Development and validation of a new patient-reported outcome measure for peripheral nerve disorders of the hand: the I-HaND© Scale.

M Ashwood[1] , C Jerosch-Herold[1*], L Shepstone[2]

[1]School of Health Sciences, University of East Anglia, Norwich

[2]Norwich Medical School, University of East Anglia, Norwich

*** corresponding author:**

Christina Jerosch-Herold, Professor of Rehabilitation Research, School of Health Sciences, University of East Anglia, Norwich, The Queen's Building, Earlham Road, Norwich, Norfolk, NR4 7TJ, UK. c.jerosch-herold@uea.ac.uk twitter: @tinajerosch

Key words

Nerve injury; nerve compression; patient-reported outcomes; psychometrics

Acknowledgments

We would like to thank the local collaborators at participating NHS trusts who assisted with patient identification and data collection: Mrs Debbie Larson, Mrs Kathryn Johnson, Mrs Caroline Miller, Mr Dominic Power, Mrs Nikki Burr, Mrs Megan Blakeway, Ms Sarah Mee, Mrs Raelene Marx, Dr Niall Fitzpatrick, Miss Sarah Turner and the members of their clinical teams.

Hard copies of the I-HaND version 2 can be obtained under a creative commons license from the corresponding author. (https://www.uea.ac.uk/health-sciences/research/research-groups/rehabilitation/musculoskeletalrehabilitation/resources-and-tools).

Conflict of interest: The authors declare that there is no conflict of interest.

Funding: Mark Ashwood was funded through a University of East Anglia, Faculty of Medicine and Health Sciences PhD Studentship. Christina Jerosch-Herold was funded through a National Institute for Health Research (NIHR) Senior Research Fellowship (NIHR-SRF-2012-05-119). This article presents independent research funded by the National Institute for Health Research (NIHR). The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health.

Ethical approval: The study received ethical approval from the North East York Research Ethics Committee (14/NE/1087/ 28 July 2014) and local NHS Research Governance approval from each participating site.

Informed consent: All participants gave fully informed written consent prior to any data collection

Development and validation of a new patient-reported outcome measure for peripheral nerve disorders of the hand, the I-HaND© Scale

Abstract

Following guidelines from the Patient-Centred Outcomes Research Institute and using a mixed methods study, a new patient reported outcome measure for both nerve trauma and compression affecting the hand, the Impact of a Hand Nerve Disorders (I-HaND) Scale was developed. Face-to-face interviews with 14 patients and subsequent pilot-testing with 61 patients resulted in the development of the 32-item patient reported outcome measure. A longitudinal validation study with 82 patients assessed the psychometric properties of the I-HaND. Content and construct validity was confirmed by cognitive interviews with patients and through Principal Components Analysis. The I-HaND has high internal consistency (α=0.98) and excellent test-retest reliability (intraclass correlation coefficient = 0.97). Responsiveness statistics showed that the I-HaND is able to detect change over three months, discriminate between improvers and non-improvers. We conclude that the I-HaND can be used as a patient reported outcome measure for people with a range of hand nerve disorders.

Introduction

The assessment of outcome following peripheral nerve lesions remains a challenge for surgeons and therapists (Wang et al., 2013). Currently there is no disorder-specific patient reported outcome measure (PROM) suitable for patients with conditions comprising both traumatic and compression nerve injuries of the hand. Two condition-specific PROMs exist for patients with single nerve compression type disorders: the Boston Carpal Tunnel Questionnaire (Levine et al., 1993) for carpal tunnel syndrome and the Patient Rated Ulnar Nerve Evaluation (MacDermid and Grewal, 2013) for ulnar nerve compression. However, neither is suitable for patients with peripheral nerve trauma. In the absence of any condition specific PROM for nerve trauma, region-specific measures designed and developed more generally for musculoskeletal disorders of the hand and the upper limb have been used instead (MacDermid, 2005; Vordemvenne et al., 2007). They include the Patient Evaluation Measure (PEM) (Macey et al., 1995), the Michigan Hand Outcome Questionnaire (MHQ) (Chung et al., 1998; Chung et al., 1999) and the Disabilities of the Arm, Shoulder and Hand (DASH) (Hudak et al., 1996). A limitation is that their content was not developed specifically for people with hand nerve disorders (MacDermid, 2005). Furthermore, these PROMs were developed around 20 years ago and do not conform to current methodological standards for the development of PROMs (FDA, 2009; Patrick et al., 2011a; Patrick et al., 2011b) namely, in-depth qualitative research methods were not used to develop their content.

Developing a PROM for nerve trauma only was one option. However, a narrative review of qualitative studies of the impact of nerve compression (Martin, 2007; Khu et al., 2011, Jerosch-Herold et al., 2008) highlighted that compression syndromes also cause a significant burden to patients' functioning and quality of life, thus justifying the inclusion of trauma and compression of nerves of the hand. The aim of this study was to develop and validate a new hand nerve disorder PROM, using current guidelines from the health measurement literature and which assesses the impact of a hand nerve disorder on body structure or function, activity and participation.

Methods

Study design and patients

A multi-centre study using mixed methods was undertaken which comprised three phases: (1) item generation (qualitative methods), (2) content validation (qualitative and quantitative methods), and (3) psychometric evaluation (quantitative methods).

NHS Research Ethics approval was obtained prior to commencement and all participants provided written and informed consent.

Phase 1: Item generation

Development of PROMs needs to have a strong conceptual basis to ensure content and construct validity and provide operational meaning (FDA 2009). Kathy Charmaz's (Charmaz, 2006) constructivist grounded theory methods were modified for this qualitative study to generate a theory about the impact of nerve disorders on activities and participation. One-to-one interviews were conducted with 14 patients with a range of hand nerve disorders. These data served as a basis for developing the items for the new measure (see Ashwood et al., 2017 for details). Transcribed interviews were coded using the International Classification of Functioning, Disability and Health (ICF) (WHO, 2001) as a conceptual model. A hand nerve disorder-specific conceptual framework was developed and criteria for questionnaire design were followed to produce an item pool (Streiner and Norman, 2008). To ensure clinical relevance, a working group of experts was consulted during the development process. This structured and methodical process was followed to establish face and content validity of the new measure (Mullin et al., 2000).

Phase 2: Content and structural validity

This phase comprised two stages: Firstly, cognitive debriefing interviews as described by Gordon Willis (Willis, 2005) were conducted to clarify how patients understood the items and responses in the I-HaND version 1. Secondly, statistical methods were used to examine the structural validity of the new PROM (de Vet et al., 2011). Prospective data were

collected in patients with hand nerve disorders to assess how the items making up the I-HaND scale interact (Fayers and Machin, 2013). This content validation and item refinement process finalised the development of the I-HaND version 2.

Phase 3: Psychometric evaluation

This phase was concerned with the evaluation of construct validity, reliability and responsiveness (Mokkink et al., 2010) (see Table 1 for definitions). These attributes are key indicators of the quality of a measure and should be considered when selecting PROMs (FDA, 2009). Patients with a range of nerve conditions were recruited across eight hand therapy centres in the UK. At baseline, participants completed the I-HaND Scale, the Quick-DASH and a global status measure. These baseline data were used to evaluate construct validity. To assess test-retest reliability participants were asked to complete the I-HaND a second time (between 7 to 14 days). This timeframe was chosen as nerve recovery would not be likely, yet was long enough to minimise recall of previous responses (Frost et al., 2007). To assess responsiveness participants were asked to complete the I-HaND, Quick-DASH and global status measures again at 12 weeks from baseline, during which a proportion of patients were likely to have experienced a change in their condition.

Outcome measures

The I-HaND version 2 comprises 32 items scored on a 5-point ordinal scale (1 to 5) giving a possible raw summed score range of 32 to 160 points transformed into 0-100 percentage score. Higher scores indicate greater disability. There is no consensus on what proportion of missing items is acceptable. De Vet et al. (2011) propose that anything greater than 15% is unacceptable. Using a similar threshold to the 30-item DASH, we suggest that a total score should not be calculated if more than 3 items have missing responses (<10%). The Quick-DASH was used as a comparator measure at baseline and 12 weeks. A global status measure was used to obtain an estimation of function at baseline and 12 weeks. The percentage of normal hand function (%NHF) score was modified from the Stanmore Percentage of Normal Shoulder Assessment (SPONSA), a validated, single-item PROM (Noorani et al., 2012)

(supplementary file I). A global rating of change (GROC) measure was also used at the 12-week follow-up. Participants were asked to rate on a three-point Likert scale whether their condition had improved, stayed the same or worsened. The %NHF and GROC were used as external anchors for the assessment of change (responsiveness) (Husted et al., 2000). A clinical record form asked patients questions about their sociodemographic status and clinicians about the patients' peripheral nerve diagnosis and their surgical history.

Statistical analysis

Data from phases 2b and 3 were explored through descriptive analysis. Inter-item correlations, range of scores, homogeneity of items, and distribution of the data and the presence of outliers were also explored. The latent structure of the scale was evaluated using principal components analysis (PCA)[1]. The internal consistency of the scale was examined using Cronbach's alpha. Construct validity was assessed by *a priori* hypotheses (Table 1). Using the Quick DASH as well as GROC and %NHF as comparators, a moderate to strong (Pearson's r ≥0.6) correlation was hypothesised, as evidence of construct validity. Test-retest reliability was calculated using intraclass correlation coefficients (ICC). Responsiveness was assessed by *a priori* hypothesis testing (Table 1). Cohen's effect size (ES) and standardised response mean (SRM) were calculated for the I-HaND and Quick-DASH. The GROC and % NHF were used to dichotomise patients into improvers and non-improvers and receiver operating characteristic (ROC) curves created and the area under the curve (AUC) calculated.

[Footnote : 1: The term Exploratory Factor Analysis is sometimes used to mean the same analysis as PCA though in general PCA and factor analysis are distinct. See deVellis (2017) for an interesting discussion of the terminology.]

Results

Phase 1: Concept-elicitation interviews and item generation

Fourteen participants recruited from a single centre took part in face to face interviews. They were aged between 25 and 74 years with diagnoses including median or ulnar nerve

trauma and median, ulnar or radial nerve compression (supplementary file II). The items for a 34-item I-HaND version 1 were generated from this framework (Figure 1) covering four domains: (1) symptoms, physical difficulties and feelings; (2) pain or discomfort; (3) activities and (4) participation. Careful consideration was given to the layout and instructions, framing of questions, response format and recall period to reduce missing or invalid responses and minimise cognitive and respondent burden (Streiner and Norman, 2008).

Phase 2: Content Validation

Eleven of the 14 participants who were involved in phase 1 also took part in the cognitive interviews. Examples of illustrative quotations from patients for the overall endorsement, content, response categories, instructions, layout and time required to complete the I-HaND Scale are provided in supplementary file III. Three rounds of cognitive interviews took place, with revisions made to the I-HaND after each round. The refinement to the content of the items for each round of interviews is presented in supplementary file III.

Fifty participants were recruited from three UK centres for the assessment of structural validity. A summary of the characteristics of the sample is provided in Table 2. Their mean I-HaND total score was 87 points (SD = 40). For all the items, each of the five available response categories was used and missing data was low (0.5%). There were no ceiling effects observed. However, floor effects were observed in five items with more than 50% of respondents selecting the lowest category for these questions.

Phase 2 - Construct (structural) validity

A principal component analysis (PCA) was carried out on the I-HaND Scale to explore its structural validity. Principal component analysis is appropriate to identify underlying domains (components) of instruments (Fayers and Machin, 2013). From the 50 participants, 42 cases were included as the analysis was based on cases with no missing values. The PCA of the I-HaND Scale identified four components with eigenvalues ≥ 1.00. However, most of the variance (72%) was explained by the first component as can be seen in Cattell's scree

plot where a sharp drop (the point of inflexion) is visible after the first component and then the line becomes more level (Figure 2). The other components individually add little to the variability explained.

Cronbach alpha for the I-HaND was 0.98, demonstrating excellent internal consistency. However, high alpha values (>0.90) can also indicate potential item redundancy (Streiner and Norman, 2008). This was explored further by item-total and inter-item correlation analysis. In addition to the statistical analysis the conceptual importance of items, as previously identified from the concept elicitation interviews, as well as their clinical relevance through discussion with experts on the PROM development group were used to determine whether items should be removed. This approach highlighted 13 potential items of which three were removed (supplementary file IV) resulting in the 32-item I-HaND Scale version 2 (Figure 3).

Phase 3: Psychometric evaluation

Eighty-two people with a range of hand nerve disorders were recruited from eight UK centres. To evaluate structural validity with a larger sample size the data from phase 2b were combined resulting in a sample size of 132 participants (Table 2). Only participants with complete data were included in the analysis (n=118). The mean raw total I-Hand score for the sample was 90 (SD=31) out of a possible 160 points. Missing responses from participants were low (<1%). There were no ceiling effects but floor effects were observed with three items [Q9: I feel self-conscious if people look at my hand/arm; Q12: I have hurt my hand and not realised it until later; and Q19: putting toothpaste on a toothbrush], with more than 40% of respondents selecting the lowest (easiest) category.

Construct (structural) validity

Of the 132 participants, 118 had complete baseline data and were included in the PCA. Components with Eigenvalues ≥ 1.00 were identified, following Kaiser's criterion. The PCA of the I-HaND Scale revealed four components, which together explained 74% of the variance. Most of the variance was explained by the first component (58%). This was higher than the minimum recommended 50% value for a stable one-factor solution, but lower than in phase

2, where the first component accounted for 72% of the total variance. The internal consistency of the I-HaND Scale was very high (Cronbach's alpha 0.98).

Hypothesis-testing construct validity

Using the Quick DASH and %NHF as comparators baseline data were available for 82 participants. Seventy-two participants provided complete data. Nine participants with some missing data (three or less missing items) were also included in the correlation analysis by substituting missing items with the scale mean. One participant who had more than 10% missing data was excluded. As hypothesised, a positive, strong correlation was found between the I-HaND and Quick DASH (r= 0.87) and a negative, strong correlation was seen with %NHF (r=-0.64).

Test-retest reliability

Sixty-one participants completed the I-HaND Scale at baseline and 7 to 14 days (mean 12 days, range 4 to 30 days). Complete data were available for 56 people and used in the analysis. Test-retest reliability for the I-HaND was excellent (ICC = 0.97; 95% CI = 0.94 to 0.98).

Responsiveness to change

Fifty participants completed the I-HaND at baseline and at the second follow-up (12 weeks) providing data for the responsiveness analysis. Forty-five participants provided complete data; five participants who had < 10% missing data (three or less missing items) were also included in the analysis, by substituting missing items with the scale mean. One participant who had more than 10% missing data was excluded. Effect sizes (ES) and standardised response means (SRM) for the I-HaND were moderate (ES=0.51; SRM=0.60) and marginally higher than the Quick DASH (Table 1).

The hypothesis that the I-HaND can discriminate between patients who reported themselves as improved and those remaining the same or worse was evaluated by constructing ROC curves and calculating the area under the curve. The larger the area under the curve (closer to 1), the greater the ability of the scale to discriminate (Husted et al, 2000). The group was dichotomised into improvers and non-improvers using the global

change (GROC) measure. The global status measure (%NHF) scores at baseline and follow-up were also converted into a change score to create an additional patient anchor with which to classify patients into improvers and non-improvers. The area under the curve was large (≥0.82) for both types of anchors (Table 1, Figures 4a and b).

Discussion

An in-depth qualitative study of the impact of hand nerve disorders including trauma and compression generated a conceptual framework from which a new PROM for hand nerve disorders was developed, the I-HaND. Cognitive interviews confirmed that patients found the I-HaND relevant, highly acceptable and quick to complete. Subsequent psychometric evaluation of the 32-item I-HaND confirmed its construct validity, high internal consistency, excellent test-retest reliability and that it is responsive over three months.

Our study took an approach to scale refinement that is recommended (FDA, 2009) but differs from the approaches adopted by others in the field of hand surgery and rehabilitation. Specifically, the I-HaND Scale was developed on the basis of patient interviews, which defined the areas for scale content (Patrick et al., 2011a; Patrick et al., 2011b). In hand surgery and rehabilitation it has been typical to develop an item pool based on expert clinicians' opinion or from the literature, followed by an item-reduction process using factor analysis (Chung et al., 1998; Hudak et al., 1996). With this approach, the content of a scale, rather than the construct intended for measurement, defines what the scale measures (Hobart et al., 2007).

Cognitive debriefing interviews with patients provided further evidence that previous steps taken to ensure trustworthiness had been effective and that the preliminary I-HaND was clear, understood and relevant for people with nerve conditions. The complementary use of statistical methods, identified strengths and weaknesses of the developing PROM. Only minor changes were made, as caution is advocated when making changes to newly developed instruments on the basis of small samples and therefore a very parsimonious approach to item reduction was taken to retain content and clinical validity.

Classical test theory methods were used to assess the psychometric properties of the I-HaND and our results provide initial evidence of this. The proportion of missing data was low, suggesting that it was acceptable to patients. Scale scores spanned the entire range of response options. There were some floor effects, however, PROMs need to be able to capture different levels of ability so the fact that some items were easy for some people but not for others was desirable. The exploratory PCA supports the notion of a unidimensional scale with high internal consistency, as demonstrated with a high alpha coefficient and item-total correlations. An alpha of 0.90 to 0.95 is desirable (Bland and Altman, 1997), although our $\alpha = 0.98$ exceeds this and may indicate some item redundancy. The high number of items making up the I-HaND scale may also inflate alpha. However, moderate to strong item-total correlations, provided further evidence that the items are measuring different aspects of the same construct and there were no correlations >0.9. Whilst there is a trend towards producing shorter versions of PROMs, this can be at the expense of patient and clinical relevance. The PCA identified that one factor explained over 58% of the score variance although this was substantially lower than in phase 2. This discrepancy may be due to smaller sample sizes used in phase 2. In phase 3, sample sizes were on the borders of acceptability for the assessment of structural validity (Mokkink et al., 2010). Although some authors argue that useful estimates can be obtained from small samples, further examination of the structure of the I-HaND in larger samples is needed (Hobart et al., 2012).

Test-retest reliability was excellent. The generated hypotheses relating to the strength of association with the Quick-DASH and %NHF were supported, thus providing evidence of construct validity. Although the correlation is stronger than hypothesised it does not indicate that these instruments measure the same constructs. The Quick-DASH is made up of 11 items compared to 32 in the I-HaND. Furthermore patients gave strong endorsement to the relevance of items in the I-HaND such as '*You would think that it was made for me to be honest*' and '*Everything in there was what actually occurred and what I have been through*'. Finally, the time required to complete the I-HaND is relatively short, with participants taking between three and seven minutes, which would be considered a minimal burden.

The use of classical test theory methods for the development of new PROMs has been criticised (Cano and Hobart, 2011) as these methods produce measures which are ordinal in nature, in that they describe order but not the relative size or degree of the difference between measurements. A more modern approach to scale development is the use of Rasch measurement methods which have the ability to construct linear, interval-level measurements from ordinal-level rating scale data (Cano et al., 2011). Further exploration of the structural validity of the I-HaND using Rasch model analysis is recommended.

The results of this study provide evidence that the I-HaND Scale can measure change over time, when change is expected. This is particularly important for condition-specific PROMs (Guyatt et al., 1987). The use of distribution and anchor-based methods to assess external responsiveness provided a more meaningful estimate of change, as patients have defined this themselves (Wyrwich et al., 2013). In addition using two patient measures - global status and global change, can help to minimise the effect of recall bias associated with global rating of change (Norman et al., 1997).

A limitation of the responsiveness study is that whilst the overall sample size was good, when the group was dichotomised into groups of improvers and non-improvers, each sub-group was small. In responsiveness studies, change is usually reported in relation to a known effective intervention, such as carpal tunnel decompression. In this study patients with a range of different nerve diagnoses were recruited, undergoing a wide range of conservative and surgical treatments and over a relatively short time span. This may explain why the effect size for the I-HaND was only modest compared to the Quick-DASH. On the other hand, a potential benefit of this approach is that the people recruited were representative of the target population. Further work is necessary to evaluate the responsiveness of the I-HaND Scale over a longer period and define minimally clinically important difference (MCID) which is an aspect of a PROM's interpretability (Mokkink et al., 2010).

Subject to further psychometric testing, including Rasch model analysis, the I-HaND Scale has the potential to be used in research as part of an agreed core outcome set for nerve disorders of the hand and in future clinical trials (Williamson et al., 2012). The I-Hand version 2 is a clinically useful instrument which patients find relevant, quick and easy to complete. It can be used for the routine evaluation of outcome for peripheral nerve

disorders of the hand, outcomes that are ultimately best judged by patients themselves and can support patient-focused decision making and goal planning. The I-HaND could be used as a complementary outcome measure to other clinician-derived impairment scores such as the validated Model Instrument for the Documentation of Outcome after Nerve Repair, also known as the Rosén score (Rosén and Lundborg, 2000).

Acknowledgement: Hard copies of the I-HaND version 2 can be obtained under a creative commons license from the corresponding author. (https://www.uea.ac.uk/health-sciences/research/research-groups/rehabilitation/musculoskeletalrehabilitation/resources-and-tools).

Figures and Tables

Figure 1. Conceptual framework derived from patient interviews of impact of hand nerve disorders using the WHO ICF.

Figure 2. Cattel's scree plot of Eigenvalues for the components of the I-HaND Scale and point of inflection (arrow).

Figure 3. Questions and response categories for the Impact of Hand Nerve Disorders (I-HaND) Scale.

Figure 4. Receiver operating characteristic (ROC) curve showing area under the curve for I-HaND score change compared with (a) global rating of change (GROC) and (b) percentage of normal hand function (%NHF) change.

Table 1. Overview of psychometric properties assessed for the I-HaND including definitions, methods used and summary of results.

Table 2. A summary of the sociodemographic and clinical characteristics of participants at baseline and follow-up.

Supplementary Files (online only)

Supplementary file I: Percentage of Normal Hand Function

Supplementary file II: Characteristics of study sample from Phase 1 study (interviews)

Supplementary file III: Examples of illustrative quotations from patients for the overall endorsement, content, response categories, instructions, layout and time required to complete the I-HaND Scale

Supplementary file IV: Development of the content of the items of the I-HaND Scale with changes made highlighted in bold

Supplementary file V: Summary of item-revision process (Phase 2), with changes highlighted in bold

References

Ashwood M. Jerosch-Herold C, Shepstone L. Learning to live with a hand nerve disorder: A constructed grounded theory. J Hand Ther. 2017, published online November 29, 2017

Bland J M and Altman DG. Cronbach's alpha. BMJ. 1997, 314:572.

Cano S J, Barrett LE, Zajicek JP, Hobart JC. Beyond the reach of traditional analyses: using Rasch to evaluate the DASH in people with multiple sclerosis. Mult Scler. 2011, 17: 214-22.

Cano S J, Hobart JC. The problem with health measurement. Patient Prefer Adherence. 2011, 5: 279-90.

Charmaz K. *Constructing grounded theory: a practical guide through qualitative analysis*. London, Sage, 2006:177.

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: 619-22.

Chung K C, Pillsbury MS, Walters MR, Hayward RA. Reliability and validity testing of the Michigan Hand Outcomes Questionnaire (MHQ). J Hand Surg Am. 1998, 23: 575-87.

DeVellis RF. *Scale Development: theory and applications*., 4th Edn. Los Angeles, Sage,. 2017.

de Vet HC, Terwee CB, Mokkink LB, Knol DL. *Measurement in Medicine: a practical guide*. Cambridge, UK, Cambridge University Press, 2011:328.

Fayers PM, Machin D. *Quality of life: the assessment, analysis and interpretation of patient-reported outcomes*. 2nd Edn. Chichester, UK, John Wiley & Sons, 2013:568.

Federal Drug Administration. *Guidance for Industry -Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labelling Claims*. Rockville, Maryland, USA, U.S. Department of Health and Human Services Food and Drug Administration, 2009: 47

Frost MH, Reeve BB, Liepa AM, Stauffer JW, Hays RD, . What is sufficient evidence for the reliability and validity of patient-reported outcome measures? Value Health. 2007, 10: S94-S105.

Guyatt G, Walter S, Norman G. Measuring change over time: assessing the usefulness of evaluative instruments. J Chronic Dis. 1987, 40: 171-8.

Hobart JC, Cano SJ, Warner TT, Thompson AJ. What sample sizes for reliability and validity studies in neurology? J Neurol. 2012, 259: 2681-94.

Hobart JC, Cano SJ, Zajicek JP, Thompson AJ. Rating scales as outcome measures for clinical trials in neurology: problems, solutions, and recommendations. Lancet Neurol. 2007, 6: 1094-105.

Hudak P, Amadio P, Bombardier C . Development of an upper extremity outcome measure: the DASH (Disabilities of the arm, shoulder, and hand). The Upper Extremity Collaborative Group (UECG). Am J Ind Med. 1996, 29: 602-8.

Husted JA, Cook RJ, Farewell VT, Gladman DD. Methods for assessing responsiveness: a critical review and recommendations. J Clin Epidemiol. 2000, 53: 459-68.

Levine D, Simmons B, Koris M 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: 1585-92.

MacDermid JC. Measurement of health outcomes following tendon and nerve repair. J Hand Ther. 2005. 18: 297-312.

MacDermid JC, Grewal R. Development and validation of the patient-rated ulnar nerve evaluation. BMC Musculoskelet Disord. 2013, 14: 146.

Macey AC, Burke FD, Abbott K et al. Outcomes of hand surgery. British Society for Surgery of the Hand. J Hand Surg Br. 1995, 20: 841-55.

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.

Mullin PA, Lohr KN, Bresnahan BW, McNulty P. Applying cognitive design principles to formatting HRQOL instruments. Qual Life Res. 2000, 9: 13-27.

Noorani AM, Roberts DJ, Malone AA et al. Validation of the Stanmore percentage of normal shoulder assessment. Int J Shoulder Surg. 2012, 6: 9-14.

Norman GR, Stratford P, Regehr G. Methodological problems in the retrospective computation of responsiveness to change: the lesson of Cronbach. J Clin Epidemiol. 1997, 50: 869-79.

Patrick DL, Burke LB, Gwaltney CJ et al (a). Content validity-establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO good research practices task force report: part 1-eliciting concepts for a new PRO Instrument. Value Health. 2011, 14: 967-77.

Patrick DL, Burke LB, Gwaltney CJ et al (b). Content validity-establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO good research practices task force report: part 2-assessing respondent understanding. Value Health. 2011, 14: 978-88.

Rosén B, Lundborg G. A model instrument for the documentation of outcome after nerve repair. J Hand Surg Am. 2000, 25:535-43.

Rothman M, Burke L, Erickson P, Leidy NK, Patrick DL, Petrie CD. Use of existing patient-reported outcome (PRO) instruments and their modification: The ISPOR Good Research Practices for Evaluating and Documenting Content Validity for the Use of Existing

Instruments and Their Modification PRO Task Force Report. Value Health. 2009, 12: 1075-83.

Streiner DL, Norman GR. *Health measurement scales: a practical guide to their development and use*. 4th Edn. New York, Oxford University Press, 2008:423.

Vordemvenne T, Langer M, Ochman S, Raschke M, Schult M. Long-term results after primary microsurgical repair of ulnar and median nerve injuries. A comparison of common score systems. Clin Neurol Neurosurg. 2007, 109: 263-71.

Wang Y, Sunitha M, Chung KC. How to measure outcomes of peripheral nerve surgery. Hand Clin. 2013, 29: 349-61.

WHO. *International Classification of Functioning, Disability and Health (ICF)*. Geneva, Switzerland, World Health Organization, 2001 [url: http://www.who.int/classifications/icf/en/] [last accessed 10/09/17].

Williamson PR, Altman DG , Blazeby JM et al. Developing core outcome sets for clinical trials: issues to consider. Trials. 2012, 13: 132.

Willis GB. *Cognitive interviewing: A Tool for Improving Questionnaire Design.* London, Sage, 2004:335.

Wyrwich K, Norquist JM, Lenderking WR, Acaster S. Industry Advisory Committee of International Society for Quality of Life. Methods for interpreting change over time in patient-reported outcome measures. Qual Life Res. 2013 22: 475-83.

Table 1: Overview of psychometric properties assessed for the I-HaND including definitions, methods used and summary of results

Domain	Measurement property	Definition (from COSMIN)	Methods	Results for I-HaND
Validity	Content validity	The degree to which the items of the PRO – instrument look like an adequate reflection of the construct to be measured	Phase 1 – qualitative study involving interviews with 14 patients from the target population	Development of a conceptual framework on the impact of a hand nerve disorder from which items for the I-HaND were generated
			Phase 2a – qualitative study involving cognitive interviews with 11 patients from the target population	Patients confirmed I-HaND scale as relevant and acceptable (see supplementary file I)
	Face validity	The degree to which the items of a PRO instrument indeed look as though they are an adequate reflection of the construct to be measured	Phase 1 - review by a PROM development group with experience in upper limb rehabilitation, outcome measurement and PROM development	Face validity established by PROM development group
Construct validity	Structural validity	The degree to which the scores of a PRO instrument are an adequate reflection of the dimensionality of the construct being measured	Phase 2b & 3 – dimensionality explored using Principal Component Analysis (PCA) of I-HaND scores	58% of variance explained by 1st component (PC1), item loading >0.5 on PC1 for all items, communalities range from 0.5 to 0.7
	Hypothesis-testing	The degree to which scores on the PRO instrument are consistent with hypotheses regarding its relations hip to scores on other instruments	*a priori* hypotheses:	
			I-HaND scores will show a positive, moderately strong correlation (>0.6) with the Quick-DASH and negative, moderately strong correlation (> -0.6) with the %NHF	Correlation with Quick-DASH r=0.87 Correlation with % NHF r= -0.64

Reliability	Test-retest	Extent to which scores for patients who have not changed are the same over time	Phase 3 study – repeated administration of I-HaND over 7 to 14 day interval, in a stable group where no change was anticipated. Quantified using Intra-class correlation coefficient (ICC)	ICC = 0.97 95%CI = 0.94 to 0.98
	Internal consistency	The degree of the interrelatedness of the items	Phase 2b & 3 – statistical examination of the interrelatedness of items using Cronbach's alpha	Cronbach alpha = 0.98
Responsiveness	Responsiveness	The ability of a PRO instrument to detect change over time in the construct being measured	*a priori* hypotheses:	
			1)the I-HaND can detect change over a 12 week period measured by effect size of >0.5, in a group where change is expected	I-HaND ES=0.51, SRM=0.60
			2) the I-HaND can discriminate between Improvers and non-improvers	AUC using GROC = 0.82 (95%CI 0.70;0.94) AUC using %NHF = 0.83 (95%CI 0.71;0.94)
			3) the I-HaND will be more responsive relative to the Quick DASH	Quick-DASH ES=0.42, SRM=0.56

AUC = area under the curve, ES = effect size GROC = global rating of outcome, CI= confidence interval, NHF= percentage of normal hand function, PRO = patient rated outcome, PC = principal component, SRM = standardised response mean

Table 2: A summary of the sociodemographic and clinical characteristics of participants at baseline and follow-up

	Phase 2b	Phase 3			
	Structural Validity (N=50)	**Structural validity (N = 132)**	**Hypothesis testing (N = 82)**	**Test-retest reliability (N = 61)**	**Responsive-ness (N = 50)**
No. (%) of men	27 (54%)	72 (55%)	49 (60%)	39 (64%)	29 (58%)
Mean age (range) in years	55 (18 to 88)	52 (18 to 93)	49 (18 to 75)	52 (21 to 93)	54 (21 to 93)
Carpal tunnel syndrome	20 (40%)	42 (32%)	22 (27%)	18 (30%)	14 (28%)
Cubital tunnel syndrome	1 (2%)	12 (9%)	11 (13%)	8 (13%)	9 (18%)
Radial nerve palsy	9 (18%)	16 (12%)	7 (9%)	4 (7%)	4 (8%)
Median nerve injury	7 (14%)	23 (17%)	16 (20%)	12 (20%)	9 (18%)
Ulnar nerve injury	7 (14%)	19 (14%)	12 (14%)	11 (18%)	8 (16%)
Combined nerve lesion	3 (6%)	17 (13%)	14 (17%)	8 (13%)	6 (12%)
Concomitant injury	21 (42%)	54 (41%)	33 (40%)	22 (36%)	16 (32%)
Treated surgically	42 (84%)	109 (83%)	67 (82%)	52 (85%)	43 (86%)
Mean duration (range) in months:	39 (2 to 367)	29 (1 to 367)	22 (1 to 179)	24 (1 to 79)	27 (1 to 79)
Mean time since surgery (range) in months	15 (1 to 88)	9 (1 to 88)	5 (1 to 24)	5 (1 to 20)	5 (1 to 19)
Dominant hand affected	22 (44%)	55 (42%)	33 (40%)	23 (38%)	19 (38%)
Living alone	8 (16%)	17 (13%)	9 (11%)	7 (12%)	6 (12%)
Caring for others	17 (34%)	36 (27%)	19 (23%)	16 (26%)	14 (28%)
Working	24 (48%)	68 (52%)	44 (54%)	31 (51%)	28 (56%)
Changed work	13 (26%)	36 (27%)	23 (28%)	16 (26%)	11 (22%)

Figure 1

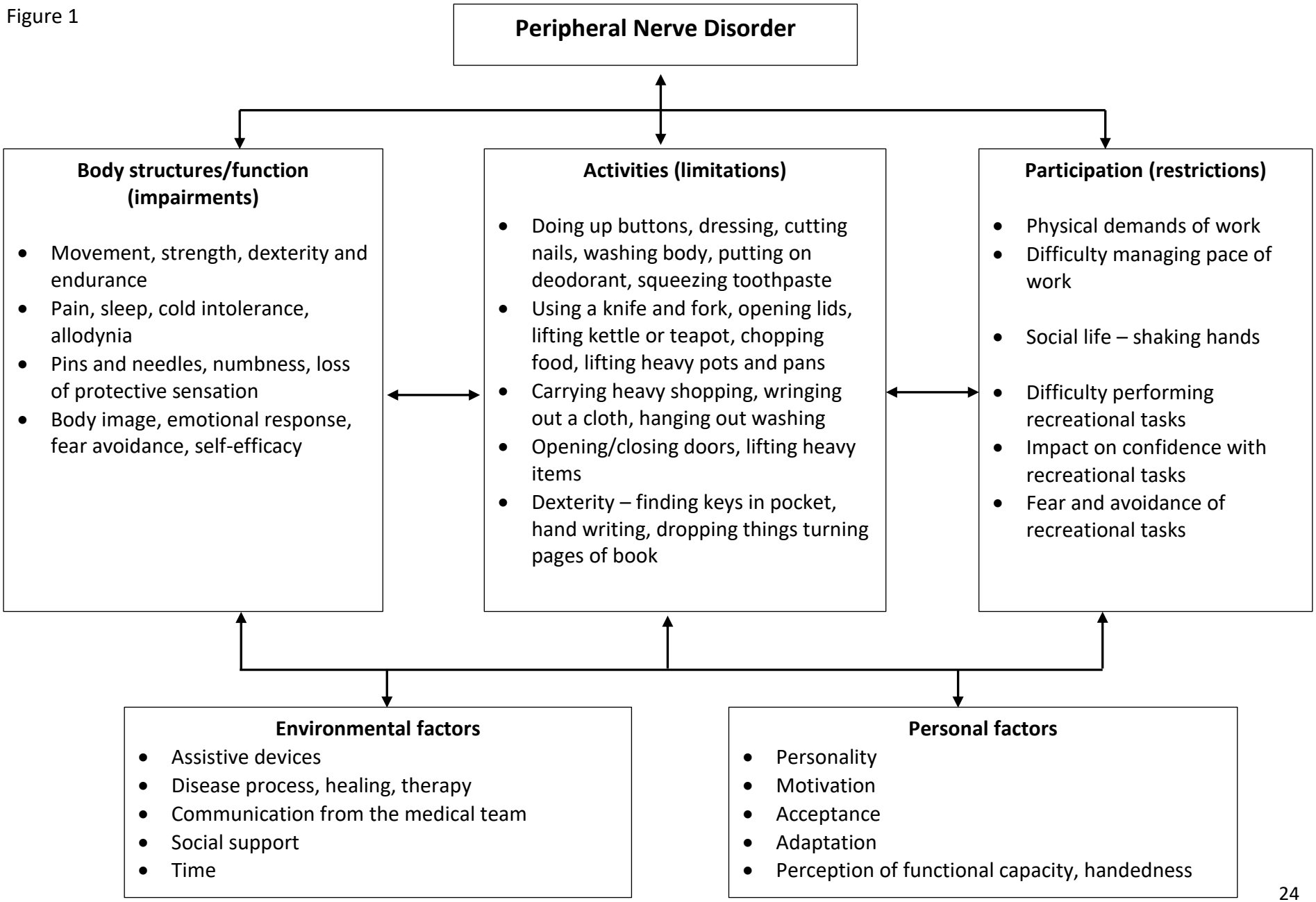
Peripheral Nerve Disorder
Body structures/function (impairments)
Movement, strength, dexterity and endurance
Pain, sleep, cold intolerance, allodynia
Pins and needles, numbness, loss of protective sensation
Body image, emotional response, fear avoidance, self-efficacy
Activities (limitations)
Doing up buttons, dressing, cutting nails, washing body, putting on deodorant, squeezing toothpaste
Using a knife and fork, opening lids, lifting kettle or teapot, chopping food, lifting heavy pots and pans
Carrying heavy shopping, wringing out a cloth, hanging out washing
Opening/closing doors, lifting heavy items
Dexterity – finding keys in pocket, hand writing, dropping things turning pages of book
Participation (restrictions)
Physical demands of work
Difficulty managing pace of work
Social life – shaking hands
Difficulty performing recreational tasks
Impact on confidence with recreational tasks
Fear and avoidance of recreational tasks
Environmental factors
Assistive devices
Disease process, healing, therapy
Communication from the medical team
Social support
Time
Personal factors
Personality
Motivation
Acceptance
Adaptation
Perception of functional capacity, handedness

Figure 2: Cattel's scree plot of Eigenvalues for the components of the I-HaND Scale and point of inflection (arrow)

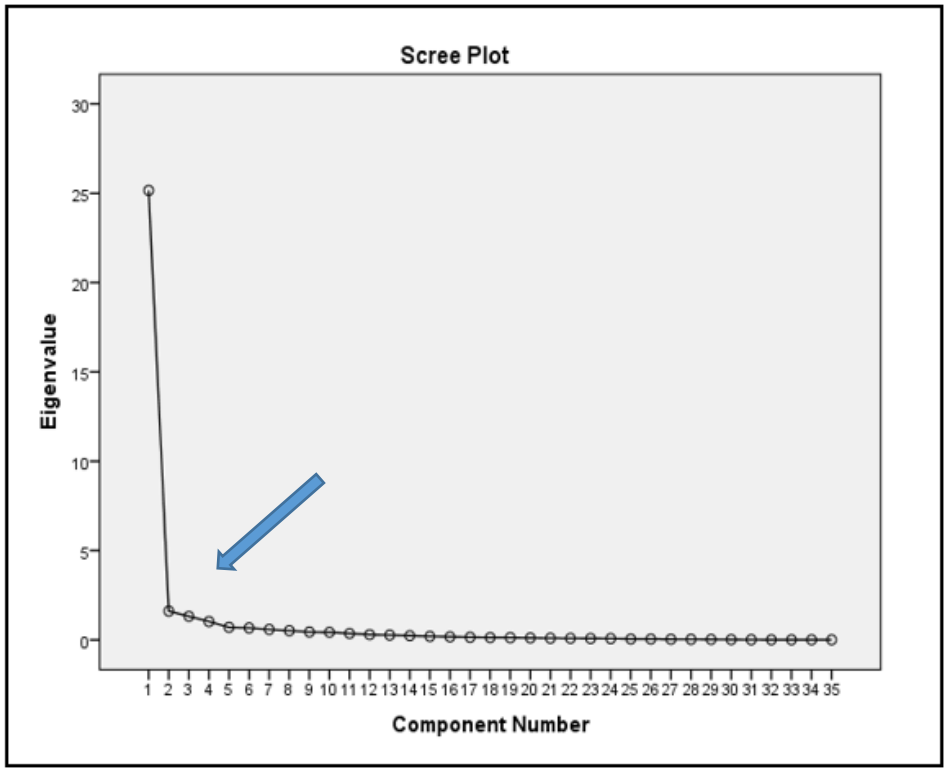

Figure 3: Questions and response categories for the *Impact of Hand Nerve Disorders (I-HaND) Scale*

Part 1: *questions ask about any* ***symptoms, physical difficulties and feelings*** *experienced as a result of your nerve disorder of the hand(s) over the past week.*

1.	*How well did your hand(s) work?*	*Very well (1) to very poorly (5)*
	How satisfied are you with the following?	
2.	The movement of your hand(s)	*Very satisfied (1) to very dissatisfied (5)*
3.	The sense of touch in your hand(s)	
4.	The strength in your hand(s)	
	Please indicate how often you have experienced the following in the past week	
5.	I can't grip or pinch for very long without my hand getting tired	*Never (1) to always (5)*
6.	When I touch certain things it causes pins and needles or tingling	
7.	When I go to pick something up it falls out of my hand	
	Please indicate how often you have experienced the following in the past week	
8.	Using my hand(s) can bring about strong emotions e.g. frustration, anger, sadness	*Never (1) to always (5)*
9.	I feel self-conscious if people look at my hand/arm	

Part 2: *The following questions ask about any* ***pain or discomfort*** *that you may have experienced as a result of your nerve disorder of the hand(s).*

10.	*The pain or discomfort* ***in my hand(s) has been***	*None (1) to very severe (5)*
11.	How often would you say that your pain or discomfort impacts on your daily routine?	*Never (1) to always (5)*
12.	I have hurt my hand and not realised it until later	
13.	My hand feels over sensitive when touched	
14.	I feel pain or discomfort when my hand is cold	
15.	It is difficult to get a good night's sleep because of the pain or discomfort in my hand/arm	

Part 3: *The following questions ask about difficulty with* ***activities*** *that you may have experienced as a result of your nerve disorder of the hand(s).*

16.	*How well have you been able to carry out your daily routine e.g. getting ready, cooking, childcare etc.*	*Very well (1) to very poorly (5)*
	How difficult has it been for you to complete the following activities?	
17.	Getting dressed or undressed	*Not at all difficult (1) to unable (5)*
18.	Doing up buttons	
19.	Putting toothpaste on a toothbrush	
20.	Cutting your nails	
21.	Cutting food using a knife & fork together	
22.	Opening lids of tight jars and bottles	
23.	Pouring from a kettle	
24.	Wringing out a cloth	
25.	Preparing a meal	
26.	Opening & closing heavy doors	
27.	Turning pages of a book, magazine or newspaper	
28.	Using electronic devices e.g. a remote control, mobile phone, tablet or computer	
29.	Carrying a heavy shopping bag	
30.	Handling small coins e.g. 5 pence or 1 pence	

Part 4: *The following questions ask about how your nerve disorder of the hand(s) has affected your ability to take part in your* ***daily work*** *(including paid work, school work or housework) and* ***recreational activities****.*

31.	*How well have you been able to manage the physical demands of your daily work?*	*Very well (1) to very poorly (5)*
32.	How well have you been able to take part in recreational activities e.g. Hobbies or sport?	

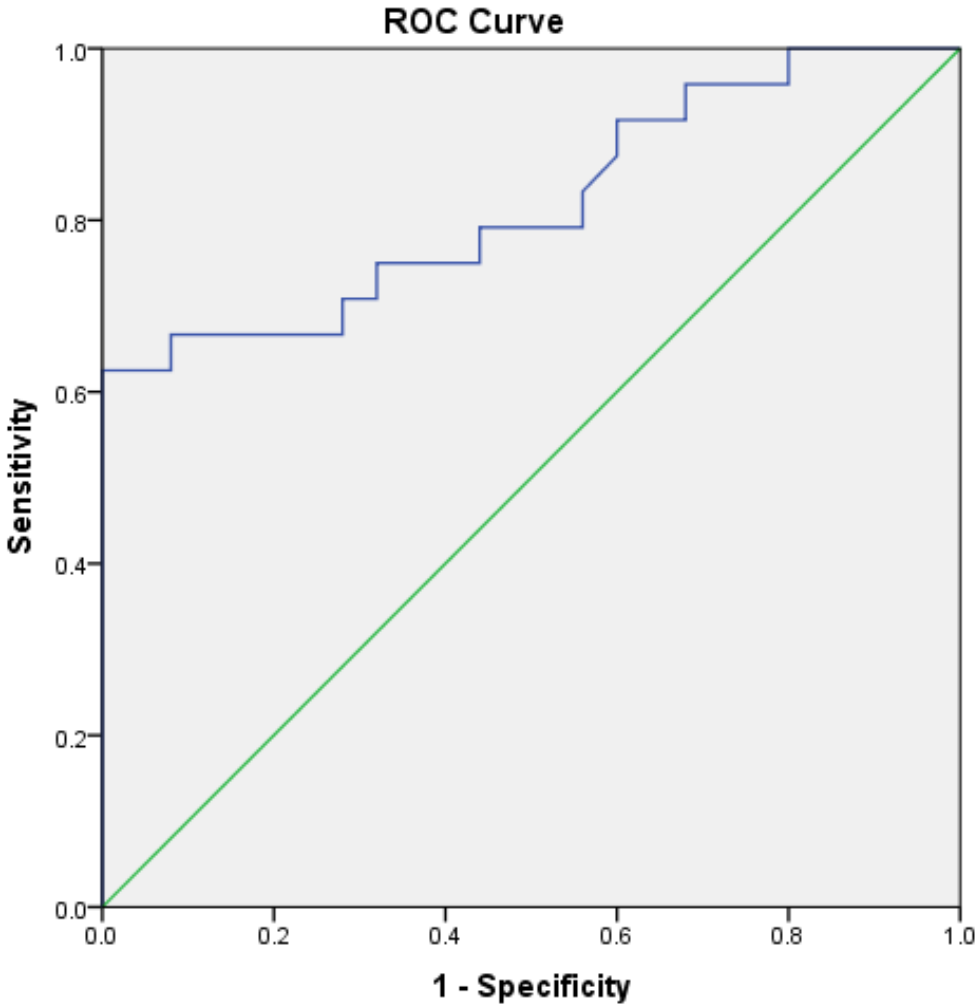

Figure 4a: Receiver operating characteristic (ROC) curve of I-HaND using GROC

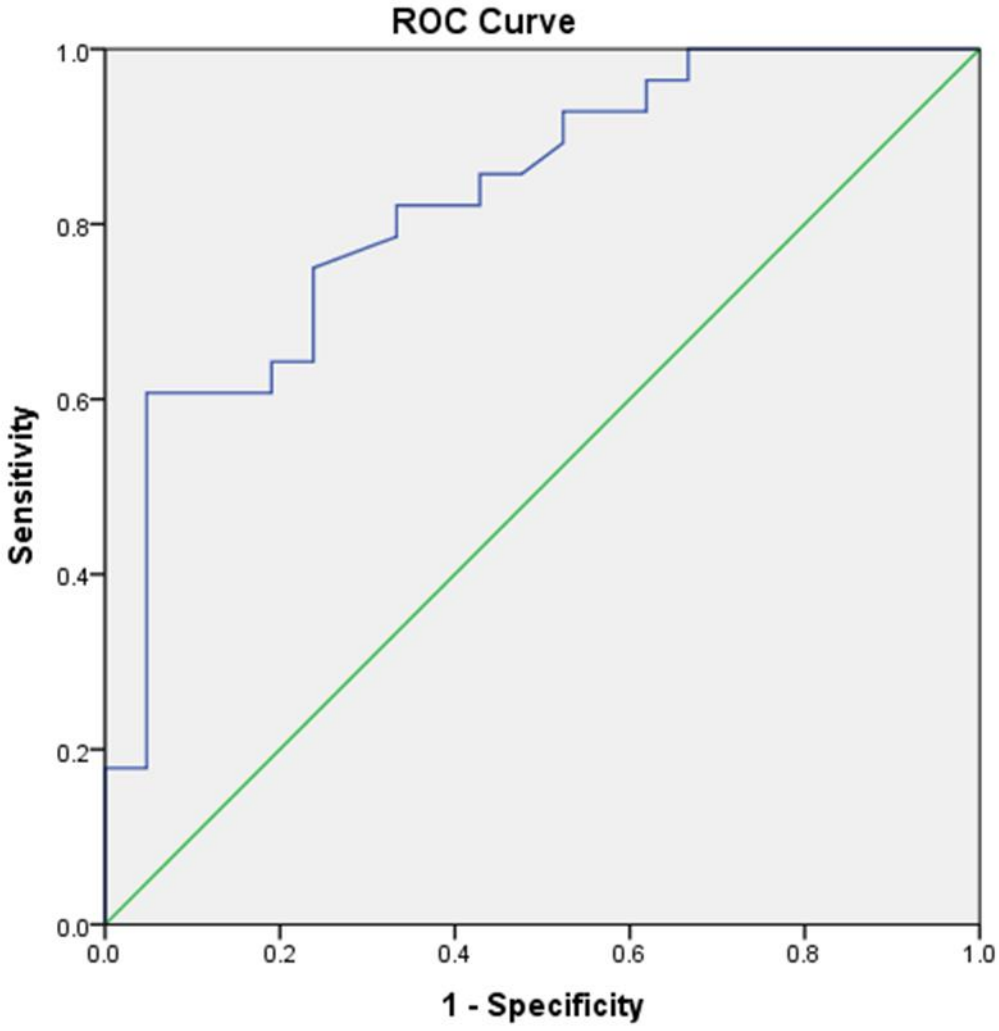

Figure 4b: Receiver operating characteristic (ROC) curve of I-HaND using %NHF

Supplementary file I

Percentage of Normal Hand Function

Participant Identification Number:	Baseline / Follow-up 1 / Follow-up 2

Please read the following statement:

"A normal hand is one which is pain-free, with a full range of movement, normal strength, dexterity and sensation, and allows you to do what you feel your hand, if normal, should allow you to do. A normal hand is scored as 100 percent, while a completely useless hand is scored as 0 percent. Overall where would you rate your hand between 0 and 100 percent, at this present time"

Percentage of Normal Hand Function	%

PLEASE PROVIDE THE DATE THAT YOU COMPLETED THIS FORM HERE: / /

Supplementary file II:

Characteristics of study sample from Phase 1 study (interviews)

Diagnosis	age	sex	Duration (months)	Work status	Intervention
Median Nerve Injury	59	M	34	Light manual	1^0 repair
	63	F	28	Volunteer	1^0 repair
	26	M	35	Skilled manual	1^0 repair
Ulnar nerve injury	74	M	47	Skilled manual	1^0 repair
	25	M	25	Heavy manual	1^0 repair
	66	M	7	retired	1^0 repair
	26	F	24	retail	graft
	62	F	72	Skilled manual	decompression
Ulnar nerve compression (UNC)	59	M	58	retired	transposition
Combined UNC and CTS	71	F	60	retired	Transp/decomp
Carpal tunnel syndrome (CTS)	71	F	108	carer	decompression
	56	F	39	retail	decompression
Radial nerve palsy	57	M	44	managerial	conservative
	61	F	52	Office work	decompression

Supplementary file III:

Examples of illustrative quotations from patients for the overall endorsement, content, response categories, instructions, layout and time required to complete the I-HaND Scale

Patient endorsement categories	Examples of illustrative quotations from patients
Overall endorsement	"It's simple to use, it's simple to understand, I don't really think it needs changing".
	"It's nicely set out, it's easy to read, it's easy to mark and it covers everything that should have been asked".
	"I didn't have any trouble answering the questions".
	"I didn't have to think twice about any of the questions".
	"I think it is more simple and straight forward than the majority of questionnaires you get at the hospital".
	"You would think that it was made for me to be honest".
	"Everything in there was what actually occurred and what I have been through".
Content	"One question I like in particular was the question about emotions".
	"Nobody asks about that and you do feel these emotions because you have lost part of you, lost part of the use of you, so you get very frustrated".
	"It seems to cover everything that affects me".
	"As I said it is more or less designed for me that one".
	"It covers everything that should be asked or should have been asked".
	"It's very impressive, I like the way it is all everyday tasks that are being asked about".
Response categories	"I thought it was really good, especially the range of answers. You've got five choices as opposed to three and you can really pin it down".
	"I think it is well thought out; the range of answers".
Instructions	"The instructions are self-explanatory".
	"It was pretty easy to follow, it was good".
Layout	"The layout is lovely, it is fine, I can't pick any holes in it really".
	"The print is a decent size which makes a change for us old people".
	"I like how you have greyed out every other line to make it easier to follow across".
Time frame	"It isn't that long; I've had a lot longer ones to complete".
	"It's quite short really".

Supplementary file IV:
Development of the content of the items of the I-HaND Scale with changes made highlighted in bold

Item at pre-test	Round 1	Round 2	Round 3
How well did your hand(s) work?	No change	No change	Retained
The movement of your hand(s)	No change	No change	Retained
The sense of touch in your hand(s)	No change	No change	Retained
The strength in your hand(s)	No change	No change	Retained
I can't grip or pinch for very long without my hand getting tired	No change	No change	Retained
When I touch certain things it causes pins and needles or tingling	No change	No change	Retained
I have hurt my hand and not realised it until later	No change	No change	Retained
When I go to grab something it just falls out of my hand	**Revised**	No change	Retained
Using my hand(s) can bring about strong emotions e.g. frustration, anger, sadness	No change	No change	Retained
I feel self-conscious if people look at my hand/arm	No change	No change	Retained
The pain in my hand(s) has been (…)	No change	No change	Retained
How often would you say that your pain impacts on your daily routine?	No change	No change	Retained
I am sensitive in my hand and do not like it to be touched	**Revised**	No change	Retained
I feel discomfort or pain in cold weather or when handling cold objects	**Revised**	No change	Retained
It is difficult to get a good night's sleep because of the pain in my hand/arm	No change	**Revised**	Retained
How well have you been able to carry out your daily routine, e.g. getting ready, cooking, childcare etc.	No change	No change	Retained
Doing up buttons	No change	No change	Retained
Cutting food using a knife & fork together	No change	No change	Retained
Cutting your nails	No change	No change	Retained
Washing your body	No change	No change	Retained
Putting toothpaste on a toothbrush	No change	No change	Retained
Getting dressed or undressed	No change	No change	Retained
Opening lids of tight jars and bottles	No change	No change	Retained
Pouring from a kettle	No change	No change	Retained
Carrying a heavy shopping bag	No change	No change	Retained
Wringing out a cloth	No change	No change	Retained
Preparing a meal	No change	No change	Retained
Opening & closing heavy doors	No change	No change	Retained
Handwriting	No change	No change	Retained
Turning pages of a book, magazine or newspaper	No change	No change	Retained
Handling small coins e.g. 5 pence or 1 pence	No change	No change	Retained
Using electronic devices e.g. a remote control, mobile phone, tablet or computer	No change	No change	Retained
How well have you been able to manage the physical demands of your daily work?	No change	No change	Retained
How well have you been able to take part in recreational tasks, e.g. hobbies or sport?	**Revised**	No change	Retained
Driving a car			**Added**

Supplementary file V: Summary of item-revision process (Phase 2), with changes highlighted in bold

Items with poor fit	Reason for selection	Decision
Q1: How well did your hand(s) work?	≥ 0.9 item-total correlation ≥ 0.9 inter-item correlation	Retained
Q2: The movement of your hand(s)	≥ 0.9 inter-item correlation	Retained
Q12: I have hurt my hand and not realised it until later	≥ 50% no. 1 responses (floor effect)	Retained
Q16: How well have you been able to carry out your daily routine e.g. Getting ready, cooking, childcare etc.	≥ 0.9 item-total correlation	Retained
Q17: Washing your body	≥ 50% no. 1 responses (floor effect) ≥ 0.9 inter-item correlation	**Removed**
Q20: Putting toothpaste on a toothbrush	≥ 50% no. 1 responses (floor effect)	Retained
Q24: Pouring from a kettle	≥ 50% no. 1 responses (floor effect)	Retained
Q33: Driving a car	≥ 50% no. 1 responses (floor effect) ≥ 5% missing item Written comments from participants	**Removed**
Q28: Handwriting	Written comments from participants	**Removed**
Q18: Getting dressed or undressed	≥ 50% no. 1 responses (floor effect) ≥ 0.9 inter-item correlation	Retained
Q26: Preparing a meal	≥ 0.9 inter-item correlation ≥ 0.9 item-total correlation	Retained
Q25: Wringing out a cloth	≥ 0.9 inter-item correlation ≥ 0.9 item-total correlation	Retained
Q27: Opening & closing heavy doors	≥ 0.9 inter-item correlation	Retained