 3D motion capture for assessing hip joint ROM. Based on the ICC threshold of 0.75 for sufficient validity, both devices were valid, with the smartphone exceeding this threshold for five of the six joint ROM and the inclinometer for all six. The use of 3D motion capture as a criterion measure may be more important when assessing dynamic rather than static joint ROM due to the inherent difficulties in maintaining correct position of the universal goniometer on the joint centre and its alignment with the proximal and distal joints during movement, especially at high movement velocities [11]. All the five 3D motion capture validity studies included in the present review assessed static ROM [31, 41, 44, 55, 63]—i.e. when range is recorded when a joint was positioned statically at its limit of motion. The lack of assessment of apps on dynamic ROM may not be surprising given that assessors need a joint to be held transiently in a static position to record the range from the app. For apps to measure dynamic ROM, it needs to sample a joint’s motion throughout the movement task, and this data needs to be post-processed to extract parameters of ROM—similar to how a 3D motion capture system quantifies ROM. Future studies are warranted to quantify the validity and reliability of smartphone apps in the assessment of dynamic ROM.

Another issue of major importance to clinicians is whether the smartphones and apps display adequate reliability and validity across all joints, joint actions and populations. It was heartening to see that most of these variables did not seem to influence the reliability and validity of the apps in measuring joint ROM. There was clear variation in the reliability and validity in different spinal joint movements, as well as a tendency for differences in reliability and validity between healthy and clinical populations and to a lesser extent smartphone models that the clinician should be aware.

When examining the 11 studies examining spinal ROM, it appeared that the assessment of flexion, extension and lateral flexion typically exhibited adequate relative reliability and/or validity [30, 31, 38–46], although not all of these studies assessed absolute reliability and validity. Compared to the assessment of spinal flexion, extension and lateral flexion, the assessment of spinal axial rotation did not exhibit adequate reliability and validity in four [30, 40, 44, 45] of the nine studies. On this basis, it would appear that while the apps used within the studies reviewed in this manuscript typically had adequate reliability and validity for measuring spinal flexion, extension and lateral flexion, they are somewhat more questionable for measuring spinal rotation. Nevertheless, a recent study by Furness et al. [5] demonstrated comparable (or slightly better) reliability of an iPhone 6 and the Compass app to the universal goniometer for assessing thoracic rotation in healthy individuals. Unfortunately, while this study also demonstrated strong correlations between the Compass app and the universal goniometer, absolute validity was again inadequate as the limits of agreement between the two devices was ~25˚ [5]. Such findings suggest that the ability to perform a valid assessment of spinal rotation using devices that are feasible in clinical practice, be it goniometers or smartphone-based apps, may...
still remain somewhat questionable. Further research and/or additional clinical training into the use of these devices in this context is therefore warranted.

The comparatively poorer reliability and validity of smartphone apps measuring ROM in axial rotation compared to flexion-extension and lateral flexion could be attributed to several factors. First is the difference in smartphone sensor performance in different Cardinal planes [67]. Using performance testing of commercial Inertial Measurement Units (IMUs) as an example, the static error of the Xsens MT9 IMU was three times greater in the yaw (axial rotation) direction, than in the other two Cardinal Planes [68]. Second, is the reliance of different components of the smartphone sensor (e.g. magnetometer vs gyroscope) when measuring ROM in different Cardinal planes. Magnetometers are required when testing axial rotation in an anti-gravity position (e.g. sitting) [44, 45]. Compared to gravity-dependent gyroscopes, magnetometers are more sensitive to signal distortion arising from the environmental magnetic fields, potentially reducing their validity and reliability. In contrast, Pourahmadi et al. [30] tested cervical rotation in supine using the gravity-dependent gyroscope component of the smartphone sensor. This could explain the better validity and reliability of Pourahmadi et al. [30] compared to two other studies who reported poor reliability and validity [44, 45]. Third, is the issue of axis mis-alignment which occurs when the sensor’s coordinate axes are not aligned with anatomically meaningful axes [69]. There may be greater potential for axis mis-alignment, during axial rotation than in other movement directions [70–72]. Given that spinal axial rotation commonly couples with secondary movement in other directions, maintaining a pure axial rotation may be difficult.

While most of the studies reviewed in this manuscript involved healthy participants, some recruited patients with joint pain. These studies included groups of individuals with neck pain [30, 31], shoulder pathology [27, 32, 33], various upper limb injuries [34, 35] or knee pain [36, 37]. The intra-rater and inter-rater reliability of the apps in these clinical populations was typically adequate in these nine studies, with the exception of Pereira et al. [37]. The validity of the apps in these populations was sufficiently high in six of the nine studies. For the three studies with insufficient validity [34, 36, 37], a variety of statistical approaches were used, with the results being CCC = 0.50–0.72, r = 0.68 and LoA ranging from -10 to +17.3° for the measured joint actions. Such results may suggest that using smartphones and apps can be quite reliable in a range of population groups, including some clinical populations presenting with musculoskeletal pathology.

The clinician should also be aware of the potential for how the make and model of the smartphone and the actual app can influence the reliability of assessment and how these two factors; as well as how the criterion test selected may influence the validity. While there was some variability between studies in the smartphone used (29 studies using iPhones, most commonly iPhone 4 or 5), there was little evidence of any effect of smartphone with the exception of one study [45]. Specifically, Tousignant-Laflamme et al. [45] reported adequate relative intra-rater reliability for an examiner with an iPhone 4, but not an examiner with an iPhone 3; with this ultimately resulting in poor inter-rater reliability. Further, the two examiners were unable to demonstrate adequate validity when compared to the CROM device, which is considered a criterion measure for measuring cervical ROM. Such results suggest that clinicians should use more recently developed smartphones, which are more likely to have improved sensor capacity than older smartphone models such as the iPhone 3.

With respect to the number of apps included in this review, there was a wide variety examined in these 37 studies. This was clearly demonstrated as only two apps were used in more than two studies, these being the Clinometer (n = 5) and Knee Goniometer (n = 3). The wide diversity of apps utilised in these studies and the general support for all of these apps’ reliability and validity demonstrated in this review, suggest that the clinician has multiple options when
selecting the most appropriate app for measuring a particular joint ROM. However, it would still be recommended that clinicians utilise apps that have been demonstrated to be reliable and valid for measuring the particular joint action they wish to measure. We would also recommend that researchers need to continue to examine the reliability and validity of more recently developed apps and smartphones to determine if they offer advantages over those previously developed and assessed in the scientific literature.

This systematic review has several strengths and limitations that need to be acknowledged. A primary strength of this review in comparison to the literature [14, 15] is that it provides more detailed reporting of key aspects of the methodology and the actual relative and absolute intra-rater, inter-rater and validity outcomes for each joint action assessed in each study within our summary tables. The current study also appears to be the first systematic review on this topic to use a validated tool to assess the included studies’ methodological quality. By performing this assessment of study quality, it was determined that only two of the 37 studies were considered to be of low quality, based on a CAT score of less than 60% [53, 55]. Further, only two of the 13 CAT criteria were achieved in less than 50% of the studies (Criteria Six: Order of Examination and Criteria 13: Statistical Methods). The low score for Criteria Six: Order of Examination reflected the lack of randomisation and the potential for a learning or fatigue effect in many of the studies. The low score for Criteria 13 (Statistical Methods) tended to reflect the fact that most studies only reported relative reliability and/or validity statistics (e.g. r or ICC) without also reporting comparable absolute reliability and/or validity statistics (e.g. SEM, MDC or MD±LOA). As each of these three CAT criteria are highly important characteristics of strong psychometric study design, improvement in these areas would further strengthen the level of evidence described in this review.

The primary limitation of our review process reflected the manner in which sufficient validity and reliability was described. Specifically, we utilised a process in which a particular app was described as suitably reliable and/or valid when recommended statistical thresholds were achieved in more than 50% of the movements examined in each study. While this approach is useful as a generalised approach to describe the reliability and/or validity of an app, it is perhaps a little bit too simplistic due to the relatively high between-study variation in populations, joints, joint actions, smartphone and app (including software updates). This potential negative influence of software updates on the reliability and validity of apps has also been recently highlighted as a major issue in the use of global positioning systems (GPS) in sport [73]. Due to the limitation of our somewhat arbitrary greater than 50% reliability and validity threshold, we would suggest that clinicians should still examine the actual data summarised in this systematic review, as it is quite possible that different joint motions may demonstrate differences in their reliability and validity, even when assessed in the same population with the same smartphone and app. The final limitation of this systematic review is that we cannot be 100% certain that all eligible articles were identified and included in this systematic review.

Conclusion

The results of this systematic review provide relatively strong evidence regarding the intra-rater, inter-rater and validity of smartphones and apps to assess joint ROM; with these results tending to be observed across multiple joints, joint actions, populations, smartphones and apps. Such results suggest that clinicians may be able to use a relatively wide variety of smartphones and apps to quantify joint ROM. However, when absolute validity was assessed, they were often reasonably large differences in the angle determined by an app compared to a criterion measure such as 3D motion capture, goniometry or inclinometers. On this basis, it is imperative that the clinician does not switch between different assessment devices (such as a
goniometer and a smartphone based apps) when assessing an individual across multiple time points. Clinical researchers should also aim to develop more reliable and valid protocols for using smartphones and apps, while continuing to collaborate with smartphone and app developers to further improve their reliability and validity for assessing joint ROM.

Supporting information
S1 Checklist. PRISMA checklist.
(DOC)
S1 Appendix. PRISMA flowchart.
(DOCX)

Acknowledgments
We would like to thank Mr David Honeyman, who is the Librarian for the Faculty of Health Sciences and Medicine, Bond University for assistance with the systematic review process.

Author Contributions
Conceptualization: Justin W. L. Keogh, Ben Schram, James Furness.
Data curation: Justin W. L. Keogh, Alistair Cox, Sarah Anderson, James Furness.
Formal analysis: Justin W. L. Keogh, Alistair Cox, Sarah Anderson, Bernard Liew, Alicia Olsen.
Investigation: Justin W. L. Keogh, Alistair Cox, Sarah Anderson.
Methodology: Justin W. L. Keogh, Alistair Cox, Sarah Anderson, Ben Schram, James Furness.
Supervision: Justin W. L. Keogh.
Writing – original draft: Justin W. L. Keogh, Alistair Cox, Sarah Anderson, Bernard Liew, Alicia Olsen, James Furness.
Writing – review & editing: Justin W. L. Keogh, Alistair Cox, Sarah Anderson, Bernard Liew, Alicia Olsen, Ben Schram, James Furness.

References
1. Milanese S, Gordon S, Buettner P, Flavell C, Ruston S, Coe D, et al. Reliability and concurrent validity of knee angle measurement: smart phone app versus universal goniometer used by experienced and novice clinicians. Man Ther. 2014; 19(6):569–74. Epub 2014/06/20. https://doi.org/10.1016/j.math.2014.05.009 PMID: 24942491.
2. Brosseau L, Tousignant M, Budd J, Chartier N, Duciaume L, Plamondon S, et al. Intratester and intertester reliability and criterion validity of the parallelogram and universal goniometers for active knee flexion in healthy subjects. Physiother Res Int. 1997; 2(3):150–66. Epub 1997/01/01. PMID: 9421820.
3. MacDermid JC, Chesworth BM, Patterson S, Roth JH. Intratester and intertester reliability of goniometric measurement of passive lateral shoulder rotation. J Hand Ther. 1999; 12(3):187–92. PMID: 10459526.
4. May S, Chance-Larsen K, Littewood C, Lomas D, Saad M. Reliability of physical examination tests used in the assessment of patients with shoulder problems: a systematic review. Physiotherapy. 2010; 96(3):179–90. Epub 2010/08/03. https://doi.org/10.1016/j.physio.2009.12.002 PMID: 20674649.
5. Furness J, Schram B, Cox AJ, Anderson SL, Keogh J. Reliability and concurrent validity of the iPhone® Compas application to measure thoracic rotation range of motion (ROM) in healthy participants. PeerJ. 2018; 6:e4431. https://doi.org/10.7717/peerj.4431 PMID: 29568701.
6. Gajdosik RL, Bohannon RW. Clinical measurement of range of motion. Review of goniometry emphasizing reliability and validity. Phys Ther. 1987; 67(12):1867–72. Epub 1987/12/01. PMID: 3685114.
7. Youdas JW, Carey JR, Garrett TR. Reliability of measurements of cervical spine range of motion—comparison of three methods. Phys Ther. 1991; 71(2):98–106. Epub 1991/02/01. PMID: 1989013.

8. Burdett RG, Brown KE, Fall MP. Reliability and validity of four instruments for measuring lumbar spine and pelvic positions. Phys Ther. 1986; 66(5):677–84. Epub 1986/05/01. PMID: 3703932.

9. Nitschke J, Nattrass C, Disler P, PhD F, Chou M, Ooi K. Reliability of the American Medical Association Guides’ Model for Measuring Spinal Range of Motion: Its Implication for Whole-Person Impairment Rating. Spine. 1999; 24(3):262–8. PMID: 10025021

10. Cronin J, Nash M, Whatman C. Assessing dynamic knee joint range of motion using siliconcoach. Phys Ther Sport. 2006; 7(4):191–4. Epub 2006/11/01. https://doi.org/10.1016/j.ptsp.2006.07.003 PMID: 21663831.

11. Piriyaprasarth P, Morris ME, Winter A, Bialocerkowski AE. The reliability of knee joint position testing using electrogoniometry. BMC Musculoskeletal Disorders. 2008; 9:6. https://doi.org/10.1186/1471-2474-9-6 PMID: 18211714

12. Keogh JWL, Espinosa HG, Grigg J. Evolution of smart devices and human movement apps: recommendations for use in sports science education and practice. Journal of Fitness Research. 2016; 5(Special Issue ASTN–Q Conference):14–5.

13. Gratton C, Jones I. Theories, concepts and variables. Research methods for sports studies. 2nd ed. London: Routledge; 2010. p. 77–99.

14. Milani P, Coccetta CA, Rabini A, Sciarra T, Massazza G, Ferriero G. Mobile smartphone applications for body position measurement in rehabilitation: a review of goniometric tools. PM R. 2014; 6(11):1038–43. Epub 2014/05/23. https://doi.org/10.1016/j.pmrj.2014.05.003 PMID: 24844445.

15. Brink Y, Louw QA. Clinical instruments: reliability and validity critical appraisal. J Eval Clin Pract. 2012; 18(6):1126–32. Epub 2011/06/22. https://doi.org/10.1111/j.1365-2753.2011.01707.x PMID: 21689217.

16. Shiel F, Persson C, Furness J, Simas V, Pope R, Climstein M, et al. Dual energy X-ray absorptiometry positioning protocols in assessing body composition: A systematic review of the literature. Journal of science and medicine in sport. 2018; 21(10):1038–44. Epub 2018/03/29. https://doi.org/10.1016/j.jsams.2018.03.005 PMID: 29588115.

17. Barrett E, McCreech K, Lewis J. Reliability and validity of non-radio graphic methods of thoracic kyphosis measurement: a systematic review. Man Ther. 2014; 19(1):10–7. Epub 2013/11/20. https://doi.org/10.1016/j.math.2013.09.003 PMID: 24246907.

18. Fleiss J. The measurement of interrater agreement: statistical methods for rates and proportions. Sons JW, editor. New York1981.

19. Fox B, Henwood T, Neville C, Keogh J. Relative and absolute reliability of functional performance measures for adults with dementia living in residential aged care. International Psychogeriatrics. 2014; 26 (10):1659–67. https://doi.org/10.1017/S1041610214001124 PMID: 24894393

20. Liaw L-J, Hsieh C-L, Lo S-K, Chen H-M, Lee S, Lin J-H. The relative and absolute reliability of two balance performance measures in chronic stroke patients. Dis Rehabil. 2008; 30(9):656–61. https://doi.org/10.1080/09638280701400698 PMID: 17852318

21. Fleiss J. The measurement of interrater agreement: statistical methods for rates and proportions. Sons JW, editor. New York1981.

22. Mukaka MM. A guide to appropriate use of Correlation coefficient in medical research. Malawi Med J. 2012; 24(3):69–71. PMC3576830. PMID: 23638278

23. McBride GB. A proposal for strength-of-agreement criteria for Lin’s Concordance Correlation Coefficient. Hamilton, New Zealand: 2005.

24. McGinley JL, Baker R, Wolfe R, Morris ME. The reliability of three-dimensional kinematic gait measurements: A systematic review. Gait Posture. 2009; 29(3):360–9. https://doi.org/10.1016/j.gaitpost.2008.09.003 PMID: 19013070

25. Shin SH, Ro du H, Lee OS, Oh JH, Kim SH. Within-day reliability of shoulder range of motion measurement with a smartphone. Man Ther. 2012; 17(4):298–304. Epub 2012/03/17. https://doi.org/10.1016/j.j.math.2012.02.010 PMID: 22421186.

26. Werner BC, Holzgrefe RE, Griffin JW, Lyons ML, Cosgrove CT, Hart JM, et al. Validation of an innovative method of shoulder range-of-motion measurement using a smartphone clinometer application. J
Reliability and validity of apps for measuring joint range of motion

29. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gatchel PC, Ioannidis JPA, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. J Clin Epidemiol. 2009; 62(10):e1–e34. https://doi.org/10.1016/j.jclinepi.2009.06.006 PMID: 19631507

30. Pourahmadi MR, Bagheri R, Taghipour M, Takamjani IE, Mohseni-Bandpei MA. A new iPhone(R) application for measuring active cranio-cervical range of motion in patients with non-specific neck pain: a reliability and validity study. Spine J. 2017. Epub 2017/09/12. https://doi.org/10.1016/j.spinee.2017.08.229 PMID: 28890223.

31. Stenneberg MS, Busstra H, Eskes M, van Trijffel E, Cattrysse E, Scholten-Peeters GGM, et al. Concurrent validity of measuring active wrist pronation and supination range of motion with a smartphone. J Knee Surg. 2017; 30(6):577–84. Epub 2017/07/19. https://doi.org/10.1055/s-0036-1593877 PMID: 27894147.

32. Werner D, Willson J, Willy R, Barrios J. Validity, Reliability, and Normative Values for Clinically-Assessed Frontal Tibial Orientation as a Measure of Varus-Valgus Knee Alignment. Int J Arti Ther Train. 2017; 22(2):29–33. https://doi.org/10.1123/iatt.2016-0047 PMID: 21526208. Language: English. Entry Date: 20170307. Revision Date: 20170309. Publication Type: Article. Note: For CE see Supplement pages 78–80. Journal Subset: Allied Health.

33. Mejia-Hernandez K, Chang A, Eardley-Harris N, Jaarsma R, McLean JM. Smartphone applications for the evaluation of pathologic shoulder range of motion and shoulder scores—a comparative study. JSES Open Access. 2018; 2(1):109–14. https://doi.org/10.1016/j.jses.2017.10.001 PMID: 30675577

34. Santos C, Pauchard N, Guilloteau A. Reliability assessment of measuring active wrist pronation and supination range of motion with a smartphone. Hand Surg Rehabil. 2017; 36(5):338–45. Epub 2017/07/30. https://doi.org/10.1016/j.hansur.2017.06.005 PMID: 28754335.

35. Modest J, Clair B, DeMasi R, Meulenaere S, Howley A, Aubin M, et al. Self-measured wrist range of motion by wrist-injured and wrist-healthy study participants using a built-in iPhone feature as compared with a universal goniometer. J Hand Ther. 2018. Epub 2018/07/19. https://doi.org/10.1016/j.jht.2018.03.004 PMID: 30017418.

36. Mehta SP, Barker K, Bowman B, Galloway H, Oliashirazi N, Oliashirazi A. Reliability, Concurrent Validity, and Minimal Detectable Change for iPhone Goniometer App in Assessing Knee Range of Motion. J Knee Surg. 2017; 30(6):577–84. Epub 2017/05/31. https://doi.org/10.1055/s-0036-1593877 PMID: 27894147.

37. Pereira LC, Rwakabayiza S, Lecureux E, Jolles BM. Reliability of the Knee Smartphone-Application Goniometer in the Acute Orthopedic Setting. J Knee Surg. 2017; 30(3):223–30. https://doi.org/10.1055/s-0036-1584184 PMID: 27218479.

38. Bedekar N, Suryawanshi M, Rairikar S, Sanchet P, Shyam A. Inter and intra-rater reliability of mobile device goniometer in measuring lumbar flexion range of motion. J Back Musculoskeletal Rehabil. 2014; 27(2):161–6. Epub 2013/09/14. https://doi.org/10.3233/BMR-130431 PMID: 24029833.

39. Furness J, Schram B, Cox AJ, Anderson SL, Keogh J. Reliability and concurrent validity of the iPhone ((R)) Compass application to measure thoracic rotation range of motion (ROM) in healthy participants. PeerJ. 2018; 6:e4431. Epub 2018/03/24. https://doi.org/10.7717/peerj.4431 PMID: 29568701; PubMed Central PMCID: PMC5845564.

40. Grondin F, Hall T, von Piekartz H. Does altered mandibular position and dental occlusion influence upper cervical movement: A cross–sectional study in asymptomatic people. Musculoskeletal Science and Practice. 2017; 27:85–90. https://doi.org/10.1016/j.math.2016.06.007 PMID: 27847242.

41. Jung SH, Kwon OY, Jeon IC, Hwang UJ, Weon JH. Reliability and criterion validity of measurements using a smart phone-based measurement tool for the transverse rotation angle of the pelvis during single-leg lifting. Physiotherapy Theory Pract. 2018; 34(1):58–65. https://doi.org/10.1080/09593985.2017.1368756 PMID: 28922042.

42. Kolber MJ, Hanney WJ. The reliability and concurrent validity of shoulder mobility measurements using a digital inclinometer and goniometer: a technical report. Int J Sports Phys Ther. 2012; 7(3):306–13. Epub 2012/06/06. PMID: 22666445; PubMed Central PMCID: PMC3362980.

43. Pourahmadi MR, Taghipour M, Jannati E, Mohseni-Bandpei MA, Ebrahim Takamjani I, Rajabzadeh F. Reliability and validity of an iPhone((R)) application for the measurement of lumbar spine flexion and extension range of motion. PeerJ. 2016; 4:e2355. Epub 2016/09/17. https://doi.org/10.7717/peerj.2355 PMID: 27635328; PubMed Central PMCID: PMC45012335.

44. Quek J, Brauer SG, Treleaven J, Pua Y-H, Mentley B, Clark RA. Validity and intra-rater reliability of an Android phone application to measure cervical range-of-motion. J Neuroeng Rehabil. 2014; 11:65–. https://doi.org/10.1186/1743-0003-11-65 PMC4021613. PMID: 24742001
45. Tousignant-Laflamme Y, Boutin N, Dion AM, Vallée C-A. Reliability and criterion validity of two applications of the iPhone™ to measure cervical range of motion in healthy participants. J Neuroeng Rehabil. 2013; 10(1):69. https://doi.org/10.1186/1743-0003-10-69 PMID: 23829201

46. Ullucci PA Jr., Tudini F, Moran MF. Reliability of smartphone inclinometry to measure upper cervical range of motion. J Sport Rehabil. 2018;1–12. Epub 2018/07/25. https://doi.org/10.1123/jsr.2018-0048 PMID: 30040023.

47. Lim JY, Kim TH, Lee JS. Reliability of measuring the passive range of shoulder horizontal adduction using a smartphone in the supine versus the side-lying position. J Phys Ther Sci. 2015; 27(10):3119–22. Epub 2015/12/09. https://doi.org/10.1589/jpts.27.3119 PMID: 26644657; PubMed Central PMCID: PMCPMC4668148.

48. Mitchell K, Gutierrez SB, Sutton S, Morton S, Morgenthaler A. Reliability and validity of goniometric iPhone applications for the assessment of active shoulder external rotation. Physiother Theory Pract. 2014; 30(7):521–5. Epub 2014/03/29. https://doi.org/10.3109/09593985.2014.900593 PMID: 24654927.

49. Ramkumar PN, Haeberle HS, Navarro SM, Sultan AA, Mont MA, Ricchetti ET, et al. Mobile technology and telerehabilitation for shoulder range of motion: validation of a motion-based machine-learning software development kit. J Shoulder Elbow Surg. 2018; 27(7):1198–204. Epub 2018/03/12. https://doi.org/10.1016/j.jse.2018.01.013 PMID: 29052011.

50. Hambly K, Sibley R, Ockendon M. Level of agreement between a novel smartphone application and a long arm goniometer for the assessment of maximum active knee flexion by an inexperienced tester. Int J Physiother Phys Rehabil. 2012; 2:1–14.

51. Jones A, Sealey R, Crowe M, Gordon S. Concurrent validity and reliability of the Simple Goniometer iPhone app compared with the Universal Goniometer. Physiother Theory Pract. 2014; 30(7):512–6. Epub 2014/03/29. https://doi.org/10.3109/09593985.2014.900835 PMID: 24666408.

52. Ockendon M, Gilbert RE. Validation of a novel smartphone accelerometer-based knee goniometer. J Knee Surg. 2012; 25(4):341–5. Epub 2012/11/15. https://doi.org/10.1055/s-0031-1299669 PMID: 23150162.
62. Romero Morales C, Calvo Lobo C, Rodriguez Sanz D, Sanz Corbalan I, Ruiz Ruiz B, Lopez Lopez D. The concurrent validity and reliability of the Leg Motion system for measuring ankle dorsiflexion range of motion in older adults. PeerJ. 2017; 5:e2820. https://doi.org/10.7717/peerj.2820 PMID: 28070457.

63. Vohralik SL, Bowen AR, Burns J, Hiller CE, Nightingale EJ. Reliability and validity of a smartphone app to measure joint range. Am J Phys Med Rehabil. 2015; 94(4):325–30. Epub 2014/10/10. https://doi.org/10.1097/PHM.0000000000000221 PMID: 25299533.

64. Williams CM, Caserta AJ, Haines TP. The TiltMeter app is a novel and accurate measurement tool for the weight bearing lunge test. J Sci Med Sport. 2013; 16(5):392–5. https://doi.org/10.1016/j.jsams.2013.02.001 PMID: 23491138

65. Lendner N, Wells E, Lavi I, Kwok YY, Ho PC, Wollstein R. Utility of the iPhone 4 Gyroscope Application in the Measurement of Wrist Motion. Hand. 2017:1558944717730604. https://doi.org/10.1177/1558944717730604 PMID: 28918662.

66. Chia K, Sangeux M. Quantifying sources of variability in gait analysis. Gait Posture. 2017; 56:68–75. Epub 2017/05/16. https://doi.org/10.1016/j.gaitpost.2017.04.040 PMID: 28505546.

67. Mourcou Q, Fleury A, Franco C, Klocpic F, Vuillerme N. Performance Evaluation of Smartphone Inertial Sensors Measurement for Range of Motion. Sensors (Basel). 2015; 15(9):23168–87. Epub 2015/09/22. https://doi.org/10.3390/s150923168 PMID: 26389900; PubMed Central PMCID: PMCPMC4610531.

68. Brodie MA, Walmsley A, Page W. The static accuracy and calibration of inertial measurement units for 3D orientation. Comput Methods Biomech Biomed Engin. 2008; 11(6):641–8. https://doi.org/10.1080/10255840802326736 PMID: 18688763

69. Seel T, Raisch J, Schauer T. IMU-based joint angle measurement for gait analysis. Sensors (Basel). 2014; 14(4):6891–909. Epub 2014/04/20. https://doi.org/10.3390/s140406891 PMID: 24743160; PubMed Central PMCID: PMCPMC4029684.

70. Favre J, Jolles BM, Alissaou R, Aminian K. Ambulatory measurement of 3D knee joint angle. J Biomech. 2008; 41(5):1029–35. Epub 2008/01/29. https://doi.org/10.1016/j.jbiomech.2007.12.003 PMID: 18222459.

71. Luinge HJ, Veitink PH, Baten CTM. Ambulatory measurement of arm orientation. J Biomech. 2007; 40 (1):78–85. https://doi.org/10.1016/j.jbiomech.2005.11.011 PMID: 16455089

72. Vargas-Valencia LS, Elias A, Rocca E, Bastos-Filho T, Frizera A. An IMU-to-Body Alignment Method Applied to Human Gait Analysis. Sensors (Basel). 2016; 16(12). Epub 2016/12/16. https://doi.org/10.3390/s16122090 PMID: 27973406; PubMed Central PMCID: PMCPMC5191070.

73. Malone JJ, Lovell R, Varley MC, Coulls AJ. Unpacking the Black Box: Applications and Considerations for Using GPS Devices in Sport. Int J Sports Physiol Perform. 2017; 12(Suppl 2):S218–s26. Epub 2016/10/14. https://doi.org/10.1123/ijspp.2016-0236 PMID: 27736244.