

Outcome Measurement in Plastic Surgery

Justin C.R. Wormald MBBS MRes MRCS¹, Jeremy N. Rodrigues BSc MBChB MSc PhD PGDip MRCS¹

1. Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Botnar Research Centre, Nuffield Orthopaedic Centre, Windmill Road, Oxford, OX3 7LD, UK

Corresponding author: Jeremy Rodrigues

Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Botnar Research Centre, Nuffield Orthopaedic Centre, Windmill Road, Oxford, OX3 7LD, UK

j.n.rodrigues@doctors.org.uk

Conflict of interest

No benefits in any form have been received or will be received from a commercial party related directly or indirectly to the subject of this article.

Funding:

Jeremy Rodrigues is an NIHR academic clinical fellow

Summary

Outcome measurement in plastic surgery is often surgeon-centred, and clinician-derived. Greater emphasis is being placed on patient-reported outcomes (PROs), in which the patient's perspective is measured directly from them. Numerous patient-reported outcome measures (PROMs) have been developed in a range of fields, with a number of good quality PROMs in plastic surgery. They can be deployed to support diagnosis, disease severity determination, referral pathways, treatment decision-making, post-operative care and in determining cost-effectiveness. In order to understand the impact of disease and health interventions, appropriate PROMs are a logical choice in plastic surgery, where many conditions involve detriment of function or cosmesis. PROMS can be classified as disease-specific, domain-specific, dimension-specific, population-specific and generic. Choosing the correct outcome and measure can be nebulous. The two most important considerations are: is it suitable for the intended purpose? and how valid is it? Measurement that combines being patient-centred and also aligns with clinicians' understanding is achievable, and can be studied scientifically. Rational design of new PROMs and considered choice of measures is critical in clinical practice and research. There are a number of tools that can be employed to assess the quality of PROMs that are outlined in this overview. Clinicians should consider the quality of measures both in their own practice and when critically appraising evidence. This overview of outcome measurement in plastic surgery provides a tool set enabling plastic surgeons to understand, implement and analyse outcome measures across clinical and academic practice.

Word count: 249

Keywords:

Patient Reported Outcome Measures; Patient Outcome Assessment; Treatment Outcome; Quality of Life

Introduction

Outcome measurement is often surgeon-centred, and clinician-derived. Greater emphasis is being placed on patient-reported outcomes (PROs), in which the patient's perspective is measured directly from them.¹ Numerous patient-reported outcome measures (PROMs) have been developed in a range of fields, and across the tiers of the World Health Organisation's International Classification of Functioning, Disability and Health (ICF): capturing physical impairment, function, participation and health status or quality of life.²⁻⁶ PROMs can be deployed across the patient care journey, to support diagnosis, disease severity determination, referral pathways, treatment decision-making, and post-operative care. They are used in determining cost-effectiveness and quality assurance.⁷ Furthermore, PROMs have been promoted politically.⁸

This does not undermine the value of clinician-derived outcomes, which are particularly appropriate where clinical expertise is required (e.g. determining postoperative wound infection).⁹ However, to understand the impact of disease and health interventions, appropriate PROMs are a logical choice in plastic surgery, where many conditions involve detriment of function or cosmesis.

A current overview of outcome measurement is provided. Some of this is PROM-specific, though much is applicable to all kinds of measures that plastics surgeons might use.

Definitions

The United States of America Food and Drugs Administration (FDA) provide detailed PROM guidance. They define a patient reported outcome as: ‘any report of the status of a patient’s health condition that comes directly from the patient, without interpretation of the patient’s response by a clinician or anyone else. The outcome can be measured in absolute terms (e.g., severity of a symptom, sign, or state of a disease) or as a change from a previous measure’.¹⁰ There is a difference between a patient reported outcome (PRO) and a PROM: the latter describes a ‘tool’ or ‘instrument’ used to quantify the former. For example, satisfaction after facial surgery is a patient reported outcome (PRO), which can be measured using FACE-Q PROM.¹¹

PROMs are usually questionnaires, comprising a series of questions or “items”. Although the FDA definition of a PROM focuses on patient completion of the tool, contemporary PROMs are typically considered fit for use if their content is also patient-centred, i.e. their items are of importance to patients. Importantly, a PROM may not be patient-centred, if the items it comprises do not reflect patients’ priorities (for example, following breast reconstruction a hand function PROM is unlikely to reflect what matters to that cohort). At the same time, clinician-derived measurement (such as grip strength) may represent what matters to patients in specific situations. The key issue is the patient’s perspective, and surgeons should avoid assumptions about patient opinion.

Often, PROM items are pre-specified, though some PROMs exist in which each individual patient defines the items, for example The Canadian Occupational Performance Measure, in which a patient specifies their priorities, and then scores them.¹² PROMs can be completed in different ways: e.g. on paper or on electronic devices. The latter can allow for selective deployment methods such as Computerised Adaptive Testing, which will be discussed.

In contrast to PROMs, Patient-Reported Experience Measures (PREMs) assess the patient’s experience of care (with items like “were you seen on time?”).¹³ Although ‘outcome’ and ‘experience’ are different constructs, there are overlaps and the outcome for the patient may retrospectively affect their experience of care.¹⁴

Types of PROM

One way to classify the range of PROMs available is by their scope. Relevant categories include disease-specific PROMs, domain-specific PROMs and generic PROMs.¹⁵ However, other groups exist, such as dimension-specific and population-specific PROMs.¹⁶

Disease-specific PROMs apply to a specific disease or health problem. An example of a disease-specific PROM is the CLEFT-Q, for cleft lip and palate.¹⁷ Disease-specific PROMs provide highly specific data about patients with specific conditions and may be expected to detect clinically relevant changes in an outcome sensitively. The narrow scope of disease-specific PROMs means that issues indirectly related to the disease, such as adverse effects of treatment, may be missed.^{18,19}

Domain-specific PROMs apply to a defined site. For example, the Disabilities of the Arm, Shoulder and Hand (DASH) covers the upper limb.²⁰ Similarly, dimension-specific PROMs cover one health-related area. They may assess symptom status (e.g. pain, stiffness), physical function (e.g. mobility, grip strength), social function (e.g. employment, activities) or psychosocial concepts (e.g. satisfaction with appearance). Such PROMs give highly specific information about an aspect of a disease or condition and can therefore detect clinically relevant changes following intervention. When used in isolation, they may miss important general and disease-specific information, including adverse effects.²¹

Population-specific PROMs measure outcomes across a population regardless of disease, such as the elderly. Population specific PROMs have a relatively wide scope and can capture health information across specific patient groups. However, they cannot be applied to the general population or other patient groups. For example, a paediatric population-specific PROM cannot be used to compare to an adult population-specific PROM.²²

Generic PROMs can be applied across all situations, and may measure constructs such as health status or health-related quality of life. A common health status instrument is EuroQol-5D, which is the preferred instrument for use in Technology Appraisal by the National Institute for Health and Care Excellence (NICE).^{23,24}

Generic PROMs may be less sensitive to change than specific ones. A systematic review of the responsiveness to change of the EQ5D suggests that it is sensitive for many conditions, though not for limb reconstruction.²⁵ In contrast, disease-, domain-, dimension- and population-specific PROMs

cannot be used to make comparisons to other groups or to the general population, unless the specific PROM has been mapped to a different measure, or has had utilities studied for its value. For example, some PROMs have been mapped to EQ-5D.²⁶

How should a PROM be developed?

Development of a new PROM is a significant undertaking. A research team with experience in mixed methods research, health measure development and psychometrics is required. It is critical that patients are involved in the process, if the tool developed is to be relevant to their priorities.

The first step is to define the scope of what is to be measured, and confirm the need for a new PROM through a systematic review. As well as specific PROMs, core outcome sets may have been defined for particular contexts.²⁷

If a new PROM is needed, the review will help develop the conceptual model for it. The second stage in the data gathering process is to undertake qualitative interviews with both clinicians and relevant patient groups. This is performed until thematic saturation occurs (i.e. no new ideas or concepts are emerging). Once the data gathering process is complete, the PROM can be fully defined and a conceptual model of the questionnaire created.

Item generation is then performed. This involves creating individual candidate items that could be included in the PROM. Each item should be based on a rigorous analysis of the data gathering process and should reflect the concepts generated in the qualitative studies. Patient and public involvement should be introduced early into the process and should continue throughout the development and implementation of the PROM. Once candidate item generation process has been completed, aspects of the item performance can be studied to determine which are best to be included in the PROM. This may involve identifying candidate items that do not fit with other items, items that are redundant, or items that are not well targeted at the population concerned. Then, item refinement will be performed. This involves testing the PROM on the patient group to ensure that it is readable and user friendly and that it is appropriate for the target patient audience. Cognitive interviews are conducted to explore patient's interpretation and ideas about the draft PROM items.²⁸ The PROM is then formally validated in the target population.

Outcome Measure Properties

Evaluation principles can be applied when developing a PROM, or when studying existing PROMs or other outcome measures. There are specific domains that need to be assessed when determining measure quality:^{19,29} These include validity, reliability and responsiveness. Interpretability is also important when using measures in practice.

Validity

This comprises several areas:

- **Content Validity**

This is concerned with whether the items in a PROM effectively cover what is being is trying to be measured. It broadly involves assessing whether the measure is both comprehensive and relevant.

- **Construct Validity**

This studies the behaviour of the PROM to the underlying concept (construct) it is measuring, including how its items relate to each other (structural validity), its relationship with other measures and its differences between groups (cross-cultural validity). Such work may be applicable to, or involve, clinician-derived measures. For example, it might be hypothesised that a new hand function PROM will correlate closely with finger range of motion, but correlate less closely with an anxiety/depression PROM.

Classical Test Theory (CTT) and Item Response Theory (IRT)

There are two approaches to studying structural validity: classical test theory (CTT) and item response theory (IRT).³⁰ Classical test theory assumes that each observed score on a PROM is a combination of an underlying true score, based on the questionnaire concept and random error. This model assumes that each person has a true score – a score that would not change over an infinite number of times the PROM is administered. An observed score is the true score with an element of error.

Item response theory (IRT) comprises a collection of analyses that assume that items measure the same underlying construct, but are not equally weighted, e.g. that hand function PROM items belong in different positions along a “ruler” of hand function. In some analyses, the “location” of the item can be studied quantitatively, such that the overall PROM summary score is interval-scaled – i.e. the scores can be summed or subtracted in a meaningful way. IRT is essentially a theory of testing a PROM item based on the relationship an individuals’ performance and that same individuals’ performance on an overall measure of the ability that item was designed to measure.³¹ Rasch analysis is one type of IRT that has been used widely in outcome measurement science.³² Some consider PROMs that comply with IRT, especially Rasch analysis, to be of higher quality than those that do not.³³

- ***Criterion Validity***

This tests a PROM against a ‘gold standard’. The only accepted use for this in PROM research is the comparison of a shortened PROM against the long version (e.g. the QuickDASH against the DASH). It is often used inappropriately to compare a PROM to an existing outcome measure considered to be the “gold standard”. This is unhelpful, as it incorrectly assumes that the existing outcome measure is perfect.

- ***Discriminative Validity***

Discriminative validity assesses whether the PROM can isolate pre-existing characteristics between patient groups. For example, a scar-related PROM should be able to distinguish between difference surgical groups that are known to have different baseline attributes, such as burn wounds compared to an abdominoplasty scar.³⁴

- ***Responsiveness***

Responsiveness is the ability to detect change over time – in surgery, this often means change in score following treatment. However, this assumes that treatment is effective and has been successful, which is not always the case.³⁵

Reliability

- ***Internal consistency***

This is measure of how well each item relates to the underlying construct it is testing, and thus that the items are measuring the same concept. Internal consistency is an important measurement property for questionnaires that intend to measure a single construct by using multiple items. Internal consistency can be adequately measured by the statistical test, Cronbach's alpha. A low alpha (e.g. <0.7) suggests that the items do not correlate, and a high alpha (e.g. >0.9) may indicate redundancy of items.

- ***Reproducibility***

Repeated measurements in a stable patient should be similar – this is reproducibility. Test-re-test reliability involves giving the same scale to the same person over a defined interval and assessing the similarity of their results. Part of reliability is agreement, which describes the absolute measurement error, i.e., how close the scores on repeated measures are.

Interpretability

This is concerned with the meaningfulness of the magnitude of change, either following treatment for an individual (e.g. how much increase in range of motion does a burn contracture patient need to experience for a surgical release to have been worthwhile?) , or between different groups of people who have undergone treatments (e.g. how great a difference needs to exist between patients who have undergone needle aponeurotomy for Dupuytren's disease and those who have undergone fasciectomy, before it becomes important?).³⁶

Outcome Measure Evaluation

It can be seen from what has been described that validity is not binary. Rather than a PROM or other outcome measure being "valid" or "not valid", it may have demonstrated good performance in terms of some domains but not others. Several tools are available for assessing the quality of a PROM. See **Box 1**, for an illustrative example of an outcome measure evaluation.

The EMPRO tool applies the Medical Outcomes Trust (MOT) guidelines for the development of measures that incorporate the patient's perspective of health.³⁹ Following a systematic review, the EMPRO tool can determine a PROM's quality, based on the domains described above, using a 39-

item assessment comprising items followed by an overall recommendation in reference to the relevant research question.

Terwee *et al.* developed another tool for assessing the measurement properties of PROMS.²⁹ This group developed quality criteria for design, methods, and outcomes of studies on the development and evaluation of health status questionnaires, also based on the domains described above. For each measurement property a criterion was defined for a positive, negative, or indeterminate rating, depending on the design, methods, and outcomes of the validation study. These criteria can be used as a quality assessment tool in systematic reviews of PROMS, to detect faults in established PROMs, and to design PROM validation studies.

The COnsensus-based Standards for the selection of health status Measurement INstruments (COSMIN) has been developed to assess the reporting of studies of PROM characteristics, including PROM validation studies.³³ Each domain has a four-point scale (poor to excellent) based on assessment of the key domains described above. There is an additional criterion that assesses generalisability. COSMIN also assesses whether Item Response Theory (IRT) has been used in the development of the PROM and awards points positively for this.

Deploying PROMs

PROMs can be deployed differently.⁴⁰ They may be paper-based or computerised, and can be completed in healthcare establishments, at home, or online. Self-administration by means of a postal questionnaire is the most common form of administration. Direct contact administration, either by interviewer or telephone produces a higher response rate but is costly, and might affect a patient's responses. Electronic PROMs (ePROMs) are popular in many patient groups and can be delivered via hand held devices as well as full computers.⁴¹ There is evidence to suggest that study participants prefer ePROMs compared to conventional methods, regardless of age.⁴²

Computerised adaptive testing (CAT) involves electronic PROM administration. However, rather than pose all items to all respondents, the minimum number of items required to find the participant's location is used. For example, if the first item that a person responds to demonstrates that he can jog, then there would no further value in posing lower function questions such as whether he can walk. Instead, the computer then chooses the next item to pose to gain more information to identify his location – the next item might establish whether he can sprint or not, for example. CAT advantages include reduced burden on the respondent, and the ability to select different questions from an item bank, so that the respondent does not become familiar with items. However, it requires interval-scaled measurement, item banks, and the ability to use a computer.

Limitations of PROMs

There is evidence that PROMs improve the doctor-patient interaction and patient satisfaction plus intuitively, measuring the patient's outcome makes sense.^{36,43} Despite this, incorporating PROMs into routine **clinical and research practice** has been difficult. One barrier is a fear that PROMs create additional administrative workload, especially when some clinicians feel they already understand the patient perspective. To overcome these issues, new PROMs must be designed in accordance with the realities of healthcare systems and both patient and clinician expectations. Another issue is the user interface. PROMs require the patient to interact with the measure, by pen, paper or electronically. This may lead to exclusion of groups, including those with dementia or blindness. Although not insurmountable, many available PROMs are limited to those able to read and write or use a computer. Using some PROMs involves paying a licence fee, which may limit their usability in lower income countries. Many PROMs are translated once finalised or during development. Language barriers however are an important consideration when developing a PROM and ties in with the issue of cross-cultural validity, i.e. that the question wording has the same meaning across different languages and cultures.⁴⁴

Choosing an outcome measure

Choosing the correct outcome and measure is not necessarily straightforward. There **are** a variety of measures applicable to plastic surgery, including some high quality and Rasch-analysed PROMs. The two most important considerations are: is it suitable for the intended purpose? and how valid is it?⁴⁵

The first consideration is what is to be measured. For example, in a trial in rheumatoid arthritis of the hand in which continuous data are desired, an interval-scaled PROM that has undergone Rasch analysis might be most appropriate, such as ABILHAND.⁴⁶ In contrast, in a study based on a threshold, then a Rasch-analysed PROM may not be required. For example, if the proportion of patients who required endotracheal intubation for reduced consciousness in trauma is being studied, then dichotomising the cohort around a threshold of a Glasgow Come Scale (GCS) of 8 can be performed, irrespective of GCS not being interval-scaled: in this case, that threshold determines the treatment required.

The choice of measure may also vary within a condition. For example, an explanatory trial of facial palsy treatment may choose to measure facial movement, but a pragmatic trial might measure health-related quality of life.

In some situations, core outcome sets have been defined, which specify the domains of importance to all stakeholders, including patients and clinicians, and identify appropriate measures to assess them.⁴⁷ Such approaches may standardise the choice of outcome measures used in a field, allowing for extrapolation of findings and comparisons of studies.

Conclusion

Measurement that combines being patient-centred yet aligns with clinicians' understanding is achievable, and can be studied scientifically. Rational design of new PROMs, and considered choice of measures is critical in clinical practice and research. Where possible, standardisation of appropriate measures is desirable. Ultimately, clinicians should consider the quality of measures both in their own practice and when critically appraising evidence.

Word count: 3078

References

1. Barry MJ, Edgman-Levitan S. Shared decision making—the pinnacle of patient-centered care. *N Engl J Med* 2012;366:780-1.
2. Marshall S, Haywood K, Fitzpatrick R. Impact of patient-reported outcome measures on routine practice: a structured review. *J Eval Clin Pract* 2006;12:559-68.
3. Hunt KJ, Hurwit D. Use of patient-reported outcome measures in foot and ankle research. *J Bone Joint Surg Am* 2013;95:e118.
4. Wang D, Jones MH, Khair MM, Miniaci A. Patient-reported outcome measures for the knee. *J Knee Surg* 2010;23:137-51.
5. Chow A, Mayer EK, Darzi AW, Athanasiou T. Patient-reported outcome measures: the importance of patient satisfaction in surgery. *Surgery* 2009;146:435-43.
6. World Health Organisation. International Classification of Functioning, Disability and Health: ICF: World Health Organization; 2001.
7. 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.
8. Lansley A. Equity and excellence: liberating the NHS. London: Department of Health 2010.
9. Basch E, Jia X, Heller G, et al. Adverse symptom event reporting by patients vs clinicians: relationships with clinical outcomes. *J Natl Cancer Inst* 2009;101:1624-32.
10. US Food and Drug Administration (FDA). Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims. *Fed Regist* 2009;74:65132-3.
11. Klassen AF, Cano SJ, Scott A, Snell L, Pusic AL. Measuring patient-reported outcomes in facial aesthetic patients: development of the FACE-Q. *JAMA Facial Plast Surg* 2010;26:303-9.
12. Law M, Baptiste S, McColl M, Opzoomer A, Polatajko H, Pollock N. The Canadian occupational performance measure: an outcome measure for occupational therapy. *Can J Occup Ther* 1990;57:82-7.
13. Black N, Jenkinson C. Measuring patients' experiences and outcomes. *BMJ* 2009;339.
14. Black N, Varaganum M, Hutchings A. Relationship between patient reported experience (PREMs) and patient reported outcomes (PROMs) in elective surgery. *BMJ Qual Saf* 2014: 2013-002707.
15. Szabo RM. Outcomes assessment in hand surgery: when are they meaningful? *J Hand Surg A* 2001;26:993-1002.
16. Garratt A, Schmidt L, Mackintosh A, Fitzpatrick R. Quality of life measurement: bibliographic study of patient assessed health outcome measures. *BMJ* 2002;324:1417.

17. Tsangaris E, Riff KWW, Goodacre T, et al. Establishing Content Validity of the CLEFT-Q: A New Patient-reported Outcome Instrument for Cleft Lip/Palate. *Plast Reconstr Surg Glob Open* 2017;5.
18. Rodrigues JN, Zhang W, Scammell BE, Davis TRC. What patients want from the treatment of Dupuytren's disease—is the Unité Rhumatologique des Affections de la Main (URAM) scale relevant? *J Hand Surg Eur* 2015;40:150-4.
19. Rodrigues J, Zhang W, Scammell B, et al. Validity of the Disabilities of the Arm, Shoulder and Hand patient-reported outcome measure (DASH) and the Quickdash when used in Dupuytren's disease. *J Hand Surg Eur* 2016;41:589-99.
20. Hudak PL, Amadio PC, Bombardier C, et al. Development of an upper extremity outcome measure: the DASH (Disabilities of the Arm, Shoulder, and Hand). *Am J Ind Med* 1996;29:602-8.
21. Ball C, Pratt AL, Nanchahal J. Optimal functional outcome measures for assessing treatment for Dupuytren's disease: a systematic review and recommendations for future practice. *BMC Musculoskelet Disord* 2013;14:131.
22. Matza LS, Patrick DL, Riley AW, et al. Pediatric patient-reported outcome instruments for research to support medical product labeling: report of the ISPOR PRO good research practices for the assessment of children and adolescents task force. *Value Health* 2013;16:461-79.
23. Rabin R, Charro Fd. EQ-SD: a measure of health status from the EuroQol Group. *Ann Med* 2001;33:337-43.
24. Earnshaw J, Lewis G. NICE guide to the methods of technology appraisal. Springer International Publishing; 2008.
25. Payakachat N, Ali MM, Tilford JM. Can the EQ-5D detect meaningful change? A systematic review. *Pharmacoeconomics* 2015;33:1137.
26. Health Economics Research Centre. HERC database of mapping studies. (Accessed 7th July, 2017, at <https://http://www.herc.ox.ac.uk/downloads/herc-database-of-mapping-studies>)
27. Williamson PR, Altman DG, Blazeby JM, Clarke M, Gargon E. The COMET (core outcome measures in effectiveness trials) initiative. *Trials* 2011;12:A70.
28. Knafl K, Deatrick J, Gallo A, et al. The analysis and interpretation of cognitive interviews for instrument development. *Res Nurs Health* 2007;30:224-34.
29. Terwee CB, Bot SD, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 2007;60:34-42.
30. Cappelleri JC, Jason Lundy J, Hays RD. Overview of classical test theory and item response theory for the quantitative assessment of items in developing patient-reported outcomes measures. *Clin Ther* 2014;36:648-62.
31. Kean J and Reilly J. "Item response theory." *Handbook for Clinical Research: Design, Statistics and Implementation*.(pp195-198). New York, NY: Demos Medical Publishing (2014).
32. Tennant A, McKenna SP, Hagell P. Application of Rasch analysis in the development and application of quality of life instruments. *Value Health* 2004;7:S22-S6.

33. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol* 2010;63:737-45.
34. Durani P, McGrouther DA, Ferguson MW. Current scales for assessing human scarring: a review. *J Plast Reconstr Aesthet Surg*. 2009 Jun;62(6):713-20.
35. Rodrigues JN, Zhang W, Scammell BE, Davidson D, Fullilove S, Chakrabarti I, Russell PG, Davis TR. Recovery, responsiveness and interpretability of patient-reported outcome measures after surgery for Dupuytren's disease. *J Hand Surg Eur* 2017 Mar;42(3):301-9.
36. Rodrigues JN, Mabvuure NT, Nikkhah D, Shariff Z, Davis TR. Minimal important changes and differences in elective hand surgery. *J Hand Surg Eur* 2015 Nov;40(9):900-12.
37. Kennedy CA, Beaton DE, Smith P, et al. Measurement properties of the QuickDASH (Disabilities of the Arm, Shoulder and Hand) outcome measure and cross-cultural adaptations of the QuickDASH: a systematic review. *Qual Life Res* 2013; 22(9):2509-47.
38. Forget NJ, Jerosch-Herold C, Shepstone L, Higgins J. Psychometric evaluation of the Disabilities of the Arm, Shoulder and Hand (DASH) with Dupuytren's contracture: validity evidence using Rasch modelling. *BMC Musculoskelet Disord* 2014;15(1):361.
39. Valderas JM, Ferrer M, Mendivil J, et al. Development of EMPRO: A tool for the standardized assessment of patient-reported outcome measures. *Value Health* 2008;11:700-8.
40. Snyder CF, Aaronson NK, Choucair AK, et al. Implementing patient-reported outcomes assessment in clinical practice: a review of the options and considerations. *Qual Life Res* 2012;21:1305-14.
41. Gwaltney CJ, Shields AL, Shiffman S. Equivalence of electronic and paper-and-pencil administration of patient-reported outcome measures: a meta-analytic review. *Value Health* 2008;11:322-33.
42. Jensen RE, Snyder CF, Abernethy AP, et al. Review of electronic patient-reported outcomes systems used in cancer clinical care. *J Oncol Pract* 2013;10:e215-e22.
43. Nelson EC, Eftimovska E, Lind C, Hager A, Wasson JH, Lindblad S. Patient reported outcome measures in practice. *BMJ* 2015;350:g7818.
44. Wild D, Grove A, Martin M, et al. Principles of good practice for the translation and cultural adaptation process for patient-reported outcomes (PRO) measures: report of the ISPOR task force for translation and cultural adaptation. *Value Health* 2005;8:94-104.
45. McClimans LM, Browne J. Choosing a patient-reported outcome measure. *Theor Med Bioeth* 2011;32:47-60.
46. Penta M, Thonnard J-L, Tesio L. ABILHAND: a Rasch-built measure of manual ability. *Arch Phys Med Rehabil* 1998;79:1038-42.
47. The International Consortium for Health Outcomes Measurement (ICHOM), available at <http://www.ichom.org/> (accessed 19th July 2017).