Patient-Centred Outcomes in Lateral Elbow Tendinopathy: a systematic review of available evidence in UK populations

<table>
<thead>
<tr>
<th>Journal:</th>
<th>Shoulder and Elbow</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manuscript ID</td>
<td>Draft</td>
</tr>
<tr>
<td>Manuscript Type:</td>
<td>Review Article</td>
</tr>
<tr>
<td>Keywords:</td>
<td>Patient reported outcome measures, Lateral elbow tendinopathy, Lateral epicondylitis, Tennis Elbow, Psychometrics, Validation</td>
</tr>
</tbody>
</table>
Patient-Centred Outcomes in Lateral Elbow Tendinopathy:  
a systematic review of available evidence in UK populations

Key Words:
Patient reported outcome measures
Lateral elbow tendinopathy
Lateral epicondylitis
Tennis Elbow
Psychometrics
Validation
Abstract

Aim

To systematically review the evidence for clinical ratings systems in the assessment of outcomes of UK patients with Lateral Elbow Tendinopathy (LET).

Methods

A systematic search was performed in Ovid MEDLINE, Embase and CINAHL. Studies were included if they reported the administration of PROMs in UK populations with LET. PROMs characteristics and the populations in which they had been used were assessed using a structured classification system. PROMs reporting in randomised controlled trials was assessed against CONSORT standards (PRO extension).

Results

A total of 16 articles were included based on eligibility criteria. Out of seven different PROMs, there was evidence of partial validation for five of them. The assessment of validity, reliability and responsiveness of all PROMs in LET UK populations extended to just 20 individual patients. No articles conformed to the CONSORT PRO extension standards.

Conclusion

There exists a huge paucity of data on the psychometrics and usability of PROMs in UK LET populations. Without these data, trial design and interpretation of health technology assessment are significantly hindered. The high prevalence of this condition allied with the significant volume of studies being conducted into novel treatments, highlight the need for this knowledge gap to be resolved.
**Background**

Lateral Elbow Tendinopathy (LET), known more commonly as Tennis Elbow, is a prevalent and potentially debilitating condition. Though the condition is regarded as benign and self-limiting, absenteeism due to LET in the UK is estimated to cost the economy £27 million per annum. With a UK prevalence between 1.5-3%, it is surprising that no clear treatment consensus exists. This treatment equipoise has driven a large volume of research activity, with over 80 registered trials currently ongoing. However, to be confident in our treatments, we must be certain that the outcome measures used in these trials truly reflect patient benefit or harm.

Successful treatment in LET can be regarded as amelioration of pain and return of function. Constructs such as these are now commonly quantified through the use of Patient-Reported Outcome Measures (PROMS). Collaborative work by academics and clinicians has crystallised in the development of systematic, robust and valid ways of collecting health outcomes from patients that purport to quantify, in a meaningful way, how the patient feels their condition affects them. In reference to musculoskeletal pathology, this has resulted in numerous PROMs used to quantify the burden of a specific disease, such as the use of the Oxford Hip and Knee scores in assessing the outcome of joint arthroplasty.

Appropriate outcome measures must demonstrate that they are acceptable to patients, reliable, valid and responsive (sensitive to change). When the outcome measure has been developed in a different clinical or geographical population, there needs to be evidence of equivalence both in a disease-specific and cross-cultural context.

A structured assessment of outcome measurement in LET in UK populations has not been undertaken. This study aimed to address this gap by systematically assessing the outcome measures used for measuring PROMs in Lateral Elbow Tendinopathy in a UK population, and to assess the reporting of randomised controlled trials using PROMs in LET. Only when
valid outcomes have been identified, can recommendations on choice of outcome measures for future research be made.

Materials and Methods

PRISMA guidelines on the reporting of systematic reviews were followed \([^{[12]}]\). All articles reporting the development, psychometric evaluation, or use of patient reported outcome measures in Lateral Elbow Tendinopathy in UK adults (≥18yrs) were included. Any measures of symptoms and functioning in LETs that involved a patient-reported outcome measurement (regardless of whether this also contained a physician-reported outcome component) were included. Studies in paediatric populations, case-reports, case-studies and conference abstracts were excluded.

A search strategy was constructed using MeSH and free-text terms (appendix 1). The search strategy development was guided by previously published search strategies for systematic reviews of interventions in elbow pathology \([^{[13]}]\) and for the identification of outcome measures \([^{[14]}]\), along with terms specifically selected in order to capture names of relevant instruments published in previous systematic reviews of elbow-specific rating scales \([^{[15-18]}]\). The strategy was further adapted to each database through the modification of thesaurus terms, wildcards, and truncation. The search was run on the 1\(^{st}\) May 2017 in Medline (Ovid MEDLINE, 1948 to 2016 & Ovid MEDLINE In-Process & Non-indexed Citations) accessed through OVIDSP, Embase (Embase 1974 to 2016) accessed through OVIDSP and CINHAL (CINHAL 1981 to 2016) accessed through EBSCO host.

The screening process was conducted in a step-wise manner. At each stage, one researcher and a further researcher reviewed each title and abstract. In cases of disagreement, the article proceeded to the next stage of review to ensure maximum sensitivity.

At the full-text level, articles were also sub-categorised in two groups to: articles reporting primary research on the development and/or psychometric evaluation of PROMs in LET in
UK populations (development); and articles reporting the use of outcome measures in clinical studies in UK populations (use).

Data synthesis

Development articles were classified according to three guiding concepts, using the structured classification system proposed by Valderas and Alonso\(^{19}\): construct (the measurement object), population (based on age, gender, condition and culture) and measurement model (dimensionality, metric and adaptability)\(^{19}\).

The assessment of construct denotes, for the purpose of this study, the range of characteristics measured by the outcome measure, which are affected by LET. The construct analysis has, at its foundation, the conceptual strengths of the Wilson and Cleary model\(^{20}\), but is also integrated with the theoretical model that underpins the International Classification of Functioning, Disability and Health (ICF). A strength of the model that is particularly pertinent in the assessment of LET outcome measures, is the systematic consideration of intended population of use. Within the axis of population consideration of culture is also made, where there is information pertaining to the dyad of language and country for which the outcome measures have been devised.

It should be noted that this system is only descriptive and does not provide any fundamental evaluation of measurement properties\(^{19}\). But in the early stages of outcome measure assessment, this approach provides the clearest method of identifying the candidate pool of measures. Only once this is undertaken and deemed to be adequate, can a systematic evaluation of measurement properties in a specific population of use be undertaken.

Articles reporting the use of PROMs (use) were peer-reviewed, published articles with outcome measure evaluation in a population of LET patients. Date of publication, outcome measure(s) chosen and population of use was extracted. For randomised control trials, the CONSORT Patient-Reported Outcome (PRO) extension\(^{21}\) was used to systematically assess
the reporting of outcome measure choice and justification. The original CONSORT statement aims to encourage transparent and complete reporting of clinical trials and is associated with improved reporting practice.\(^{22}\)

An a priori hypothesis was formulated with regard to informed choice of outcome measures in UK populations. We hypothesised that articles reporting the use of PROMs would more frequently use PROMs for which there would be evidence from studies of validation of such measures in UK populations.

**Results**

We identified 7,261 records from the electronic database search. A total of 16 articles met the inclusion criteria: five articles reporting the development and/or psychometric evaluation of outcome measures in LET-specific patients and 11 articles reporting their use in a UK population (fig 1)(Appendix 2).

**Measures**

Five outcome measures were identified that were developed or had undergone psychometric evaluation, on UK populations that at least, in part, contain patients with LET (Table 1). They were all fully standardised measures that had all been developed for measuring symptoms (mainly pain) and functioning in English speaking UK adults of either gender. However, only one of them, the Patient-Rated Tennis Elbow Evaluation (PRTEE) was LET specific, the remaining instruments were developed as elbow-specific tools designed for varying pathologies, but including in their validation a sub-sample of LET patients. Two outcome measures were originally developed for UK populations: the Oxford Elbow Score (OES) and the Liverpool Elbow Score (LES). The remaining three outcome measures were developed in the English language outside of the UK (US, Canada and Australia), but had undergone some level of psychometric evaluation in UK populations. Of note, no modification was deemed necessary in the wording or description of the symptoms or activities measured for any of those instruments.
Only the PRTEE has had its metric properties assessed in a UK cohort that was exclusively diagnosed with LET. This was conducted on 57 patients to quantify the Minimally Important Difference (MID) of the PRTEE. This study formed part of a larger prospective trial assessing microcurrent therapy in LET and analysed data from 57 individuals with clinically and sonographically diagnosed LET who all underwent microcurrent therapy. They report a weak correlation between the PRTEE and global change scale, but no assessment of construct validity or any other metric assessment is undertaken. For the four remaining outcome measures, the proportion of patients included within their study cohorts who were diagnosed of LET ranged from 11% to 12.7% (Table 1). None were evaluated in more than 12 patients, and as multiple measures were reported on the same patients cohorts, when all individual patients from these studies were tallied it reveals that this equates to 20 UK LET patients in total.

Eleven additional articles reported using PROMs to evaluate disease impact in UK populations with LET. These studies were published between 2003 and 2014 (Table 2). Out of the five outcome measures for which there had been a previous psychometric evaluation, only three were subsequently applied to evaluate LET outcomes (DASH, OES and PRTEE). The outcome measures that were not utilised were the LES and the Mayo Elbow Performance Score (MEPS). Perhaps more surprisingly, two additional measures were used, namely the Nirschl score and the Patient-Rated Wrist Evaluation (PRWE), although no evidence on the psychometric properties or even their cross-cultural equivalence was available. Overall, the PRTEE (and precursor PRFEQ) was reported six times, the DASH four times, the Nirschl score twice, the OES once and the PRWE once. Seven of the 11 studies stated that the outcome measure was their study’s primary outcome.

Four of these 11 studies were randomised controlled trials (RCTs). The level of adherence to CONSORT standards for reporting PROMs outcomes for RCTS for the four trials suggested substantial room for improvement (Table 3). No information was available for three
standards for any RCT and only partial information was available for the other two standards in a minority of studies.

Discussion
This study has identified a lack of evidence with which to inform outcome measure choice in Lateral Elbow Tendinopathy in the UK. Future validation of outcome measures in UK populations is required in order to be able to ground any recommendations on a firm evidence base. Furthermore, some outcome measures are currently being used as primary outcomes in UK-based studies in the absence of any evidence for their cross-cultural appropriateness and psychometric properties.

We were able to retrieve at least some evidence of the evaluation of the psychometric properties of five outcome measures. The PRTEE is the only measure specifically designed for the evaluation of a LET population. All measures attempt to measure the domains of function and symptoms in adults. All but the DASH have been designed to assess these domains in reference to the elbow exclusively.

The total reporting of validity, reliability or reproducibility of outcome measures in UK LET patients is limited to 20 patients. All of these patients have been embedded in larger cohorts containing a heterogeneous group of elbow pathology. Due to the limited size of this LET sample, it has been unfeasible for the reporting authors to conduct a standardised psychometric assessment of the outcome measures using methods such as COSMIN or EMPRO.

The largest assessment outcome measure utility in UK LET patients was published by Poltawski et al and included 57 patients. Although this is by far the largest sample of LET patients of any of the studies included here, outcome interpretability through derivation of MCID score was undertaken with no evaluation of other relevant psychometric characteristics. The PRTEE was not originally designed for a UK population and no evidence of formalised cross-cultural evaluation is presented. This would always be necessary when
applying a new instrument to a different population, as the use of language across
continents, though English in origin, confers both linguistic and cultural differences. But in
this case the need was additionally increased by the fact that items in the PRTEE had been
altered prior to administration (the words coffee and milk were removed from the item
“Lift a full coffee cup or glass of milk to your mouth”, “pants” were replaced by “trousers”
and “washcloth or wet towel” by “wet cloth”). The authors acknowledge that the altering of
the outcome measure wording may have altered its measurement properties.

In many circumstances it will be completely appropriate and even highly advisable to alter
the wording of outcome measures. However, it should be undertaken under the principles
of cross-cultural adaptation. It is widely recognised that if a measure is to be used
across cultures, the items must be both linguistically translated and culturally adapted to
maintain the content validity of the outcome measure at a conceptual level. Guillemin et
al have proposed scenarios that should alert authors to situations where translation or
adaptation should be undertaken. In the situation of an outcome measure being used in
another country, but in the same language, cultural adaptation is required. For LET in UK
populations, this would be the case for the DASH, MEPS, PRTEE and Nirschl outcome
measures. Of note, the DASH and quickDASH score have been culturally adapted to UK
English since 2015. To the best of our knowledge, this score had not been utilised in any
of the identified studies.

The process of cross-cultural adaptation has been well reported. A 10-stage process
proposed by the International Society for Pharmacoeconomics and Outcomes Research
(ISPOR) involves forward and backwards adaptation by multiple reviewers, cognitive
interviewing with patient populations and pre-testing of the final questionnaire. Though
this may be seen as a laborious process, users of measures that have not been rigorously
adapted must also be aware that language alterations may alter measurement properties.
Therefore, reference values for group comparison, minimally important difference data or
power calculations may not be valid in the new cultural context of use.
This study has identified that the reporting of outcome measures in UK LET randomised controlled trials does not conform to the CONSORT-PRO guidance. Though two of the studies were published prior to the guidance publication in 2010, the stark paucity of reporting of outcome measure detail is concerning. This lack of reporting is in line with the deficits in outcome measure validity highlighted through the Valderas classification system. Though we hypothesised that there would be a preference for outcome measures with published validity in the target population, we have identified that with the current level of evidence this is not possible. This lack of suitable outcome measures has been identified by other authors (29, 30). Long et al (2015) reported in their National Institute of Health Research, Health Technology Assessment review of systematic reviews of conservative treatments in LET, that a lack of standardised outcome measures hindered interpretation and synthesis of results. They recommend that the inclusion of a patient-reported measure of upper extremity function in interventional trials would ease results synthesis. However, we have identified that the lack of a clear choice within the UK population is likely to significantly hinder a researcher’s ability to undertake this.

The authors acknowledge the inherent limitations of this study. The search strategy may have failed to identify all outcome measures used, and the identification of the study populations’ nationality in interventional trials can be prone to error. However, attempts were made to ensure that the strategy was as robust as possible. Outcomes in LET can be measured in numerous ways, including grip strength, pain provocation tests and visual analogue scales to mention a few, this may be a highly legitimate method and was not assessed as part of this study. The authors feel that this approach is justified owing to the increasing view that the ultimate measure of success in health care is whether it helps patients from their own point of view (31). Outcome measures, that quantify patient’s health-related quality of life, with particular reference to PROMs, are recommended by National bodies across the world, including the NIHR in the UK and FDA in the USA (7). Furthermore, the use of condition-specific PROMs is increasingly common in
musculoskeletal medicine and are collected as part of the English NHS PROMs programme\(^7\). With the increasing use of PROMs used as primary outcomes in clinical trials, it is, therefore, relevant that their use is rigorously assessed.

This study has identified that, with current levels of evidence, it is not appropriate to recommend any PROMs for LET studies in UK populations. Though the OES, PRTEE and DASH show potential as patient-reported measures, with domains likely to be appropriate in LET, further assessment is required in UK populations to quantify their validity, reliability, responsiveness and patient acceptability.

**Take home messages**

There is some evidence for the psychometric properties of OES, PRTEE and DASH PROMs in the assessment of patients with Lateral Elbow Tendinopathy. Robust evidence on the validity, reliability and responsiveness of any PROM in UK populations of Lateral Elbow Tendinopathy patients is lacking.

**References**

7. Devlin NJ, Appleby J. Getting the most out of PROMS. Putting health outcomes at the heart of NHS decision making. London: King’s Fund. 2010.
Fig 1: PRISMA Flowchart of the systematic literature review.

Database Searching (n=7261)
(CINHAL =1403, Embase =199, Medline =5607)
Duplicates removed =1076
Review of titles (n=6165)
Articles excluded as non-topic related = 4271
Review of abstracts (n=1914)
Excluded (n= 1688)
Non-Elbow/upper limb specific outcome, study in those aged <18yrs, or non-elbow specific condition = 825
Non-LET elbow pathology = 836
Physical examination only = 0
PROMs Review article = 16
Editorial comment = 5
Full text review (n=256)
Articles reporting metric properties of elbow/upper limb outcome measure in Lateral Epicondylar Tendinopathy (Internationally)
Articles excluded as non-UK population = 242
Included studies (n=16)
Lateral Epicondylar Tendinopathy outcome measure in UK populations
Articles reporting instrument development = 6
Articles reporting instrument use = 11 (7 Outcome measures)
### Table 1: Outcome measures for the assessment in Lateral Elbow Tendinopathy (LET) with psychometric evaluation in UK population

<table>
<thead>
<tr>
<th>Outcome measure</th>
<th>Country of origin</th>
<th>Exclusively Patient Reported (no. items)</th>
<th>Construct (no. items)</th>
<th>Population*</th>
<th>Measurement model $</th>
<th>UK LET assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disabilities of the Arm Shoulder and Hand (DASH) $^{[35]}$</td>
<td>US, Canada, Australia</td>
<td>Yes (30)</td>
<td>A1. Symptoms Pain (5) / A2. Function Physical function (21) / Psychosocial (4)</td>
<td>B1. Adults / B2. All genders / B3. Applied to multiple elbow pathologies (36)</td>
<td>C1. Index / C2. Psychometric / C3. Completely Standardised</td>
<td>Surgically treated LET patients make up 11.2% (n=12/107) of the total development and validation cohort (31, 32, 33) / Tertiary care patients with LET make up 12.7% (n=8/63) of the total development and validation cohort (24)</td>
</tr>
</tbody>
</table>

* All measures were developed for English-speaking adults of either gender. $ All measures were fully standardised.
Table 2: Studies reporting the use of PROMs in patients with Lateral Elbow Tendinopathy.

<table>
<thead>
<tr>
<th>Author</th>
<th>Year</th>
<th>Title</th>
<th>Study Type and Population</th>
<th>Outcome measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dunkow, P. D., et al.</td>
<td>2004</td>
<td>A comparison of open and percutaneous techniques in the surgical treatment of tennis elbow</td>
<td>RCT</td>
<td>DASH*</td>
</tr>
<tr>
<td>Connell, D. A., et al.</td>
<td>2006</td>
<td>Ultrasound-guided autologous blood injection for tennis elbow</td>
<td>Prospective Cohort</td>
<td>Nirshi*</td>
</tr>
<tr>
<td>Alizadehkhaiyat, O., et al.</td>
<td>2007</td>
<td>Pain, functional disability, and psychologic status in tennis elbow</td>
<td>Cross-sectional</td>
<td>DASH PRWE PRFEQ</td>
</tr>
<tr>
<td>Connell, D., et al.</td>
<td>2009</td>
<td>Treatment of lateral epicondylitis using skin-derived tenocyte-like cells</td>
<td>Prospective Pilot Study (Not Randomised)</td>
<td>PRTEE*</td>
</tr>
<tr>
<td>Clarke, A. W., et al.</td>
<td>2010</td>
<td>Lateral elbow tendinopathy: correlation of ultrasound findings with pain and functional disability</td>
<td>Prospective Cohort</td>
<td>PRTEE*</td>
</tr>
<tr>
<td>Creaney, L., et al.</td>
<td>2011</td>
<td>Growth factor-based therapies provide additional benefit beyond physical therapy in resistant elbow tendinopathy: a prospective, single-blind, randomised trial of autologous blood injections versus platelet-rich plasma injections</td>
<td>RCT</td>
<td>PRTEE*</td>
</tr>
<tr>
<td>Nazar, M., et al.</td>
<td>2012</td>
<td>Percutaneous Tennis Elbow Release Under Local Anaesthesia</td>
<td>Prospective Cohort</td>
<td>DASH* OES</td>
</tr>
<tr>
<td>Tonks, J. H., et al.</td>
<td>2007</td>
<td>Steroid injection therapy is the best conservative treatment for lateral epicondylitis: a prospective randomised controlled trial.</td>
<td>RCT</td>
<td>PRTEE</td>
</tr>
</tbody>
</table>

* Primary outcome
Table 3: Adherence to CONSORT reporting standards (PRO extension) of UK-based Lateral Elbow Tendinopathy RCTs.

<table>
<thead>
<tr>
<th>CONSORT 2010 statement</th>
<th>PRO Extension</th>
<th>Studies meeting the requirements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Structured summary of trial design, methods, results, and conclusions</td>
<td>The PRO should be identified in the abstract as a primary or secondary outcome</td>
<td>1/4</td>
</tr>
<tr>
<td>Specific objectives or hypotheses</td>
<td>The PRO hypothesis should be stated and relevant domains identified, if applicable</td>
<td>0/4</td>
</tr>
<tr>
<td>Completely defined pre-specified primary and secondary outcome measures, including how and when they were assessed</td>
<td>Evidence of PRO Instrument validity and reliability should be provided or cited if available, including the person completing the PRO and methods of data collection (paper, telephone, electronic, other)</td>
<td>0/4 (validity of PROM in UK population) vs 2/4 (validity of PROM in another LET population) 1/4 (data collection method)</td>
</tr>
<tr>
<td>Statistical methods used to compare groups for primary and secondary outcomes</td>
<td>Statistical approaches for dealing with missing data are explicitly stated</td>
<td>0/4</td>
</tr>
<tr>
<td>Trial limitations addressing sources of potential bias, imprecision, and, if relevant multiplicity of analyses</td>
<td>PRO-specific limitations and implications for generalisability and clinical practice should be discussed</td>
<td>0/4</td>
</tr>
</tbody>
</table>
Appendix 1:

Search Strategy – MEDLINE – Run 1/5/2017

Medline
1. exp Elbow/
2. elbow.tw.
3. exp Elbow joint/
4. exp Tennis Elbow/
5. epicondylitis.tw.
6. common extensor origin.tw.
7. epicondylalgia.tw.
8. 1 or 2 or 3 or 4 or 5 or 6 or 7
9. exp "Outcome Assessment (Health Care)"/
11. patient reported outcome?.tw.
12. outcome? measure?.tw.
13. exp health status/
14. health status.tw.
15. exp "quality of life"/
16. quality of life.tw.
17. (QL or QoL or HRQL or HRQoL).tw.
18. (function* adj2 (status or psychological or mental or physical or social)).tw.
19. disabilit*.tw.
20. exp "Activities of Daily Living"/
21. activities of daily living.tw.
22. (wellbeing or well being).tw.
23. exp happiness/
24. (happi* or happy).tw.
25. 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24
26. assessment.tw.
27. index.tw.
28. indices.tw.
29. instrument?.tw.
30. measure?.tw.
31. profile?.tw.
32. rating?.tw.
33. report*.tw.
34. scale?.tw.
35. schedule?.tw.
36. scor*.tw.
37. exp health surveys/
38. survey?.tw.
39. 26 or 27 or 28 or 29 or 30 or 31 or 32 or 33 or 34 or 35 or 36 or 37 or 38
40. (symptom? adj2 (assessment or index or indices or instrument? or measure? or profile? or rating? or report* or scale? or schedule? or scor* or survey?)).tw.
41. 25 or 40
42. exp Self-Assessment/
43. self-assess*.tw.
44. exp Questionnaires/
45. questionnaire?.tw.
46. self report*.tw.
47. 42 or 43 or 44 or 45 or 46
48. (Validation Studies or Comparative Study),pt. or exp psychometrics/ or psychometr*.tw. or clinimetr*.tw. or exp observer variation/ or observer variation.tw. or exp Health Status Indicators/ or exp reproducibility of results/ or reproducib*.tw. or exp discriminant analysis/ or relia*b.tw. or unrelia*b.tw. or valid*.tw. or coefficient.tw. or homogeneity.tw. or homogeneous.tw. or internal consistency.tw. or (cronbach* and (alpha or alphas)).tw. or (item and (correlation* or selection* or reduction*)).tw. or agreement.tw. or precision.tw. or imprecision.tw. or precise values.tw. or test-retest.tw. or (test and retest).tw. or (reliab* and (test or retest)).tw. or stability.tw. or interrater.tw. or inter-rater.tw. or intrarater.tw. or intra-rater.tw. or intertester.tw. or inter-tester.tw.
or intratester.tw. or intra-tester.tw. or interobserver.tw. or inter–observer.tw. or intraobserver.tw. or intratechnician.tw. or intertechnician.tw. or intra-technician.tw. or interexaminer.tw. or inter-examiner.tw. or intraexaminer.tw. or interindividual.tw. or intra-individual.tw. or interindividual.tw. or intra-individual.tw. or interparticipant.tw. or inter-participant.tw. or intraparticipant.tw. or intra-participant.tw. or kappa.tw. or kappa*.tw. or kappas.tw. or repeat.tw. or ((replicab* or repeated) and (measure or measures or findings or result or results or test or tests)).tw. or concordance.tw. or (intraclass and correlation*).tw. or discriminative.tw. or known group.tw. or factor analysis.tw. or factor analyses.tw. or dimension*.tw. or subscale*.tw. or (multitrait and scaling and (analysis or analyses)).tw. or item discriminant.tw. or interscale correlation*.tw. or error.tw. or errors.tw. or individual variability.tw. or (variability and (analysis or values)).tw. or (uncertainty and (measurement or measuring)).tw. or standard error of measurement.tw. or sensitiv*.tw. or responsive*.tw. or ((minimal or minimally or clinical or clinically) and (important or significant or detectable) and (change or difference)).tw. or (small* and (real or detectable) and (change or difference)).tw. or meaningful change.tw. or ceiling effect.tw. or floor effect.tw. or Item response model.tw. or IRT.tw. or Rasch.tw. or Differential item functioning.tw. or DIF.tw. or computer adaptive testing.tw. or item bank.tw. or cross-cultural equivalence.tw.

49. 39 or 47 or 48
50. 41 and 49
51. (Oxford elbow score or Liverpool Elbow Score or Elbow Self-Assessment Score or Elbow Function Assessment or (American Shoulder and Elbow Surgeons-elbow) or (Modified American Shoulder and Elbow Surgeons) or Mayo Elbow Performance Score or Hospital for Special Surgery score or Hospital for Special Surgery short version or patient-rated elbow evaluation or Patient-Rated Tennis Elbow Evaluation or Elbow Functional Assessment or (Disabilities of the Arm, Shoulder and Hand questionnaire) or subjective elbow value or (Broberg and Morrey) or Ewald).mp. or Pritchard.tw.
[mp=ti, ab, ot, nm, hw, kf, px, rx, ui, tn, dm, mf, dv, kw]
52. (OES or LES or ESAS or ASES or ASES-e or MEP or PREE or PRTEE or EFA or DASH or quickDASH).mp.
[mp=ti, ab, ot, nm, hw, kf, px, rx, ui, tn, dm, mf, dv, kw]
53. 8 and 52
54. 8 and 50
55. 51 or 53 or 54
56. exp ANIMALS/ not humans.sh.

Appendix 2 : Included reference from systematic search