Elsevier Editorial System(tm) for Journal of Shoulder and Elbow Surgery

Manuscript Draft

Manuscript Number:

Title: Clinical Rating Systems in Elbow Research - A Systematic Review Exploring Trends and Distributions of Use

Article Type: Review Article

Keywords: Clinical rating systems
Elbow
Quality of Life
Patient-reported outcome measure
Validity

Corresponding Author: Mr. Jonathan Peter Evans, MBChB, BSc (Hons), MSc, MRCS

Corresponding Author's Institution: Royal Devon and Exeter Hospital

First Author: Jonathan Peter Evans, MBChB, BSc (Hons), MSc, MRCS

Order of Authors: Jonathan Peter Evans, MBChB, BSc (Hons), MSc, MRCS; Chris Smith, MD FRCS; Nicola Fine, MBChB MRCS; Ian Porter, PhD; Jaheeda B Gangannagaripalli3 , PhD; Vicki Goodwin, PhD, MBE; Jose M Valderas, BMBS, MPH, PhD

Abstract: Background: Clinical rating systems are used as outcome measures in clinical trials and attempt to gauge the patient's views of their own health. The choice of clinical rating system should be supported by its performance against established quality standards.

Methods: A search strategy was developed to identify all studies reporting the use of clinical rating systems in the tennis elbow literature. The strategy was run from inception in Medline Embase and CINHAL. Data extraction identified the date of publication, country of data collection, pathology assessed and outcome measure used.

Results: 980 studies were identified that reported clinical rating system use. 72 separate rating systems were identified. 41% of studies used two or more separate measures. Overall 54% of studies used the Mayo Elbow Performance Score (MEPS). For Arthroplasty 82% used MEPS, 17% used Disabilities of Arm, Shoulder and Hand (DASH), 7% used quickDASH. For Trauma 66.7% used MEPS, 32% used DASH, 23% used the Morrey Score. For Tendinopathy, 31% used DASH, 23% used Patient-Rated Tennis Elbow Evaluation, 13% used MEPS. Over time there is increased proportional use of the MEPS, DASH, qDASH, Patient Rated Tennis Elbow Evaluation (PRTEE) and Oxford Elbow Score (OES).

Conclusions: This study has identified the wide choice and usage of clinical rating systems in the elbow literature. Numerous studies report measures without a history of either pathology specific or cross-cultural validation. Interpretability and comparison of outcomes is dependent on the unification of outcome measure choice. This is not currently demonstrated.

Royal Devon and Exeter Hospital

Barrack Rd

Exeter, UK

EX2 4RU

25/08/17

Dear JSES editorial team,

We wish to submit an original review article entitled “Clinical Rating Systems in Elbow Research – A Systematic Review Exploring Trends and Distributions of Use” for consideration by the Journal of Shoulder and Elbow Surgery.

We confirm that this work is original and has not been published elsewhere, nor is it currently under consideration for publication elsewhere.

This work was presented at the British Elbow and Shoulder Society Annual Conference 2017 (Coventry, UK 21st-23rd June)) and has been accepted for presentation at the British Orthopaedic Association Annual Congress 2017 (Liverpool, UK 19th-22nd September). The abstracts from these conferences are not published in peer reviewed journals.

The manuscript has been read and approved by all authors.

For onward correspondence please contact
Jonathan Evans
Smeall Building, JS03,
St Lukes Campus,
Exeter, UK
EX1 2LU
Email: j.evans3@exeter.ac.uk
Phone (+44) 07974987989

Thank you for your consideration of this manuscript.

Sincerely,

Jonathan Evans

***Conflict of Interest**

Royal Devon and Exeter Hospital

Barrack Rd

Exeter, UK

EX2 4RU

25/08/17

Dear JSES editorial team,

Regarding the manuscript:

"Clinical Rating Systems in Elbow Research – A Systematic Review Exploring Trends and Distributions of Use"

The submitting authors do not have any conflicts of interest to disclose:

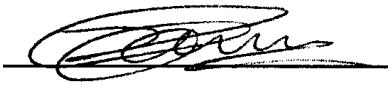
This study did not receive industry sponsorship of grant funding.

Thank you for your consideration of this manuscript.

Sincerely,

All signed on 12/9/17

Jonathan P Evans MBChB MRCS



Chris Smith MD FRCS

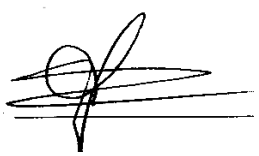
Nicola Fine MBChB MRCS

Ian Porter PhD

Jaheeda Gangannagaripalli PhD

Vicki Goodwin PhD MBE

Jose Valderas BMBS MPH PhD



Clinical Rating Systems in Elbow Research –

A Systematic Review Exploring Trends and Distributions of Use

Short title: Clinical Rating Systems in Elbow Research

Jonathan P Evans[1,2,3] MBChB MRCS
Chris Smith[1] MD FRCS
Nicola Fine[1] MBChB MRCS
Ian Porter[2] PhD
Jaheeda Gangannagaripalli[2] PhD
Vicki Goodwin[3] PhD
Jose Valderas[2] BMBS MPH PhD

[1] Royal Devon and Exeter NHS Foundation Trust
[2] Health Services and Policy Research Group, University of Exeter
[3] National Institute for Health Research (NIHR) Collaboration for Leadership in Applied Health Research and Care (CLAHRC) South West Peninsula, University of Exeter Medical School

Corresponding Author:

Jonathan P Evans

Health Services and Policy Research, Smeall Building, JS03, St Lukes Campus, Exeter, EX1 2LU, United Kingdom

j.evans3@exeter.ac.uk

Disclaimer: None

Figures 3 and 4 to be reproduced in color to improve clarity.

Abstract

Clinical Rating Systems in Elbow Research –
A Systematic Review Exploring Trends and Distributions of Use

Background: Clinical rating systems are used as outcome measures in clinical trials and attempt to gauge the patient's views of their own health. The choice of clinical rating system should be supported by its performance against established quality standards.

Methods: A search strategy was developed to identify all studies reporting the use of clinical rating systems in the tennis elbow literature. The strategy was run from inception in Medline Embase and CINHAL. Data extraction identified the date of publication, country of data collection, pathology assessed and outcome measure used.

Results: 980 studies were identified that reported clinical rating system use. 72 separate rating systems were identified. 41% of studies used two or more separate measures. Overall 54% of studies used the Mayo Elbow Performance Score (MEPS). For Arthroplasty 82% used MEPS, 17% used Disabilities of Arm, Shoulder and Hand (DASH), 7% used quickDASH. For Trauma 66.7% used MEPS, 32% used DASH, 23% used the Morrey Score. For Tendinopathy, 31% used DASH, 23% used Patient-Rated Tennis Elbow Evaluation, 13% used MEPS. Over time there is increased proportional use of the MEPS, DASH, qDASH, Patient Rated Tennis Elbow Evaluation (PRTEE) and Oxford Elbow Score (OES).

Conclusions: This study has identified the wide choice and usage of clinical rating systems in the elbow literature. Numerous studies report measures without a history of either pathology specific or cross-cultural validation. Interpretability and comparison of outcomes is dependent on the unification of outcome measure choice. This is not currently demonstrated.

Background

The ultimate measure of success in health care is whether it helps patients as they see it [8]. In an effort to capture the effect of health interventions on patients, there has been a considerable investment of resources by academics and clinicians to develop systematic, robust and valid ways of collecting health data from patients [24]. It is now the current standard that treatment evaluation includes the use of clinical rating systems [52].

Current clinical rating systems in elbow research utilize both physician and patient completed measures. They aggregate various attributes of interest such as elbow pain, range of motion and ability to perform specific tasks [46]. Though there was a historical focus on physician-administered tools, recent emphasis has been on the patient-rated outcome measurement (PROM), whereby information is gathered pertaining to the patients' perception of their elbow function [52].

The rise in the use of clinical rating systems has accompanied a fundamental shift in how we measure health. Traditional measurements of treatment effect, such as length of hospital stay, radiographic markers or range of motion, are increasingly accompanied by, or indeed replaced by rating systems, with a particular emphasis on PROMs [24]. In the United States, the Food and Drug Administration (FDA) recommends the use of PROMs in clinical trials [10]. Within the UK, the use of PROMs is commonplace in assessing the effectiveness and cost-effectiveness of healthcare technologies [40]. Furthermore, the UK's National PROMs programme [56] has led the world in the standardized collection of PROMs for hip and knee arthroplasty.

The increasing popularity of patient-focused outcome measurement has accompanied a consequent rise in the production of numerous rating systems. When choosing the appropriate rating system, to be applied to either clinical or research purposes, it is necessary

to identify existing systems that measure the outcome of interest in the target population [58]. To aid in this selection, databases such as ePROVIDE (https://eprovide.mapi-trust.org/) catalog potential rating systems, though by their own admission, their database is supplied exponentially with new tools.

Careful consideration must be given to the selection of the clinical rating system. An appropriate measure should be supported by published evidence demonstrating that it is acceptable to patients, reliable, valid and responsive (sensitive to change) [18]. Furthermore, these properties should have been tested on similar reference groups of patients to those being studied, thereby ensuring the validity of a tool from a language and cultural perspective [2]. Within the domain of orthopedics, particular emphasis has been placed on the use of clinical rating systems for particular anatomical locations (predominantly joints) rather than generic health measures. More recently this has evolved to concentrate on condition-specific tools, where, in certain groups or in certain conditions, generic or region specific tools miss important aspects of health status [24]. For the appropriate interpretation, it is, therefore, vital that the clinical rating system selected is validated for use in the population of interest and for the specific condition being investigated.

Heterogeneity of outcome selection has been reported in systematic reviews of elbow related controlled trials where there is consistent comment that this heterogeneity hampers effective evidence synthesis [12, 14, 34]. Initiatives to combat this include the Core Outcome Measures in Effectiveness Trials (COMET) and the U.S. National Institutes of Health (NIH) Patient Reported Outcomes Measurement Information System (PROMIS®), who aim to bring standardization to outcome measure selection. By adopting common standards and metrics clinical researchers will be able to directly compare patients' evaluations of interventional effects across countries, thereby increasing the relevance of results and enabling International syntheses (such as meta-analyses) of research findings [2].

Systematic reviews assessing elbow-specific clinical rating systems have concluded that a paucity of quality measures exist [21, 30, 52, 53]. The most recent review by The et al [52] included the assessment of 12 rating systems using the Consensus-Based Standards for the Selection of health Measurement Instruments (COSMIN) checklist, the authors conclude that the Oxford Elbow Score (OES) is the only system that has been developed using high-quality methodology.

The distribution of use of elbow-specific rating systems across different elbow pathologies is not known. Riedel et al [46] reviewed 65 articles, which used elbow specific aggregate scores specifically in elbow arthroplasty published between 2004 - 2011. They report the predominant use of the Mayo Elbow Performance score in 75% of the literature they identified. They criticise the use of this physician administered score that was not developed with a formal methodology and is frequently inconsistently applied.

This study aims to assess the use of clinical rating systems in elbow related interventional studies. The assessment of the appropriation of rating systems to specific elbow pathologies and across populations has not been undertaken. Furthermore, the change in trends of use over time, with the recent increased emphasis on PROMs use, has not been evaluated. Only when armed with the knowledge of either the conformity or heterogeneity of rating systems, can compelling arguments be made for the need for standardization.

Method

A comprehensive review of elbow specific clinical rating systems in the elbow literature was conducted. This review aimed to identify all articles reporting the use of both physician and patient-reported rating systems. Both rating systems designed specifically for use in elbow pathology and generic upper limb rating systems with a history of validation and in elbow pathology were included. The presented report has been written following PRISMA guidelines[35]. A search strategy was constructed using MeSH and free-text terms (appendix 1).

The strategy was modeled to each database through the modification of thesaurus terms, wildcards, and truncation. The search was run on 1st May 2017 in Medline (Ovid MEDLINE, 1948 to 2016 & Ovid MEDLINE In-Process & Non-indexed Citations) accessed through OVIDSP, Embase (Embase 1974 to 2017) accessed through OVIDSP and CINHAL (CINHAL 1981 to 2017) accessed through EBSCO host.

The search strategy development was guided by previously published search strategies for systematic reviews of interventions in elbow pathology [11] and for the identification of outcome measures [26], along with terms specifically selected in order to capture names of relevant instruments published in previous systematic reviews of elbow specific rating scales [21, 30, 52, 53].

The review was conducted in a step-wise manner. At each stage, dual review was employed with the lead author and a further co-author reviewer. In cases of disagreement between reviewers, the article proceeded to the next stage of review to ensure maximum sensitivity. Initial title review was used to exclude duplicates, studies in pediatric populations, non-elbow based studies, case-reports, case-studies, surgical technique papers and conference abstracts. Abstract review used the above criteria and also excluded studies that did not report the use

of rating systems designed specifically for elbow measurement, or generic rating systems with no history of validation in elbow measurement.

Data extraction was conducted by [redacted] and [redacted]. Publication date, geographical location of lead author or publishing institution, elbow pathology investigated and elbow specific clinical rating systems reported was extracted.

The elbow specific pathology or intervention of interest was grouped into the following categories for ease of interpretation: arthritis interventions (non-arthroplasty), arthroplasty (trauma and elective), arthroscopy, distal biceps intervention, neuropathy intervention, sports-specific population, tendinopathy (non-sports specific population) and trauma interventions (non-arthroplasty).

References were retrieved and imported into reference management software (Endnote X7, © 2017 Clarivate Analytics, PA, USA). Database management was conducted in Excel (Microsoft® Excel® 2013, Redmond, WA, USA).

Results

The review identified 980 articles reporting the use of elbow-specific clinical rating systems (fig 1). Articles from 52 countries were included. 72 separate instruments were identified (appendix 2).

The 980 articles reported the use of 1383 outcomes. 322 (32%) of articles reported the use of two separate elbow-specific clinical rating systems, 77 (8%) reported the use of three, 4 (0.4%) reported the use of four separate elbow-specific clinical rating systems.

The number of articles reporting elbow specific rating systems has increased over time (fig 2) reaching 106 published articles in 2016.

Overall, from database inception, the Mayo Elbow Performance (MEP) score was reported in 54% of articles, the Disabilities of Arm Shoulder and Hand (DASH) in 29%, the Morrey Score 12%, the abbreviated DASH (quickDASH) in 8%, the Patient-Rated Tennis Elbow Evaluation (PRTEE) in 5%, the American Shoulder and Elbow Society-Elbow score (ASES-e) in 4%, the Oxford Elbow Score (OES) in 4%. All other scores were reported in less than 2% of articles.

Since 2000 and 2010 respectively, the proportionate use within the literature for the above rating systems are: MEPS 55% & 61%, DASH 30% & 34%, Morrey 12% & 9%, quickDASH 9% & 13%, PRTEE 5% & 7%, ASES-e 4% & 3%, and OES 4% & 6% (fig 3).

The top five clinical rating systems for the individual pathology or intervention group are outlined in fig 13.

For the three largest groups; arthroplasty, tendinopathy, and trauma, the most popular clinical rating systems are further grouped by time periods; database inception, since 2000 and since 2010 (table 1).

Geographic distribution is shown in table 2, with data grouped into three broad localities; North America, Europe, and Rest of the World.

Discussion

The elbow has long been thought of as the forgotten joint, with pathologies that are difficult to treat and surgical procedures that carry higher complication rates than any other major joint [7]. However, modern diagnostic and treatment practices have shown great promise, and clinical effectiveness research has sought to accurately quantify the benefits patients are experiencing. In keeping with modern research reporting practice, the ultimate goal has been to demonstrate the ability of an intervention to restore or preserve functioning and well-being related to health, that is health-related quality of life (HRQoL) [42].

This study has demonstrated that the use of elbow-related clinical rating systems, that aim, in some form, to demonstrate patient related benefit following an intervention, is rapidly expanding year on year. Though previously published systematic reviews of elbow rating systems have highlighted the deficits in many of the tools [30-32, 52], trend data have failed to show large shifts in choice towards tools produced with high quality methodology.

Global data across pathologies and interventions of 980 articles have identified the Mayo Elbow Performance Score (MEPS) Score as the predominant rating system. The MEPS was developed by Morrey and Adams in 1992 [37], for outcome assessment in total elbow arthroplasty. It consists of a physician assessment of pain, arc of motion and stability, with a patient rating of daily function. It has a history of validation in elective elbow surgery patients with mixed pathology[17, 19], arthroplasty [37], trauma [38] and rheumatoid arthritis [21]. Assessment under the COSMIN checklist rated all its development and validation domains as fair to poor [52]. The Disabilities of Arm Shoulder and Hand (DASH) is also commonly employed. This patient-reported outcome measure (PROM) was introduced in 1996 [28]. It consists of a 31 core item questionnaire with 8 additional questions for sport and work assessment. It was designed to evaluate the entire upper limb but has a history of validation in

elbow-specific pathology including; arthrolysis [27], arthroplasty [6, 5], lateral epicondylar tendinopathy [3, 13, 29, 39, 41, 47, 50, 55], rheumatoid arthritis [44], neuropathy [33, 59], elective elbow surgery [17, 19, 25, 49], biceps tendon repair and radial head post-surgery [57]. It has not undergone systematic evaluation and head to head comparison with other elbow-specific rating systems using recognized techniques such as COSMIN [36] or EMPRO [54]. Of the other scores, large heterogenicity of application was demonstrated, astoundingly 72 separate instruments were identified across the literature, since 2010, 45 of these separate instruments are continuing to be used.

Assessment of the use of rating systems in the predefined criteria groups showed some element of preference for specially designed scales. Neuropathy, sports specific population, and tendinopathy groups showed the utilization of scores specifically designed for population or pathology use. Examples include the Dellon score [23] in neuropathy, Andrews-Carson score [4] in sports population and the Patient-Rated Tennis Elbow Evaluation (PRTEE)[47] in the tendinopathy group. Of note, within all the above-mentioned groups and distal biceps group, the DASH score remains the first or second score of preference.

The three largest subgroups were assessed for a change of rating scale use over time. The recent emergence and promotion of patient-rated over physician-rated evaluation would lead most to the hypothesis of increased proportional representation within these groups over time. Within all groups, the use of DASH and quickDASH is rising, the tendinopathy group also demonstrated a particular emergence of the PRTEE. Yet, whilst the Morrey and HSS are declining in use, of interest is the progressive rise of the MEPS across all groups.

Trends in rating-systems in differing geographical areas since 2010 was also assessed. Though the sub-division of areas is rather crude, this sub-division yielded groups of a size substantial enough to interpret broad distribution trends. The MEPS, though developed in the

USA, has a higher total proportion of use in Europe and Rest of the World groups. The MEPS has only been formally assessed for cross-cultural validity in Turkish [16] [15] UK English [17, 19] and Dutch [20]. The DASH score is proportionally more popular in North America, though it has been cross-culturally adapted to multiple languages. Interestingly, the abbreviated quickDASH is twice as commonly employed in Europe when compared to the USA. Again the quickDASH is available in multiple languages, but it is important to note that in terms of elbow-specific cross-cultural adaptation and validation, this has only been conducted in Turkish, Italian and Dutch [3, 22, 25].

Standardization of outcome evaluation, together with consensus in the scientific community, is an essential component of the future of comparative effectiveness research. Only then will we be able to compare results between different groups, hospitals, and protagonists [51]. The shift in focus from physician to patient-reported outcomes is well documented, with support both within the literature and from a governmental/health service level [1]. Within elbow specific literature Dawson et al [17] reported that patient-reported results are more likely than clinically assessed outcome measures to reflect patient satisfaction with elbow surgery. Furthermore, they also provide support that condition-specific measures are more likely than generic measures to be more closely aligned with patient satisfaction. Yet, we have shown that within the literature there remains a persistent reticence to embrace PROMs more fully. As Snyder et al [48] comment, though PROMs have the potential to improve the quality of patient-centeredness medical care, there is a great deal of research to be done before they are fully embraced by all stakeholders. Within elbow-specific literature, it may be the consensus that the literature is, as yet, uncompelling and lacking clear recommendations. Recent review evidence, that systematically assess the development and psychometric properties of elbow specific rating systems, has only emerged since 2013 [46, 52] and it may be that the trickle-down effect may simply not have been felt.

It is vitally important to recognize that inappropriate rating system choice can have a great impact on the interpretation of results, particularly where they are used as the primary endpoint in clinical studies [52]. The choice of a rating system should be optimally aligned with a conceptual framework that defines the health condition and will meet the performance requirements of the clinical context and measurement needs [32]. Therein, a score developed for the elbow may not be valid across all populations and all pathologies. Quantification of health-related quality of life in an elderly rheumatoid arthritis patient undergoing total elbow arthroplasty may require an evaluation of very different domains to a middle-aged manual laborer with tennis elbow. The clinical rating scale must have demonstrated its validity, reliability, responsiveness, and interpretability for the pathology being investigated to justify its choice. This must also be the case for the population of question where the cross-cultural validation of an outcome measures is a vital component in ensuring its interpretability, with clearly described methodological and reporting requirements [45]. Consequently, for example, an American developed measure in the English language, does not have automatic validity in other English speaking populations.

The future of clinical rating systems in elbow pathology is fluctuating with the same uncertainty that pervades the whole orthopedic research community. Though we have identified numerous rating scales, new measures continue to be produced. Though this highlights the expanding, and exciting, growth in this field, where there are multiple choices, this can lead to greater uncertainty and create barriers to uptake [9]. The use of registries may force some level of conformity in data collection. In the UK the National PROMs programme has collected Oxford hip and knee scores since 2009. With the inclusion of elbow arthroplasty into the National Joint Registry in 2012, it remains likely that an outcome measure will be added to this dataset. The use of PROMIS (Patient-Reported Outcome Measurement Information System), to provide a set of common metrics to which PROMs that

assess comparable constructs can be scaled [9], has shown great utility, but has not been applied to elbow pathology [43].

The authors accept that this study has limitations. As with all systematic reviews, this study is limited by the search strategy used, however, considerable care was taken to produce a strategy that was as sensitive as possible. The subclassifications of data into pathology and population groups were derived to give the best impression possible of rating scale use. The use of arthroplasty, for example, was kept as a single group, though a case can be made that rheumatoid and trauma patients may respond differently and require different rating systems, under the recommendations outlined above. Equally, the trauma group could easily be further sub-classified. However, we feel that the strength of the data is the representation of the three large sub-classification groups.

Conclusion

This study is the first to identify the true magnitude of choice of clinical rating systems for the elbow. From 980 manuscripts we identified 72 individual clinical rating systems. Though we are seeing a small advance in the use of validated condition-specific PROMs, such as the PRTEE, the overwhelming key players in outcome measurements remain the historic or generic measures, such as the MEPS and DASH score. The co-administration of multiple scores may be seen as a panacea, but there is little justification for ever increasing the patient burden. Though the rapid progression of outcomes research may provide computational models of comparison between measurements, in the immediate term, we would call for the clear, systematic evaluation of condition-specific elbow related rating systems, using well-recognized methods such as the COSMIN checklist [36] or EMPRO tool [54]. Only then can clinicians and researchers make informed decisions on the appropriate tool for the elbow pathology and population of interest.

References

1. Ahmed S, Berzon RA, Revicki DA, Lenderking WR, Moinpour CM, Basch E, et al. The use of patient-reported outcomes (PRO) within comparative effectiveness research: implications for clinical practice and health care policy. Medical Care. 2012;50(12):1060-70. DOI: 10.1097/MLR.0b013e318268aaff.

2. Alonso J, Bartlett SJ, Rose M, Aaronson NK, Chaplin JE, Efficace F, et al. The case for an international patient-reported outcomes measurement information system (PROMIS®) initiative. Health and Quality of Life Outcomes. 2013;11(1):1-5 DOI: 10.1186/1477-7525-11-210.

3. Altan L, Ercan I, Konur S. Reliability and validity of Turkish version of the patient rated tennis elbow evaluation. Rheumatology International. 2010;30(8):1049-54 DOI: 10.1007/s00296-009-1101-6.

4. Andrews JR, St Pierre RK, Carson Jr WG. Arthroscopy of the elbow. Clinics in Sports Medicine. 1986;5(4):653-62.

5. Angst F, John M, Pap G, Mannion AF, Herren DB, Flury M, et al. Comprehensive assessment of clinical outcome and quality of life after total elbow arthroplasty. Arthritis & Rheumatism: Arthritis Care & Research. 2005;53(1):73-82 10p.

6. Angst F, Goldhahn J, Drerup S, Kolling C, Aeschlimann A, Simmen BR, et al. Responsiveness of five outcome measurement instruments in total elbow arthroplasty. Arthritis Care Res (Hoboken). 2012;64(11):1749-55 DOI: 10.1002/acr.21744.

7. Antuna S, Barco R. Essential Techniques in Elbow Surgery: Springer; 2016. ISBN 978-3-319-31575-1

8. Berwick D, Hiatt H, Janeway P, Smith R. An ethical code for everybody in health care. BMJ. 1997;315(7123):1633-4.

9. Black N, Burke L, Forrest CB, Sieberer UR, Ahmed S, Valderas J, et al. Patient-reported outcomes: pathways to better health, better services, and better societies. Quality of Life Research. 2016;25(5):1103-12. DOI: 10.1007/s11136-015-1168-3

10. Bren L. The importance of patient-reported outcomes... it's all about the patients. FDA consumer. 2006;40(6):26.

11. Buchbinder R, Johnston RV, Barnsley L, Assendelft W, Bell SN, Smidt N. Surgery for lateral elbow pain. Cochrane Database Syst Rev. 2011;3(3). DOI: 10.1002/14651858.CD003525.pub2

12. Buchbinder R, Maher C, Harris IA. Setting the research agenda for improving health care in musculoskeletal disorders. Nature Reviews Rheumatology. 2015;11(10):597-605 DOI: 10.1038/nrrheum.2015.81.

13. Cacchio A, Necozione S, MacDermid JC, Rompe JD, Maffulli N, di Orio F, et al. Cross-cultural adaptation and measurement properties of the Italian version of the Patient-Rated Tennis Elbow Evaluation (PRTEE) questionnaire. Phys Ther. 2012;92(8):1036-45 DOI: 10.2522/ptj.20110398.

14. Caliandro P, La Torre G, Padua R, Giannini F, Padua L. Treatment for ulnar neuropathy at the elbow. Cochrane Database Syst Rev. 2011;2. DOI: 10.1002/14651858.CD006839.pub4

15. Celik D, Atalar AC, Demirhan M, Dirican A. Translation, cultural adaptation, validity and reliability of the Turkish ASES questionnaire. Knee Surg Sports Traumatol Arthrosc. 2013;21(9):2184-9 DOI: 10.1007/s00167-012-2183-3.

16. Celik D. Psychometric properties of the Mayo Elbow Performance Score. Rheumatol Int. 2015;35(6):1015-20 DOI: 10.1007/s00296-014-3201-1.

17. Dawson J, Doll H, Boller I, Fitzpatrick R, Little C, Rees J, et al. Factors associated with satisfaction in patients undergoing elbow surgery: a prospective study. J Shoulder Elbow Surg. 2010;19(5):635-44 DOI: 10.1016/j.jse.2010.02.003.

18. Dawson J, Doll H, Fitzpatrick R, Jenkinson C, Carr AJ. The routine use of patient reported outcome measures in healthcare settings. BMJ. 2010;340:c186. DOI: 10.1038/tp.2016.61

19. Dawson J, Doll H, Boller I, Fitzpatrick R, Little C, Rees J, et al. Specificity and responsiveness of patient-reported and clinician-rated outcome measures in the context of elbow surgery, comparing patients with and without rheumatoid arthritis. Orthop Traumatol-Sur. 2012;98(6):652-8 DOI: 10.1016/j.otsr.2012.05.011.

20. De Boer YA, Van Den Ende CHM, Eygendaal D, Jolie IMM, Hazes JMW, Rozing PM. Clinical reliability and validity of elbow functional assessment in rheumatoid arthritis. Journal of Rheumatology. 1999;26(9):1909-17 9p.

21. de Boer YA, Hazes JM, Winia PC, Brand R, Rozing PM. Comparative responsiveness of four elbow scoring instruments in patients with rheumatoid arthritis. The Journal of Rheumatology. 2001;28(12):2616-23.

22. de Haan J, Goei H, Schep NW, Tuinebreijer WE, Patka P, den Hartog D. The reliability, validity and responsiveness of the Dutch version of the Oxford elbow score. J Orthop Surg Res. 2011;6:39 DOI: 10.1186/1749-799X-6-39.

23. Dellon AL, Hament W, Gittelshon A. Nonoperative management of cubital tunnel syndrome An 8-year prospective study. Neurology. 1993;43(9):1673-.

24. 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. ISBN978 1 85717 591 2

25. Franchignoni F, Vercelli S, Giordano A, Sartorio F, Bravini E, Ferriero G. Minimal clinically important difference of the disabilities of the arm, shoulder and hand outcome measure (DASH) and its shortened version (QuickDASH). Journal of Orthopaedic & Sports Physical Therapy. 2014;44(1):30-9. DOI: 10.2519/jospt.2014.4893

26. Gonçalves Bradley DC, Gibbons C, Ricci-Cabello I, Bobrovitz NJ, Gibbons EJ, Kotzeva A, et al. Routine provision of information on patient-reported outcome measures to healthcare providers and patients in clinical practice. The Cochrane Library. 2015. DOI: 10.1002/14651858.CD011589

27. Gosling T, Blauth M, Lange T, Richter M, Bastian L, Krettek C. Outcome assessment after arthrolysis of the elbow. Arch Orthop Traum Su. 2004;124(4):232-6. DOI: 10.1007/s00402-003-0524-x

28. Hudak PL, Amadio PC, Bombardier C, Beaton D, Cole D, Davis A, et al. Development of an upper extremity outcome measure: the DASH (disabilities of the arm, shoulder, and hand). American Journal of Industrial Medicine. 1996;29(6):602-8.

29. Janssen S, De Smet L. Responsiveness of the DASH questionnaire for surgically treated tennis elbow. Acta Chir Belg. 2008;108(5):583-5. DOI:10.1080/00015458.2008.11680290

30. Longo UG, Franceschi F, Loppini M, Maffulli N, Denaro V. Rating systems for evaluation of the elbow. British Medical Bulletin. 2008;87:131-61 DOI: 10.1093/bmb/ldn023.

31. MacDermid JC, Michlovitz SL. Examination of the elbow: Linking diagnosis, prognosis, and outcomes as a framework for maximizing therapy interventions. Journal of Hand Therapy. 2006;19(2):82-97 DOI: 10.1197/j.jht.2006.02.018.

32. Macdermid JC, Silbernagel KG. Outcome Evaluation in Tendinopathy: Foundations of Assessment and a Summary of Selected Measures. Journal of Orthopaedic & Sports Physical Therapy. 2015;45(11):950-64 DOI: doi:10.2519/jospt.2015.6054.

33. Malay S, Chung KC, Grp SS. The minimal clinically important difference after simple decompression for ulnar neuropathy at the elbow. Journal of Hand Surgery-American Volume. 2013;38(4):652-9. DOI: 10.1016/j.jhsa.2013.01.022

34. Matar H, Ali A, Buckley S, Garlick N, Atkinson H. Surgical interventions for treating fractures of the olecranon (an elbow bone) in adults. Health. 2014. DOI: 10.1002/14651858.CD010144.pub2

423 35. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews
424 and meta-analyses: the PRISMA statement. Annals of Internal Medicine. 2009;151(4):264-9.

425 36. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN
426 checklist for assessing the methodological quality of studies on measurement properties of health
427 status measurement instruments: an international Delphi study. Qual Life Res. 2010;19 DOI:
428 10.1007/s11136-010-9606-8.

429 37. Morrey B, Adams R. Semiconstrained arthroplasty for the treatment of rheumatoid arthritis of
430 the elbow. J Bone Joint Surg Am. 1992;74(4):479-90.

431 38. Morrey BF. Fracture of the radial head. The elbow and its disorders 2nd ed Philadelphia: WB
432 Saunders. 1993:383-404.

433 39. Newcomer KL, Martinez-Silvestrini JA, Schaefer MP, Gay RE, Arendt KW. Sensitivity of
434 the Patient-rated Forearm Evaluation Questionnaire in lateral epicondylitis. Journal of Hand Therapy.
435 2005;18(4):400-73 8p. DOI: 10.1197/j.jht.2005.07.001

436 40. NICE. National Institute for Healthcare and Clinical Excellence: Guide to the Methods of
437 Technology Appraisal 20132013 [Available from:
438 https://www.nice.org.uk/process/pmg9/chapter/foreword.

439 41. Nilsson P, Baigi A, Marklund B, Mansson J. Cross-cultural adaptation and determination of
440 the reliability and validity of PRTEE-S, a questionnaire for patients with lateral epicondylalgia, in a
441 Swedish population. BMC Musculoskelet Disord. 2008;9:79 DOI: 10.1186/1471-2474-9-79.

442 42. Osoba D, King M. Meaningful differences. Assessing quality of life in clinical trials.
443 2005;2:243-57. ISBN: 9780198527695

444 43. PROsettaStone. Linking Patient-Reported Outcome Measures. 2014. [Available from
445 http://www.prosettastone.org/Pages/default.aspx] Last accessed 20/8/17

446 44. Raven EEJ, Haverkamp D, Sierevelt IN, van Montfoort DO, Poll RG, Blankevoort L, et al.
447 Construct validity and reliability of the disability of arm, shoulder and hand questionnaire for upper

448 extremity complaints in rheumatoid arthritis. Journal of Rheumatology. 2008;35(12):2334-8 DOI:
449 10.3899/jrheum.080067.

450 45. Reeve BB, Wyrwich KW, Wu AW, Velikova G, Terwee CB, Snyder CF, et al. ISOQOL
451 recommends minimum standards for patient-reported outcome measures used in patient-centered
452 outcomes and comparative effectiveness research. Quality of Life Research. 2013;22:1889-905. DOI:
453 10.1007/s11136-012-0344-y

454 46. Riedel K, Beaton DE. Update on the state of outcome measurement in total elbow
455 arthroplasty research: identifying a need for consensus. J Bone Joint Surg Am. 2013;95(14):e97 1-8
456 DOI: 10.2106/JBJS.K.01420.

457 47. Rompe JD, Overend TJ, MacDermid JC. Validation of the Patient-rated Tennis Elbow
458 Evaluation Questionnaire. Journal of Hand Therapy 2007;20(1):3-10; quiz 1 DOI:
459 10.1197/j.jht.2006.10.003.

460 48. Snyder CF, Jensen RE, Segal JB, Wu AW. Patient-reported outcomes (PROs): putting the
461 patient perspective in patient-centered outcomes research. Medical Care. 2013;51(8 0 3):S73. DOI:
462 10.1097/MLR.0b013e31829b1d84

463 49. Sorensen AA, Howard D, Tan WH, Ketchersid J, Calfee RP. Minimal clinically important
464 differences of 3 patient-rated outcomes instruments. The Journal of Hand Surgery. 2013;38(4):641-9.
465 DOI: 10.1016/j.jhsa.2012.12.032

466 50. Stasinopoulos D, Papadopoulos C, Antoniadou M, Nardi L. Greek adaptation and validation
467 of the Patient-Rated Tennis Elbow Evaluation (PRTEE). J Hand Ther. 2015;28(3):286-90; quiz 91
468 DOI: 10.1016/j.jht.2014.12.005.

469 51. Sussmann PS, Simmen BR, Goldhahn J, Sculco TP. Challenges for large orthopaedic
470 hospitals worldwide—an ISOC position statement. HSS journal. 2010;6(1):57-60. DOI:
471 10.1007/s11420-009-9144-5

472 52. The B, Reininga IH, El Moumni M, Eygendaal D. Elbow-specific clinical rating systems:
473 extent of established validity, reliability, and responsiveness. J Shoulder Elbow Surg.
474 2013;22(10):1380-94 DOI: 10.1016/j.jse.2013.04.013.

475 53. Turchin DC, Beaton DE, Richards RR. Validity of observer-based aggregate scoring systems
476 as descriptors of elbow pain, function, and disability. J Bone Joint Surg Am. 1998;80(2):154-62.

477 54. Valderas JM, Ferrer M, Mendivil J, Garin O, Rajmil L, Herdman M, et al. Development of
478 EMPRO: a tool for the standardized assessment of patient-reported outcome measures. Value Health.
479 2008;11(4):700-8 DOI: 10.1111/j.1524-4733.2007.00309.x.

480 55. van Ark M, Zwerver J, Diercks RL, van den Akker-Scheek I. Cross-cultural adaptation and
481 reliability and validity of the Dutch Patient-Rated Tennis Elbow Evaluation (PRTEE-D). BMC
482 Musculoskelet Disord. 2014;15:270 DOI: 10.1186/1471-2474-15-270.

483 56. Vincent JI, MacDermid JC, King G, Grewal R. Psychometric evaluation of self-reported pain
484 and disability measures for elbow pathologies. Journal of Hand Therapy. 2011;24(4):381-2 2p. DOI:
485 http://dx.doi.org/10.1016/j.jht.2011.07.012

486 57. Vincent JI, MacDermid JC, King GJ, Grewal R. Validity and sensitivity to change of patient-
487 reported pain and disability measures for elbow pathologies. Journal of Orthopaedic & Sports
488 Physical Therapy. 2013;43(4):263-74. DOI: 10.2519/jospt.2013.4029

489 58. Wu AW, Kharrazi H, Boulware LE, Snyder CF. Measure once, cut twice—adding patient-
490 reported outcome measures to the electronic health record for comparative effectiveness research.
491 Journal of Clinical Epidemiology. 2013;66(8):S12-S20. DOI: 10.1016/j.jclinepi.2013.04.005

492 59. Zimmerman NB, Kaye MB, Wilgis EFS, Zimmerman RM, Dubin NH. Are standardized
493 patient self-reporting instruments applicable to the evaluation of ulnar neuropathy at the elbow? J
494 Shoulder Elbow Surg. 2009;18(3):463-8. DOI: 10.1016/j.jse.2009.02.010

495

496

Tables and Figures:

Figure 1: PRISMA flowchart

Figure 2: Number of studies published per year that report use of elbow/region-specific clinical rating systems.

Figure 3: Proportional prevalence of the most common rating systems in articles since database inception, 2000 and 2010.

Figure 4: Top five clinical rating systems in each pathology group. (MEPS=Mayo Elbow Performance Score, DASH=Disabilities of Arm Shoulder and Hand, QDASH=quick Disabilities of Arm Shoulder and Hand, ASES-e=American Shoulder and Elbow Society-elbow, PRTEE=Patient Reported Tennis Elbow Evaluation, OES=Oxford Elbow Score, HSS=Hospital for Special Surgery Score, KJOC=Kerlan Jobe Orthopaedic Clinic overhead athlete score)

Table 1: Change in use of clinical rating system over time (no. of articles using a clinical rating system (percentage of articles using clinical rating system).

Table 2: Geographical use of clinical rating systems (no. of articles using a clinical rating system and percentage of articles using clinical rating system).

Appendix 1: Medline Search strategy (run 1/5/2017)

Appendix 2: List or Clinical Ratings Systems Identified – Ordered by Prevalence of Use

Figure (No.1)

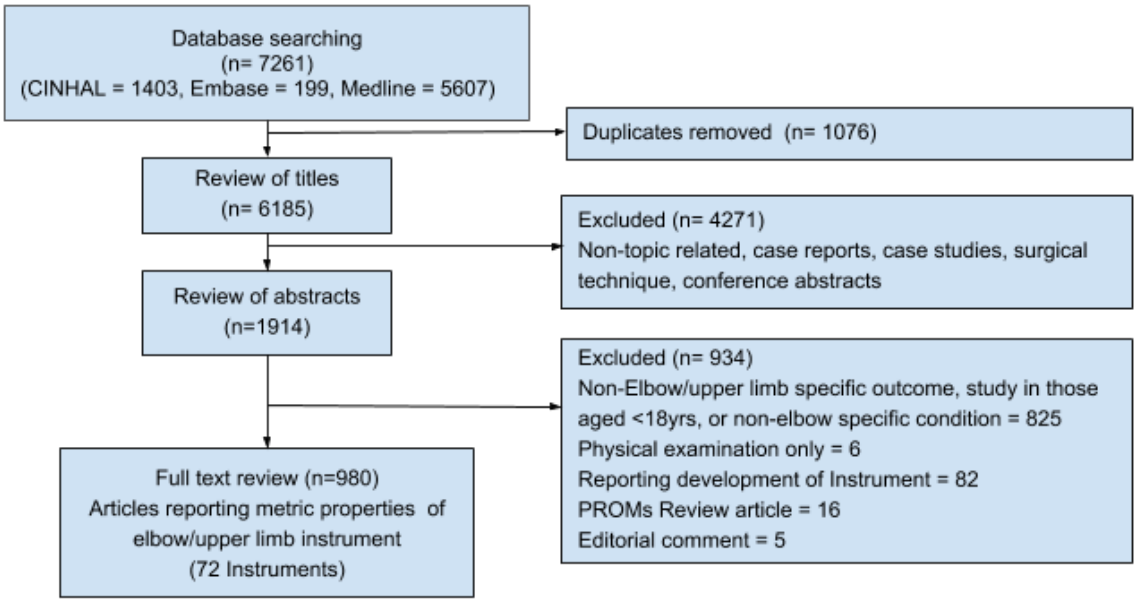


Figure 1

Figure (No.2)

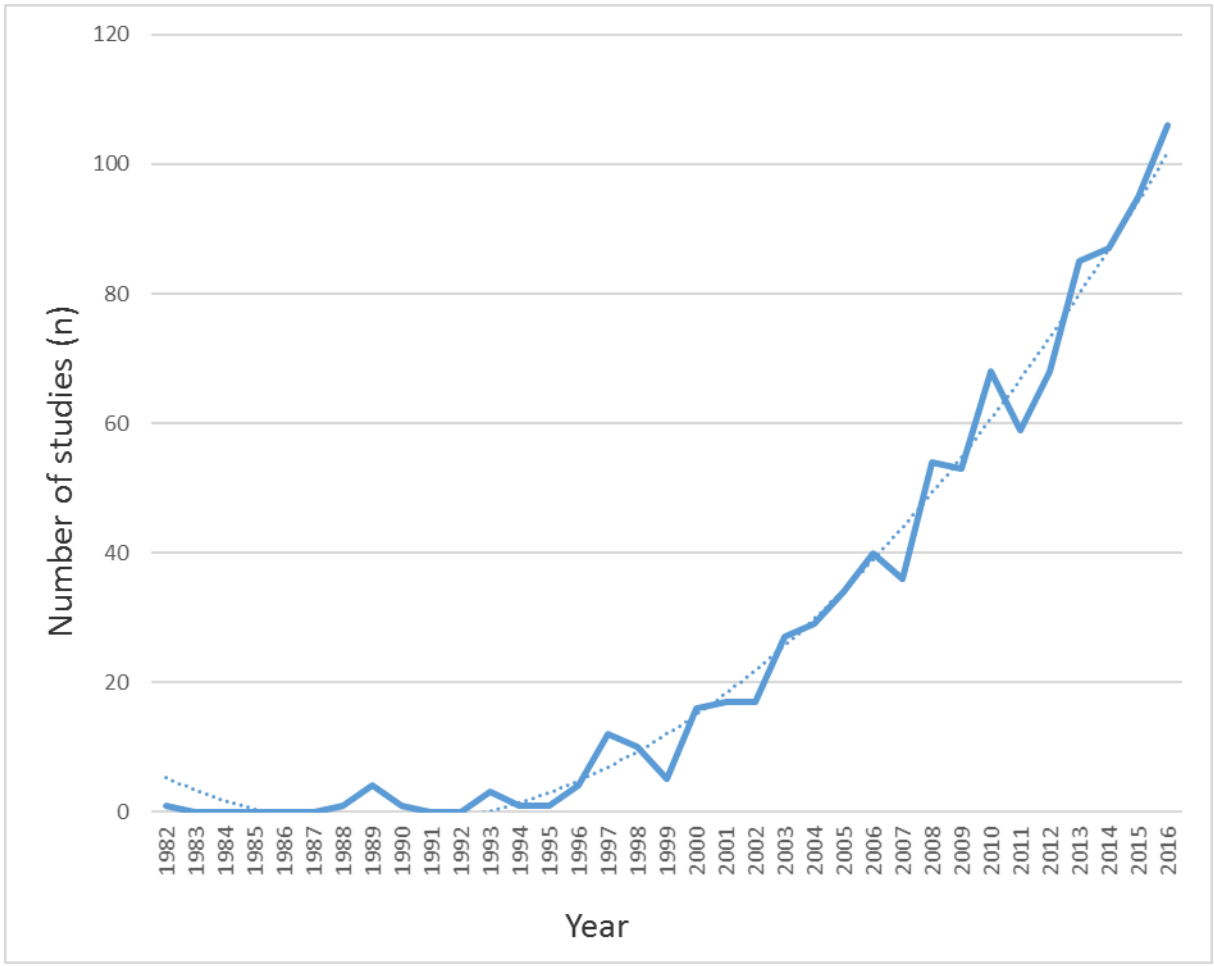


Figure 1

Figure (No.3)

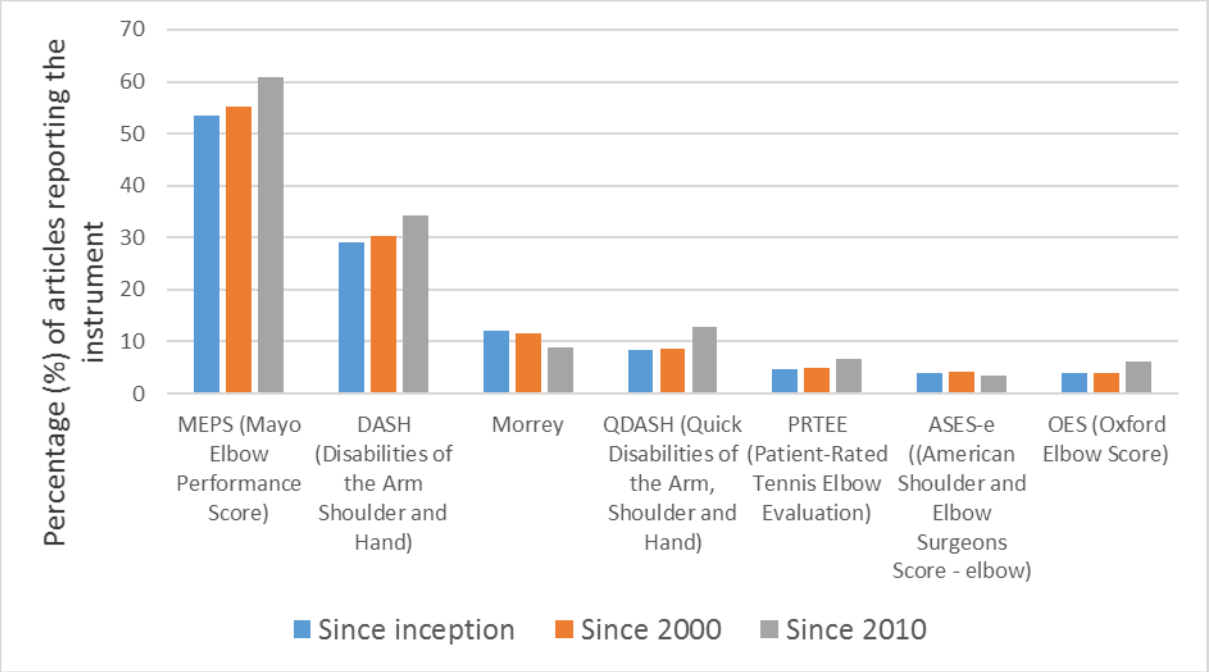


Figure 1

Figure (No.4)

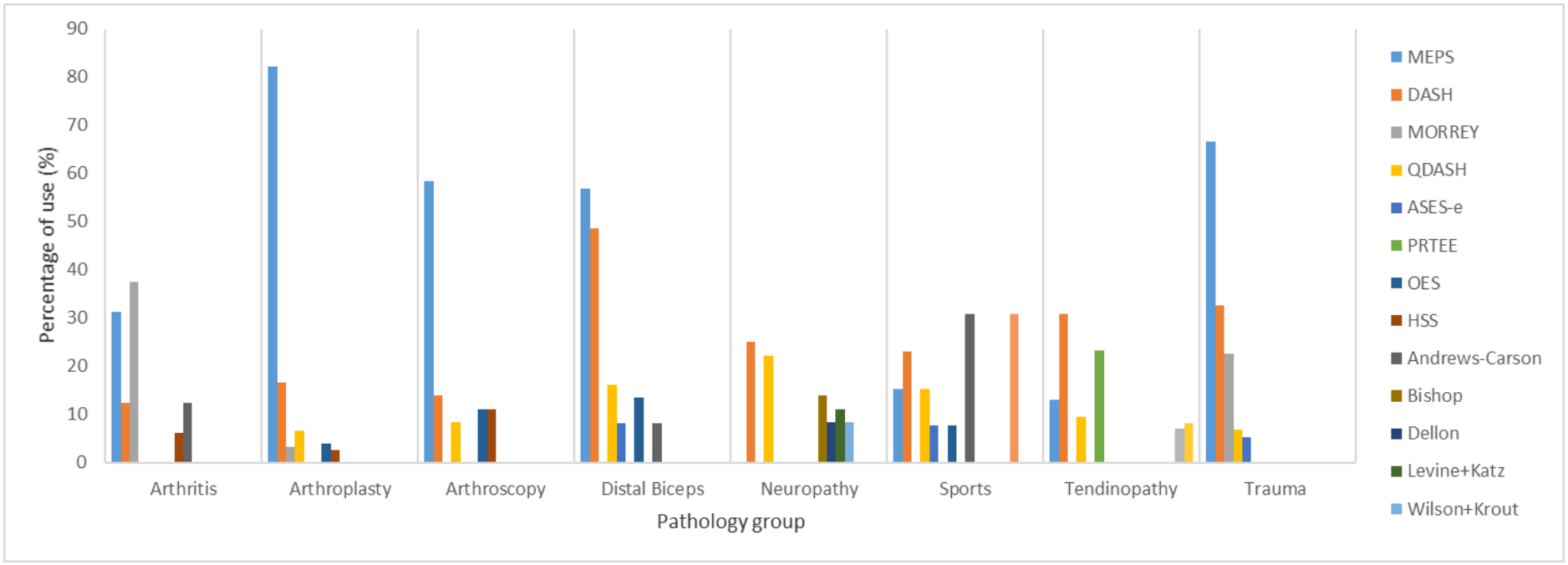


Figure 1

Tables (No. 1)

	Arthroplasty						Tendinopathy						Trauma					
	Inception		Since 2000		Since 2010		Inception		Since 2000		Since 2010		Inception		Since 2000		Since 2010	
Articles (n)	151	%	133	%	74	%	198	%	190	%	128	%	405	%	365	%	128	%
Total Outcomes (n)	199		180		109		244		235		162		621		580		162	
MEPS	124	82.1	116	87.2	66	89.2	26	13.1	25	13.2	22	17.2	270	66.7	270	74.0	200	74.6
DASH	25	16.6	24	18.0	17	23.0	61	30.8	61	32.1	46	35.9	132	32.6	132	36.2	100	37.3
MORREY	5	3.3	3	2.3	1	1.4							92	22.7	85	23.3	46	17.2
QDASH	10	6.6	10	7.5	9	12.2	19	9.6	19	10.0	19	14.8	28	6.9	28	7.7	27	10.1
ASES-e													21	5.2	21	5.8	12	4.5
PRTEE							46	23.2	46	24.2	39	30.5						
OES	6	4.0	6	4.5	6	8.1												
Roles+Maudsley							14	7.1	13	6.8	2	1.6						
Nirschl							17	8.6	16	8.4	10	7.8						

Table 1

Tables (No. 2)

	North America								Europe								Rest of the World							
	Total		Arthroplasty		Tendinopathy		Trauma		Total		Arthroplasty		Tendinopathy		Trauma		Total		Arthroplasty		Tendinopathy		Trauma	
Articles (n)	264	%	50	%	47	%	86	%	370	%	76	%	66	%	151	%	319	%	26	%	77	%	155	%
Total No. of Outcomes (n)	370		58		57		140		558		109		78		254		411		32		98		204	
MEPS	117	44.3	37	74.0	8	17.0	47	54.7	198	53.5	63	82.9	6	9.1	101	66.9	195	61.1	25	96.2	13	16.9	114	73.5
DASH	97	36.7	6	12.0	16	34.0	43	50.0	109	29.5	15	19.7	19	28.8	50	33.1	73	22.9	3	11.5	24	31.2	35	22.6
MORREY	20	7.6	1	2.0			16	18.6	59	15.9	3	3.9			42	27.8	35	11.0	1	3.8			31	20.0
QDASH	15	5.7			2	4.3	4	4.7	22	5.9	11	14.5	5	7.6	15	9.9	18	5.6	1	3.8	9	11.7	5	3.2
ASES-e	23	8.7	2	4.0			12	14.0	10	2.7							5	1.6					5	3.2
PRTEE	9	3.4			9	19.1			10	2.7			9	13.6			26	8.2			21	27.3		
OES	4	1.5							29	7.8	6	7.9			15	9.9	3	0.9						
HSS	2	0.8	1	2.0					6	1.6							6	1.9						
Roles+ Maudsley	1	0.4							11	3.0			10	15.2			3	0.9						
Nirschl	5	1.9			6	12.8			3	0.8			3	4.5			8	2.5			8	10.4		

Table 1

Appendix 1 - Search strategy

Click here to download Supplemental File: Appendix 1.doc

Appendix 2 - All rating systems

Click here to download Supplemental File: Appendix 2.doc