Psychometric Properties of the iHandy Level Smartphone Application for Measuring Lumbar Spine Range of Motion and Lordosis: A Systematic Review of the Literature

Mohammadreza Pourahmadi, Hamid Hesarikia, Ali Ghanjal, and Alireza Shamsoddini

Context: Advent of smartphones has brought a wide range of clinical measurement applications (apps) within the reach of most clinicians. The vast majority of smartphones have numerous built-in sensors such as magnetometers, accelerometers, and gyroscopes that make the phone capable of measuring joint range of motion (ROM) and detecting joint positions. The iHandy Level app is a free app which has a visual display alike with the digital inclinometer in regard to numeric size. *Objective:* The purpose of this systematic review was to evaluate available evidence in the literature to assess the psychometric properties (ie, reliability and validity) of the iHandy Level app in measuring lumbar spine ROM and lordosis. *Methods:* PubMed/MEDLINE, Scopus, Ovid, Google Scholar, and ScienceDirect were searched from inception to September 2018 for single-group repeated-measures studies reporting outcomes of lumbar spine ROM or lordosis in adult individuals without symptoms of low back pain (LBP) or patients with LBP. The quality of each included study was assessed using the Quality Appraisal of Reliability Studies checklist. *Results:* A total of 4 studies with 273 participants were included. Two studies focused on measuring active lumbar spine ROM, and 2 studies evaluated lumbar spine lordosis. Three studies included asymptomatic subjects, and one study recruited patients with LBP. The results showed that the iHandy Level app has sufficient psychometric properties for measuring standing thoraco-lumbo-sacral flexion, extension, lateral flexion, isolated lumbar spine flexion ROM, and lumbar spine lordosis in asymptomatic subjects. One study reported poor concurrent validity with a bubble inclinometer ($r = .19–.53$), poor intrarater reliability (intraclass correlation coefficient = $.19–.39$), and poor to good interrater reliability (intraclass correlation coefficient = $.24–.72$) for the measurement of active lumbar spine ROM using the iHandy Level app in patients with LBP. *Conclusions:* This review provided a valuable summary of the research to date examining the psychometric properties of the iHandy Level app for measuring lumbar spine ROM and lordosis.

Keywords: iHandy Level application, lumbar vertebrae, physiological movements, lumbar curvature

Smartphones have become central to our lives and provide users direct and instant access to a wealth of electronic media (ie, the Internet, e-mail, and instant messaging) and numerous applications (apps). The number of smartphone users has been increased exponentially with over 1.91 billion users in 2015 and predicted to be approximately 6.1 billion users by 2020, which is almost 70% of the world population.[1] Using smartphone technology in medicine is becoming more prevalent, and its impact on physical therapy and medicine has already been significant.[2] The smartphones have given health care providers the capability to assess medical references, clinical measurement apps,[3,4] research tools and patient information,[5] medical consultation,[6] documentation,[7] and patient education.[8] The smartphones and other mobile devices are equipped with a mass of technology including cameras, touchscreen displays, wireless internet capability, a set of powerful embedded sensors (accelerometers, magnetometers, and gyroscopes), and audio/video media storage, along with a unique ability to run various downloadable apps.[3] These technologies enable the smartphones to detect joint position and measuring joint range of motion (ROM).[9] A continuously growing number of iOS or android apps for measuring joint ROM are currently available on the Internet which transform the smartphones into a medical device. Many of these apps are free, relatively inexpensive, or have a free trial version. The main purpose of clinical measurement apps is to provide easy, quick, precise, and convenient measurement of ROM.[4] Some apps directly investigate the joint angle, while others require the user to take a photo of the joint position prior to superimposing a virtual goniometer onto the image.[10]

The iHandy Level is a smartphone app, which was first developed in December 2008, and designed for measuring joint position and ROM. The last modified version of this app (version 1.70.3) was released in May 2017. This measurement app uses the smartphone's built-in accelerometer, gyroscope, and a digital display to show the measured angle. To date, the iHandy Level app has been often used in the literature to evaluate human joints flexibility and some papers have studied its psychometric properties (reliability and validity). This study aims to critically review the published literature in order to assess the psychometric properties of the iHandy Level app in measuring lumbar spine lordosis and ROM.

Methods

Search Strategy

This systematic review was planned and reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement.[11] The following databases

Pourahmadi is with the Department of Physiotherapy, School of Rehabilitation Sciences, Iran University of Medical Sciences, Tehran, Iran. Hesarikia is with the Department of Orthopedic Surgery, Baqiyatallah University of Medical Sciences, Tehran, Iran. Ghanjal is with Health Management Research Center, Life Style Institute, Baqiyatallah University of Medical Sciences, Tehran, Iran. Shamsoddini is with Exercise Physiology Research Center, Life Style Institute, Baqiyatallah University of Medical Sciences, Tehran, Iran. Shamsoddini (alirezaot@bmsu.ac.ir) is corresponding author.

were searched from inception to September 1, 2018 for relevant literature which included: MEDLINE (via PubMed; https://www.ncbi.nlm.nih.gov/pubmed/), Scopus (http://www.elsevier.com/solutions/scopus), Ovid (http://ovidsp.tx.ovid.com/sp-3.31.1b/ovidweb.cgi), Google Scholar (http://scholar.google.com/), and ScienceDirect (https://www.sciencedirect.com/search/advanced). Two reviewers (M.R.P. and A.R.S.) independently conducted the literature review. Both Medical Subject Heading terms searches and searches using free text-word were performed. Details of PubMed/Medline database search syntax are as follows: (((Smart AND Phone*) OR iHandy* OR iPhone OR Android) AND ((Lumbosacral AND Region*) OR (Lumbar AND Region*) OR Lumbar* OR "Lumbar Vertebrae")). Wildcards and truncations were used, and words were combined with Boolean operators such as "OR" and "AND" when appropriate. The search strategy was slightly modified for searches of other databases. Furthermore, citation tracking and reference lists scanning of the included studies and relevant reviews were searched for eligible studies.

Eligibility Criteria

At the completion of the search, the studies retrieved were imported into the EndNote software (version X8; Thomson Reuters, New York, NY) and duplicates were removed. The decision to include an article was primarily made based on title and abstract review, followed by full-text screening. Eligible studies included in this review were assessed by the 2 independent reviewers (M.R.P. and A.R.S.). The initial articles selected by each reviewer for inclusion were further evaluated by 2 additional reviewers (H.H. and A.G.), and the final studies to be included in this systematic review were confirmed. Studies were screened for selection according to the review objectives and participants, diagnosis, comparison, outcomes, and study design criteria[12]:

(1) Participants and diagnosis: Studies in which participants were adult individuals (≥18 y) without symptoms of low back pain (LBP) or patients with LBP. The LBP was defined as pain or discomfort on the posterior aspect of the trunk from the lower margin of the 12th ribs to the lower gluteal folds with or without pain referred into one or both lower limbs that lasts for at least 1 day.[13]

(2) Comparison: At least a golden standard instrument (eg, bubble inclinometer) should be considered as a comparator in the eligible studies in which the iHandy Level app validity was assessed.

(3) Outcomes: The primary outcomes of this systematic review were ROM including flexion, extension, lateral flexion, rotation, and lumbar spine lordosis.

(4) Study design: Single-group repeated-measures design (reliability and validity) studies[14] published in peer-reviewed journals with full text available in English; results obtained from theses/dissertations, conference proceedings, abstracts, and websites were excluded.

Quality Assessment

The quality of each eligible study was assessed by 2 reviewers (M.R.P. and A.R.S.) using the Quality Appraisal of Reliability Studies (QAREL) checklist.[15] This scoring system includes 11 items that investigate 7 principles: (1) the representativeness of participants and examiners, (2) examiner blinding, (3) order of examination, (4) suitability of the time interval, (5) applied and interpreted appropriately, and (6) statistical analysis.[16] Each item on QAREL can be answered with 3 categorical responses: "Yes," "No," and "Unclear." Yes response shows a good quality aspect of the study, whereas a No response shows a poor quality aspect.[15] The maximum score of this tool is 11.[15] A total score >60% was considered as the cutoff point for high quality.[16,17] When an item was not applicable to a study, its value was not taken into account in the total percentage. The level of interrater agreement was measured using Cohen kappa coefficient, with a method developed for comparing the level of agreement with categorical data along with their respective 95% confidence intervals ($\kappa = 0$–0.20, poor agreement; 0.21–0.40, fair agreement; 0.41–0.60, moderate agreement; 0.61–0.80, good agreement; and 0.81–1, very good agreement).[18] A third reviewer (H.H.) was available to resolve any disagreements that could not be resolved by consensus of the 2 reviewers.

Level of Evidence

The level of evidence for each primary outcome (ie, lumbar ROM and lordosis) was categorized as "strong" when consistent findings of at least one high-quality study was present and the total sample size of eligible combined studies was ≥100, "moderate" when consistent findings of at least one high-quality study was present and the total sample size ≥50, "limited" when findings of at least one high-quality study was present and the total sample size between 25 and 49, and "unknown" when findings were of indeterminate rating, in studies with poor methodological quality or with a sample of ≤25.[19]

Data Extraction

The following data were extracted from all the included articles by 2 reviewers independently (M.R.P. and A.R.S.): (1) population (healthy participants or patients with LBP), (2) number of participants included, (3) movement measured (flexion, extension, lateral flexion, rotation, or lumbar lordosis), (4) active or passive ROM, (5) if bony landmarks were used or defined prior to the measuring, (6) validity, (7) intrarater and interrater reliability and any other relevant detail, and (8) information about the examiner (profession and/or level of experience in goniometry).

Data Analysis

Data analysis was conducted by one author (M.R.P.) using Microsoft Excel (version 365; Microsoft Corp, Washington, DC). Intraclass correlation coefficient (ICC) <.40 was considered poor, .40 to .59, fair; .60 to 0.74, good; and >.75, excellent.[20] SEM was used to examine the precision of the instruments and was computed as pooled $SD \times \sqrt{1 - ICC}$.[21] Minimum detectable change at the 95% confidence level (MDC_{95}) was calculated as $\sqrt{2} \times 1.96 \times SEM$, which represents the magnitude of change required to provide confidence that a change is not the result of random variation or measurement error.[21] Furthermore, the value of Pearson correlation coefficient (r) was classified according to Hopkins extension of Cohen guidelines (.00–.09, nonexistent; .10–.29, small; .30–.49, medium; .50–.69, large; .70–.89, very large; .90–.99, nearly perfect; and 1.00, perfect).[22]

Results

Study Selection

The search resulted in 197 articles. A manual search of cross-references did not result in additional studies. After screening titles

and abstracts, 9 full-text articles were screened. Of these studies, 2 studies were conference proceedings[23,24] and 3 studies were review articles.[25–27] Therefore, 4 studies met our inclusion criteria and were included in this systematic review.[28–31] Figure 1 illustrates the preferred reporting items for systematic reviews and meta-analyses flowchart.

Quality Assessment

The QAREL checklist indicated a high quality (score >60%) in 3 out of 4 studies[29–31] and one study was of low quality.[28] The order of examination was not varied in Jayavel et al[28] and Kolber et al[29] studies, and it was unclear in Koumantakis et al[30] study. Moreover, it was unclear whether raters were blinded to the results of the reference standard in 3 studies.[29–31] The level of interrater agreement of quality assessment was moderate ($\kappa = 0.50$ [0.43]). In terms of quality assessment, a median QAREL score of 6.5 (interquartile range = 5–7) indicated a low methodological quality of the included studies. An overview of all the QAREL scores is presented in Table 1.

Level of Evidence

The level of evidence for this study primary outcomes is reported in Table 2.

Overview of Participant Characteristics

A total of 273 participants were included in the eligible studies. Three studies included asymptomatic subjects,[29–31] while one study recruited patients with LBP.[28] However, Jayavel et al[28] did not provide detailed information on the LBP subclassification. Three studies enrolled both female and male participants,[28–30] and

one study did not report the gender of participants.[31] The mean age of population of included studies at baseline ranged from 25.6 to 35.2 years. It seems that the participants in the study by Kolber et al[29] were the same as those in the study by Salamh and Kolber[31] because similar demographic characteristics were reported by both the studies. Finally, calculation of sample size was apparent in only one study.[31]

Methodology Considerations and Outcome Measures

Two of the 4 selected studies were conducted in the United States,[29,31] while the remaining studies were from Greece[30] and India.[28] In a study conducted by Kolber et al,[29] a standard plinth, new Baseline® bubble inclinometer (model 12-1056; Fabrication Enterprises; White Plains, NY) and the iHandy Level app installed on an iPhone® model 4 were used for measuring active lumbar ROM of 30 asymptomatic subjects in the standing position. Two raters each assessed the participants twice using the 2 instruments, and the second session was held after 24 to 48 hours. The mean thoraco-lumbo-sacral flexion, isolated lumbar spine flexion, thoraco-lumbo-sacral extension, and thoracolumbar right and left lateral flexion measured by the raters using the iHandy Level app was 107.5°, 49.5°, 27.5°, 31.5°, and 29.5°, respectively, and using the bubble inclinometer was 107°, 45.5°, 28°, 30.5°, and 29°, respectively.[29] To measure thoraco-lumbo-sacral flexion or extension, the instruments were placed at the level of T_{12}-L_1, and the participants were then asked to bend forward or backward as far as possible while maintaining knee extension as the angle was recorded.[29] For calculating thoraco-lumbo-sacral flexion or extension, the starting angle was recorded and subtracted from the final
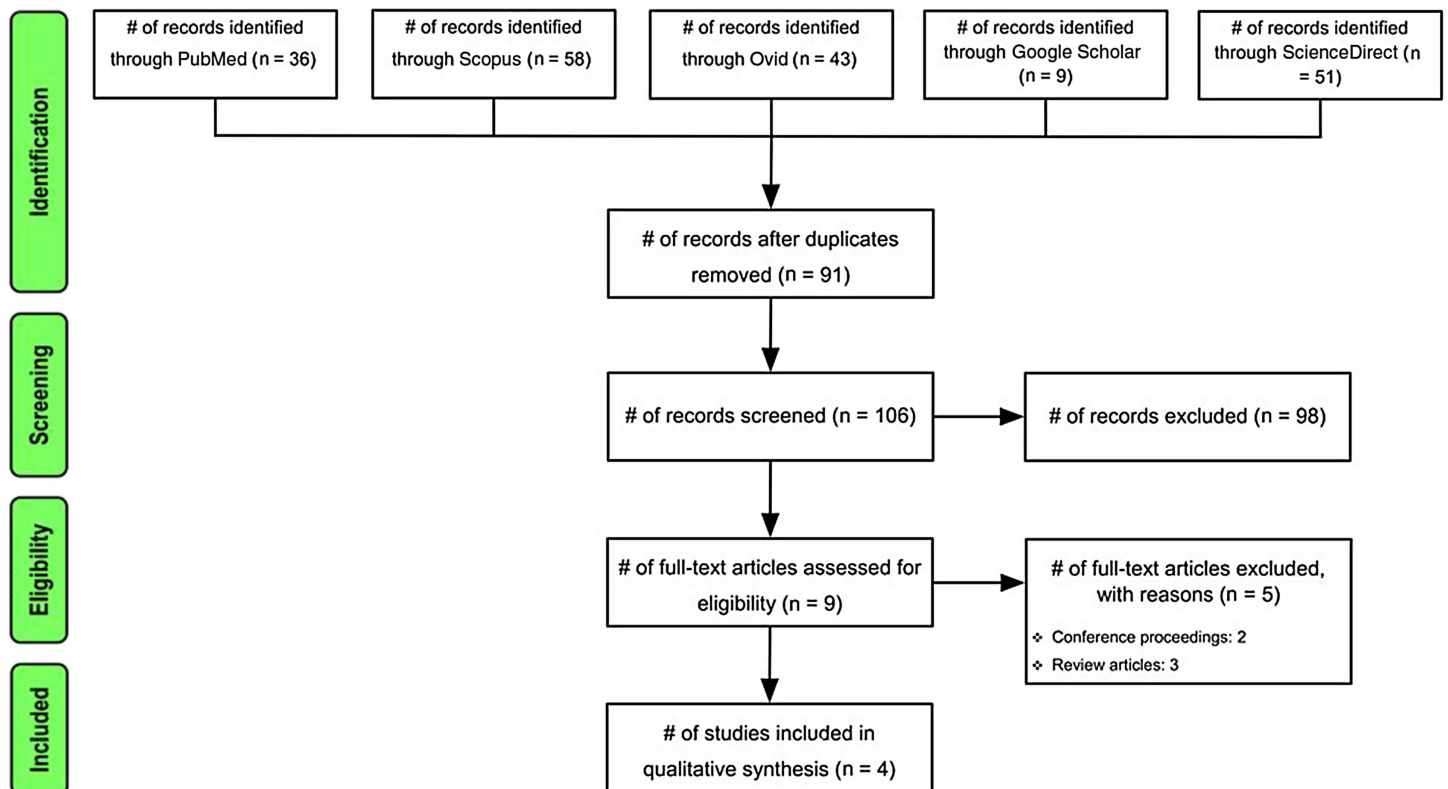

Figure 1 — Flow diagram of studies included in the systematic review.

Table 1   QAREL Scores for Each Included Study

	Study			
QAREL items	Jayavel et al[28]	Kolber et al[29]	Koumantakis et al[30]	Salamh and Kolber[31]
1. Was the test evaluated in a sample of subjects who were representative of those whom the authors intended the results to be applied?	1	1	1	1
2. Was the test performed by raters who were representative of those to whom the authors intended the results to be applied?	0	1	0	1
3. Were raters blinded to the findings of other raters during the study	1	1	1	1
4. Were raters blinded to their own prior findings of the test under evaluation?	1	1	1	1
5. Were raters blinded to the results of the accepted reference standard or disease status for the targeted disorder (or variable) being evaluated?	0	Unclear	Unclear	Unclear
6. Were raters blinded to clinical information that was not intended to be provided as part of the testing procedure or study design?	N/A	N/A	N/A	N/A
7. Were raters blinded to additional cues that were not part of the test?	N/A	N/A	N/A	N/A
8. Was the order of examination varied?	0	0	N/A	0
9. Was the stability of the variable being measured taken into account when determining the suitability of the time interval between repeated measures?	1	1	1	1
10. Was the test applied correctly and interpreted appropriately?	Unclear	1	1	1
11. Were appropriate statistical measures of agreement used?	0	1	1	1
Total	4	7	6	7
Score, %	44 (low quality)	78 (high quality)	75 (high quality)	78 (high quality)

Abbreviation: N/A, not applicable.

Table 2   Level of Evidence for Primary Outcomes

Primary outcome	Study	Level of evidence
Lumbar spine ROM	Kolber et al[29]—high quality	Moderate
	Jayavel et al[28]—low quality	
Lumbar spine lordosis	Salamh and Kolber[31]—high quality	Strong
	Koumantakis et al[30]—high quality	

angle at the end of the motion. Isolated flexion involved 2 measurements, the first at T_{12}-L_1 and the second at S_1-S_2.[29] To calculate isolated flexion, the measurement from S_1-S_2 was subtracted from T_{12}-L_1. Finally, thoracolumbar lateral flexion was evaluated with the participant standing with feet shoulder width apart and arms resting at their side. Then, the instruments were placed at T_9-T_{12} prior to initiating the motion. The participants were requested to slide their hand down the side of the leg as far as possible while maintaining trunk and head facing forward and both feet on the ground. To calculate thoracolumbar lateral flexion, the starting angle was recorded and subtracted from the final angle at the end of the motion.[29] The results of Kolber et al[29] study showed that there were no significant differences between the 2 instruments for measuring active ROM. ICC models (3,k) and (2,k) were used to determine the intrarater and interrater reliability, respectively. Their findings revealed that both the bubble inclinometer and iHandy Level app possess excellent intrarater reliability (ICC = .83–.96 and .80–.97, respectively) and interrater reliability (ICC = .84–.97 and .81–.98, respectively) as well as concurrent validity (ICC = .86–.98) (Tables 3 and 4).[29]

In another study, Jayavel et al[28] assessed the reliability and validity of the iHandy android app installed on a Samsung smartphone for measuring active lumbar ROM in 30 patients with LBP. A bubble inclinometer was used as a reference standard in their

study.[28] The testing procedure of this study was the same as the testing procedure used in the Kolber et al[29] study. The mean thoraco-lumbo-sacral flexion, isolated lumbar spine flexion, thoraco-lumbo-sacral extension, and thoracolumbar right and left lateral flexion measured by the raters using the iHandy Level app was 80.1°, 16.8°, 25.87°, 30.9°, and 33.6°, respectively, and using the bubble inclinometer was 78.2°, 24.8°, 23.7°, 26.1°, and 25.5°, respectively.[28] In addition, they reported poor intrarater reliability (ICC = .19–.39) and poor to good interrater reliability (ICC = .24–.72) for the measurement of active lumbar spine ROM using the iHandy Level app (Tables 3 and 4).[28] They did not provide information on ICC models used for assessing intrarater and interrater reliability.[28] Moreover, validity analysis indicated that there were poor relationships between the results of the Samsung app and bubble inclinometer.[28]

Two studies evaluated the standing lumbar spine curve sagittal alignment in asymptomatic participants. Salamh and Kolber[31] used a gravity-based bubble inclinometer and the iHandy Level app installed on an iPhone® model 4 for measuring lumbar spine lordosis of 30 participants. To measure lumbar spine lordosis, each instrument was placed at the landmarks T_{12}-L_1 and S_1-S_2, the angles recorded and added together.[31] The mean lumbar spine lordosis measured by 2 raters using the iHandy Level app and bubble inclinometer was 32.5° and 31.5°, respectively. ICC models

Table 3 Study Characteristics

Study	Population			Measurements						Intrarater		Interrater			Validity
	N	Healthy/symptomatic	Age, mean (SD), y	Active/passive	Flexion	Extension	Lateral flexion	Rotation	Lordosis	Trials (n)	Interval	Raters (n)	Expert/nonexpert	Interval, h	Golden standard
Kolber et al[29]	30 (12 male and 18 female)	Healthy	25.6 (2.1)	Active	✓ (TLPF, ILF)	✓ (TLSE)	✓ (TLRLF, TLLLF)	–	–	2	Same day	2	Expert	24–48	Bubble inclinometer
Salamh and Kolber[31]	30	Healthy	25.6 (2.1)	–	–	–	–	–	–	2	Same day	2	Expert	24–48	Bubble inclinometer
Koumantakis et al[30]	183 (83 male and 100 female)	Healthy	26.1 (10)	–	–	–	–	–	✓	3	Consecutively	–	–	–	–
Jayavel et al[28]	30	LBP	35.2 (NR)	Active	✓ (TLSF, ILF)	✓ (TLSE)	✓ (TLRLF, TLLLF)	–	–	2	Same day	2	NR	24–48	Bubble inclinometer

Abbreviations: ILF, isolated lumbar flexion; LBP, low back pain; NR, not reported; TLSE, thoracolumbo-sacral extension; TLPF, thoraco-lumbo-sacral flexion; TLLLF, thoracolumbar left lateral flexion; TLRLF, thoracolumbar right lateral flexion.

(3,k) and (2,k) were used to examine intrarater and interrater reliability. Excellent interrater and intrarater reliability was present for the bubble inclinometer with ICC of .90 and .85, respectively, and the iHandy Level app with ICC values of .96 and .81 (Tables 3 and 4).[31] The concurrent validity between the 2 instruments was very large with a *r* value of .86 for both raters (Tables 3 and 4).[31]

Koumantakis et al[30] investigated the intrarater reliability of sagittal lumbar spine lordosis in quiet standing and the validity of the method in differentiating between male and female subjects. The method used to calculate lumbar lordosis was the same that defined in Salamh and Kolber study.[31] Measurements were performed on the same day and time with the markings kept on participants' skin. Mean total lumbar spine lordosis, lumbar lordosis in males and females were 31.6°, 28.4°, and 34.2°, respectively.[30] Female participants generally demonstrated greater lumbar spine lordosis (mean difference = 5.8°; $P < .001$) compared with male participants.[30] The reliability of measurements was determined by the ICC model (2,1) absolute agreement for the intrarater component of analysis. Intrarater reliability was high for lumbar curve sagittal alignment (ICC = .96) (Tables 3 and 4).[30] Table 3 summarizes the included studies characteristics, and Table 4 represents the intrarater and interrater reliability and validity results of the selected studies.

Discussion

This systematic review qualitatively investigated the psychometric properties of the iHandy Level app for measuring lumbar spine lordosis and ROM. The results indicated that the iHandy Level app has sufficient reliability validity for measuring active lumbar ROM as well as lumbar spine lordosis in asymptomatic (healthy) subjects. However, the psychometric properties of the app were not acceptable for measuring active lumbar ROM in patients with LBP. Although we aimed to perform a formal meta-analysis, this was not feasible because the number of studies for each outcome was too small to obtain stable estimates.

The results of the Kolber et al[29] study indicated that there were no significant differences between the iHandy Level app and bubble inclinometer for measuring active thoraco-lumbo-sacral flexion, extension, lateral flexion, and isolated lumbar spine flexion ROM. However, isolated lumbar spine extension and lumbar spine rotation were not assessed in the present study. One possible reason could be due to the difficulty in finding a proper position to evaluate lumbar spine rotation very well. Excellent intrarater and interrater reliability of Kolber et al[29] study may be attributed to several factors such as accurate anatomical landmark identification, accurate and firm placement of the 2 instruments and controlling their slippage on participants' skin during movement, calibration of the instruments, 2 repetitions of each measurement, and the familiarity of the participants with the testing procedures.[29] There is evidence that increased familiarity could improve the reliability.[32] The results of the Jayavel et al[28] study were not consistent with the results of the Kolber et al[29] study. They recruited patients with LBP; therefore, pain may be a confounder which could adversely contribute to poor reliability of the measurements. In addition, Jayavel et al[28] mentioned that poor to moderate reliability may be because of skin movement and slippage of the instruments during extreme trunk motions especially lateral flexion and extension.

Koumantakis et al[30] also demonstrated that the iHandy Level app is a reliable and valid tool in the measurement of lumbar spine posture in the sagittal plane during a quiet standing position. They have improved the reliability of the measurement by keeping the

Table 4   Intrarater and Interrater Reliability and Validity Results of Included Studies *(continued)*

Psychometric property	Kolber et al[29]	Salamh and Kolber[31]	Koumantakis et al[30]	Jayavel et al[28]
Intrarater reliability				
TLSF				
ICC	.97 (.93–.98)	–	–	.39 (NR)
SEM	3.12°	–	–	?
MDC$_{95}$	8.65°	–	–	?
ILF				
ICC	.88 (.75–.94)	–	–	.39 (NR)
SEM	3.81°	–	–	?
MDC$_{95}$	10.56°	–	–	?
TLSE				
ICC	.80 (.58–.90)	–	–	.39 (NR)
SEM	3.13°	–	–	?
MDC$_{95}$	8.68°	–	–	?
TLRLF				
ICC	.82 (.61–.91)	–	–	.19 (NR)
SEM	2.97°	–	–	?
MDC$_{95}$	8.23°	–	–	?
TLLLF				
ICC	.84 (.67–.92)	–	–	.30 (NR)
SEM	2.40°	–	–	?
MDC$_{95}$	6.65°	–	–	?
Lordosis				
ICC	–	.81 (.61–.91)	.96 (.94–.97)	–
SEM	–	3.49°	2.13°	–
MDC$_{95}$	–	9.67°	5.90°	–
Interrater reliability				
TLSF				
ICC	.98 (.95–.99)	–	–	.73
SEM	2.57°	–	–	4.94°
MDC$_{95}$	7.12°	–	–	13.69°
ILF				
ICC	.88 (.76–.95)	–	–	.39
SEM	3.43°	–	–	10.08°
MDC$_{95}$	9.51°	–	–	27.94°
TLSE				
ICC	.81 (.60–.91)	–	–	.55
SEM	3.86°	–	–	8.65°
MDC$_{95}$	10.70°	–	–	23.98°
TLRLF				
ICC	.93 (.86–.97)	–	–	.25
SEM	1.71°	–	–	5.89°
MDC$_{95}$	4.74°	–	–	16.33°
TLLLF				
ICC	.90 (.77–.96)	–	–	.59
SEM	1.71°	–	–	6.59°
MDC$_{95}$	4.74°	–	–	18.27°
Lordosis				
ICC	–	.96 (.92–.98)	–	–
SEM	–	1.50°	–	–
MDC$_{95}$	–	4.16°	–	–
Validity				
TLSF				
r	–	–	–	–
ICC	.98 (.97–.99)	–	–	.53
ILF				
r	–	–	–	–
ICC	.87 (.45–.95)	–	–	.19
TLSE				
r	–	–	–	–
ICC	.91 (.81–.96)	–	–	.42
TLRLF				
r	–	–	–	–
ICC	.94 (.80–.98)	–	–	.29
TLLLF				
r	–	–	–	–
ICC	.91 (.82–.96)	–	–	.42
Lordosis				
r	–	.86	–	–
ICC	–	–	–	–

Abbreviations: ? indicates not possible to calculate SEM; ICC, intraclass correlation coefficient; ILF, isolated lumbar flexion; MDC$_{95}$, minimum detectable change at the 95% confidence interval; NR, not reported; TLSE, thoraco-lumbo-sacral extension; TLSF, thoraco-lumbo-sacral flexion; TLLLF, thoracolumbar left lateral flexion; TLRLF, thoracolumbar right lateral flexion.

(continued)

markings on participants' skin.[30] In addition, Koumantakis et al[30] indicated that the app has a good ability of measuring lumbar spine lordosis differences in standing between male and female participants.

From this review, a moderate level of evidence existed for psychometric properties of the app for measuring lumbar spine ROM. Several important factors can explain this moderate level of evidence. First, none of the included studies[28,29] performed a power analysis to estimate the number of participants needed to prevent type II statistical errors. Therefore, the generalizability of their results is limited due to low external validity. Second, in the Jayavel et al[28] study, it was unclear whether the raters were expert in measuring lumbar spine lordosis and ROM using the app and bubble inclinometer. This may adversely affect the results of the present study.[28] Third, the order of examination was not varied in the selected studies.[28, 29] According to Lucas et al,[15] in studies of inter-rater reliability, it may be important for the order of raters to be varied. By varying the order in which raters assess subjects, the potential for systematic error is reduced.[15] Fourth, a language bias is possible as only those studies that were available as full text in English were included. Finally, because of a low number of published studies using the iHandy Level app, meta-analysis was not performed, so only a qualitative synthesis of results is presented in this review.

Conclusions

In summary, of the 4 studies included in our systematic review, 3 studies were considered "high-quality studies," whereas 1 study was considered a "low-quality study." Because the number of included studies was limited, therefore, a definitive conclusion cannot be drawn from the results of those studies. However, the results of this review indicated that the iHandy Level app possesses adequate psychometric properties for measuring standing thoraco-lumbo-sacral flexion, extension, lateral flexion, isolated lumbar spine flexion ROM, and lumbar spine lordosis in asymptomatic (healthy) subjects. Further high-quality studies with low risk of bias are warranted to evaluate the psychometric properties of the iHandy Level app for measuring isolated active lumbar spine extension, lateral flexion, and rotation in asymptomatic individuals and patients with LBP. The results of this review suggest that the iHandy Level app must be used with caution in clinical settings for evaluating patients with LBP.

Acknowledgments

References

1. Tobing RDH. Mobile tourism application for Samosir Regency on Android platform. Paper presented at: International Symposium on Technology Management and Emerging Technologies (ISTMET); August 25–27, 2015. Langkawai Island, Malaysi.

2. van Osch M, Rövekamp A, Bergman-Agteres SN, et al. User preferences and usability of iVitality: optimizing an innovative online research platform for home-based health monitoring. *Patient Prefer Adherence*. 2015;9:857. doi:10.2147/PPA.S82510

3. Pourahmadi MR, Ebrahimi Takamjani I, Sarrafzadeh J, et al. Reliability and concurrent validity of a new iPhone® goniometric application for measuring active wrist range of motion: a cross-sectional study in asymptomatic subjects. *J Anat*. 2017;230(3):484–495. PubMed ID: 27910103 doi:10.1111/joa.12568

4. Pourahmadi MR, Taghipour M, Jannati E, Mohseni-Bandpei MA, Ebrahimi Takamjani I, Rajabzadeh F. Reliability and validity of an iPhone® application for the measurement of lumbar spine flexion and extension range of motion. *PeerJ*. 2016;4:e355.

5. Baumgart DC. Smartphones in clinical practice, medical education, and research. *Arch Intern Med*. 2011;171(14):1294–1296. PubMed ID: 21788549 doi:10.1001/archinternmed.2011.320

6. Divall P, Camosso-Stefinovic J, Baker R. The use of personal digital assistants in clinical decision making by health care professionals: a systematic review. *Health Informatics J*. 2013;19(1):16–28. PubMed ID: 23486823 doi:10.1177/1460458212446761

7. Yeung MS, Lapinsky SE, Granton JT, Doran DM, Cafazzo JA. Examining nursing vital signs documentation workflow: barriers and opportunities in general internal medicine units. *J Clin Nurs*. 2012;21(7–8):975–982. doi:10.1111/j.1365-2702.2011.03937.x

8. Shepherd JD, Badger-Brown KM, Legassic MS, Walia S, Wolfe DL. SCI-U: e-learning for patient education in spinal cord injury rehabilitation. *J Spinal Cord Med*. 2012;35(5):319–329. PubMed ID: 23031169 doi:10.1179/2045772312Y.0000000044

9. Pourahmadi MR, Bagheri R, Taghipour M, Ebrahimi Takamjani I, Sarrafzadeh J, Mohseni-Bandpei MA. A new iPhone application for measuring active craniocervical range of motion in patients with non-specific neck pain: a reliability and validity study. *Spine J*. 2018;18(3):447–457. PubMed ID: 28890223 doi:10.1016/j.spinee.2017.08.229

10. Wellmon RH, Gulick DT, Paterson ML, Gulick CN. Validity and reliability of 2 goniometric mobile apps: device, application, and examiner factors. *J Sport Rehabil*. 2016;25(4):371–379. PubMed ID: 27632853 doi:10.1123/jsr.2015-0041

11. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Ann Intern Med*. 2009;151(4):264–269. PubMed ID: 19622511 doi:10.7326/0003-4819-151-4-200908180-00135

12. Pourahmadi MR, Ebrahimi Takamjani I, Jaberzadeh S, et al. Kinematics of the spine during sit-to-stand movement using motion analysis systems: a systematic review of literature. *J Sport Rehabil*. 2019;28(1):77–93. doi:10.1123/jsr.2017-0147

13. Hoy D, March L, Brooks P, et al. The global burden of low back pain: estimates from the Global Burden of Disease 2010 study. *Ann Rheum Dis*. 2014;73(6):968–974. PubMed ID: 24665116 doi:10.1136/annrheumdis-2013-204428

14. Linek P, Klepek A, Wolny T, Mikołajowski G. Reliability of the lateral abdominal muscle thickness measurements in idiopathic scoliosis patients. *Musculoskelet Sci Pract*. 2018;38:151–154. PubMed ID: 29776776 doi:10.1016/j.msksp.2018.05.001

15. Lucas NP, Macaskill P, Irwig L, Bogduk N. The development of a quality appraisal tool for studies of diagnostic reliability (QAREL). *J Clin Epidemiol*. 2010;63(8):854–861. PubMed ID: 20056381 doi:10.1016/j.jclinepi.2009.10.002

16. van Rijn SF, Zwerus EL, Koenraadt KL, Jacobs WC, van den Bekerom MP, Eygendaal D. The reliability and validity of goniometric elbow measurements in adults: a systematic review of the literature. *Shoulder Elbow*. 2018;10(4):274–284. PubMed ID: 30214494 doi:10.1177/1758573218774326

17. Barrett E, McCreesh K, Lewis J. Reliability and validity of non-radiographic methods of thoracic kyphosis measurement: a systematic review. *Man Ther*. 2014;19(1):10–17. PubMed ID: 24246907 doi:10.1016/j.math.2013.09.003

18. Pourahmadi MR, Mohsenifar H, Dariush M, Aftabi A, Amiri A. Effectiveness of mobilization with movement (Mulligan concept techniques) on low back pain: a systematic review. *Clin Rehabil*. 2018;32(10):1289–1298. PubMed ID: 29843520 doi:10.1177/0269215518778321

19. Kroman SL, Roos EM, Bennell KL, Hinman RS, Dobson F. Measurement properties of performance-based outcome measures to assess physical function in young and middle-aged people known to be at high risk of hip and/or knee osteoarthritis: a systematic review. *Osteoarthritis Cartilage*. 2014;22(1):26–39. doi:10.1016/j.joca.2013.10.021

20. Fleiss JL, Levin B, Paik MC. *Statistical Methods for Rates and Proportions*. 3rd ed. Hoboken, NJ: John Wiley & Sons; 2003.

21. Pourahmadi MR, Ebrahimi Takamjani I, Jaberzadeh S, et al. Test–retest reliability of sit-to-stand and stand-to-sit analysis in people with and without chronic non-specific low back pain. *Musculoskelet Sci Pract*. 2018;35:95–104. doi:10.1016/j.msksp.2017.11.001

22. Hopkins WG. Research designs: choosing and fine-tuning a design for your study. *Sportscience*. 2008;12(1):1–3.

23. Bruyneel AV. Smartphone applications for range of motion measurement in clinical practice: a literature review of inclinometer and goniometric tools. *Ann Phys Rehabil Med*. 2018;61:e454. doi:10.1016/j.rehab.2018.05.1059

24. Mason J. Using a smartphone to measure joint range—attractive for clinical use, but are they valid and reliable? *Int J Osteopath Med*. 2015;18(2):151–152. doi:10.1016/j.ijosm.2015.04.008

25. 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): 1038–1043. doi:10.1016/j.pmrj.2014.05.003

26. Mosa ASM, Yoo I, Sheets L. A systematic review of healthcare applications for smartphones. *BMC Med Inform Decis Mak*. 2012; 12(1):67. doi:10.1186/1472-6947-12-67

27. Mourcou Q, Fleury A, Diot B, Franco C, Vuillerme N. Mobile phone-based joint angle measurement for functional assessment and rehabilitation of proprioception. *Biomed Res Int*. 2015;2015:328142.

28. Jayavel A, Misra P, Sivakumar V. Reliability and validity of I Handy android application on measurement of lumbar spine movement in patients with low back pain. *Int J Clin Skills*. 2017;11(3):1–3. doi:10. 4172/Clinical-Skills.1000118

29. Kolber MJ, Pizzini M, Robinson A, Yanez D, Hanney WJ. The reliability and concurrent validity of measurements used to quantify lumbar spine mobility: an analysis of an iPhone® application and gravity based inclinometry. *Int J Sports Phys Ther*. 2013;8(2):129.

30. Koumantakis GA, Nikoloudaki M, Thacheth S, et al. Reliability and validity measurement of sagittal lumbosacral quiet standing posture with a smartphone application in a mixed population of 183 college students and personnel. *Adv Orthop*. 2016;2016:3817270. doi:10. 1155/2016/3817270

31. Salamh PA, Kolber M. The reliability, minimal detectable change and concurrent validity of a gravity-based bubble inclinometer and iPhone application for measuring standing lumbar lordosis. *Physiother Theory Pract*. 2014;30(1):62–67. doi:10.3109/09593985. 2013.800174

32. Dankaerts W, O'sullivan P, Straker L, Burnett A, Skouen J. The inter-examiner reliability of a classification method for non-specific chronic low back pain patients with motor control impairment. *Man Ther*. 2006;11(1):28–39. doi:10.1016/j.math.2005.02.001