RESEARCH/Original Article

Journal of Telemedicine and Telecare
0(0) 1–13

DOI: 10.1177/1357633X16642369
jtt.sagepub.com

Validity and reliability of Internet-based physiotherapy assessment for musculoskeletal disorders: A systematic review

Suresh Mani[1], Shobha Sharma[2], Baharudin Omar[3], Aatit Paungmali[4] and Leonard Joseph[1]

Abstract

Purpose: The purpose of this review is to systematically explore and summarise the validity and reliability of telerehabilitation (TR)-based physiotherapy assessment for musculoskeletal disorders.

Method: A comprehensive systematic literature review was conducted using a number of electronic databases: PubMed, EMBASE, PsycINFO, Cochrane Library and CINAHL, published between January 2000 and May 2015. The studies examined the validity, inter- and intra-rater reliabilities of TR-based physiotherapy assessment for musculoskeletal conditions were included. Two independent reviewers used the Quality Appraisal Tool for studies of diagnostic Reliability (QAREL) and the Quality Assessment of Diagnostic Accuracy Studies (QUADAS) tool to assess the methodological quality of reliability and validity studies respectively.

Results: A total of 898 hits were achieved, of which 11 articles based on inclusion criteria were reviewed. Nine studies explored the concurrent validity, inter- and intra-rater reliabilities, while two studies examined only the concurrent validity. Reviewed studies were moderate to good in methodological quality. The physiotherapy assessments such as pain, swelling, range of motion, muscle strength, balance, gait and functional assessment demonstrated good concurrent validity. However, the reported concurrent validity of lumbar spine posture, special orthopaedic tests, neurodynamic tests and scar assessments ranged from low to moderate.

Conclusion: TR-based physiotherapy assessment was technically feasible with overall good concurrent validity and excellent reliability, except for lumbar spine posture, orthopaedic special tests, neurodynamic testa and scar assessment.

Keywords

Telerehabilitation, musculoskeletal disorders, physiotherapy, assessment, validity, reliability

Date received: 17 November 2015; Date accepted: 7 March 2016

Background

Musculoskeletal disorders (MSDs) are the second greatest cause of chronic pain and physical disabilities in the world.[1–5] The resultant disabilities have a significant impact on the individual and family, and this subsequently increases the expenditures on healthcare and social resources.[6,7] The incidence of MSDs is noted to be higher in rural populations due to the lack of available healthcare and rehabilitation services.[8,9] Telerehabilitation (TR) may be a potential solution to deliver remote rehabilitation services using information and communication technology to address this rural healthcare disparity.[10] Recent advancement in computer software, biosensors and communication technology has allowed clinicians to administer various TR-based applications in the field of health sciences. TR-based services have been shown to be feasible in evaluating and treating MSDs and, as such, are emerging as adjuvant in current physiotherapy practice.[11–14]

An objective physiotherapy assessment is a key component in diagnosing and treating MSDs. However, the primary challenge in integrating TR-based physiotherapy services in regular clinical practice, is to perform objective physiotherapy assessments which are valid and reliable.[15] Validity and reliability are two essential components for

[1]Physiotherapy Program, Universiti Kebangsaan Malaysia, Malaysia
[2]Speech Sciences Program, Universiti Kebangsaan Malaysia, Malaysia
[3]Department of Biomedical Science, Universiti Kebangsaan Malaysia, Malaysia
[4]School of Physical Therapy, Chiang Mai University, Thailand

Corresponding author:
Shobha Sharma, School of Rehabilitation Sciences, Faculty of Health Sciences, Universiti Kebangsaan Malaysia, Jalan Raja Muda Abdul Aziz, 50300 Kuala Lumpur, Malaysia.
Email: shobha.ukm@gmail.com

clinicians to determine baseline data, monitor progress and guide appropriate implementation of intervention.[16]

Validity and reliability of physical tests are the primary concerns in diagnostic research. Poor validity and reliability would have negative influences on the test results that could be misleading in determining the effectiveness of TR-based intervention outcomes.[17] Although, validity has been recommended as an important component in any diagnostic test, reliability also plays an essential role. Hence, evidence on validity and reliability of the TR-mediated physical assessment should be established prior to widescale adoption in regular physiotherapy practice. In addition, physiotherapy assessment results are the vital components that determine the therapeutic decisions made by the physiotherapist. These measures of validity and reliability are the crucial factor in assessment and the diagnostic-treatment process.[18]

The validity and reliability of TR-based physiotherapy assessment against conventional face-to-face (FTF) assessment were explored in the MSD population, such as in low back pain, ankle disorders, elbow disorders, total knee replacement (TKR) and non-articular lower limb disorders.[19–26] However, the reviewed literature has technological difficulties, methodological differences and analytical variations in the physiotherapy assessment procedures and reporting methods. Therefore, the purpose of the research was to systematically review the literature: (a) regarding the validity and (b) the intra- and inter-rater reliability of TR-based physiotherapy assessment for MSDs.

Methods

Search strategy

TR literature was searched using the following electronic databases, PubMed, EMBASE, PsycINFO, Cochrane Library and CINAHL. The search was conducted with the following primary key search terms: 'telerehabilitation OR tele-rehabilitation OR teleassessment'. The search was further carried out with the individual search terms (joint pain, musculoskeletal, physical examination, physical assessment, physical test, manual examination, shoulder, elbow, wrist, neck pain, back pain, low back pain, hip, knee, ankle, foot, reliability, validity, reproducibility, inter tester, inter examiner, intra tester, intra examiner and test retest) along with primary search terms with the combination of Boolean logic (AND). The same search terms were used in each electronic database. In addition, the literature was searched manually from the reference list of the articles found in the electronic search.

Inclusion and exclusion criteria

Literature published in English that measured the validity and/or reliability (inter-rater/and/or intra-rater) of TR-based physiotherapy assessment for MSDs from January 2000–May 2015 were included. Unpublished manuscripts, letters, guidelines, conference proceedings, theses, and other descriptive publications, and published literature on TR-based physiotherapy assessment on neurological and other chronic conditions such as stroke, multiple sclerosis, Parkinson's disease, speech disorders, swallowing impairment and lymphoedema were excluded.

Selection method

Two reviewers were involved in the selection and screening process of accessible literature (SM and LJ). Based on the set criteria, the title and abstract of articles were systematically screened and categorised into two groups (to include or exclude). Each reviewer worked independently and was blinded to the decision of the other reviewer. The two groups of literature were then counterchecked by the reviewers in consultation with a third independent reviewer to avoid potential disagreement (SS).

Methodological quality checking

The Quality Appraisal Tool for studies of diagnostic Reliability (QAREL) was used to assess the methodological quality of the reliability studies. QAREL[27] is a reliable instrument which comprises of 11 items that measures sampling bias, representativeness of participants and raters, rater's blinding, order of examination, time interval between repeated measures, application and interpretation of tests, and appropriateness of statistical analysis. The methodological qualities of the included studies were independently rated by two reviewers. Prior to using QAREL, the reviewers formed a consensus concerning the application and interpretation of each of the items in order to improve the reliability.[28] The study quality was classified based on QAREL scores: 67% or more indicating high quality; 50–66% indicating moderate quality, and less than 50% indicating low quality.[29] In addition, the Quality Assessment of Diagnostic Accuracy Studies (QUADAS) tool was used to assess the methodological quality of validity studies. The QUADAS[30] tool consists of 14 items scored as 'yes', 'no' or 'unclear', based on the QUADAS guidelines. A validity study is considered high quality if the QUADAS score is $\geqslant 60\%$.[18,31–33]

Statistical checklist criteria

The selected literature was further explored, based on the statistical tests used to analyse the validity and reliability of the TR-based physiotherapy assessment. The Guidelines for Reporting Reliability and Agreement Studies[34] (GRRAS) recommend appropriate statistical methods for analysing intra-/inter-rater reliability and validity studies based on the levels of measurements such as nominal, ordinal and continuous data. The kappa and weighted kappa were recommended for nominal and ordinal data respectively, and the interclass correlation coefficient (ICC) was recommended for continuous data.[34] For validity measures, a percentage of agreement (PA) and

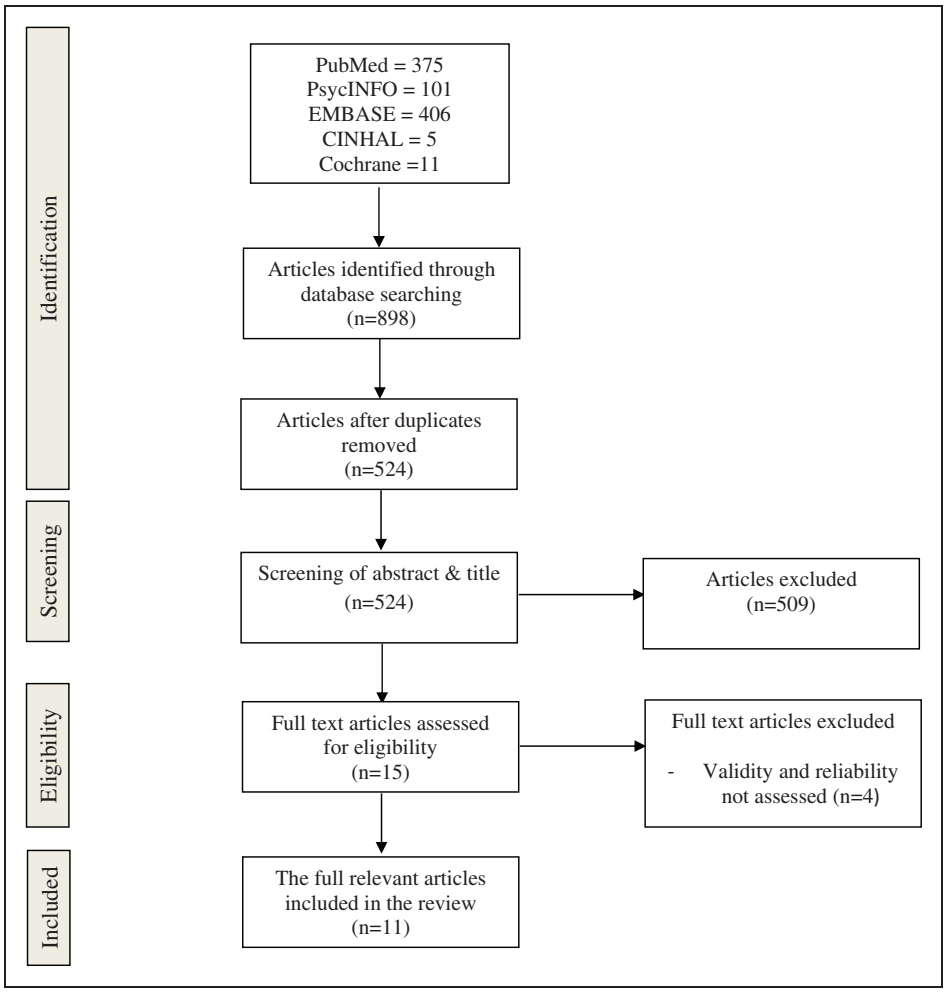

Figure 1. Flowchart of the articles reviewed.

percentage of exact agreement (PEA) were recommended for all three levels of measurement. Similarly, the standard error measurement (SEM), coefficient of variation (CV) and Bland-Altman plot were recommended to measure the limit of agreements between the two methods of assessment or raters for continuous data.[35] However, employing any one of these statistical methods alone is not sufficient to estimate the reliability and validity.[34]

Results

Study identification

A total number of 898 articles were identified through electronic searching, 375 in PubMed, 101 in PsycINFO, 406 in EMBASE, five in CINAHL and 11 in the Cochrane Database. After exclusion of duplicated articles, 524 articles were systematically screened based on the title and abstract. Finally, 11 studies were identified and included in the review (Figure 1). Characteristics of included studies are shown in Table 1. Two studies were conducted on TR-based physiotherapy assessment in low back pain population,[21,23] and one study each on shoulder pain,[36] elbow pain,[20] non-articular lower limb musculoskeletal disorders,[25] total knee arthroplasty (TKA)[23] and the ankle joint.[26] There were four technical papers which discussed the feasibility of using the Internet in assessing the physiotherapy tests.[16,19,24,37] Out of 11 studies, two studies were distinctive in assessing concurrent validity without reliability measures.[21,23]

Quality assessment of reliability and validity studies

The results of methodological quality of reliability studies were appraised using the QAREL checklist (Table 2). The methodological quality of reliability studies varied from 3/11 ratings to 9/11 ratings. Out of nine reliability studies, five studies demonstrated high quality, three studies were of medium quality and one study was of low quality. Of importance is that various statistical tests were used to analyse reliability and validity, and that raters were not blinded or were insufficiently blinded to study methodologies. Agreement on methodological quality assessment among the reviewers using the QAREL resulted in a Cohen's kappa of 0.86. The methodological characteristics of the validity studies were appraised using the QUADAS checklist (Table 3).[30] The methodological

Table 1. Characteristic of included studies.

Author, year, country	Population	Study design	Objectives summary	Setting	Technology
Palacin-Marin et al.[22] 2013 Spain	Low back pain	Descriptive repeated-measures crossover	To determine the level of agreement between FTF and TR methods of assessment	Primary care centre	TPLUFIB-WEB, a web-based system for the treatment.
Truter et al.[21] 2013 Australia	Low back pain	Single blind validation	To establish the validity of performing an effective physical examination of the lumbar spine via TR	Clinical setting	eHAB, a telerehabilitation system
Steele et al.[36] 2012 Australia	Shoulder disorders	NA	To establish the validity and reliability of the individual physical examination finding via TR	Clinical setting	eHAB, a telerehabilitation system
Russell et al.[25] 2010 Australia	Lower limb MSD	NA	To establish the concurrent validity and reliability of remote physical assessment and diagnosis of lower limb MSDs	Clinical setting	eHAB, a telerehabilitation system
Lade et al.[20] 2012 Australia	Elbow disorders	NA	To determine the validity and reliability of remote physical examination and diagnosis of elbow MSDs	Clinical setting	eHAB, a telerehabilitation system
Cabana et al.[23] 2010 Canada	TKR	NA	To explore the reliability of TR-based evaluation of patient with post TKR	Clinical setting	Purpose-built teleconferencing system
Russell et al.[26] 2010 Australia	Ankle disorders	NA	To examine the validity and reliability of remote physical examination of the ankle joint complex via TR	Clinical setting	eHAB, a telerehabilitation system
Russell et al.[37] 2003 Australia	Healthy	NA	Examine the reliability and validity of Internet-based goniometer for knee joint ROM assessment	Clinical setting	eHAB, a telerehabilitation system
Russell et al.[19] 2003 Australia	Healthy	NA	To investigate the accuracy and reliability of observational gait assessment via TR	Clinical setting	eHAB, a telerehabilitation system
Russell et al.[24] 2002 Australia	Healthy	NA	To assess the reliability of physiotherapy assessment measured using two Internet speeds	Clinical setting	eHAB, a telerehabilitation system
Durfee et al.[16] 2007 USA	Healthy with simulated impairment	Repeated measure design	To assess the usability of the technology and the feasibility of teleassessment methods	Lab setting	Purpose-built teleconferencing system

FTF: face-to face; MSD: musculoskeletal disorder; ROM: range of motion; TKR: total knee replacement; TR: telerehabilitation.

Table 2. Overview of bias assessment based on Quality Appraisal of Reliability Studies (QAREL) checklist.[28]

QAREL Items	Palacin et al., 2013	Steele et al., 2012	Russell et al., 2010	Lade et al., 2012	Russell et al., 2010	Russell et al., 2003	Russell et al., 2003	Russell et al., 2002	Durfee et al., 2007
1	Y	Y	Y	Y	Y	N	N	N	N
2	Y	Y	Y	Y	Y	Y	Y	Y	Y
3	Y	Y	Y	Y	Y	U	Y	NA	N
4	Y	Y	Y	Y	Y	U	Y	Y	U
5	U	Y	Y	U	Y	NA	NA	NA	NA
6	U	Y	U	U	Y	NA	NA	NA	NA
7	U	U	U	U	U	Y	U	N	U
8	Y	Y	Y	Y	Y	Y	NA	Y	N
9	Y	Y	Y	Y	Y	U	Y	Y	Y
10	Y	U	Y	Y	U	Y	Y	Y	Y
11	Y	Y	Y	Y	Y	Y	Y	Y	N
Total	**8/11**	**9/11**	**9/11**	**8/11**	**9/11**	**5/11**	**6/11**	**6/11**	**3/11**

N: no; N/A: not applicable; U: unclear; Y: yes.

Table 3. Methodological quality of validity studies using Quality Assessment of Diagnostic Accuracy Studies (QUADAS).[30]

No.	QUADAS items	Truter et al., 2013	Cabana et al., 2009
1	Representative sample	Y	Y
2	Selection criteria	Y	N
3	Appropriate reference standard	Y	Y
4	Stability of target condition	Y	Y
5	Appropriate sample received reference standard	Y	Y
6	Same reference standard to all	Y	Y
7	Reference standard independent of the index	U	U
8	Index test detailed	N	N
9	Reference standard detailed	N	N
10	Independent interpretation of index test	Y	Y
11	Independent interpretation of reference standard	Y	Y
12	Clinical data available similar to that in practice	N	N
13	Uninterruptable/intermediate test results reported	U	U
14	Withdrawals explained	Y	N
	Total	**9/14**	**7/14**

N: no; U: unclear; Y: yes.

quality of the two reviewed studies that evaluated validity alone demonstrated high quality (>60%).

Common statistical method

Different types of statistical tests were employed within the reviewed studies. In five studies, the physiotherapy assessment data were converted into binary and categorical variables, hence the percentage of agreement, chi square (x^2) and quadratically weighted kappa statistics were administered.[20,21,25,26,36] Chi square (x^2) was used in four studies to determine the significance of association between the TR and FTF methods of assessments. For continuous data, the Bland-Altman plot,[19,22,24,37] Pearson correlation (r)[21] and Krippendorff's α[23] were employed. The level of statistical significance was set at $p < 0.05$ for all studies.

Outcome measures

Range of motion (ROM). The results of concurrent validity, intra- and inter-rater reliabilities of TR based physiotherapy assessment are shown in Table 4. There were differences in the method of measuring the joint ROM via TR. The knee ROM was measured by placing a universal goniometer directly on to a computer screen, and demonstrated good concurrent validity (Krippendorff's $\alpha = 0.80$).[23] No significant difference was found in the joint ROM whether documented directly by the TR therapist or with assistance from the care giver ($t = 1.15$,

Table 4. Result of validity, inter-rated and intra-rated reliability.

Study	Participants	Sample size	Outcome measures	Instrument/test	Result		
					Reliability		
					Intra-rater	Inter-rater	Validity
Palacin at al., 2013	Low back pain	Male = 6 Female = 9	ROM	Lumbar lateral flexion	ICC = 0.95	ICC = 0.92	$\alpha = 0.751^{a}$
				Finger-floor distance	ICC = 0.94	ICC = 0.92	$\alpha = 0.991^{a}$
				Finger-floor lateral	ICC = 0.96	ICC = 0.93	$\alpha = 0.972^{a}$
Agreement between tele-rehabilitation and face-to-face clinical outcome assessment for low back pain in primary care,			Endurance	Sorensen test	ICC = 0.94	ICC = 0.92	$\alpha = 0.796^{a}$
			Motor control	Anterior straight leg raising	ICC = 0.95	ICC = 0.93	$\alpha = 0.968^{a}$
			Disability	Oswestry Disability Index			$\alpha = 0.994^{a}$
			Pain	Visual analog scale			$\alpha = 0.940^{a}$
			Health survey	Physical component SF-12			$\alpha = 0.971^{a}$
				Mental component SF-12			$\alpha = 0.973^{a}$
			Kinesiophobia	Tampa Scale			$\alpha = 0.977^{a}$
Truter et al., 2013	Low back pain	Male = 11 Female = 15	Posture	Scoliosis			25–75%[b]
				Pelvic tilt			*kappa* = 0.07–0.20
The validity of physical therapy assessment of low back pain via tele-rehabilitation in a clinical setting.			ROM	Flexion			$r = 0.89^{c}$ $p < 0.001$
				Extension			$r = 0.83^{c}$ $p < 0.001$
				Lateral flexion: Right/ Left			$r = 0.69/0.67^{c}$ $p < 0.001$
			SLR				$r = 0.64^{c}$ $p < 0.001$
Lade et al., 2012	Elbow injury or pain	Male = 9 Female = 1	ROM		95%[d] $p < 0.001$	93%[d] $p < 0.000$	55%[d] $p < 0.001$
Validity and reliability of the assessment and diagnosis of musculos-keletal elbow disorders using telerehabilitation			Neurodynamic	ULNT	98%[d] $p < 0.001$	68%[d] $p = 0.003$)	46%[d] $p = 0.616$
			Special test		94%[d] $p < 0.001$	91%[d] $p < 0.001$	75%[d] $p < 0.003$
			Pain response		97%[d] $p < 0.001$	98%[d] $p < 0.001$	82%[d] $p < 0.001$
			Joint assessment		81%[d] $p < 0.001$	97%[d] $p < 0.001$	47%[d] $p = 0.386$
			Strength		98%[d] $p < 0.001$	96%[d] $p < 0.001$	90%[d] $p < 0.006$

(continued)

Table 4. Continued

Study	Participants	Sample size	Outcome measures	Instrument/test	Result		
					Reliability		
					Intra-rater	Inter-rater	Validity
			Limiting factor		86%[d] $p < 0.001$	84%[d] $p < 0.001$	68%[d] $p < 0.001$
			Pain		88%[e] *kappa* = 0.95	85%[e] *kappa* = 0.95	86%[e] *kappa* = 0.45
			Severity Scale		97[e] kappa = 0.83	98[e] kappa = 0.82	77[e] kappa = 0.69
Russell et al., 2009	Lower limb injury	Male = 5, Female = 14	Posture Gait analysis Swelling		**Categorical data** 99.2%[e] *kappa* = 0.99	**Categorical data** 97.7%[e] *kappa* = 0.98	**Categorical data** 90.3%[e] *kappa* = 0.76
The diagnostic accuracy of telerehabilitation for non-articular lower limb musculoskeletal disorders.			ROM Special test Strength Neural test		**Binary data** 97.4%[d] $X^2 = 969.81$ $p < 0.001$	**Binary data** 95.1%[d] $X^2 = 827.23$ $p < 0.001$	**Binary data** 82.9%[d] $X^2 = 227.69$ $p < 0.001$
Steele et al., 2012	Shoulder pain	Male = 16 Female = 6	Pain		97.2%[e] *kappa* = 0.95	97.2%[e] *kappa* = 0.95	76.8%[e] *kappa* = 0.50
Assessment and diagnosis of musculoskeletal shoulder disorders over the Internet			Severity scale		97.7%[e] *kappa* = 0.83	99.2%[e] *kappa* = 0.83	96.0%[e] *kappa* = 0.66
			ROM		95.8 %[d] $X^2 = 393.950$ $p < 0.0001$	92.1%[d] $X^2 = 298.492$ $p < 0.001$	87.4%[d] $X^2 = 30.782$ $p < 0.001$
			Orthopedic test		88.7%[d] $X^2 = 185.337$ $p < 0.001$	88.1%[d] $X^2 = 209.515$ $p < 0.001$	75.9%[d] $X^2 = 54.765$ $p < 0.001$
			Pain response in static muscle testing		88.7%[d] $X^2 = 185.337$ $p < 0.001$	88.1%[d] $X^2 = 209.515$ $p < 0.001$	75.9%[d] $X^2 = 54.765$ $p < 0.001$
			Neurodynamic		96.8%[d] $X^2 = 510.596$ $p < 0.001$	98.3%[d] $X^2 = 618.832$ $p < 0.001$	81.7%[d] $X^2 = 70.867$ $p < 0.001$
			Strength				

(continued)

Table 4. Continued

Study	Participants	Sample size	Outcome measures	Instrument/test	Result		
					Reliability		
					Intra-rater	Inter-rater	Validity
					97.3%[d] $X^2 = 585.732$ $p < 0.001$	95.4%[d] $X^2 = 476.739$ $p < 0.001$	87.1%[d] $X^2 = 31.546$ $p < 0.001$
			Joint assessment		85.9%[d] $X^2 = 51.004$ $p < 0.001$	90.9%[d] $X^2 = 43.990$ $p < 0.001$	64.4%[d] $X^2 = 0.762$ $p = 0.383$
			Limiting factor		88.9%[d] $X^2 = 1795.945$ $p < 0.001$	87.0%[d] $X^2 = 1549.903$ $p < 0.001$	68.9%[d] $X^2 = 320.182$ $p < 0.001$
Cabana et al., 2010	Total knee arthroplasty	Male = 8 Female = 7	ROM	Knee flexion			$\alpha = 0.80$[f]
				Knee extension			$\alpha = 0.85$[f]
Inter-rater agreement between telerehabilitation and face-to-face clinical outcome measures for total knee arthroplasty			Scar	Quantities scale			$\alpha = 0.34$[f]
			Swelling				$\alpha = 0.85$[f]
			Mobility	TUG			$\alpha = 0.86$[f]
			Locomotion	Tinetti test			$\alpha = 0.79$[f]
			Balance	Berg balance scale			$\alpha = 0.76$[f]
Russell et al., 2010	Ankle disorders	Male = 5 Female = 10	Posture		**Categorical data** 99.2%[e] *kappa* = 0.99	**Categorical data** 97.3%[e] *kappa* = 0.98	**Categorical data** 89.3%[e] *kappa* = 0.98
Telerehabilitation mediated physiotherapy assessment of ankle disorders			ROM Gait Self-palpation		**Binary data**	**Binary data**	**Binary data**
			Strength Orthopaedic test Neurodynamic		99.2%[d] $X^2 = 694.45$ $p < 0.001$	99.9%[d] $X^2 = 579.68$ $p < 0.001$	99.3%[d] $X^2 = 234.41$ $p < 0.001$

(continued)

Table 4. Continued

Study	Participants	Sample size	Outcome measures	Instrument/test	Result		
					Reliability		
					Intra-rater	Inter-rater	Validity
Russell et al., 2003 Can the Internet be used as a medium to evaluate knee angle?	Knee joint	1 healthy	ROM	Knee flexion Knee extension	$ICC_{(2,1)} = 1.00$ $ICC_{(2,1)} = 1.00$	$ICC_{(2,2)} = 1.00$ $ICC_{(2,1)} = 0.96$	$ICC_{(2,1)} = 1.00$ $ICC_{(2,1)} = 0.99$
Russell et al. 2003 The diagnostic reliability of Internet-based observational kinematic gait analysis	Gait	24 healthy	Gait	GARS	94%[d] ICC = 0.99 (0.952–0.996)	92%[d] ICC = 0.97 (0.901–0.992)	92%[d] ICC = 0.99 (0.967–0.997)
Russell et al., 2002 Physical outcome measures via the Internet: reliability of two Internet speeds	Physical measures	20 healthy	ROM Knee flexion Knee extension Strength Limb girth				$r = 0.97$[c] $r = 0.99$[c] $r = 0.99$[c] $r = 0.97$ (upper)[c] $r = 0.90$ (middle)[c] $r = 0.97$ (lower)[c]
Durfee et al., 2007	Physical measures	10 healthy with simulated impairment	ROM Strength Balance Mobility	Shoulder abduction Shoulder external rotation Knee flexion Biceps Quadriceps Berg balance scale: items 1 and 8 only TUG test			Shown agreement between universal (FTF) and virtual goniometry methods. Statistics not available. 0.21[g] $p = 0.83$ 0.18[g] $p = 0.85$ 1.37[g] $p = 0.20$

FTF: face-to-face; GARS: gait assessment rating scale; ROM: range of motion; SF-12: Short Form Health Survey; TUG: Timed Up and Go test; ULNT: upper limb neural tension test.
[a]Cronback α; [b]exact agreement; [c]Pearson correlation coefficient; [d]percentage of agreement; [e]exact and close agreement; [f]Kirpendorff's α; [g]paired *t* test.

$p = 0.25$).[16] The lateral flexion of the lumbar spine assessed using a virtual goniometer demonstrated a moderate correlation ($r = 0.69$ and 0.67, $p < 0.001$) between TR and FTF methods of assessment.[21] However, the Internet goniometer ROM measurement demonstrated a good-to-excellent concurrent validity in seven reviewed studies.[20,21,24–26,36,37] In addition, the intra-and inter-rater reliabilities of TR-based ROM assessment were reported as being good-to-excellent.

Posture assessment. Assessment of posture via TR showed conflicting evidence. Two studies have reported on postural analysis.[21,26] TR-based postural analysis for the lumbar spine demonstrated slight to substantial concurrent validity (PA = 25–75%, kappa $\leqslant 0.19$) and good intra- and inter-rater reliabilities (kappa = 0.83).[26]

Strength, endurance and motor control. A variation existed in reporting strength, endurance and motor control. When lumbar spine endurance and motor control (anterior straight leg raising (SLR) test) among the low back pain population were studied, the results demonstrated that there was good validity in endurance (Cronbach's $\alpha = 0.80$) and excellent validity in motor control (Cronbach's $\alpha = 0.97$) between TR and FTF methods of assessment. The intra-rater and intra- reliabilities of TR based endurance and motor control assessment were excellent (ICC > 0.90).[22]

Different techniques were employed to assess muscle strength via TR. Four studies measured static muscle strength by applying patients' self-resistance.[20,25,26,36] These four studies demonstrated good-to-excellent agreement in muscle strength assessment. Strong correlation ($r = 0.99$, $p = 0.512$) was demonstrated in quadriceps muscle strength assessment via TR using progressive load weights on the distal part of the limb.[24] Furthermore, lower limb functional muscle strength assessment among the TKA population demonstrated good validity (Krippendorff's $\alpha = 0.85$).[23]

Special orthopaedic tests (SOTs). Four reviewed studies reported the SOTs. Patients were asked to perform modified self-administered SOTs under the guidance of the TR physiotherapist. Patients also viewed high quality demonstration videos. The acquired data were converted into binary form. The reported PA between TR and FTF was 75% ($p = 0.003$) for elbow conditions,[20] 76% ($x^2 = 54.765$, $p < 0.001$) for shoulder disorders,[36] 99.3% ($x^2 = 234.41$, $p < 0.001$) for ankle conditions,[26] and 82.9% ($x^2 = 227.69$, $p < 0.001$) for non-articular lower limb musculoskeletal injuries.[25] The PA for intra- and inter-rater reliabilities ranged between 88% and 99% ($p < 0.001$).

Neurodynamic tests (NDTs). Four studies reported the TR-based NDTs.[20,21,26,36] In the low back pain population, the SLR test showed moderate correlation ($r = 0.64$; $p < 0.001$) between TR and FTF methods with a mean difference of -6° (standard deviation (SD) = 15°).[21] There was low level of agreement in the NDTs between TR and FTF methods among elbow (46% agreement; $p = 0.62$)[20] and shoulder (56.1% agreement; $x^2 = 6.291$; $p = 0.012$)[36] disorders. The result of the concurrent validity of ankle disorders was not reported. Furthermore, the intra- and inter-rater reliabilities of NDTs in shoulder and elbow disorders demonstrated a high level of agreement (>68%, $p < 0.001$). However, there was moderate inter-rater reliability in ankle disorders (68% agreement; $p = 0.003$).[26]

Pain, swelling, and scarring. Pain was assessed as an independent variable or as an associated factor in static muscle testing, limiting factor in range of motion, functional examination, self-palpation and gait analysis. The analysis of agreement of pain rating was fair to moderate agreement in the ankle, shoulder and elbow musculoskeletal conditions (kappa < 0.8) with good inter-and intra-rater reliabilities, although Palacin et al. demonstrated high level of agreement ($\alpha = 0.94$) among patients with low back pain.[22] Quantitative scar measurement demonstrated low concurrent validity (Krippendorff's $\alpha = 0.34$).[23] The concurrent validity for swelling was moderate (similar agreement = 90.3%, kappa = 0.76) with excellent inter-rater (similar agreement = 99.2%, kappa = 0.99) and intra-rater (similar agreement = 97.7%, kappa = 0.98) reliability.[25]

Gait and balance. There was a difference in the tools for measurement for gait and balance assessment in the reviewed studies. One study which assessed gait and balance in TKA used the Tinetti tool[38] with good concurrent validity (Krippendorff's $\alpha = 0.79$) between TR and FTF.[23] The kinematic observational gait analysis,[19] however, had excellent concurrent validity (ICC = 0.96), intra-rater reliability (ICC = 0.96) and inter-rater reliability (ICC = 0.92).[19]

Functional outcome measures. Three studies reported TR-based functional outcome measurement.[16,22,23] The concurrent validity between TR and FTF exhibited substantial-to-good levels of agreement for the Timed Up and Go test (Krippendorff's $\alpha = 0.86$), Tinetti test (Krippendorff's $\alpha = 0.79$) and Berg test (Krippendorff's $\alpha = 0.76$) for TKR patients.[23] Similarly, excellent reliability was reported for the Oswestry Disability Index (Cronbach's $\alpha = 0.99$), Short Form Health survey (SF-12) (Cronbach's $\alpha = 0.97$) and Tampa Kinesiophobia Scale (Cronbach's $\alpha = 0.97$) in patients with low back pain.[22]

Discussion

This systematic review identified 11 studies, which examined concurrent validity and reliability of TR-based physiotherapy assessments. Overall, the methodological quality of the included reliability studies was moderate-to-high based on the QAREL checklist.[28] Although

blinding is an important factor in reliability studies, QAREL revealed variations in blinding in the reviewed studies. Methodological qualities of two validity studies were high[21] in one study and medium[38] in another study based on QUADAS.

Statistical considerations

The reviewed studies displayed variations in the statistical methods employed in the analysis of validity and reliability. Continuous data were converted into categorical data and analysed with kappa statistics along with the PA in most studies. While the PA is simple, intuitive and easy to calculate, one major limitation of the PA is that it could potentially be affected by chance or inflation of reliability by adding new categories, whereby it may not give any values to raters.[39] In these instances, the categorical data fail to represent the actual meaning of original continuous data and may impose the risk of decreasing the statistical power (type II error).[16,40–42]

Use of chi square to assess reliability, while generally not recommended, was demonstrated in several studies to evaluate the association for binary data. Unfortunately, no studies had employed SEMs and CV in analysing the absolute reliability in TR-based physiotherapy assessments. SEMs reflect the extent of expected error in a measurement between the raters,[43] while CV reflects the extent of variability of the measurement score between the TR and FTF methods or raters.[43,44] Smaller CV values demonstrate less dispersion in the measurement values.[45] Previous literature has shown that a measurement tool is considered reliable if the value CV is 15% and SEMs is 5%.[43,46,47] Therefore, future studies should consider appropriate statistical methods when reporting validity and reliability.

Study issues

Various types of physiotherapy assessments such as pain, range of motion, strength, endurance, balance, posture, SOTs, NDTs and functional outcome measures were measured via the TR method. The results of concurrent validity and reliability of TR-based physiotherapy assessment clearly demonstrate the feasibility of objective physiotherapy assessments for MSDs. However, there is a scarcity of literature on TR-based musculoskeletal assessments. This indicates that there is a need to understand the field of TR for MSDs specifically to test the validity and reliability for physiotherapy assessments. Additionally, it also revealed the lack of adoption of TR applications in the rehabilitation sciences. Most reviewed articles also used simulated patients in a laboratory setting to address the validity and reliability of their research. The QAREL recommends the recruitment of real target populations and study locations to test the reliability of intended measures.[28] Subjects in the reviewed studies did not represent real populations who would include those unable to travel due to chronic disability or aging or those who lived in rural areas with insufficient access to rehabilitation services.

Validity and reliability

One prerequisite to adopt TR in current clinical practice is to test the new technology against the gold standard (FTF). If the magnitude of difference between TR and traditional FTF assessment results is small and clinically negligible, TR may be used as an alternative modality.[35] According to the current review, objective physiotherapy assessments of MSDs via TR were found to be technically feasible with good-to-excellent concurrent validity and reliability. Therefore, TR may be a potential platform for physiotherapists to perform remote evaluations of several components of physical assessment, including observation, range of motion,[16,20,22,23,26] muscle strength,[16,24–26] gait analysis,[19] posture,[21,26] NDTs[20,21] and SOTs.[20,26]

Within the reviewed studies, various methods to measure the joint ROM were employed, such as measurement by direct observation, use of a universal goniometer on the digital form of images, and using sophisticated image-processing software.[15,16] The lack of difference in TR joint angle measurements and universal goniometer measures can conclude that joint range of motion assessment is highly feasible via TR. In clinical practice, Quantitative Voluntary Muscle Testing (QVMT) can be assessed by using manual resistance by the physiotherapist to muscles and grading them using the 0–5 Medical Research Council (MRC) grade.[48] The QVMT via TR however would be challenging as the TR physiotherapist lacks physical contact with patients. To overcome this limitation, use of a trained caregiver/family member or allied health assistant, or use of the patient's own resistance or the dynamometer may be possible. The use of a purpose-trained allied healthcare assistant has been shown to bring about a positive result in other areas of rehabilitation.[49] The results of this review showed the feasibility of remote QVMT assessment.

The technological practicality of the remote physiotherapy assessment is a crucial step before the TR-based intervention can be tested for clinical efficacy and subsequent adoption in regular clinical practice. Some physiotherapy assessments, however, demonstrated a low concurrent validity in lumbar spine posture assessment, and moderate concurrent validity in SOTs, NDTs and scar assessment. Hence, it may difficult to adopt these tests in routine clinical practice. With the self-administered nature of the SOTs, NDTs, and self-palpation under the guidance of the TR physiotherapist, patients reported extreme difficulty in self-execution of the task and reporting of the measurements to the TR physiotherapist. Future studies should consider technological advancements to improve the concurrent validity of TR-based assessment of such physiotherapy tests.

There are many other factors associated with poor concurrent validity, including bandwidth limitations,[20,25,26,50] low camera resolution,[23] bad lighting,[23] complexity of the

tests administered, inexperienced raters,[20] lack of video conference etiquette[25] and poor rapport. Low image resolution due to poor bandwidth has made it difficult for the TR physiotherapist to guide the patient to palpate the anatomical location of the source of pain and tenderness. This barrier may be overcome by providing a body chart via the TR system ahead of time.[20] In addition, the validity of SOTs and NDTs via the TR method might be improved by guiding and training the patients or caregiver through real-time feedback, supplemented by high-quality video or a video weblink.[26] Poor rapport during the TR session may have a negative influence on clinical reasoning in the diagnosis of MSDs via TR.[20,25] Future studies should focus on technological innovation and strategies to overcome these barriers in TR.

Limitation of this review

The reviewed studies have heterogeneity in areas of study population, physiotherapy assessment and statistical reporting of outcome variables, which makes it difficult to directly compare the results across chosen studies. Therefore, a meta-analysis was not viable.

Implication for future research

It is recommended that future reliability studies follow GRRAS guidelines to enhance the methodological quality and to ensure proper judgement.[51] TR-based physiotherapy assessments must be objective and measured as a continuous variable. Additionally, measures such as the ICC, SEM and CV could be used to enhance the analytical quality of the outcome data. Similar attention should be given to the utilisation of an optimal sample size, randomisation procedure, data on real clinical populations and real target environments. Further studies are warranted to extend this TR-based physiotherapy assessment to other clinical populations of interest. In addition, there is a need for technological advancement in the development of assessment tools pertaining to special tests in neuromusculoskeletal disorders and joint and posture disorders to enhance diagnostic and treatment accuracy in MSD populations.

Conclusion

This systematic review revealed that TR-based physiotherapy assessment is technically feasible to measure pain, swelling, ROM, muscle strength, balance, gait and functional outcomes with overall good concurrent validity. A low level of concurrent validity was demonstrated for lumbar spine posture assessment, whereas SOTs, NDTs and scar assessments demonstrated moderate concurrent validity. Inter-rater and intra-rated reliabilities showed good to excellent levels for TR-based physiotherapy assessment for MSDs. More research is required to consider appropriate statistical tests and the optimal sample size of a real target population, while assessing the concurrent validity and reliability of TR-based physiotherapy assessment for other MSDs.

Acknowledgement

The authors would like to extend sincere gratitude to Senthil Kumar, Deepashini, Ambusam Subramaniam, Vinothini Padmanathan, Fairus Fariza, and Saravanan Rajadurai for support and advice in preparing the manuscript. The authors wish to thank Heather Newton who assisted in the proofreading and English language correction of the manuscript.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

References

1. Hoy D, Brooks P, Blyth F, et al. The epidemiology of low back pain. *Best Pract Res Clin Rheumatol* 2010; 24: 769–781.
2. Hoy DG, Protani M, De R, et al. The epidemiology of neck pain. *Best Pract Res Clin Rheumatol* 2010; 24: 783–792.
3. Cross M, Smith E, Hoy D, et al. The global burden of hip and knee osteoarthritis: Estimates from the Global Burden of Disease 2010 study. *Ann Rheum Dis* 2014; 73: 1323–1330.
4. Woolf AD and Pfleger B. Burden of major musculoskeletal conditions. *Bull World Health Organ* 2003; 81: 646–656.
5. Storheim K and Zwart J-A. Musculoskeletal disorders and the Global Burden of Disease study. *Ann Rheum Dis* 2014; 73: 949–950.
6. Slim ZN, Chaaya M, Habib RR, et al. High burden of musculoskeletal conditions: A problem that has only recently come to recognition. *Chronic Illn* 2011; 7: 311–320.
7. Punnett L, Pruss-UstUn A, Nelson DI, et al. Estimating the global burden of low back pain attributable to combined occupational exposures. *Am J Ind Med* 2005; 48: 459–469.
8. Sheppard L. Work practices of rural and remote physiotherapists. *Aust J Rural Health* 2001; 9: 84–90.
9. Grimmer K and Bowman P. Differences between metropolitan and country public hospital allied health services. *Aust J Rural Health* 1998; 6: 181–188.
10. Russell TG. Telerehabilitation: A coming of age. *Aust J Physiother* 2009; 55: 5–6.
11. Eriksson L, Lindström B, Gard G, et al. Physiotherapy at a distance: A controlled study of rehabilitation at home after a shoulder joint operation. *J Telemed Telecare* 2009; 15: 215–220.
12. Russell TG, Buttrum P, Wootton R, et al. Internet-based outpatient telerehabilitation for patients following total knee arthroplasty: A randomized controlled trial. *J Bone Joint Surg Am* 2011; 93: 113–120.
13. Bernard M-M, Janson F, Flora PK, et al. Videoconference-based physiotherapy and tele-assessment for homebound older adults: A pilot study. *Act Adapt Aging* 2009; 33: 39–48.
14. Russell TG. Physical rehabilitation using telemedicine. *J Telemed Telecare* 2007; 13: 217–220.
15. Russell T. Goniometry via the internet. *Aust J Physiother* 2007; 53: 136.

16. Durfee WK, Savard L and Weinstein S. Technical feasibility of teleassessments for rehabilitation. *IEEE Trans Neural Syst Rehabil Eng* 2007; 15: 23–29.
17. Kimberlin CL and Winterstein AG. Validity and reliability of measurement instruments used in research. *Am J Health Syst Pharm* 2008; 65: 2276–2284.
18. May S, Chance-Larsen K, Littlewood C, et al. Reliability of physical examination tests used in the assessment of patients with shoulder problems: A systematic review. *Physiotherapy* 2010; 96: 179–190.
19. Russell TG, Jull G and Wootton R. The diagnostic reliability of Internet-based observational kinematic gait analysis. *J Telemed Telecare* 2003; 9: S48–S51.
20. Lade H, McKenzie S, Steele L, et al. Validity and reliability of the assessment and diagnosis of musculoskeletal elbow disorders using telerehabilitation. *J Telemed Telecare* 2012; 18: 413–418.
21. Truter P, Russell T and Fary R. The validity of physical therapy assessment of low back pain via telerehabilitation in a clinical setting. *Telemed J E Health* 2014; 20: 161–167.
22. Palacín-Marín F, Esteban-Moreno B, Olea N, et al. Agreement between telerehabilitation and face-to-face clinical outcome assessments for low back pain in primary care. *Spine (Phila Pa 1976)* 2013; 38: 947–952.
23. Cabana F, Boissy P, Tousignant M, et al. Interrater agreement between telerehabilitation and face-to-face clinical outcome measurements for total knee arthroplasty. *Telemed J E Health* 2010; 16: 293–298.
24. Russell TG, Wootton R and Jull G. Physical outcome measurements via the Internet: Reliability at two Internet speeds. *J Telemed Telecare* 2002; 8: S50–S52.
25. Russell T, Truter P, Blumke R, et al. The diagnostic accuracy of telerehabilitation for nonarticular lower-limb musculoskeletal disorders. *Telemed J E Health* 2010; 16: 585–594.
26. Russell TG, Blumke R, Richardson B, et al. Telerehabilitation mediated physiotherapy assessment of ankle disorders. *Physiother Res Int* 2010; 15: 167–175.
27. Lucas N, Macaskill P, Irwig L, et al. The reliability of a quality appraisal tool for studies of diagnostic reliability (QAREL). *BMC Med Res Methodol* 2013; 13: 1.
28. Lucas NP, Macaskill P, Irwig L, et al. The development of a quality appraisal tool for studies of diagnostic reliability (QAREL). *J Clin Epidemiol* 2010; 63: 854–861.
29. Onyewu O, Manchikanti L, Falco FJE, et al. An Update of the appraisal of the accuracy and utility of cervical discography in chronic neck pain. *Pain Physician* 2012; 15: 777–806.
30. Whiting P, Rutjes AWS, Reitsma JB, et al. The development of QUADAS: A tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol* 2003; 3: 25.
31. van der Wurff P, Hagmeijer RH and Meyne W. Clinical tests of the sacroiliac joint. A systematic methodological review. Part 1: Reliability. *Man Ther* 2000; 5: 30–36.
32. May S, Littlewood C and Bishop A. Reliability of procedures used in the physical examination of non-specific low back pain: A systematic review. *Aust J Physiother* 2006; 52: 91–102.
33. Adhia DB, Bussey MD, Ribeiro DC, et al. Validity and reliability of palpation-digitization for non-invasive kinematic measurement - A systematic review. *Man Ther* 2013; 18: 26–34.
34. Kottner J, Audige L, Brorson S, et al. Guidelines for reporting reliability and agreement studies (GRRAS) were proposed. *Int J Nurs Stud* 2011; 48: 661–671.
35. Bland JM and Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; 1: 307–310.
36. Steele L, Lade H, McKenzie S, et al. Assessment and diagnosis of musculoskeletal shoulder disorders over the internet. *Int J Telemed Appl* 2012; 2012(ID 945745): 8.
37. Russell TG, Jull G and Wootton R. Can the Internet be used as a medium to evaluate knee angle? *Man Ther* 2003; 8: 242–246.
38. Cabana F, Boissy P, Tousignant M, et al. Interrater agreement between telerehabilitation and face-to-face clinical outcome measurements for total knee arthroplasty. *Telemed J E Health* 2010; 16: 293–298.
39. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas* 1960; 20: 37–46.
40. Sim J, Sim J, Wright CC, et al. The kappa statistic in reliability studies: Use, interpretation, and sample size requirements. *Phys Ther* 2005; 85: 257–268.
41. Bartfay E and Donner A. The effect of collapsing multinomial data when assessing agreement. *Int J Epidemiol* 2000; 29: 1070–1075.
42. Donner A and Eliasziw M. Statistical implications of the choice between a dichotomous or continuous trait in studies of interobserver agreement. *Biometrics* 1994; 50: 550–555.
43. Atkinson G and Nevill A. Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sport Med* 1998; 26: 217–238.
44. Joseph LH, Hussain RI, Pirunsan U, et al. Clinical evaluation of the anterior translation of glenohumeral joint using ultrasonography: An intra- and inter-rater reliability study. *Acta Orthop Traumatol Turc* 2014; 48: 169–174.
45. Senthil KNS, Omar B, Htwe O, et al. Reliability, agreement, and validity of digital weighing scale with MatScan in limb load measurement. *J Rehabil Res Dev* 2014; 51: 591–598.
46. Gallagher AG, Ritter EM and Satava RM. Fundamental principles of validation, and reliability: Rigorous science for the assessment of surgical education and training. *Surg Endosc* 2003; 17: 1525–1529.
47. Khamwong P, Nosaka K, Pirunsan U, et al. Reliability of muscle function and sensory perception measurements of the wrist extensors. *Physiother Theory Pract* 2010; 26: 408–415.
48. O'Brien M. *Aids to the Examination of the Peripheral Nervous System*. Vol 5th ed. UK: Elsevier Health Sciences, 2010.
49. Sharma S, Ward EC, Burns C, et al. Training the allied health assistant for the telerehabilitation assessment of dysphagia. *Dysphagia* 2013; 28: 273–274.
50. Lemaire ED, Boudrias Y and Greene G. Technical evaluation of a low-bandwidth, Internet-based system for teleconsultations. *J Telemed Telecare* 2000; 6: 163–167.
51. Kottner J, Audige L, Brorson S, et al. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. *Int J Nurs Stud* 2011; 48: 661–671.