Published in final edited form as:

J Hand Surg Am. 2012 April ; 37(4): 851–855. doi:10.1016/j.jhssa.2012.01.043.

Clinical Research and Patient-Rated Outcome Measures in Hand Surgery

Ryan P. Calfee, MD and Amelia A. Adams, MD

Washington University School of Medicine, Department of Orthopaedic Surgery, St. Louis, Missouri

Abstract

High quality clinical research incorporating standardized outcome assessments is necessary to advance the field of hand surgery. Although such research can be conducted with little direct cost, effectively answering clinical questions requires thoughtful study design. Relevant concepts to consider include sample size determination, study end-points, data management, and choice of outcome measures. Provided the emphasis on, and proliferation of, patient-rated outcome measures, the clinician-researcher should consider the unique aspects of commonly referenced outcome measures when initiating an investigation.

Key Terms

hand; outcome measure; patient-rated; surgery

General Considerations for Conducting Clinical Outcome Studies

Prior to study enrollment, investigators must obtain approval for their proposed study from a local Institutional Review Board (IRB) to ensure ethical research design. Considerations relevant to both study design and study conduct that are assessed by the IRB include sample size calculations (statistical power), study endpoint, and data management.

Sample Size

Accurate sample size estimations are necessary to prevent falsely accepting the null hypothesis (i.e., failing to identify statistically significant differences when a true difference exists either between groups or as a result of treatment) and prevent burdening an unnecessary number of patients with participation when fewer would have been sufficient. When calculating necessary sample sizes based on patient-rated outcome measures, the target for study enrollment should be sufficient to achieve appropriate statistical power (generally >80%) to detect a significant ($p < 0.05$) difference between groups based on a minimally important clinical difference for the outcome measure used. In general, estimates for the number of subjects to be screened for enrollment should exceed the calculated sample size by 10–20% to account for patients failing to complete study protocols.

© 2012 The American Society for Surgery of the Hand. Published by Elsevier Inc. All rights reserved.

Corresponding Author: Ryan Calfee, M.D., Washington University School of Medicine, Department of Orthopaedic Surgery, 660 South Euclid Avenue, Campus Box 8233, St. Louis, MO 63110, 314-747-4705, calfeer@wudosis.wustl.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Reputable free software to assist in sample size estimations include PS Power and Sample Size Calculator from Vanderbilt University⁽¹⁾ and G*Power.⁽²⁾

The concept of a minimally important clinical difference (MCID) is critical when considering use of patient-rated assessments as primary outcomes. The MCID represents the minimal necessary change in scoring on an outcome measure that would be perceived by the patient as either beneficial or harmful. This value likely varies based on the population of interest and the disease or injury being studied. The MCID becomes the basis for estimating necessary sample size as it is compared to the variance of the outcome measure. As the reported scores on an outcome measure become more widely dispersed (i.e., increased variance) or the amount of change needing to be detected becomes smaller (i.e., you are trying to identify more subtle differences), then one must recruit more subjects to achieve appropriate power. The MCID also is important when assessing published outcomes. Comparing reported treatment effects to the MCID allows a determination of clinically meaningful change which, unlike statistical significance, is independent of the number of data points collected. Minimally clinically important differences are not universally defined for current patient-rated outcomes measures of the upper extremity but have been examined in select populations.^(3–5)

Study End-Point

Similar to sample size calculations, proposed patient follow up should be sufficient, but not excessive, to test the research hypothesis. In general, follow up should not conclude before function plateaus following treatment. For example, Rozental et al documented improvements in motion and patient-rated function over a 12 month period after surgical fixation of distal radius fractures.⁽⁶⁾ Therefore, it would not be recommended to examine functional outcomes for a novel distal radius fracture treatment with final assessment at less than 1 year. If assessing for complications that develop slowly (e.g., arthritis) then one may need to consider continuing longitudinal examinations over several years.

Data Management

Data management is becoming increasingly scrutinized in regard to security of files to ensure that patient confidentiality is not breached and protected health information is secured. Files should only be maintained on password protected devices and servers and all data should be as de-identified as possible. We would recommend discussing required data safeguards with IRB officials before establishing any online or remote server based data repository.

Missing Data

Nearly all clinical studies suffer from missing data. The amount of missing data may affect the validity of an outcome measure. There is no consensus amount of missing data that is considered unacceptable on patient-rated outcome measures. Instead, each outcome measure generally provides recommendations for score calculation when items are missed. For example, the MHQ subscales are only scored if greater than 50% of relevant items are completed.⁽⁷⁾ In general, increased survey brevity is associated with less tolerance for missing data elements.

For clinical trials, entire data time-points may be missing when patients are lost to follow up. This is one source of bias as patients completing the study may differ in unmeasured ways from those failing to complete the study. For this reason, one randomized control trial study may be assessed as a Level I study (80% patient follow up) and another identically designed trial reach only Level II status (<80% patient follow up).

Patient-Rated Outcome Measures

Standardized patient-rated outcome measures are available as indicators of general health, extremity-specific function, or disease-specific impact. For any given topic of investigation, no consensus outcome instrument exists. Over the past 12 months in the *Journal of Hand Surgery*, the most commonly used measures included: Disabilities of the Arm Shoulder and Hand (DASH, n=31), Michigan Hand Questionnaire (MHQ, n=8), Boston Carpal Tunnel Questionnaire (BCTQ, n=3), Patient-Rated Wrist Evaluation (PRWE, n=2), Cold Intolerance Severity Scale (n=5). Use of standardized outcome instruments improves the comparability of data and often capitalizes on pre-existing determinations of survey validity, reliability, and responsiveness.

Finding Outcome Measures

There is no single repository of patient-rated outcome tools for hand surgeons although a recent work by Smith et al details many of the most commonly referenced outcome measures for the upper extremity from the shoulder to the hand.⁽⁸⁾ One method for selecting the appropriate outcome measure is to review related, peer-reviewed publications and catalog the outcome measures employed. Relevant literature is generally identified through PubMed⁽⁹⁾ or by searching the intended journal's website. This method will identify commonly used outcome measures. The use of measures employed in prior studies on a topic of interest will often simplify data comparison. Several of the most commonly referenced measures are available for download on the Internet (DASH;⁽¹⁰⁾ MHQ;⁽⁷⁾ PRWE.⁽¹¹⁾

Validation of Outcome Measures

An instrument is considered valid if it truly measures what it is intended to measure. Validity is a concept that is not necessarily specific to the outcome measure, but more so to the population being studied and the hypothesis tested. Validity is not an "all or nothing" phenomenon but is based on amassing supporting evidence and has several components (e.g., face validity, criterion validity, construct validity, and content validity). Validity of an outcome measure is often established by comparisons between the given outcome measure and objective examination findings or alternative outcome measures. For example, the PRWE was evaluated against a gold-standard measure of health (Short Form 36) for correlation between the measures in patients recovering from distal radius fractures and scaphoid fractures as tests of its validity.⁽¹²⁾

Common Patient-Rated Outcome Measures for the Hand and Wrist

Table 1 presents several commonly referenced patient-rated outcome measures for the hand and wrist.

Disability of the Arm, Shoulder and Hand (DASH)

The DASH was designed as a measure of composite upper-extremity disability.⁽¹³⁾ The DASH consists of 30 base questions (5-point Likert scales) with supplemental modules adding 8 questions regarding work and sport (or performing arts) activity. A total score is calculated (range 0–100) with higher scores representing greater disability. Subjects rate task difficulty regardless of hand dominance or affected side. Thus, the DASH score reflects total upper-extremity disability predicated on bilateral upper-extremity ability. Use of the DASH in studies of high functioning athletes has revealed a ceiling effect limiting its responsiveness.⁽¹⁴⁾

Quick DASH—The Quick DASH is an 11 question subset from the DASH minimizing both completion time and data entry burden. Scoring is highly correlated with traditional DASH scores.⁽¹⁵⁾ Ten out of 11 questions must be answered to score the Quick DASH.

Michigan Hand Questionnaire (MHQ)

The MHQ assesses multiple inter-related aspects of hand function. Twenty five questions are duplicated for right and left hands and 16 questions provide bilateral assessment. Six hand sub-scale scores (range 0–100) are produced: hand function, activities of daily living, pain, satisfaction, aesthetics, and work performance. A total combined score for each hand is also calculated. Higher scores represent greater function (except for pain subscale: higher score represents greater pain).

Brief MHQ—This 12 item survey drawn from the MHQ is intended for clinical application. This abbreviated questionnaire produces a final score and does not differentiate between right and left hands.

Patient-rated Wrist Evaluation (PRWE)

The 15 question PRWE examines wrist pain (n=5) and function (n=10). Scores range from 0 (no pain, full function) to 10 (worst pain or greatest impairment). Pain and function scores are converted to scales of 0–50 and are summed for a total score (0–100 pain and function weighted equally). The PRWE differs from the Patient-rated Wrist/Hand Evaluation (PRWHE) through a change in question stems from the “wrist” to the “wrist/hand”.

Boston Carpal Tunnel Questionnaire (BCTQ)

The BCTQ evaluates the severity of symptoms (i.e., pain, altered sensibility, and weakness: n=11) and functional status (n=8) associated with carpal tunnel syndrome.⁽¹⁶⁾ Scores for each dimension range from 1–5 with higher scores representing greater symptom severity and functional difficulty. Scores may only weakly correlate with objective sensibility testing, and median nerve conduction parameters.⁽¹⁶⁾

Choosing Among Outcome Measures

Clinical studies utilizing patient-rated outcome measures can often be performed with little direct cost. Patient-rated evaluations are generally appropriate for self-administration and then entered into an electronic database by the researcher. However, simply collecting as much data as possible is not recommended. More detailed outcome questionnaires and the use of multiple inter-related survey measures allow for a greater variety of analyses but data quality may suffer from responder fatigue. Patients are both less likely to skip questions on a 10 question survey than on a 100 question assessment and may give more thought to each answer on the shorter survey. Additionally, in busy practice settings, the time required for administration and scoring of an outcome measure may be a limiting factor. There is no single right answer when deciding how much data to collect. The choice always involves compromise and should be thoughtfully considered prior to initiating data collection to prevent later remorse during data analysis.

When making the final decision on using outcome measure before embarking on a clinical study, the choice between measures should be based on the primary aim of the proposed study. Provided the necessary resources, identifying the optimal outcome measure for a particular study on a topic such as distal radius fractures may vary (Examples follow).

Aim: Quantify final function in the affected wrist compare to the contralateral wrist.

We recommend the MHQ for independent scores of right and left hands as opposed to DASH or PRWE.

Aim: Determine the overall functional limitations when completing daily activities during the period of post-operative wrist immobilization.

We recommend the DASH as an assessment of composite upper-extremity function. The MHQ includes only 12/66 questions regarding bilateral function and with half of the PRWE total score depending on pain (unrelated to our aim) we would not choose those measures.

Aim: Determine if subjective paresthesias in the median nerve distribution at pre-operative evaluation is best addressed by simultaneous fracture fixation and carpal tunnel release.

We recommend the BCTQ due to the focus on sensibility changes and night symptoms. While other surveys query about sensibility (1/30 questions DASH, 4/66 questions MHQ), they would require collecting a substantial amount of information extraneous to this aim.

Aim: Compare the time-course of pain resolution to that of functional recovery in the affected wrist.

We recommend the PRWE as it would efficiently provide relevant data in its two scales. The DASH would only estimate function of both arms and the MHQ would require a longer survey (25 questions on opposite hand and 4 questions on appearance) to gather similar data.

Acknowledgments

Calfee Support: Research support by Grant Number UL1 RR024992 from the NIH-National Center for Research Resources (NCRR).

References

1. PSPower. <http://biostat.mc.vanderbilt.edu/wiki/Main/PowerSampleSize>
2. GPower. <http://www.psych.uni-duesseldorf.de/aap/projects/gpower/>
3. Shauver MJ, Chung KC. The minimal clinically important difference of the Michigan hand outcomes questionnaire. *J Hand Surg Am.* 2009; 34:509–514. [PubMed: 19258150]
4. Chung KC, Hamill JB, Walters MR, Hayward RA. The Michigan Hand Outcomes Questionnaire (MHQ): assessment of responsiveness to clinical change. *Ann Plast Surg.* 1999; 42:619–622. [PubMed: 10382797]
5. Ozyurekoglul T, McCabe SJ, Goldsmith LJ, LaJoie AS. The minimal clinically important difference of the Carpal Tunnel Syndrome Symptom Severity Scale. *J Hand Surg Am.* 2006; 31:733–738. discussion 739–740. [PubMed: 16713833]
6. Rozental TD, Blazar PE, Franko OI, Chacko AT, Earp BE, Day CS. Functional outcomes for unstable distal radial fractures treated with open reduction and internal fixation or closed reduction and percutaneous fixation. A prospective randomized trial. *J Bone Joint Surg Am.* 2009; 91:1837–1846. [PubMed: 19651939]
7. MHQ. <http://sitemaker.umich.edu/mhq/overview>
8. Smith MV, Calfee RP, Baumgarten KM, Brophy RH, Wright RW. Upper extremity specific measures of disability and outcomes in orthopaedic surgery. *J Bone Joint Surg Am.* 2011; 94:277–285. [PubMed: 22298061]
9. PubMed. <http://www.ncbi.nlm.nih.gov/pubmed>
10. DASH. <http://www.dash.iwh.on.ca>
11. PRWE. <http://www.srs-mcmaster.ca/ResearchResources/ResearchThemes/Musculoskeletal/UpperLimbNeck/tabid/2723/Default.aspx>

12. MacDermid JC, Turgeon T, Richards RS, Beadle M, Roth JH. Patient rating of wrist pain and disability: a reliable and valid measurement tool. *J Orthop Trauma*. 1998; 12:577–586. [PubMed: 9840793]
13. Hudak PL, Amadio PC, Bombardier C. Development of an upper extremity outcome measure: the DASH (disabilities of the arm, shoulder and hand) [corrected]. The Upper Extremity Collaborative Group (UECG). *Am J Ind Med*. 1996; 29:602–608. [PubMed: 8773720]
14. Hsu JE, Nacke E, Park MJ, Sennett BJ, Huffman GR. The Disabilities of the Arm, Shoulder, and Hand questionnaire in intercollegiate athletes: validity limited by ceiling effect. *J Shoulder Elbow Surg*. 2010; 19:349–354. [PubMed: 20303460]
15. Beaton DE, Wright JG, Katz JN. Development of the QuickDASH: comparison of three item-reduction approaches. *J Bone Joint Surg Am*. 2005; 87:1038–1046. [PubMed: 15866967]
16. Levine DW, Simmons BP, Koris MJ, Daltroy LH, Hohl GG, Fossel AH, et al. A self-administered questionnaire for the assessment of severity of symptoms and functional status in carpal tunnel syndrome. *J Bone Joint Surg Am*. 1993; 75:1585–1592. [PubMed: 8245050]

Table 1

Widely referenced patient-rated outcome measures for the hand and wrist.

Scale	Anatomic Region	Measures	Scores	Responder Burden [*]	Populations Commonly Tested
Boston Carpal Tunnel Questionnaire (Levine and Katz)	Hand	Pain, sensibility, weakness and function	Symptom severity, functional status	19 q	Carpal tunnel syndrome
Michigan Hand Questionnaire	Hand	Hand function, daily activities, work activity, pain, appearance, satisfaction	Total, ADL, work, pain, aesthetics, satisfaction for Right and Left	71 q	General hand and wrist disorders
Brief MHQ	Hand	Hand function, daily activities, work activity, pain, appearance, satisfaction	Total	12 q	General hand and wrist disorders
Patient Rated Wrist/Hand Evaluation	Hand	Pain, daily activities, recreation and work activities	Total, pain, function	15 q	General hand and wrist disorders
Cold Intolerance Severity Scale	Hand	Cold intolerance frequency, duration, severity, impact on activity	Total	6 q	Patient with post-traumatic cold intolerance
DASH	Upper extremity	Composite bilateral disability	Total	38 q	General upper extremity disorders
Quick DASH	Upper extremity	Composite bilateral disability	Total	11 q	General upper extremity disorders

^{*} Number of questions for completion