Seminars in Arthritis and Rheumatism 51 (2021) 1311–1319

Contents lists available at ScienceDirect

Seminars in Arthritis and Rheumatism

journal homepage: www.elsevier.com/locate/semarthrit

Core outcome measurement instrument selection for physical function in hand osteoarthritis using the OMERACT Filter 2.1 process

Féline P.B. Kroon[a,*], Désirée van der Heijde[b], Lara J. Maxwell[c], Dorcas E. Beaton[d], Abishek Abishek[e], Francis Berenbaum[f], Francisco J. Blanco[g], Philip G. Conaghan[h], Krysia Dziedzic[i], Catherine L. Hill[j], Ida K. Haugen[k], Mariko Ishimori[l], Valentin Ritschl[m], Tanja A. Stamm[n], Ruth Wittoek[o], Margreet Kloppenburg[p]

[a] *Department of Rheumatology, and Department of Rheumatology, Zuyderland Medical Center, Leiden University Medical Center, Leiden, Heerlen, the Netherlands*
[b] *Department of Rheumatology, Leiden University Medical Center, Leiden, the Netherlands*
[c] *University of Ottawa, and Centre for Practice-Changing Research, Ottawa Hospital Research Institute, Ottawa, Canada*
[d] *Musculoskeletal Health and Outcomes Research, and Department of Occupational Science and Occupational Therapy, St. Michael's Hospital and Institute for Work and Health, Rehabilitation Sciences Institute, and the Institute for Health Policy Management and Evaluation, University of Toronto, Toronto, Canada*
[e] *Academic Rheumatology, University of Nottingham, and National Institute for Health Research, Nottingham Biomedical Research Centre, Nottingham, United Kingdom*
[f] *Department of Rheumatology, Sorbonne Université, INSERM, Saint-Antoine hospital AP-HP, Paris, France*
[g] *Grupo de investigación en Reumatología. Instituto de Investigación Biomédica de A Coruña (INIBIC). Complexo Hospitalario Universitario de A Coruña (CHUAC), Sergas. Universidade da Coruña (UDC). As Xubias, 15006. A Coruña, España*
[h] *Leeds Institute of Rheumatic and Musculoskeletal Medicine, University of Leeds, and National Institute for Health Research, Leeds Biomedical Research Centre, Leeds, United Kingdom*
[i] *Impact Accelerator Unit, School of Medicine, Keele University, Keele, United Kingdom*
[j] *Discipline of Medicine, Adelaide, and Rheumatology Department, Queen Elizabeth Hospital, University of Adelaide, Woodville South, South Australia*
[k] *Division of Rheumatology and Research, Diakonhjemmet Hospital, Oslo, Norway*
[l] *Department of Medicine, Division of Rheumatology, Cedars-Sinai Medical Center, Los Angeles, California, United States of America*
[m] *Institute of Outcomes Research, Centre for Medical Statistics, Informatics and Intelligent Systems, Medical University of Vienna, and Ludwig Boltzmann Institute for Arthritis and Rehabilitation, Vienna, Austria*
[n] *Institute of Outcomes Research, Centre for Medical Statistics, Informatics and Intelligent Systems, Medical University of Vienna, and Ludwig Boltzmann Institute for Arthritis and Rehabilitation, Vienna, Austria*
[o] *Department of Rheumatology, Ghent University Hospital, Ghent University, Ghent, Belgium*
[p] *Departments of Rheumatology and Clinical Epidemiology, Leiden University Medical Center, Leiden, the Netherlands*

ARTICLE INFO

Keywords:
OMERACT
Hand osteoarthritis
Outcome measurement
Function
Clinical trials

ABSTRACT

Objective: Physical function is one of the Outcome Measures in Rheumatology (OMERACT) core outcome domains for hand osteoarthritis studies. Our aim was to select appropriate instrument(s) to measure this domain, as part of the development of a core outcome measurement set.
Methods: Following the OMERACT Filter 2.1 instrument selection process, the (function subscale of) the Australian/Canadian Hand Osteoarthritis Index (AUSCAN), Functional Index for Hand Osteoarthritis (FIHOA) and Michigan Hand Outcomes Questionnaire (MHQ) were assessed for domain match, feasibility, truth and discrimination. Data gathered from available literature, working group and patient surveys, and additional analyses in two hand osteoarthritis cohorts were used to inform a consensus process. Results were summarized in Summary of Measurements Properties tables and reviewed by the OMERACT technical advisory group.
Results: MHQ passed the assessment of domain match and feasibility by the working group and patient research partners. For AUSCAN important limitations in feasibility were noted, but domain match was good. FIHOA did not pass the assessment and was not taken through the follow-up assessment. Based on published literature, reliability and construct/longitudinal validity of both MHQ and AUSCAN fulfilled OMERACT standards. While clinical trial discrimination and thresholds of meaning were good for AUSCAN, results for MHQ were ambiguous.

* Corresponding author: Department of Rheumatology, Leiden University Medical Center, PO Box 9600, 2300 RC, Leiden, the Netherlands.
E-mail address: fpbkroon@gmail.com (F.P.B. Kroon).

https://doi.org/10.1016/j.semarthrit.2021.08.014

Conclusion: MHQ was provisionally endorsed as OMERACT core outcome measure for the core domain physical function. While AUSCAN may have better metric properties than MHQ, it received provisional endorsement as a second measure of function due to important feasibility issues. A research agenda to merit full endorsement was set.

Introduction

Physical function is an important outcome domain to evaluate in patients with hand osteoarthritis (OA) as it is often impaired, impacts daily living and reduces quality of life [1,2]. A variety of instruments is available to measure hand function [3].

In 2014, the Outcome Measures in Rheumatology (OMERACT) Hand OA Working Group endorsed a core domain set for clinical trials of symptom and structure modification and observational studies [4]. It includes six domains for all settings (pain, physical function, patient global assessment, health-related quality of life, joint activity, and hand strength), and two additional domains for trials of structure modification and observational studies (hand mobility, structural damage). At the same time, a preliminary core instrument set was proposed, based on available literature at the time. Since then, the Working Group worked on the development of a core outcome measurement set, according to the OMERACT Filter 2.1 instrument selection process [5-7]. Preliminary work was presented and discussed at a Special Interest Group (SIG) at OMERACT 2018 [8]. Here we report the final instrument selection for the first OMERACT hand OA core domain: physical function.

Material and methods

OMERACT hand OA core outcome domain set. Physical function was the first core outcome domain for which candidate instruments were evaluated through the OMERACT Filter 2.1 using the OMERACT Instrument Selection Workbook templates [9]. The domain involves self-reported limitations in physical functioning of the hands due to OA in any of the hand joints as observed in clinical trials of symptom or structure modification and in observational studies.

Identification and description of candidate instruments. Candidate instruments were identified in two previously published systematic literature reviews (SLR) of instruments measuring pain, physical function or patient global assessment in hand OA [10,11]. Furthermore, input from the Working Group was collected at meetings at European League Against Rheumatism (EULAR) and American College of Rheumatology (ACR) conferences. Instruments that assess performance-based hand function, as opposed to self-reported function, were not considered.

Based on this, three instruments measuring physical function of the hand were selected to be taken through the OMERACT Filter 2.1: the Australian/Canadian Hand Osteoarthritis Index (AUSCAN) function subscale [12], the Functional Index for Hand Osteoarthritis (FIHOA) [13] and the Michigan Hand Outcomes Questionnaire (MHQ) overall function and activities of daily living subscales [14]. A short description of each instrument is provided in the Appendix.

Domain match and feasibility. Domain match (including face validity and numeric performance) and feasibility of the three instruments were assessed by the working group and in a survey among patient research partners (PRP).

Assessment of face validity and feasibility was done by the working group, using standardized questionnaires from the OMERACT Instrument Selection Workbook. Assessment of numeric performance (e.g., missing data patterns, floor/ceiling effects, score distribution) was done in baseline data from the HOSTAS cohort, a Dutch observational cohort of hand OA patients ($n = 383$) [15]. The assessment of domain match and feasibility was presented and discussed among working group members in a SIG at OMERACT 2018 [8].

Patients with hand OA ($n = 48$) from eight centers in different settings (primary and secondary care) in seven countries worldwide (Adelaide (Australia), Vienna (Austria), Ghent (Belgium), Leiden (The Netherlands), Oslo (Norway), A Coruña (Spain), Keele and Nottingham (United Kingdom)) were asked for their opinion on domain match and feasibility of the instruments, using a standardized questionnaire developed by OMERACT for this purpose, and translated by working group members to the local language if applicable. The surveys were approved by the local medical ethics committees.

Based on these data, a web-based vote was held among working group members to assess which of the instruments were suitable to be assessed in more detail in the second part of Filter 2.1 (Fig. 1).

Colored letters represent the working group's assessment (green (G), good; amber (A), noncritical limitation; red (R), unsuitable). AUSCAN, Australian/Canadian Hand Osteoarthritis Index; FIHOA, Functional Index for Hand Osteoarthritis; MHQ, Michigan Hand Outcomes Questionnaire.

Assessment of construct validity and discrimination. Instruments that passed the domain match and feasibility criteria were assessed for construct validity and discrimination. For this purpose, published studies with information on construct validity, test-retest reliability, longitudinal construct validity, clinical trial discrimination and thresholds of meaning of each of the instruments in patients with hand OA were gathered. According to OMERACT guidelines, at least two pieces of evidence that support the performance of the instrument from studies with good methods (green or amber ratings on methods) are needed for each measurement property [7]. Where possible, we made use of existing SLRs in the field of hand OA. In case not enough studies (at least two) were available from the existing SLRs to assess a certain measurement property, a complementary literature search was performed to screen for additional studies. Search terms for the population (patients with hand OA) were taken from an SLR for the EULAR hand OA management recommendations [16]. For the measurement properties, search terms proposed by the COSMIN group were used, as advised in the OMERACT Instrument Selection workbook. The search strategies can be found in the Appendix. For AUSCAN, data from a previous SLR of the OMERACT hand OA working group [10] (databases: PubMed, Embase, Web of Science, COCHRANE, CINAHL, Academic Search Premier, ScienceDirect; date of search from inception up to January 2014) and from a complementary SLR focusing on thumb base OA [11] (databases: PubMed, EMBASE, Web of Science, COCHRANE, CINAHL, Academic Search Premier, ScienceDirect, PEDro; date of search from inception up to November 2010) were complemented with a literature search for studies of thresholds of meaning (database: PubMed; date of search: January 2014 to November 2019). For MHQ, no SLR was available, so a literature search for all metric properties was performed (database: PubMed; date of search: inception up to October 2019). Clinical trials to extract data on clinical trial discrimination for both instruments were identified from a recent SLR for the EULAR hand OA management recommendations (databases: PubMed/MEDLINE, Embase, Cochrane CENTRAL; up to June 2017) [16].

One reviewer (FK) screened studies retrieved in the literature searches to determine eligibility for inclusion according to predefined inclusion criteria. Relevant data on study characteristics and the above-mentioned measurement properties was extracted. Two

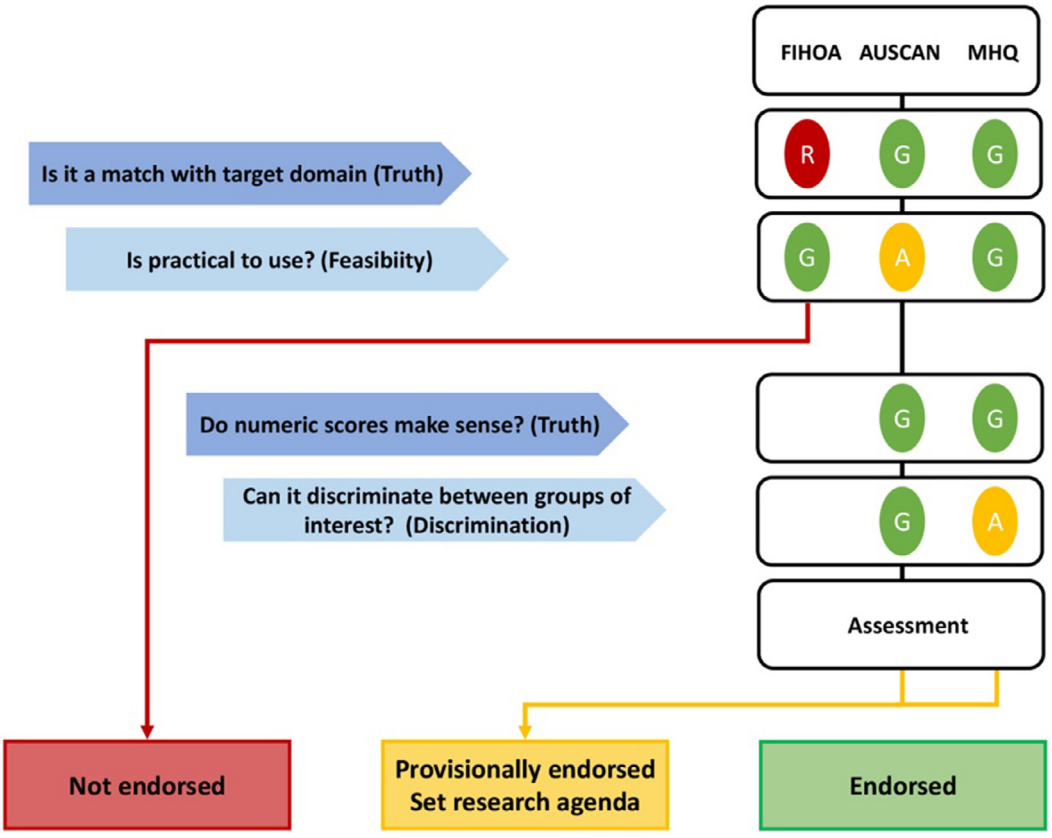

Fig. 1. The OMERACT Filter 2.1 instrument selection process for the hand OA core domain physical function.

reviewers (FK, DvdH) individually assessed risk of bias of all studies using the COSMIN—OMERACT Good Methods Checklist and sought consensus. In case of conflict of interest through authorship a third reviewer was consulted (IKH). Green indicates good methods (i.e., good avoidance of risk of bias), amber a noncritical limitation (i.e., some concern of risk of bias), and red unsuitable evidence (i.e., a warning for risk of bias).

Evidence of each metric property of the instruments was reviewed in studies with good methods (green or amber rating). We used the OMERACT provisional standards for adequacy, outlined in the OMERACT instrument selection workbook, to judge whether evidence was supportive of adequate performance of the metric property [7]. Evidence was rated as + (positive support), ± (ambivalent support), or – (does not reach performance standards for metric property). Not all studies evaluating (longitudinal) construct validity presented pre-specified hypotheses. To be able to assess whether the study results were supportive of the instrument, the working group compared the results to pre-specified hypotheses on (longitudinal) construct validity from a previously published study [17].

Since only one study was identified that provided information on thresholds of meaning of MHQ, an analysis to fill this knowledge gap was performed by the working group. This analysis concerns unpublished data. In short, we used data from a short-term clinical trial [18] (six-week follow-up) and a long-term observational cohort study [15] (two-year follow-up) to estimate thresholds of minimal clinically important improvement (MCII; clinical trial) and patient acceptable symptom state (PASS; clinical trial and observational cohort) of AUSCAN and MHQ in patients with hand OA. To determine these thresholds, we used an anchoring method. We used a two-step anchor question for estimation of the MCII: "Compared to when you started the study, how has your hand function been during the last 48 h?" (improved, no change, worse), and "If you answered *improved* at the previous question, how much is this improvement?" (very much, moderately, slightly, not at all improved). We defined the MCII as the minimal improvement in function achieved by 75% of participants who stated to have had a slight or moderate improvement during the trial, in line with previous studies of thresholds of meaning in OA [19]. It was calculated by taking the 75th percentile of the distribution of AUSCAN/MHQ change scores from baseline in participants who indicated to be 'slightly' or 'moderately' improved. We used the following anchor question to estimate the PASS: "If you were to remain for the rest of your life as you were during the last 48 h, would this be acceptable or unacceptable for you?" (acceptable or unacceptable). The PASS was defined as the minimal score considered acceptable for 75% of participants, and was calculated by taking the 75th percentile of the distribution of the AUSCAN/MHQ scores in participants who rated their health 'acceptable'. Percentiles were calculated using Stata V15.1 software. The quality of the methods used was independently reviewed by the OMERACT technical advisory group (TAG).

Final rating. Evidence on each metric property from studies using good or amber methods was extracted and summarized in Summary of Measurement Properties (SOMP) tables. Each measurement property was given a final rating based on the gathered evidence according to OMERACT guidance. A green rating indicates consistently good performance from multiple studies identified as having good methods; amber indicates a noncritical limitation in the evidence, which merits a research plan. Finally, an overall rating across all the measurement properties for each instrument was proposed by the working group, evaluated by the TAG and finally brought to a broader group of the OMERACT community for final approval of our proposed level of endorsement.

Results

Domain match and feasibility. A comparison of the aspects of the three instruments and the working group's assessment of domain

Table 1
Overview of assessment of domain match and feasibility to measure physical function in hand osteoarthritis.

	AUSCAN	FIHOA	MHQ
Domain match	Good match	Some items may be culturally challenging (accepting a handshake) or outdated (writing for >10 minutes; gender-specific question (sewing for women, using screwdriver for men)).	Good match
Numeric performance			
Completion rate	Good	Good	Good
Missing data	No missing data pattern	No missing data pattern	No missing data pattern
Floor/ceiling effects	No (1.8% with lowest score, 0.3% with highest score)	No (4.2% with lowest score, 0% with highest score)	No (overall hand function/ ADL: 0%/0% with lowest score, 1.3%/3.1% with highest score).
Number of items	9 items	10 items	Overall hand function: 10 items; ADL: 17 items
Score distribution	Normal distribution	Slightly right skewed	Overall hand function: normal distribution; ADL: somewhat left skewed
Feasibility	The questionnaire is copyrighted and may only be used with permission and after payment of a considerable fee	Feasible	Feasible

Colors represent the working group's assessment (green, good; amber, noncritical limitation; red, unsuitable). ADL, activities of daily living; AUSCAN, Australian/Canadian Hand Osteoarthritis Index; FIHOA, Functional Index for Hand Osteoarthritis; MHQ, Michigan Hand Outcomes Questionnaire.

match and feasibility of (the function subscales of) AUSCAN, FIHOA and MHQ was previously published in a report of the hand OA working group SIG at OMERACT 2018 [8]. Table 1 provides an overview.

Most PRPs voted green for the three instruments (AUSCAN 79%, FIHOA 66%, MHQ 71%) in the survey. PRP who voted amber for AUSCAN (21%) or MHQ (29%) suggested various additional questions to be included, particularly more common daily tasks (AUSCAN) or more specific questions (MHQ). While some PRP preferred FIHOA for its brevity, MHQ was favored by others because they felt it covered areas that AUSCAN and FIHOA did not. Generally, PRP were least positive about FIHOA (26% amber, 8% red), because they felt it missed important questions, gender-specific questions were not appreciated and it was felt to have items that were dated and no longer relevant in today's context.

In the web-based vote among the working group ($n = 26$; 22 clinicians, 19 researchers), domain match was accepted for AUSCAN (81% green, 15% amber) and MHQ (65% green, 31% amber), but concern was raised for FIHOA (62% amber, 12% red), mainly because voters felt the instrument includes some outdated, culturally challenging and gender-specific questions that are not acceptable in this day and age. Feasibility of MHQ was accepted (46% green, 46% amber), though important caveats are that the questionnaire is long, calculation of the scores is not straightforward, and that it should be translated to more languages. Feasibility of FIHOA was also accepted (81% green, 15% amber), the only concern being whether all translations are fully validated. Feasibility of AUSCAN was problematic (58% amber, 19% red), because the questionnaire is copyrighted and not freely available.

The concerns raised for FIHOA in its current form seriously limit its usefulness, and would require the instrument to be modified and subsequently revalidated. Therefore, it was decided to move forward only the AUSCAN and MHQ through the OMERACT Filter 2.1.

Assessment of construct validity and discrimination. Fig. 2 presents a flowchart of the literature search for the detailed assessment of the metric properties of AUSCAN and MHQ (PRISMA flowcharts are additionally provided in the Appendix).

For AUSCAN, the SLR by Visser et al. reported 34 studies that applied this instrument, of which 20 studies contained relevant quantitative data on its metric properties [10]. The SLR by Marks et al. reported one study that applied AUSCAN [11]. The additional search for studies on thresholds of meaning of AUSCAN yielded 2 additional inclusions. Furthermore, one additional clinical trial was identified in the SLR by Kroon et al. [16]. In total, 24 studies provided evidence for construct validity ($n = 9$) [12,20-27], test-retest reliability ($n = 4$) [12,21,23,28], longitudinal construct validity ($n = 5$) [22,29-32], clinical trial discrimination ($n = 8$) [12,18,33-38], and thresholds of meaning ($n = 3$) [19,39,40]. Most studies on (longitudinal) construct validity had amber methods, because the authors had not clearly pre-specified their hypotheses. Data from two studies could not be used based on the good methods check: one construct validity study was unsuitable because it used AUSCAN as the external standard [22], and one test-retest reliability study used unsuitable analysis methods [28]. Clinical trials that were judged amber had no control group [37], were not blinded [35,38] or had no expected change in the intervention group [34]. Two studies on thresholds of meaning had low risk of bias [19,40], and one had amber methods [39]. Generally, studies supported adequate performance of the measurement properties of the AUSCAN.

For MHQ, our literature search yielded 5 studies which met inclusion criteria, and 3 additional clinical trials were identified (Fig. 2). In total, 8 studies provided evidence for construct validity ($n = 4$) [17,41-43], test-retest reliability ($n = 2$) [42,44], longitudinal construct validity ($n = 3$) [17,42,43], clinical trial discrimination ($n = 3$) [18,45,46], and thresholds of meaning ($n = 1$) [42]. Three studies on (longitudinal) construct validity had amber methods, because they lacked pre-specified hypotheses. One construct validity study was unsuitable because MHQ was used as the external standard [43]. One clinical trial had amber methods because it was an open study [46]. Test-retest reliability and threshold of meaning studies had low risk of bias. Though less studies were available compared to AUSCAN, overall, evidence supported adequate performance of the measurement properties of the MHQ. Only for clinical trial discrimination

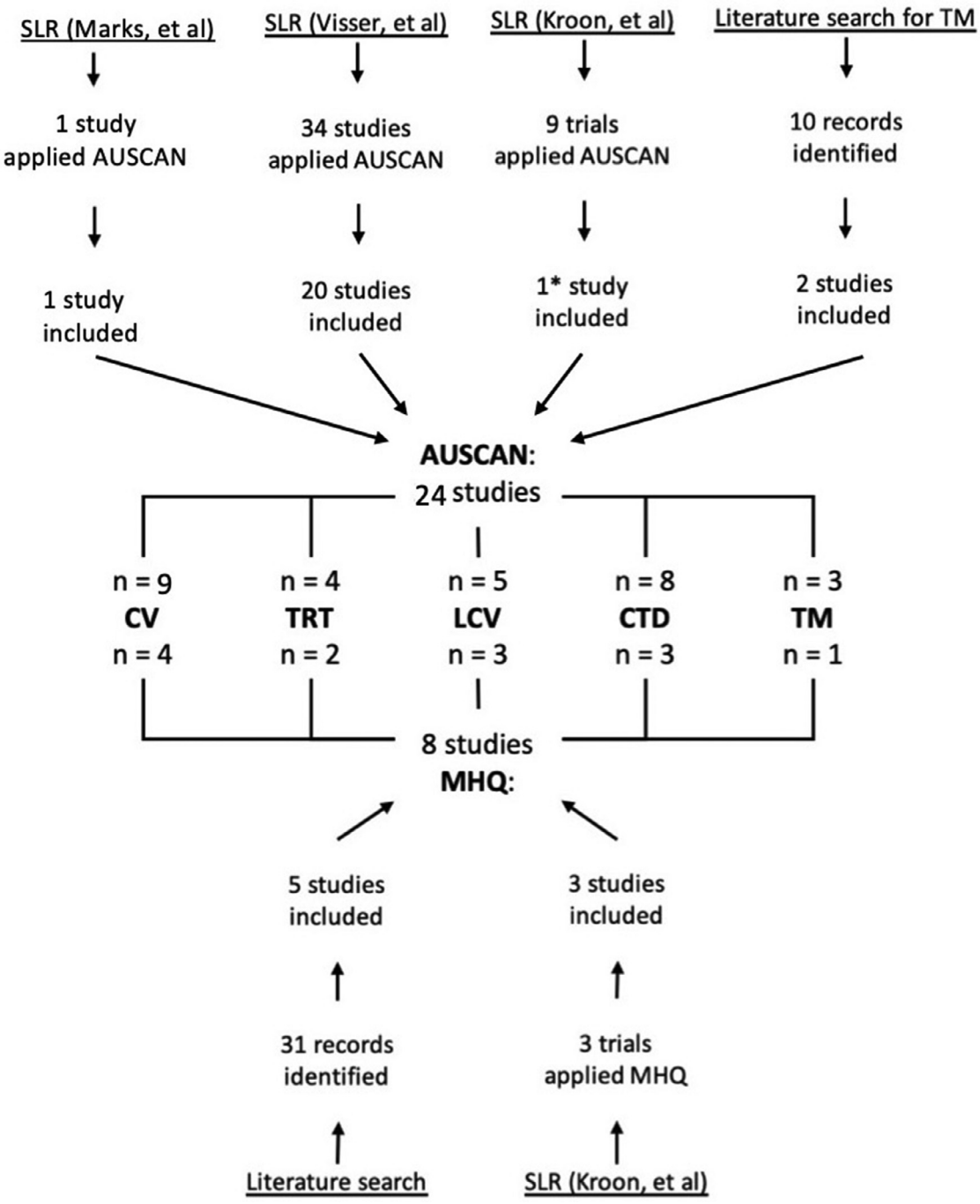

Fig. 2. Flowchart of studies for assessment of construct validity and discrimination of AUSCAN (upper) and MHQ (lower).

evidence was inconsistent, with two studies being supportive (green [45] and amber methods [46]) while in a third study (green methods [18]) MHQ was less sensitive than other function instruments to measure clinical response to treatment.

*8 trials were also identified through the SLR by Visser et al. and therefore excluded as duplicates. AUSCAN, Australian/Canadian Hand Osteoarthritis Index; CTD, clinical trial discrimination; CV, construct validity; LCV, longitudinal construct validity; MHQ, Michigan Hand Outcomes Questionnaire; SLR, systematic literature review, TM, thresholds of meaning; TRT, test-retest reliability.

Only one published study was available to provide evidence on thresholds of meaning of the MHQ. Therefore, we conducted a separate analysis to estimate MCII and PASS of both instruments. While AUSCAN had a credible MCII (e.g., relative percentage improvement of 9.8%), comparable to previous literature, the MCII estimates of MHQ were small values in the wrong direction of effect, which would indicate that even a small worsening was rated as a functional improvement. PASS values lie around 50% of the possible maximum score for AUSCAN and MHQ overall hand function, while for MHQ ADL a relatively high PASS was indicative of a floor effect. All in all, these results seem more supportive of AUSCAN than MHQ when it comes to measurement of improvement in a clinical trial setting, in line with the results for clinical trial discrimination that we found in the literature.

Final rating. Fig. 3 presents the SOMP table of AUSCAN (A) and MHQ (B). The combined rating of the evidence was supportive of a provisional endorsement of both MHQ subscales as core outcome measurement instrument (Fig. 3B). The working group noted the

A. OMERACT Summary of Measurement Properties table: AUSCAN function

Author	Truth	Feasibility	Truth	Discrimination			
	Domain match		Construct validity	Test retest reliability	Longitudinal construct validity	Clinical trial discrimination	Thresholds of meaning
Working Group Appraisal	+	±					
Allen, 2006			+				
Bellamy, 2002			+	+		+	
Dziedzic, 2007			+	+			
Fernandes, 2012					+		
Haugen, 2011							
MacDermid, 2007			+				
Moe, 2010			+	+			
Sautner 2009			+				
Stamm, 2007			+				
Wittoek, 2009			+				
Altman, 2009						+	
Grifka, 2004						+	
Keen, 2010						+	
Bijsterbosch, 2011					+		
Botha-Scheepers, 2009					+		
Haugen, 2013					+		
Marshall, 2013					+		
Dilek, 2013						+	
Brosseau, 2005						+	
Kjeken, 2011						±	
Bellamy, 2007							+
Bellamy, 2015							+
Siviero, 2019							+
Kroon, 2019						+	
Working Group, unpublished							+
Studies for each property (n)			9	4	5	8	4
Studies for synthesis (n)			8	3	5	8	4
Synthesis rating	GREEN	AMBER	AMBER	GREEN	AMBER	GREEN	GREEN
OMERACT endorsement	AMBER: Provisionally endorsed Evidence supports provisional endorsement of AUSCAN function as a core measurement instrument of function.						

B. OMERACT Summary of Measurement Properties table: MHQ overall function / activities of daily living

Author	Truth	Feasibility	Truth	Discrimination			
	Domain match		Construct validity	Test retest reliability	Longitudinal construct validity	Clinical trial discrimination	Thresholds of meaning
Working Group Appraisal	+[a,b]	+[a,b]					
Marks, 2014			±[a,b]	+[a,b]	+[a,b]		+[a,b]
Bijsterbosch, 2010			+[a,b]				
Kroon, 2018			+[a,b]		+[a,b]		
Poole, 2010				+[a,b]			
Sears, 2010					+[a,b]		
Adams, 2014						+[a]	
Arazpour, 2017						+[a,b]	
Kroon, 2019						±[a,b]	
Working Group, unpublished							±[a,b]
Studies for each property (n)			4	2	3	3	2
Studies for synthesis (n)			3	2	3	3	2
Synthesis rating	GREEN	GREEN	GREEN	GREEN	AMBER	AMBER	AMBER
OMERACT endorsement	AMBER: Provisionally endorsed Evidence supports provisional endorsement of both MHQ overall function and MHQ activities of daily living subscales as core outcome measurement instrument of function.						

Fig. 3. OMERACT Summary of Measurement Properties (SOMP) tables of AUSCAN (function subscale) (A) and MHQ (overall function[a] and activities of daily living[b] subscales) (B). The color represents the good methods assessment (green, good; amber, noncritical limitation; red, unsuitable evidence). The sign represents the adequacy of the data in support of the instrument (+, positive support; ±, ambivalent support; −, does not reach performance standards for metric property). For MHQ, superscript letters denote whether the evidence in that cell pertains to the overall function[a] or the activities of daily living[b] subscale. AUSCAN, Australian/Canadian Hand Osteoarthritis Index; MHQ, Michigan Hand Outcomes Questionnaire.

need to re-assess clinical trial discrimination in future clinical trials on their research agenda. AUSCAN received a provisional endorsement to serve as a second measure of function (Fig. 3A). While AUSCAN function may have better metric properties than MHQ, the working group felt that due to important feasibility issues (i.e., not available in public domain, costs associated with use of questionnaire), this instrument could not be recommended as a mandatory instrument to measure function in all hand OA trials.

Discussion

We evaluated appropriate instruments to measure the domain physical function in hand OA, as part of the development of an OMERACT core outcome measurement set. We identified the MHQ and AUSCAN as the first patient-reported instruments provisionally endorsed as core outcome measurement instruments for hand OA clinical trials, based on the OMERACT Filter 2.1 process.

Whilst at the moment the AUSCAN is more commonly used to measure physical function in hand OA studies, this instrument has important limitations in feasibility, i.e. the questionnaire is copyrighted and may only be used with permission and after payment of a considerable fee. This led to the decision to not recommend AUSCAN as a primary measure of function. Such limitations in feasibility were not present for MHQ. However, particularly evidence for clinical trial discrimination and thresholds of meaning was inconsistent, and further research is therefore warranted. Full endorsement of the MHQ as the core outcome measurement instrument for physical function depends on the results of future studies into the remaining amber metric properties (longitudinal construct validity, clinical trial discrimination and thresholds of meaning). It would require supportive evidence ("+") for these metric properties provided by studies with good methods or a low risk of bias ("green"). Full endorsement of AUSCAN is not possible in its current form given the identified limitations in terms of feasibility.

Whilst the FIHOA was proposed as a candidate instrument in the preliminary core instrument set [4], it was rejected as a core outcome measure for hand OA trials, after careful assessment of domain match and feasibility.

Notably, many of the studies on (longitudinal) construct validity for AUSCAN and MHQ were rated as having amber methods due to lack of (reporting of) pre-specified hypotheses, yet the results of the studies were generally supportive of the measurement property. Reporting standards for future studies of measurement properties of instruments could be useful to improve this.

The working group also aims to publish the additional analyses on thresholds of meaning in a peer-reviewed journal.

Our study is not the first to review and appraise different measurement instruments of hand function for patients with hand OA. In 2005, Dziedzic et al. published a narrative review describing and critically appraising five instruments used to measure disability in patients with hand OA, including the AUSCAN and FIHOA [47]. While they also used the OMERACT constructs of truth, discrimination and feasibility, their work differs from ours in several aspects. They did not focus on instruments measuring physical function of the hand specifically (i.e., the core domain for which we wanted to select a suitable measurement instrument), but also assessed instruments that measure 'disability' in a broader sense, such as the Health Assessment Questionnaire. Besides, their objective was to identify and appraise instruments for a specific application (general population surveys in primary care setting), while our work expands to instruments suitable for studies in all settings. Moreover, their search was performed until 2002, they used the old OMERACT framework (Filter 1.0), did not critically appraise the quality of the studies providing data on the measurement properties of instruments, and no voting in the OMERACT community was done or level of endorsement presented. A recent narrative review by van de Stadt et al. provides an overview of all available instruments to measure hand function, including an overview of evidence for their measurement properties, without focusing on hand osteoarthritis specifically [3].

Physical function is the first core domain from the OMERACT hand OA core outcome set for which core outcome measurement instruments have been endorsed. The working group will proceed with the assessment of instruments for the other core domains, with the ultimate goal of developing a complete hand OA core outcome measurement set.

Author contribution

FK, DvdH and MK were responsible for conceptualization and methodology of the study. FK, AA, FBl, KD, CH, IKH, VR, TS and RW were responsible for data curation. FK, DB, LM, DvdH and MK were responsible for formal analysis and interpretation. FK was responsible fot. All authors critically revised the manuscript and approved the final version.

Declaration of Competing Interest

AA reports grants from AstraZeneca, Oxford Immunotech; royalties from Uptodate; consulting fees from inflazome, NGM Biopharmaceuticals; honoraria from Menarini; and support from Pfizer; all outside the submitted work. FBe reports personal fees from Boehringer, Bone Therapeutics, CellProthera, Expanscience, Galapagos, Gilead, GSK, Merck Sereno, MSD, Nordic, Novartis, Pfizer, Regulaxis, Roche, Sandoz, Sanofi, Servier, UCB, Peptinov, 4P Pharma, 4Moving Biotech, and grants from TRB Chemedica; all outside the submitted work. FBl reports grants from Abbvie, grants and personal fees from Pfizer, grants from UCB, grants from Bristol-Mayers Squibb, grants from Roche, grants from Servier, grants from Bioiberica, grants from Sanofie, grants from Grunenthal, grants from GlaxoSmithKline, grants from Lilly, grants from Janssen, grants from Regeneron, grants from Amgen, and grants from TRB Chemedica; all outside the submitted work. DB reports to be on the executive of OMERACT and Technical Advisory Group, and is involved in supporting upper limb outcome measures (none that were considered here). PC reports personal fees from Eli Lilly, personal fees from EMD Serono, personal fees from Flexion Therapeutics, personal fees from Galapagos, personal fees from Novartis, personal fees from Pfizer; all outside the submitted work. KD reports grants from the National Institute for Health Research (NIHR). TS reports personal fees from AbbVie, Sanofi Genzyme, Roche, and Takeda; all outside the submitted work. IKH reports personal fees from AbbVie, grants from Pfizer, and personal fees from Novartis; all outside the submitted work. DvdH reports personal fees from AbbVie, Amgen, Astellas, AstraZeneca, Bayer, BMS, Boehringer Ingelheim, Celgene, Cyxone, Daiichi, Eisai, Galapagos, Gilead, GlaxoSmith-Kline, Janssen, Lilly, Merck, Novartis, Pfizer, Regeneron, Roche, Sanofi, Takeda, and UCB Pharma; all outside the submitted work; and is Director of Imaging Rheumatology BV. MK reports grants from Dutch Arthritis Association, during the conduct of the study, fee for consultancy (Abbvie, Pfizer, Levicept, GlaxoSmithKline, Merck-Serono, Kiniksa, Flexion, Galapagos, CHDR) and local investigator of industry-driven trial (Abbvie), from Wolters Kluwer (UptoDate), Springer Verlag (Reumatologie en klinische immunologie), grants from Pfizer, grants from IMI-APPROACH; all outside the submitted work. LM reports to be a paid staff member of OMERACT and helped develop the methodology applied within this manuscript. FK, CH, MI, VR, and RW have nothing to disclose.

Acknowledgments

We would like to thank the patient research partners for their valuable contributions.

We thank the other OMERACT hand OA working group members who completed the survey for domain match and feasibility (R. Altman, M. Boers, M. Englund, H. Jonsson, T. Karjalainen, I. Kjeken, M. Kortekaas, M. Marshall, B.F. Leeb, Y-Y. Leung, P. Steen Pettersen, R. Ramonda, Z. Szekanecz, A-L Tan, P. Taylor, T. Woodworth). We are grateful for the OMERACT hand OA working group members who translated the patient research partner questionnaire to the local language.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

KD is part funded by the National Institute for Health Research (NIHR) Applied Research center (ARC) West Midlands. KD was also part funded by an NIHR Knowledge mobilisation Research Fellowship (KMRF-2014-03-002) and is an NIHR Senior Investigator. The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care.

PGC is supported in part by the UK National Institute for Health Research (NIHR) Leeds Biomedical Research center. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR, or the Department of Health.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.semarthrit.2021.08.014.

References

[1] Zhang Y, Niu J, Kelly-Hayes M, Chaisson CE, Aliabadi P, Felson DT. Prevalence of symptomatic hand osteoarthritis and its impact on functional status among the elderly: the Framingham Study. Am J Epidemiol 2002;156(11):1021–7.
[2] Kwok WY, Vliet Vlieland TP, Rosendaal FR, Huizinga TW, Kloppenburg M. Limitations in daily activities are the major determinant of reduced health-related quality of life in patients with hand osteoarthritis. Ann Rheum Dis 2011;70(2):334–6.
[3] van de Stadt LA, Kroon FPB, Kloppenburg M. Measures of Hand Function. Arthritis Care Res (Hoboken) 2020;72(Suppl 10):321–41.
[4] Kloppenburg M, Bøyesen P, Visser A, et al. Report from the OMERACT hand osteoarthritis working group: set of core domains and preliminary set of instruments for use in clinical trials and observational studies. J Rheumatol 2015;42:2190–7.
[5] Beaton D, Maxwell L, Shea B, et al. Instrument Selection Using the OMERACT Filter 2.1: the OMERACT methodology. J Rheumatol 2019;46:1028.
[6] Boers M, Beaton D, Shea B, et al. OMERACT Filter 2.1: elaboration of the conceptual framework for outcome measurement in health intervention studies. J Rheumatol 2019;46:1021.
[7] OMERACT Handbook. Chapter 5: instrument selection for core outcome measurement sets: available at: https://omeracthandbook.org/handbook; 2019.
[8] Wittoek R, Kroon F, Kundakci B, et al. Report from the hand osteoarthritis working group at OMERACT 2018: update on core instrument set development. J Rheumatol 2019;46:1183–7.
[9] The OMERACT Handbook. https://omeracthandbook.org/handbook; 2019.
[10] Visser AW, Bøyesen P, Haugen IK, et al. Instruments measuring pain, physical function, or patient's global assessment in hand osteoarthritis: a systematic literature search. J Rheumatol 2015;42(11):2118–34.
[11] Marks M, Schoones JW, Kolling C, Herren DB, Goldhahn J, Vliet Vlieland TP. Outcome measures and their measurement properties for trapeziometacarpal osteoarthritis: a systematic literature review. J Hand Surg Eur 2013;38:822–38.
[12] Bellamy N, Campbell J, Haraoui B, et al. Clinimetric properties of the AUSCAN Osteoarthritis Hand Index: an evaluation of reliability, validity and responsiveness. Osteoarthritis Cartilage 2002;10(11):863–9.
[13] Dreiser RL, Maheu E, Guillou GB, Caspard H, Grouin JM. Validation of an algofunctional index for osteoarthritis of the hand. Rev Rhum Engl Ed 1995;62(6):43s–53s Suppl 1.
[14] Chung KC, Pillsbury MS, Walters MR, Hayward RA. Reliability and validity testing of the Michigan Hand Outcomes Questionnaire. J Hand Surg Am 1998;23(4):575–87.
[15] Damman W, Liu R, Kroon FPB, et al. Do comorbidities play a role in hand osteoarthritis disease burden? Data from the hand osteoarthritis in secondary care cohort. J Rheumatol 2017;44(11):1659–66.
[16] Kroon FPB, Carmona L, Schoones JW, Kloppenburg M. Efficacy and safety of non-pharmacological, pharmacological and surgical treatment for hand osteoarthritis: a systematic literature review informing the 2018 update of the EULAR recommendations for the management of hand osteoarthritis. RMD Open 2018;4(2):e000734.
[17] Kroon FPB, Boersma A, Boonen A, et al. Performance of the Michigan Hand Outcomes Questionnaire in hand osteoarthritis. Osteoarthritis Cartilage 2018;26(12):1627–35.
[18] Kroon FPB, Kortekaas MC, Boonen A, et al. Results of a 6-week treatment with 10mg prednisolone in patients with hand osteoarthritis (HOPE): a double-blind, randomised, placebo-controlled trial. Lancet 2019;394(10213):1993–2001.
[19] Bellamy N, Hochberg M, Tubach F, et al. Development of multinational definitions of minimal clinically important improvement and patient acceptable symptomatic state in osteoarthritis. Arthritis Care Res (Hoboken) 2015;67(7):972–80.
[20] Allen KD, Jordan JM, Renner JB, Kraus VB. Validity, factor structure, and clinical relevance of the AUSCAN Osteoarthritis Hand Index. Arthritis Rheum 2006;54(2):551–6.
[21] Dziedzic KS, Thomas E, Myers H, Hill S, Hay EM. The Australian/Canadian osteoarthritis hand index in a community-dwelling population of older adults: reliability and validity. Arthritis Rheum 2007;57(3):423–8.
[22] Fernandes L, Grotle M, Darre S, Nossum R, Kjeken I. Validity and responsiveness of the Measure of Activity Performance of the Hand (MAP-Hand) in patients with hand osteoarthritis. J Rehabil Med 2012;44(10):869–76.
[23] Moe RH, Garratt A, Slatkowsky-Christensen B, et al. Concurrent evaluation of data quality, reliability and validity of the Australian/Canadian Osteoarthritis Hand Index and the Functional Index for Hand Osteoarthritis. Rheumatology (Oxford) 2010;49(12):2327–36.
[24] Sautner J, Andel I, Rintelen B, Leeb BF. A comparison of the modified score for the assessment of chronic rheumatoid affections of the hands and the australian/canadian osteoarthritis hand index in hand osteoarthritis patients. Int J Rheumatol 2009:249096 2009.
[25] Stamm T, Mathis M, Aletaha D, Kloppenburg M, Machold K, Smolen J. Mapping hand functioning in hand osteoarthritis: comparing self-report instruments with a comprehensive hand function test. Arthritis Rheum 2007;57(7):1230–7.
[26] Wittoek R, Cruyssen BV, Maheu E, Verbruggen G. Cross-cultural adaptation of the Dutch version of the Functional Index for Hand Osteoarthritis (FIHOA) and a study on its construct validity. Osteoarthritis Cartilage 2009;17(5):607–12.
[27] MacDermid JC, Wessel J, Humphrey R, Ross D, Roth JH. Validity of self-report measures of pain and disability for persons who have undergone arthroplasty for osteoarthritis of the carpometacarpal joint of the hand. Osteoarthritis Cartilage 2007;15:524–30.
[28] Haugen IK, Moe RH, Slatkowsky-Christensen B, Kvien TK, van der Heijde D, Garratt A. The AUSCAN subscales, AIMS-2 hand/finger subscale, and FIOHA were not unidimensional scales. J Clin Epidemiol 2011;64(9):1039–46.
[29] Bijsterbosch J, Watt I, Meulenbelt I, Rosendaal FR, Huizinga TW, Kloppenburg M. Clinical and radiographic disease course of hand osteoarthritis and determinants of outcome after 6 years. Ann Rheum Dis 2011;70(1):68–73.
[30] Botha-Scheepers S, Riyazi N, Watt I, et al. Progression of hand osteoarthritis over 2 years: a clinical and radiological follow-up study. Ann Rheum Dis 2009;68(8):1260–4.
[31] Haugen IK, Slatkowsky-Christensen B, Bøyesen P, van der Heijde D, Kvien TK. Cross-sectional and longitudinal associations between radiographic features and measures of pain and physical function in hand osteoarthritis. Osteoarthritis Cartilage 2013;21(9):1191–8.
[32] Marshall M, Peat G, Nicholls E, van der Windt D, Myers H, Dziedzic K. Subsets of symptomatic hand osteoarthritis in community-dwelling older adults in the United Kingdom: prevalence, inter-relationships, risk factor profiles and clinical characteristics at baseline and 3-years. Osteoarthritis Cartilage 2013;21(11):1674–84.
[33] Altman RD, Dreiser RL, Fisher CL, Chase WF, Dreher DS, Zacher J. Diclofenac sodium gel in patients with primary hand osteoarthritis: a randomized, double-blind, placebo-controlled trial. J Rheumatol 2009;36(9):1991–9.
[34] Brosseau L, Wells G, Marchand S, et al. Randomized controlled trial on low level laser therapy (LLLT) in the treatment of osteoarthritis (OA) of the hand. Lasers Surg Med 2005;36(3):210–9.
[35] Dilek B, Gözüm M, Şahin E, et al. Efficacy of paraffin bath therapy in hand osteoarthritis: a single-blinded randomized controlled trial. Arch Phys Med Rehabil 2013;94(4):642–9.
[36] Grifka JK, Zacher J, Brown JP, et al. Efficacy and tolerability of lumiracoxib versus placebo in patients with osteoarthritis of the hand. Clin Exp Rheumatol 2004;22(5):589–96.
[37] Keen HI, Wakefield RJ, Hensor EM, Emery P, Conaghan PG. Response of symptoms and synovitis to intra-muscular methylprednisolone in osteoarthritis of the hand: an ultrasonographic study. Rheumatology (Oxford) 2010;49(6):1093–100.
[38] Kjeken I, Darre S, Smedslund G, Hagen KB, Nossum R. Effect of assistive technology in hand osteoarthritis: a randomised controlled trial. Ann Rheum Dis 2011;70(8):1447–52.
[39] Bellamy N, Wilson C. International estimation of minimally clinically important improvement (MCII75): the Reflect study. Austr Rheumatol Assoc Ann Sci Meeting 2007;37(Suppl 2):A36.
[40] Siviero P, Limongi F, Gesmundo A, et al. Minimal clinically important decline in physical function over one year: EPOSA study. BMC Musculoskelet Disord 2019;20(1):227.
[41] Bijsterbosch J, Watt I, Meulenbelt I, Rosendaal FR, Huizinga TW, Kloppenburg M. Clinical burden of erosive hand osteoarthritis and its relationship to nodes. Ann Rheum Dis 2010;69(10):1784–8.

[42] Marks M, Audigé L, Herren DB, Schindele S, Nelissen RG, Vliet Vlieland TP. Measurement properties of the German Michigan Hand Outcomes Questionnaire in patients with trapeziometacarpal osteoarthritis. Arthritis Care Res (Hoboken) 2014;66(2):245–52.
[43] Sears ED, Chung KC. Validity and responsiveness of the Jebsen-Taylor hand function test. J Hand Surg Am 2010;35(1):30–7.
[44] Poole JL, Lucero SL, Mynatt R. Self-reports and performance-based tests of hand function in persons with osteoarthritis. Phys Occup Ther Geriatr 2010;28:249–58.
[45] Adams J, Bouças SB, Hislop K. The effectiveness and efficacy of splints for thumb base osteoarthritis: a pilot randomized controlled trial. Rheumatology 2014;53(Suppl 1) i41-2.
[46] Arazpour M, Soflaei M, Ahmadi Bani M, et al. The effect of thumb splinting on thenar muscles atrophy, pain, and function in subjects with thumb carpometacarpal joint osteoarthritis. Prosthet Orthot Int 2017;41:379–86.
[47] Dziedzic KS, Thomas E, Hay EM. A systematic search and critical review of measures of disability for use in a population survey of hand osteoarthritis (OA). Osteoarthritis Cartilage 2005;13:1–12.