Article:
Oo, WM, Linklater, JM, Daniel, M et al. (6 more authors) (2018) Clinimetrics of ultrasound pathologies in osteoarthritis: systematic literature review and meta-analysis. Osteoarthritis and Cartilage, 26 (5). pp. 601-611. ISSN 1063-4584

https://doi.org/10.1016/j.joca.2018.01.021

Clinimetrics of Ultrasound Pathologies in Osteoarthritis: Systematic Literature Review and Meta-analysis

Introduction

Osteoarthritis (OA) is the ubiquitous joint disease, predisposing to severe disability and economic burden on the community [1], with its prevalence surging world-wide due to an increase in ageing population [2]. Pathophysiology of OA is complex and involves multiple tissue pathologies; there is currently no consensus on which manifestations should be measured in OA clinical studies. In attempting to objectively evaluate OA structural components, X-ray and MRI have been commonly employed as they visualize constructs related to cartilage. Ultrasound has been less well studied, but does provide certain advantages such as real-time assessment of multiple joints, sensitive visualisation of synovitis without the need for contrast agents [3-5], its detection of pathologies such as meniscus extrusion [6-9], osteophytes [10-12], degeneration of femoral trochlear cartilage [13-16], and effusions (which might be missed on clinical examination or plain radiography) [5, 17-19]. As a result of these attributes, and likely because of widespread uptake in the rheumatology community, ultrasound has increasingly been applied as an outcome tool in OA clinical studies over the last decade.

Since Keen et al. reported its clinimetrics, mainly with a focus on validity, in a systematic review in 2009, based on PubMed and Medline database searches [20], many ultrasound OA studies have been published according to recent narrative reviews [21, 22], with most papers having sound methodology, utilizing more advanced technology such as high-

frequency probes, and use of definitions and techniques from Outcome Measures in Rheumatology (OMERACT) [23] and European League Against Rheumatism (EULAR) Ultrasound Working Groups [24]. The increase in knowledge base in this area, therefore, warrants an update of the previous review in terms of clinimetrics (clinical measurement) such as reliability, validity, responsiveness [25]. Moreover, there is no published meta-analysis on these clinimetric of commonly assessed ultrasound pathologies in OA.

Therefore, the purposes of this study were: (1) to systematically review the performance metrics of ultrasound as applied to the detection of commonly assessed pathologies in people with OA with a focus on knee, hand and hip jointsand (2) to conduct a meta-analysis of each clinimetric propertyfor the ultrasound findings if feasible.

Methodology

Selection criteria

Manuscripts were included if 1) they reported clinimetrics of commonly assessed ultrasound pathologies in knee or hand or hip OA in adults, and 2) separate clinimetrics for OA were recorded if the sample included different rheumatic diseases. Articles were excluded if 1) they were not related to the use of B-mode or color/power Dopplerultrasound, 2) they utilized ultrasound only for injection guidance, 3) they did not provide any ultrasound clinimetrics, or 4) they were review or editorial articles, non-human or non-English publications. The study protocol was registered in PROSPERO database with CRD42016039954.

Information source and selection process

One reviewer (WMO) searched MEDLINE via Ovid, EMBASE, and Cochrane Library databases from their respective inception to September 2016. The search strategy for each database was developed in consultation with an experienced librarian (**supplementary data 1**). The same reviewer implemented the secondary searching in reference lists of included articles, ultrasound chapters in reference books, and conference abstracts of Osteoarthritis Research Society International (OARSI), EULAR and American College of Rheumatology (ACR) from 2014 to 2016.

The retrieved articles were imported into Covidence systematic review software [26], and two reviewers (WMO and MD) screened the titles and abstracts independently. Subsequently, the full texts of the selected articles were retrieved and judged against the inclusion and exclusion criteria. Any disagreement was resolved with a third reviewer (DJH). When the included studies referred to a previous paper for methodology or reliability, it was obtained, and appraised if it met the selection criteria.

Data extraction and quality assessment

According to the OMERACT Instrument Selection Algorithm [27], the same two reviewers conducted data extraction with a standardized excel template including: 1) characteristics of studies such as study design, setting, sample size, participants selection and

diagnostic criteria; 2) technical features such as ultrasound mode (i.e. B-mode, Power Doppler), machine settings,scanning methods, the particular joints and structures scanned; 3) pathological findings such as ultrasound definitions of pathologies and scoring methods; 4) types of clinimetrics.

For reliability, imaging and operator characteristics were recorded. Construct validity was defined if the study correlated ultrasound findings with clinical assessment, plain radiography or MRI. Criterion/predictive validity was defined when ultrasound findings were concurrently or predictively compared with the gold standard, i.e. histopathology, arthroscopy. Discriminative validity was also assessed in two aspects: internal responsiveness (the ability of ultrasound measure to change over a pre-specified time frame) or external responsiveness (the extent to which changes in ultrasound measure relate to corresponding changes in a reference measure of health status) for interventional studies. Feasibility was calculated in scanning time required for the whole ultrasound examination. One reviewer (WMO) appraised the methodological quality, using the modified 19-item version (**supplementary data 2**) derived from Downs and Black score system [28, 29] for all included papers, and 11-item Quality Appraisal of Diagnostic Reliability (QAREL) score for reliability papers [30].

Pooling Criteria for Meta-analysis

For meta-analysis, data were pooled if the paper reported sufficient data to calculate 1) kappa or ICC for reliability, 2) Pearson and Spearman correlation coefficients for validity, 3)

standardized mean difference for internal responsiveness, 4) correlation coefficient for external responsiveness. For validity, all types of regression coefficients (β) were omitted from pooling due to controversy in combining them [31].

Statistical analysis.

Qualitative analysis

Frequencies and percentages were computed for categorical variables of included papers.

Meta-analysis and Meta-regression

Unit of analysis: Each sample of subjects from studies was assumed as one unit of analysis. When two or more articles documented reliability/correlation coefficients, using the same sample, the coefficient was included only once as the unit of analysis. When one article reported more than one reliability/correlation coefficients of the same clinimetric measurement from the same sample, the mean coefficient was calculated, and then analyzed in the meta-analysis. If the study comprised independent subgroups, the subgroups were pooled as a separate unit of analysis [32].

Pooling data: Separate meta-analyses were performed for each type of clinimetrics: 1) kappa or ICC for inter-rater or intra-rater reliability 2) construct validity against healthy control, pain, functional assessment, conventional X-rays, MRI, or biomarkers, 3) internal or external responsiveness. These data were pooled, based on each ultrasound pathology (synovitis/effusion/

osteophyte/etc.) to be clinically meaningful. For reliability statistics, pooling was stratified for each grading method (binary/semi-quantitative/quantitative) of the same ultrasound pathology.

For weighted meta-analysis of kappa estimates, when the standard error (SE) was unavailable, it was calculated from 95% confidence interval (CI) bounds [33]. If both SEs and CIs were not reported, the largest observed SE from the included studies was used. For ICC statistics of reliability and Pearson or Spearman correlation coefficients of validity, effect sizes were first obtained through the z-transformations, and then the resulting pooled effect sizes were back-transformed (z to r transformation) to the level of original coefficients for easier interpretation [34]. For merging odd ratios in validity studies, the log odds ratio and the standard error of the log odds ratio were determined [35]. The standardized mean difference (SMD), using Hedges' g due to inclusion of small studies (<30 patients/joints), was calculated for internal responsiveness [36], and correlation coefficients were pooled for external responsiveness through the z-transformations [37].

For assessment of heterogeneity, Cochran Q test was computed [34]. The I^2 was used to quantify how much of the total variability can be attributed to heterogeneity [38]. To scrutinize possible publication bias, it was intended to evaluate with funnel plot techniques [39], Begg's rank test [40] and Egger's regression test [41], as appropriate, given the known limitations of these methods, if the minimum number of studies could be pooled. All analyses for calculating the estimates from primary studies, and for pooling data were carried out by using the SPSS, Excel and Comprehensive Meta-analysis software.

Results

Identification of included studies

Our search identified 1246 records (468 Medline, 774 Embase and 4 Cochrane library) with 120 duplicates. After screening the titles and abstracts, 195 articles remained. Furthermore, 9 articles were retrieved from the reference lists, totalling 204 articles eligible for full-text review. Of these, 100 articles were selected as shown in the PRISMA flow diagram (**Figure. 1**).

Study characteristics

One hundred articles (listed in **supplementary data 3**), having a total of 8542 patients and 32373 OA joints, and published between 1982 and 2016, were included in the systematic review. The studies' characteristics were summarized in **supplementary data 4**. Majority of studies (79%) were documented after 2008. Knee OA was the most widely investigated (n=64), followed by hand OA (n=28), and hip OA (n=8).

According to Oxford Centre for Evidence-Based Medicine guidelines (www.cebm.net/), 42 papers utilized a cross-sectional design (42%) and 28 papers applied a cohort design (28%). The participants were recruited from out-patient rheumatology clinics in 46 papers; the setting was not mentioned in 23 papers. The selection method was not described in half of the studies, followed by a consecutive method (n=40), convenience (n=5) and random methods (n=5). ACR criteria was employed for diagnosis in most of studies (n=81); 14 papers did not disclose

diagnostic criteria. The mean age of included studies ranged from 50.1±9.2 to 71.9±5.9 years; female participants varied from 37% to 100%; the mean BMI from 22.2±2.6 to 33.5±4.6 kg/m^2. Eight studies recruited mixed samples with different diseases, but delineated separate clinimetrics of OA sub-group.

Ultrasound scanning techniques and definition

For simplicity, the EULAR scanning method [42] and OMERACT definitions [23] were assumed as the standard criteria to identify respective OA pathologies. Out of 100 papers, power Doppler was inverstigated in 31 (**supplementary data 5).** Doppler specifications were detailed in 19 papers: Doppler frequency was reported in 9 (from 12 MHz to 6.3 MHz); pulse repetition frequency (PRF) in 10 (from 13.2KHz to 3 Hz); wall filter and gain in 17. One paper examined contrast ultrasound.

Eighty-eight papers defined ultrasound pathology; 26 papers referred the EULAR scanning protocol; 59 papers administered their own methods or modification from previous papers; 13 papers did not delineate the specific scanning method. Thirty-nine studies applied the OMERACT definitions, which were found to be increasingly used across the years from 1 paper in 2008, and then 5 papers in 2012 to 10 papers in 2016 (**Supplementary data 6**).

Ultrasound lesions and scoring system

Overall, synovial pathologies were more extensively examined, i.e, effusion (52%),

synovial hypertrophy (37%), Doppler activity (31%), Baker's cyst (25%), compared to structural lesions, i.e, osteophyte (29%), cartilage thinning (28%). A variety of grading systems was evaluated [binary (n=49,49%), semi-quantitative (n=42, 42%), and quantitative (n=40,40%)].

Qualification of ultrasound operator

Only twenty papers declared the number of operator's training years in musculoskeletal ultrasound, ranging from 3 months to 24 years. The operator/readers were also of diverse academic backgrounds: rheumatologist (27% of all papers), ultrasonographer (16%), radiologist (11%), others such as physiatrist, surgeon, fellow-in-training (26%), and no report (20%).

Methodological quality

The average quality score across the studies assessed with the modified Downs and Black instrument was 13.01 out of 19 items (taking into account the questions that were not applicable for certain studies). **The chart in supplementary data 7** outlined the proportion of the 100 studies that met each of the quality assessment items. The papers, in general, had a good rating (>60%) on the 13 items. However, most papers fell short severely on some items such as reporting of sample size calculation and sufficient power (10%).

The average QAREL score was 5.93 out of 11 items across all reliability studies (n=43). Blindness to other raters, own prior findings, clinical information and non-clinical clues were described in 40% (n=17), 28% (n=12), 56% (n=24) and 5% (n=2), respectively (**supplementary**

data 8). Randomization of patients/raters was found only in 53% (n=23). As there was no definite consensus related to time interval for stability of ultrasound findings between repeated measurements, only evaluation of stored images was given as yes (n=17), and rating of the acquired image as unclear (n=26). Overall, the regression plot displayed the significant improvement of QAREL quality score across the years (β=0.40, P=0.01) (**Supplementary data 9**).

Clinimetric properties

Among the 100 studies, 32 papers were identified for the intra-rater reliability, 25 for inter-rater reliability, 57 for construct validity, 5 for criterion validity in knee, 10 for clinical predictive validity, 6 for structural predictive validity, 21 for intrinsic responsiveness, 8 for extrinsic responsiveness and 7 for feasibility.

Meta-analysis

The meta-analysis was conducted only for knee OA. Pooling could not be performed for hand and hip OA due to a paucity of reported clinimetric data for ultrasound, and so descriptive analysis was presented. Publication bias was not examined due to inadequate numbers of included papers for a specific OA pathology, which did not allow proper assessment of funnel plots or more advanced regression-based assessments.

Knee OA

Reliability:

Inter-rater reliability: According to the pooling criteria, **s**tratified kappa meta-analysis was conducted across 11 knee studies, including 38 kappa estimates and 556 joints of 506 patients. ICC estimates was pooled across 7 knee studies with a total of 19 ICC estimates in 340 joints of 308 participants. Kappa coefficients were interpreted according to Landis and Koch (0:poor; 0.01-0.20:slight; 0.21-0.40:fair; 0.41-0.60:moderate; 0.61-0.80:substantial; 0.81-1.00:almost perfect) [43].

The pooled kappa of binary score (**Table 1**) was almost perfect for Baker's cyst [0.92(0.83-1)], and substantial for effusion [0.75(0.41,1)] (**Figure 2**), with nearly all pathologies revealing considerable heterogeneity (I^2=70 to 99). For semi-quantitative score, pooled kappa values were moderate for cartilage thinness [0.44(0.15-0.74)], and substantial for all pathologies, with high heterogeneity (I^2= 78-98). For quantitative scores, all pathologies provided almost perfect reliability for pooled ICC estimate.

Intra-rater reliability: Stratified kappa meta-analysis was peformed from 8 knee studies, including a total of 23 kappa estimates for 502 joints of 465 patients. For ICC values, data were pooled from 9 knee studies with a total of 21 ICC estimates for 566 joints of 490 participants.

The pooled kappa of semi-quantitative score (**Table 2**) was varied from moderate for cartilage thinness [0.55(0.45-0.66)], substantial for synovitis [0.69(0.60-0.78)] and osteophyte

[0.74(0.67-0.81)] to almost perfect for meniscal extrusion [0.81(0.66-0.96)], exhibiting low heterogeneity (I^2=7 to 51). For quantitative scores, reliability was almost perfect in all pathologies.

Validity

Meta-analysis was stratified for each comparator such as asymptomatic controls, pain, function, X-rays, MRI or blood biomarkers or histology or arthroscopy. Correlation coefficients were interpreted according to the Evans' classification [44], <0.20:very weak; 0.20-0.39:weak; 0.40-0.59:moderate; 0.60-0.79;strong and >0.80:very strong.

Construct validity against asymptomatic controls: Six studies, including 643 joints from 582 participants, provided 23 odd ratios. In symptomatic patients (**Table 3**), the pooled odd ratio demonstrated a very strong association with effusion [7.46(2.56,21.70)], and a strong association with Baker's cyst [3.23(1.57,6.67)] and meniscal extrusion [3.08(1.06,8.92)]. Heterogeneity was generlly moderate (I^2=41 to 61).

Construct validity against pain: Pooling 37 estimates out of 16 studies, including 2577 joints from 2085 patients, revealed weak correlation with trivial heterogeneity [I^2=0] (**Table 4**).

Construct validity against function: Meta-analysis of 15 estimates out of 9 studies,

including 1333 joints and 802 patients, resulted in weak correlation, and mild heterogeneity [I^2=20-38] (**supplementary data 10**). Six studies used WOMAC [45].

Construct validity against X-rays: Pooling across a total of 49 estimates from 11 studies (1956 joints, and 1530 patients) indicated strong correlation with osteophyte [0.60(0.45,0.71)], moderate correlation with effusion [0.54(0.37,0.68)] and meniscal extrusion [0.48(0.34,0.60)], and weak association with cartilage thickness [0.35(0.12,0.55)]. Heterogeneity was moderate [I^2=34-52] (**Table 5**). Kellgren Lawrence score [46] was applied in 10 studies.

Construct validity against MRI: Strong correlation (r>0.60) was detected on pooling 29 estimates across 4 studies examining 306 knee joints in 230 patients, using 0.2T to 1.5 T MRI with dedicated knee coils (**supplementary data 10**).

Construct validity against biomarkers: Twenty-three estimates of serum cartilage oligomeric matrix protein (COMP) were pooled across 4 studies involving 95 knee joints from 95 patients, generating weak correlation [r=0.003 to 0.21] with trivial heterogeneity [I^2=0] (**supplementary data 10**).

Criteria validity against histology: Pooling of four estimates from 2 studies, evaluating histological cartilage thickness in 190 knee joints from 113 patients, produced a

moderate correlation [r=0.66(-0.05-0.93)], and considerable heterogeneity [I^2=90] (**supplementary data 10**).

Criteria validity against arthroscopy: Ultrasound pathologies focused by three arthroscopic studies, using Noyes' grading scale [47], were not the same among the papers, and so pooling could not be executed.. Generally, arthroscopic gradings correlated strongly with osteophyte [11], moderately with cartilage grading [14]and weakly with subchondral bone [48].

Responsiveness

According to Cohen [49], values of 0.0, 0.20, 0.50, and 0.80 or greater represented trivial, small, moderate, and large responsiveness, respectively.

Internal responsiveness: Pooling 31 estimates across 10 studies, comprising 480 joints from 393 patients, produced a moderate effect size for Baker's cyst [0.58(0.40,0.77)], and small effect size for synovial hypertrophy [0.30(0.05,0.56)], effusion [0.28(0.00,0.56)] and cartilage thickness [0.20(0.04,0.36)] (**Table 6**). The interventions included injections of different steroids (n=6), platelet rich plasma (n=2), glucosamine (n=1), and exercises (n=1). The study duration ranged from 2 weeks to 6 months.

External responsiveness: Pooling 7 estimates across 4 studies with a total of 121 joints

and 121patients, provided moderate correlation for synovial hypertrophy [0.43(-0.02,0.73)], and weak correlation for Baker's cyst [0.35(-0.11,0.69)]. Substantial heterogeneity was detected [I^2=68-74] (**supplementary data 10**). The interventions were intra-articular steroid injections (n=3), and shortwave diathermy (n=1).

(**Tables for stratified meta-analysis, and figures for forest plots were also described as supplementary data 10 and 11**).

Feasibility

Five studies reported the scanning time for complete examination, which varied from 5 min to 15 min depending on how many pathologies were scanned (**supplementary data 10**).

Hand OA

Reliability

There were 4 inter-rater reliability studies for binary scores [50-53], 3 for semi-quantitative scores [5, 12, 51] and 1 for quantitative scores [54].The binary scoring system provided the kappa ranging from slight in cartilage thickness [51] to excellent in synovitis, effusion and osteophyte [52]. For semi-quantitative score, the kappa values varied from slight in cartilage thickness [51] to substantial in osteophyte and synovitis [5, 12]. For quantitative score, ICC was excellent in synovial hypertrophy [54].

Among intra-reliability studies, 7 studies applied binary scores [5, 10, 12, 50, 51, 55, 56];

five studies used semi-quantitative scores [5, 12, 51, 57, 58]; one study examined quantitative scores [59]. Similar findings of kappa values were reported for different pathologies but with a higher actual kappa values.

Validity

Only two studies reported construct validity of ultrasound with pain, disclosing very weak correlation [57, 60]. Four studies documented ultrasound data for functional correlation which varied from very weak to weak in most pathologies [55, 57, 60, 61]. Validity of ultrasound with X-rays was investigated in two studies, providing very weak correlation [56, 60]. However, ultrasound provided moderate correlation with MRI for osteophyte (r=0.49) and synovitis (r=0.43) on semi-quantitative scale [62].

Responsiveness

Two studies supplied sufficient information to calculate the internal responsiveness. One study revealed trivial effect size for synovitis and power Doppler outcomes at 12 weeks after intramuscular methylprednisolone injection [63], and small effect size was detected at 4 weeks for the same pathologies in another study, using intra-articular injections of hyaluronic acid as an intervention [64].

For external responsiveness, one study reported strong correlation of synovial thickening and power Doppler with VAS pain at 4 weeks [64].

Hip OA

Reliability

Inter-rater reliability of binary score ranged from fair in effusion to moderate for osteophyte in one study [65] while another study recorded excellent reliability for the same pathologies [66].

Intra-rater reliability of binary score was moderate in joint effusion and substantial in osteophyte [65] while the other revealed the excellent kappa [66]. For semi-quantitative scores by radiologists, excellent kappa was reported for the synovial thickness [67].

Validity

Ultrasound synovitis and osteophyte scores demonstrated a strong association with pain on activity [65]. Weak correlation was documented between effusion and Lequesne index [68], and between osteophyte and KL grading (r=0.26) [65].

Responsiveness

One study applied ultrasound synovial hypertrophy and effusion as outcome measure to evaluate internal responsiveness, providing moderate effect size (SMD=0.44) at 3 months after intra-articular injection of 8 mg betamethasone [69].

Discussion

Overall, the main findings of our meta-analysis suggest various (weak to very strong) construct validity with patients findings and other imaging modalities, depending on pathologies and comparators, moderate to substantial reliability, strong criterion validity with cartilage histology, and small to moderate responsiveness to interventions. On qualitative analysis, this systematic review revealed substantial clinical, technical and methodological heterogeneity of ultrasound within OA literature, requiring caution in interpreting these meta-analytic results. However, on quantitative analysis, I^2, which denotes statistical heterogeneity, was only low or moderate for most of clinimetrics.

Although ultrasound possesses promising potential in OA clinical trials, fewer studies in hand and hip joints were detected in the literature, compared to the knee. Although utilization/reporting of OMERACT definitions has gained a significantly positive trend over last decade, a marked variability of ultrasound scanning characteristics was noted, highlighting the necessity of following/reporting international consensus protocols in future studies.

In the context of methodological quality, a modified Downs and Black quality assessment score [28] was administered to identify the potential bias and display the summary of these bias. All studies, which documented the clinimetric data for each pathology, were pooled without applying exclusion on the basis of study quality scale because the threshold for exclusion reduced the precision [70], and was necessarily subjective [71]. According to Detsky et al, it seemed highly unlikely that these quality scores would generate a linear or monotonically

increasing association with true quality, and no objective reference standard simply existed for determining the "true" scientific rigour of a trial [72]. Moreover, due to a limited number of papers which documented clinimetric data for each ultrasound pathology, the sensitivity analysis, based on study quality score, could not be examined (i.e. there were some pathologies for each of which only one paper existed as a unit of analysis.).

In addition, definitions in OA are difficult in terms of what is normal, and what is defined for OA (radiographic OA or ACR criteria, which means totally different things), making validity research not easy.

Our meta-analysis results indicated moderate to substantial reliability [minimum kappa≥0.44(0.15,0.74) and minimum ICC≥0.82(0.73-0.89)] for ultrasound pathologies of knee OA. Generally, the binary and quantitative scores produced higher reliability statistics than semi-quantitative score. Some papers calibrated the semi-quantitative scores by utilizing the atlas-based grading methods [11, 73] while some defined the grading by quantitative cut-offs [6]. The reliability of Baker's cyst, meniscal extrusion, osteophyte, synovitis and effusion were at least substantial for the semi-quantitative scores.

The musculoskeletal experience of ultrasound operators ranged from those with short-course training to very experienced specialist, and so the meta-analysis results represented the generalizability of reliability statistics across different levels of ultrasound experience. However, it should be noted that operator-dependent nature of ultrasound measurement and quality of US machines could largely influence on the performance of the reliability statistics, especially when

smaller joints are addressed.

The limited data for criterion validity of OA ultrasound features focused predominantly on cartilage histology, with overall strong correlation. Conflicting reports were found for correlations of synovitis/Doppler signals with synovial vascularity in a mixed sample of inflammatory arthritis and osteoarthritis [74-77]. Semi-quantitative grading scores currently applied for OA synovitis were adopted from those validated for inflammatory rheumatoid arthritis, assuming that synovitis was only quantitatively but not qualitatively different between the inflammatory arthritis and osteoarthritis [78]. However, replication of these semi-quantitative scoring systems in osteoarthritis might require consideration due to the low degree of inflammation, sustained in osteoarthritis compared to rheumatoid arthritis [18], which is likely to contribute to floor effects, and thereby impairs the capability to detect improvement changes in interventional studies.

Pooling construct validity of ultrasound findings in case-control studies (OA versus healthy population) exhibited strong discrimination in some pathologies, suggesting that ultrasound might be a potential tool for developing ultrasonographic OA propositions, similar to preliminary OA propositions with MRI [79]. Furthermore, ultrasound demonstrated a strong correlation with MRI in principal OA features, indicating the promising usefulness of ultrasound in clinical care where MRI is not readily accessible.

Generally, ultrasound, as expected, had a very weak association with pain, function and blood biomarker (COMP). Almost all individual studies incorporated in the meta-analysis

consistently denoted weak correlation between ultrasound features and pain ($r \leq 0.40$). This finding may be attributed to a number of reasons such as complex causes of symptoms in OA, multi-factorial subjective experience of pain (biopsychosocial factors), and that the ultrasound outcomes used in individual studies might not captured the multi-dimensional nature of pain (measurement issues) [80]. In contrast, relationship of ultrasound with X rays produced various values ranging from weak to strong correlation, depending on ultrasound pathologies.

At least small effect size (SMD≥0.2) was documented in most of interventional studies, and the low I^2 in pooled meta-analysis was detected. Generally, the inflammatory features such as Baker's cyst, synovial hypertrophy provides greater internal responsiveness, compared to cartilage changes, perhaps due to short follow-up duration (maximum 24 weeks). However, this result should be interpreted with caution as the included studies for sensitivity to change were all small studies with some limitations. Combining external responsiveness of inflammatory pathologies revealed a moderate correlation with pain while no studies examined external responsiveness for structural pathologies.

Ultrasound scanning duration largely depended on the number of joints and pathologies assessed and the scoring systems employed, which were varied across studies. Development of international consensus guidelines for feasible composite scoring methods is essential, and still undergoing.

It should be noted that several papers included in the validity assessment of previous systematic review [20] had to be excluded as our inclusion criteria was focused only on knee,

421 hand and hip, not other joints such as foot, shoulder, cervical spine, etc and some papers did not
422 publish the comparator for validity assessment, clinimetric data, etc. However, more than
423 additional 60 papers were included in this updated review.
424 Our review had several potential limitations. The first was the considerable clinical and
425 methodological heterogeneity of included studies, requiring caution in interpreting the pooled
426 results. However, I^2 was low for validity and responsiveness measures. The second limitation
427 was that we could not rule out some publication bias although a thorough literature search was
428 attempted. The third limitation is the application of SMD for internal responsiveness instead of
429 calculating standardized response mean (SRM),as most interventional studies did not describe
430 standard deviation of mean change [81]. However, in the literature, the best statistics for
431 treatment responsiveness and interpretation is still controversial, and according to mathematical
432 formulae proposed by Norman et al.[36], SRMs tend to be higher than SMDs. The fourth
433 limitation is that we could not appropriately analyze the confounding effects over technology
434 changes over the years because there were numerous confounders such as machine model, probe
435 frequency, operator's clinical background, qualification, training period, the severity of the
436 sample, the sensitivity of comparator machine models in examining construct validity against X
437 rays and MRI, while a limited number of papers with clinimetric data for each pathology existed,
438 causing a lack of power to examine the impact of these confounders on the clinimetrics by
439 regression analysis.
440 To our knowledge, this is the first meta-analytic systematic review comprehensively

examining clinimetrics of ultrasound utilized to evaluate common features of OA, covering the original OMERACT filter components. Stratified meta-analysis demonstrated moderate to substantial reliability, various construct validity with several clinical and imaging comparators, strong criterion validity with cartilage histology and small to moderate responsiveness. Future studies should improve the conduct and reporting of clinimetric studies especially for the areas of several poor quality-items. As most of individual studies were of small sample size and just focused on some individual pathologies, larger studies with comprehensive ultrasound outcomes in future would provide more clear insight into the clinimetrics of commonly assessed ultrasound pathologies in OA.

Contributions: WMO and DJH conceived and designed the study. JML, PGC, HK, SS, JS and LAD were also involved in the design of the study. WMO, MD and DJH contributed to acquisition of the main clinimetric data of included papers. WMO had full access to all the data and analysis, drafted the manuscript and takes responsibility for the integrity of the work from inception to finished article. All authors critically revised the manuscript and gave final approval of the article for submission.

Acknowledgement: WMO is supported for his PhD project by a Presidential Scholarship of Myanmar. PGC is supported in part by the National Institute for Health Research (NIHR) Leeds Biomedical Research Centre. The views expressed are those of the authors and not necessarily

those of the NHS, the NIHR, or the Department of Health. DJH is supported by an NHMRC Practitioner Fellowship.

Conflict of interest: None

References

1. Hunter D.J., Schofield D., and Callander E., *The individual and socioeconomic impact of osteoarthritis.* Nat Rev Rheumatol, 2014. **10**(7): p. 437-41.
2. Lawrence R.C., Felson D.T., Helmick C.G., Arnold L.M., Choi H., Deyo R.A., et al., *Estimates of the prevalence of arthritis and other rheumatic conditions in the United States. Part II.* Arthritis and Rheumatism, 2008. **58**(1): p. 26-35.
3. Abraham A.M., Goff I., Pearce M.S., Francis R.M., and Birrell F., *Reliability and validity of ultrasound imaging of features of knee osteoarthritis in the community.* BMC Musculoskeletal Disorders, 2011. **12**.
4. Bevers K., Vriezekolk J.E., Bijlsma J.W.J., van den Ende C.H.M., and den Broeder A.A., *Ultrasonographic predictors for clinical and radiological progression in knee osteoarthritis after 2 years of follow-up.* Rheumatology (United Kingdom), 2015. **54**(11): p. 2000-2003.
5. Mathiessen A., Slatkowsky-Christensen B., Kvien T.K., Hammer H.B., and Haugen I.K., *Ultrasound-detected inflammation predicts radiographic progression in hand osteoarthritis after 5 years.* Annals of the Rheumatic Diseases, 2016. **75**(5): p. 825-830.
6. Nogueira-Barbosa M.H., Gregio-Junior E., Lorenzato M.M., Guermazi A., Roemer F.W., Chagas-Neto F.A., et al., *Ultrasound assessment of medial meniscal extrusion: A validation study using MRI as reference standard.* American Journal of Roentgenology, 2015. **204**(3): p. 584-588.
7. Podlipská J., Guermazi A., Lehenkari P., Niinimäki J., Roemer F.W., Arokoski J.P., et al., *Comparison of Diagnostic Performance of Semi-Quantitative Knee Ultrasound and Knee Radiography with MRI: Oulu Knee Osteoarthritis Study.* Scientific Reports, 2016. **6**: p. 22365.
8. Yanagisawa S., Ohsawa T., Saito K., Kobayashi T., Yamamoto A., and Takagishi K., *Morphological evaluation and diagnosis of medial type osteoarthritis of the knee using ultrasound.* Journal of Orthopaedic Science, 2014. **19**(2): p. 270-274.
9. Acebes C., Romero F.I., Contreras M.A., Mahillo I., and Herrero-Beaumont G., *Dynamic ultrasound assessment of medial meniscal subluxation in knee osteoarthritis.* Rheumatology (United Kingdom), 2013. **52**(8): p. 1443-1447.
10. Keen H.I., Wakefield R.J., Grainger A.J., Hensor E.M., Emery P., and Conaghan P.G., *Can ultrasonography improve on radiographic assessment in osteoarthritis of the hands? A comparison between radiographic and ultrasonographic detected pathology.* Annals of the Rheumatic Diseases, 2008. **67**(8): p. 1116-20.
11. Koski J.M., Kamel A., Waris P., Waris V., Tarkiainen I., Karvanen E., et al., *Atlas-based knee osteophyte assessment with ultrasonography and radiography: Relationship to arthroscopic degeneration of articular cartilage.* Scandinavian Journal of Rheumatology, 2016. **45**(2): p. 158-164.
12. Mathiessen A., Haugen I.K., Slatkowsky-Christensen B., Boyesen P., Kvien T.K., and Hammer H.B., *Ultrasonographic assessment of osteophytes in 127 patients with hand osteoarthritis:*

exploring reliability and associations with MRI, radiographs and clinical joint findings. Annals of the Rheumatic Diseases, 2013. **72**(1): p. 51-6.

13. Chen Y.J., Chen C.H., Wang C.L., Huang M.H., Chen T.W., and Lee C.L., *Association between the severity of femoral condylar cartilage erosion related to knee osteoarthritis by ultrasonographic evaluation and the clinical symptoms and functions.* Archives of Physical Medicine & Rehabilitation, 2015. **96**(5): p. 837-44.
14. Saarakkala S., Waris P., Waris V., Tarkiainen I., Karvanen E., Aarnio J., et al., *Diagnostic performance of knee ultrasonography for detecting degenerative changes of articular cartilage.* Osteoarthritis Cartilage, 2012. **20**(5): p. 376-81.
15. Yoon C.H., Kim H.S., Ju J.H., Jee W.H., Park S.H., and Kim H.Y., *Validity of the sonographic longitudinal sagittal image for assessment of the cartilage thickness in the knee osteoarthritis.* Clinical Rheumatology, 2008. **27**(12): p. 1507-16.
16. Riecke B.F., Christensen R., Torp-Pedersen S., Boesen M., Gudbergsen H., and Bliddal H., *An ultrasound score for knee osteoarthritis: A cross-sectional validation study.* Osteoarthritis and Cartilage, 2014. **22**(10): p. 1675-1691.
17. Hall M., Doherty S., Courtney P., Latief K., Zhang W., and Doherty M., *Synovial pathology detected on ultrasound correlates with the severity of radiographic knee osteoarthritis more than with symptoms.* Osteoarthritis and Cartilage, 2014. **22**(10): p. 1627-1633.
18. Kortekaas M.C., Kwok W.Y., Reijnierse M., and Kloppenburg M., *Inflammatory ultrasound features show independent associations with progression of structural damage after over 2 years of follow-up in patients with hand osteoarthritis.* Annals of the Rheumatic Diseases, 2015. **74**(9): p. 1720-4.
19. Mancarella L., Addimanda O., Pelotti P., Pignotti E., Pulsatelli L., and Meliconi R., *Ultrasound detected inflammation is associated with the development of new bone erosions in hand osteoarthritis: A longitudinal study over 3.9 years.* Osteoarthritis and Cartilage, 2015. **23**(11): p. 1925-1932.
20. Keen H.I., Wakefield R.J., and Conaghan P.G., *A systematic review of ultrasonography in osteoarthritis.* Annals of the Rheumatic Diseases, 2009. **68**(5): p. 611-619.
21. Oo W.M. and Bo M.T., *Role of Ultrasonography in Knee Osteoarthritis.* J Clin Rheumatol, 2016. **22**(6): p. 324-9.
22. Oo W.M., Linklater J.M., and Hunter D.J., *Imaging in knee osteoarthritis.* Curr Opin Rheumatol, 2017. **29**(1): p. 86-95.
23. Wakefield R.J., Balint P.V., Szkudlarek M., Filippucci E., Backhaus M., D'Agostino M.A., et al., *Musculoskeletal ultrasound including definitions for ultrasonographic pathology.* J Rheumatol, 2005. **32**(12): p. 2485-7.
24. D'Agostino M.A., Conaghan P., Le Bars M., Baron G., Grassi W., Martin-Mola E., et al., *EULAR report on the use of ultrasonography in painful knee osteoarthritis. Part 1: prevalence of inflammation in osteoarthritis.* Ann Rheum Dis, 2005. **64**(12): p. 1703-9.
25. Boers M., Kirwan J.R., Wells G., Beaton D., Gossec L., d'Agostino M.A., et al., *Developing core*

outcome measurement sets for clinical trials: OMERACT filter 2.0. J Clin Epidemiol, 2014. **67**(7): p. 745-53.

26. *Covidence systematic review software.* Veritas Health Innovation: Melbourne, Australia.

27. D'Agostino M.A., Boers M., Kirwan J., Van Der Heijde D., Østergaard M., Schett G., et al., *Updating the OMERACT filter: Implications for imaging and soluble biomarkers.* Journal of Rheumatology, 2014. **41**(5): p. 1016-1024.

28. Downs S.H. and Black N., *The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions.* Journal of Epidemiology and Community Health, 1998. **52**(6): p. 377-384.

29. Langham S., Langham J., Goertz H.P., and Ratcliffe M., *Large-scale, prospective, observational studies in patients with psoriasis and psoriatic arthritis: A systematic and critical review.* BMC Medical Research Methodology, 2011. **11**.

30. Lucas N.P., Macaskill P., Irwig L., and Bogduk N., *The development of a quality appraisal tool for studies of diagnostic reliability (QAREL).* J Clin Epidemiol, 2010. **63**(8): p. 854-61.

31. Becker B.J. and Wu M.-J., *The Synthesis of Regression Slopes in Meta-Analysis.* Statistical Science, 2007. **22**(3): p. 414-429.

32. Borenstein M., Hedges L.V., Higgins J.P.T., and Rothstein H.R., *Introduction to Meta-Analysis*. 2009, John Wiley & Sons, Ltd.

33. McHugh M.L., *Interrater reliability: the kappa statistic.* Biochemia Medica, 2012. **22**(3): p. 276-282.

34. Hedges LV O., I. , *Statistical Methods for Meta-Analysis*. 1985: Orlando, FL: Academic Press.

35. *introduction to meta analysis book.*

36. Norman G.R., Wyrwich K.W., and Patrick D.L., *The mathematical relationship among different forms of responsiveness coefficients*. Qual Life Res, 2007. **16**(5): p. 815-22.

37. Husted J.A., Cook R.J., Farewell V.T., and Gladman D.D., *Methods for assessing responsiveness: a critical review and recommendations.* J Clin Epidemiol, 2000. **53**(5): p. 459-68.

38. Higgins J.P. and Thompson S.G., *Quantifying heterogeneity in a meta-analysis.* Stat Med, 2002. **21**(11): p. 1539-58.

39. Sterne J.A. and Egger M., *Funnel plots for detecting bias in meta-analysis: guidelines on choice of axis.* J Clin Epidemiol, 2001. **54**(10): p. 1046-55.

40. Begg C.B. and Mazumdar M., *Operating characteristics of a rank correlation test for publication bias.* Biometrics, 1994. **50**(4): p. 1088-101.

41. Egger M., Smith G.D., Schneider M., and Minder C., *Bias in meta-analysis detected by a simple, graphical test.* BMJ, 1997. **315**(7109): p. 629-634.

42. Backhaus M., Burmester G.R., Gerber T., Grassi W., Machold K.P., Swen W.A., et al., *Guidelines for musculoskeletal ultrasound in rheumatology.* Ann Rheum Dis, 2001. **60**(7): p. 641-9.

43. Landis J.R. and Koch G.G., *The measurement of observer agreement for categorical data.*

582 Biometrics, 1977. **33**(1): p. 159-74.
583 44. Evans J.D., *Straightforward Statistics for the Behavioral Sciences*. 1996: Brooks/Cole
584 Publishing; Pacific Grove, Calif.
585 45. Bellamy N., Buchanan W.W., Goldsmith C.H., Campbell J., and Stitt L.W., *Validation study of*
586 *WOMAC: a health status instrument for measuring clinically important patient relevant*
587 *outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee.* J
588 Rheumatol, 1988. **15**(12): p. 1833-40.
589 46. Kellgren J.H. and Lawrence J.S., *Radiological assessment of osteo-arthrosis.* Ann Rheum Dis,
590 1957. **16**(4): p. 494-502.
591 47. Noyes F.R. and Stabler C.L., *A system for grading articular cartilage lesions at arthroscopy.*
592 Am J Sports Med, 1989. **17**(4): p. 505-13.
593 48. Podlipska J., Koski J.M., Pulkkinen P., and Saarakkala S., *In vivo quantitative ultrasound*
594 *image analysis of femoral subchondral bone in knee osteoarthritis.* Thescientificworldjournal,
595 2013. **2013**: p. 182562.
596 49. Cohen J., *Statistical Power Analysis for the Behavioural Sciences.* 1977: New York: Academic
597 Press
598 50. Iagnocco A., Conaghan P.G., Aegerter P., Moller I., Bruyn G.A., Chary-Valckenaere I., et al.,
599 *The reliability of musculoskeletal ultrasound in the detection of cartilage abnormalities at the*
600 *metacarpo-phalangeal joints.* Osteoarthritis & Cartilage, 2012. **20**(10): p. 1142-6.
601 51. Hammer H.B., Iagnocco A., Mathiessen A., Filippucci E., Gandjbakhch F., Kortekaas M.C., et
602 al., *Global ultrasound assessment of structural lesions in osteoarthritis: a reliability study by*
603 *the OMERACT ultrasonography group on scoring cartilage and osteophytes in finger joints.*
604 Annals of the Rheumatic Diseases, 2016. **75**(2): p. 402-7.
605 52. Wittoek R., Carron P., and Verbruggen G., *Structural and inflammatory sonographic findings*
606 *in erosive and non-erosive osteoarthritis of the interphalangeal finger joints.* Annals of the
607 Rheumatic Diseases, 2010. **69**(12): p. 2173-2176.
608 53. Wittoek R., Jans L., Lambrecht V., Carron P., Verstraete K., and Verbruggen G., *Reliability*
609 *and construct validity of ultrasonography of soft tissue and destructive changes in erosive*
610 *osteoarthritis of the interphalangeal finger joints: a comparison with MRI.* Annals of the
611 Rheumatic Diseases, 2011. **70**(2): p. 278-83.
612 54. Damman W., Kortekaas M.C., Stoel B.C., van 't Klooster R., Wolterbeek R., Rosendaal F.R., et
613 al., *Sensitivity-to-change and validity of semi-automatic joint space width measurements in*
614 *hand osteoarthritis: A follow-up study.* Osteoarthritis and Cartilage, 2016. **24**(7): p. 1172-1179.
615 55. Koutroumpas A.C., Alexiou I.S., Vlychou M., and Sakkas L.I., *Comparison between clinical*
616 *and ultrasonographic assessment in patients with erosive osteoarthritis of the hands.* Clinical
617 Rheumatology, 2010. **29**(5): p. 511-6.
618 56. Vlychou M., Koutroumpas A., Malizos K., and Sakkas L.I., *Ultrasonographic evidence of*
619 *inflammation is frequent in hands of patients with erosive osteoarthritis.* Osteoarthritis and
620 Cartilage, 2009. **17**(10): p. 1283-1287.

57. Keen H.I., Wakefield R.J., Grainger A.J., Hensor E.M.A., Emery P., and Conaghan P.G., *An ultrasonographic study of osteoarthritis of the hand: Synovitis and its relationship to structural pathology and symptoms.* Arthritis Care and Research, 2008. **59**(12): p. 1756-1763.
58. Kortekaas M.C., Kwok W.Y., Reijnierse M., Watt I., Huizinga T.W.J., and Kloppenburg M., *Pain in hand osteoarthritis is associated with inflammation: The value of ultrasound.* Annals of the Rheumatic Diseases, 2010. **69**(7): p. 1367-1369.
59. Mancarella L., Magnani M., Addimanda O., Pignotti E., Galletti S., and Meliconi R., *Ultrasound-detected synovitis with power Doppler signal is associated with severe radiographic damage and reduced cartilage thickness in hand osteoarthritis.* Osteoarthritis and Cartilage, 2010. **18**(10): p. 1263-1268.
60. Naguib A., Mohasseb D., Sultan H., Hamimi A., and Fawzy M., *Hand osteoarthritis: Clinical and imaging study.* Alexandria Journal of Medicine, 2011. **47**(3): p. 237-242.
61. Mallinson P.I., Tun J.K., Farnell R.D., Campbell D.A., and Robinson P., *Osteoarthritis of the thumb carpometacarpal joint: correlation of ultrasound appearances to disability and treatment response.* Clinical Radiology, 2013. **68**(5): p. 461-5.
62. Kortekaas M.C., Kwok W.Y., Reijnierse M., Wolterbeek R., Boyesen P., van der Heijde D., et al., *Magnetic Resonance Imaging in Hand Osteoarthritis: Intraobserver Reliability and Criterion Validity for Clinical and Structural Characteristics.* J Rheumatol, 2015. **42**(7): p. 1224-30.
63. Keen H.I., Wakefield R.J., Hensor E.M.A., Emery P., and Conaghan P.G., *Response of symptoms and synovitis to intra-muscular methylprednisolone in osteoarthritis of the hand: An ultrasonographic study.* Rheumatology, 2010. **49**(6): p. 1093-1100.
64. Klauser A.S., Faschingbauer R., Kupferthaler K., Feuchnter G., Wick M.C., Jaschke W.R., et al., *Sonographic criteria for therapy follow-up in the course of ultrasound-guided intra-articular injections of hyaluronic acid in hand osteoarthritis.* European Journal of Radiology, 2012. **81**(7): p. 1607-11.
65. Qvistgaard E., Torp-Pedersen S., Christensen R., and Bliddal H., *Reproducibility and inter-reader agreement of a scoring system for ultrasound evaluation of hip osteoarthritis.* Annals of the Rheumatic Diseases, 2006. **65**(12): p. 1613-9.
66. Tormenta S., Sconfienza L.M., Iannessi F., Bizzi E., Massafra U., Orlandi D., et al., *Prevalence study of iliopsoas bursitis in a cohort of 860 patients affected by symptomatic hip osteoarthritis.* Ultrasound in Medicine & Biology, 2012. **38**(8): p. 1352-6.
67. Robinson P., Keenan A.M., and Conaghan P.G., *Clinical effectiveness and dose response of image-guided intra-articular corticosteroid injection for hip osteoarthritis.* Rheumatology, 2007. **46**(2): p. 285-91.
68. Rennesson-Rey B., Rat A.C., Chary-Valckenaere I., Bettembourg-Brault I., Juge N., Dintinger H., et al., *Does joint effusion influence the clinical response to a single Hylan GF-20 injection for hip osteoarthritis?* Joint Bone Spine, 2008. **75**(2): p. 182-188.
69. Micu M.C., Bogdan G.D., and Fodor D., *Steroid injection for hip osteoarthritis: efficacy under*

ultrasound guidance. Rheumatology, 2010. **49**(8): p. 1490-4.

70. Higgins J.P.T., Altman D.G., Gøtzsche P.C., Jüni P., Moher D., Oxman A.D., et al., *The Cochrane Collaboration's tool for assessing risk of bias in randomised trials.* The BMJ, 2011. **343**: p. d5928.

71. Carroll C. and Booth A., *Quality assessment of qualitative evidence for systematic review and synthesis: Is it meaningful, and if so, how should it be performed?* Research Synthesis Methods, 2015. **6**(2): p. 149-154.

72. Detsky A.S., Naylor C.D., O'Rourke K., McGeer A.J., and L'Abbe K.A., *Incorporating variations in the quality of individual randomized trials into meta-analysis.* J Clin Epidemiol, 1992. **45**(3): p. 255-65.

73. Bruyn G.A., Naredo E., Damjanov N., Bachta A., Baudoin P., Hammer H.B., et al., *An OMERACT reliability exercise of inflammatory and structural abnormalities in patients with knee osteoarthritis using ultrasound assessment.* Ann Rheum Dis, 2016. **75**(5): p. 842-6.

74. Schmidt W.A., Volker L., Zacher J., Schlafke M., Ruhnke M., and Gromnica-Ihle E., *Colour Doppler ultrasonography to detect pannus in knee joint synovitis.* Clin Exp Rheumatol, 2000. **18**(4): p. 439-44.

75. Walther M., Harms H., Krenn V., Radke S., Faehndrich T.P., and Gohlke F., *Correlation of power Doppler sonography with vascularity of the synovial tissue of the knee joint in patients with osteoarthritis and rheumatoid arthritis.* Arthritis Rheum, 2001. **44**(2): p. 331-8.

76. Karim Z., Wakefield R.J., Quinn M., Conaghan P.G., Brown A.K., Veale D.J., et al., *Validation and reproducibility of ultrasonography in the detection of synovitis in the knee: a comparison with arthroscopy and clinical examination.* Arthritis Rheum, 2004. **50**(2): p. 387-94.

77. Koski J.M., Saarakkala S., Helle M., Hakulinen U., Heikkinen J.O., and Hermunen H., *Power Doppler ultrasonography and synovitis: correlating ultrasound imaging with histopathological findings and evaluating the performance of ultrasound equipments.* Annals of the Rheumatic Diseases, 2006. **65**(12): p. 1590-1595.

78. Kuryliszyn-Moskal A., *Comparison of blood and synovial fluid lymphocyte subsets in rheumatoid arthritis and osteoarthritis.* Clinical Rheumatology, 1995. **14**(1): p. 43-50.

79. Hunter D.J., Arden N., Conaghan P.G., Eckstein F., Gold G., Grainger A., et al., *Definition of osteoarthritis on MRI: results of a Delphi exercise.* Osteoarthritis Cartilage, 2011. **19**(8): p. 963-9.

80. Neogi T., *The Epidemiology and Impact of Pain in Osteoarthritis.* Osteoarthritis and cartilage / OARS, Osteoarthritis Research Society, 2013. **21**(9): p. 1145-1153.

81. Tordrup D., Mossman J., and Kanavos P., *Responsiveness of the EQ-5D to clinical change: is the patient experience adequately represented?* Int J Technol Assess Health Care, 2014. **30**(1): p. 10-19.