Patient-Reported Outcome Measurement in Patients with Hand Conditions

by

Dr Kyra Linda Sierakowski

MD, BApp Sci, MS

Thesis submitted to Flinders University

for the degree of

Doctor of Philosophy

College of Medicine and Public Health

Flinders University

Adelaide, South Australia, Australia

25th February 2020

Dedication

Content removed for privacy reasons.

Summary

This thesis begins with a broad literature review to explore outcome measurement in the field of hand surgery. Of primary interest is the measurement of patient-reported outcomes and the instruments available in the field to do so. Chapter 2 reports the systematic literature review of the available patient-reported outcome measures (PROMs) for hand conditions, which was performed to establish the direction of further investigation. All relevant PROMs were identified and the development methodology used for each instrument was compared with the guidelines set forth by the Scientific Advisory Committee of the Medical Outcomes Trust (Lohr 2002). Chapter 3 reports on a clinical study performed with existing PROMs: the Michigan Hand Questionnaire (MHQ) (K. Chung et al. 1998), the Patient-Rated Wrist/Hand Evaluation (PRWHE) (MacDermid 1996) and the Disabilities of the Arm, Shoulder and Hand questionnaire (DASH) (Hudak et al. 1996). This study examined the acceptability of PROMs to hand clinic patients waiting for their clinical review appointment. Chapter 4 uses the participants' responses to the DASH (from the previous study in Chapter 3) to explore whether the DASH complies with the Rasch measurement model.

The thesis then progresses with the development of a new PROM for hand conditions, the HAND-Q. Chapter 5 documents an international qualitative study involving 62 in-depth patient interviews with Canadian and Australian patients. The approach used was interpretive description (Thorne, Kirkham and MacDonald-Emes 1997), which acknowledges the current clinical knowledge that forms a framework for the qualitative work. Interviews were audio recorded and transcribed. The line-by-line analysis resulted in the development of a conceptual framework, which was used to guide the development of each of the scales of the HAND-Q. Chapter 6 details the process of item generation, which used quotes from participants, preserving their phrases and wording as much as possible to create the items for each scale of

the HAND-Q. The resulting instrument is composed of 20 independently functioning scales, of which 10 are outcome scales and 10 patient experience/process of care scales.

The initial drafts of the HAND-Q underwent a process of content validation using cognitive interviews with an international sample of patients. Feedback was gained from an international sample of professionals in the field of hand conditions or psychometrics. On the basis of this feedback, the HAND-Q was further refined, with all changes discussed with patient participants. The HAND-Q is currently being translated and culturally adapted to allow for international field testing, as detailed in Chapter 7. The HAND-Q field test is to be carried out in nine countries speaking seven different languages. To confirm the HAND-Q scales were performing as intended, a preliminary Rasch analysis was performed on scales of the HAND-Q with adequate data from the Australian and Canadian field-test sites. Chapter 8 shows this preliminary analysis. The results suggest that seven scales are supported by the Rasch model. Further Rasch analysis will be performed at the completion of the international field test to finalise the HAND-Q scales. Chapter 9 shows the planned Phase III development to establish the psychometrics of the HAND-Q further. This work is ongoing and will be published separately to this thesis.

Declaration

I certify that this thesis:

1. does not incorporate without acknowledgment any material previously submitted for a degree or diploma in any university; and
2. to the best of my knowledge and belief, does not contain any material previously published or written by another person except where due reference is made in the text.

Dr Kyra Sierakowski

Acknowledgements

I sincerely thank my long-time mentor and primary supervisor Dr Nicola Dean for her encouragement, unwavering support and timely responses. Without her encouragement, I would never have embarked on this journey. Dr Dean has been instrumental to the development of the HAND-Q and the completion of my doctorate. My thanks to my adjunct supervisor Dr Philip Griffin who has been a great supporter of my research, skills as a clinician and career development in general. I would also like to thank my associate supervisor Professor Gregory Bain for his encouragement to focus on the goal and avoid distraction and his valued input from the orthopaedic perspective. Thank you for consistently challenging me to improve. Thanks to Professor David Watson for his oversight and valued input.

I am incredibly grateful to have had the opportunity to work with the Q-team who are responsible for the Q-PROM series. Thank you to Dr Andrea Pusic, who I originally approached to collaborate with developing the HAND-Q. Thank you to Dr Stefan Cano for helping me to understand many of the complexities of Rasch Analysis, particularly for your guidance in Chapter 4. Special thanks to Professor Anne Klassen for hosting me in her home in Canada to work one-on-one with me to develop the item bank for the HAND-Q (Chapter 6), and later meeting with me in Dublin to perform the preliminary Rasch Analysis together (Chapter 8).

Another driver of this research has been Dr Don Lalonde; thank you for your support for this work. I appreciate you hosting me in Saint John and allowing me to interview your patients, participating in the field test and making many of the introductions that allowed for the HAND-Q field test to truly be an international study.

Kathleen Evans Sanchez has been a valued contributor to the development of the HAND-Q; thank you for your efforts with the systematic review (Chapter 2), transcribing hours of interviews and performing initial coding of the qualitative data (Chapter 5). Thank you also for your efforts with participant recruitment for the field test at Noarlunga Hospital and HAND-Q data entry.

Also involved in the HAND-Q field-test data entry were Shabnam Torabiardakani, Nicola Foura, Susan Aldhous and Karen Burford. Thank you for your time and effort ensuring that the field-test data were entered in an accurate and timely fashion.

Chapter 3 documents a study that would not have been possible without the efforts of Dr Mehka John and Dr Riche Mohan. Mehka and Riche helped with the collection of patient responses to the questionnaires and entry of this data. Also integral to this study was Philippa van Essen, who provided support with the development of an access database and scoring algorithms for the patient reported outcomes included in the study.

Much of the data collected for the studies in Chapters 3, 4 and 8 were collected at the Plastic and Reconstructive Surgery and Orthopaedic Hand Clinics at Flinders Medical Centre and Noarlunga Hospital. My thanks also go to the nursing staff and administrative staff of these clinics. Of note is Tracy Smart, who assisted with recruitment of participants for the HAND-Q field test. Thanks also to the staff and volunteers at the Plastic and Reconstructive Surgery Hand Clinic at the Royal Adelaide Hospital who were involved in the HAND-Q field test.

All the staff at the Department of Plastic and Reconstructive Surgery at Flinders Medical Centre have been incredibly helpful in this work. Thank you to Suzanne Patten, Andrea Smallman, Tamara Crittenden and the rest of the team for your support.

International collaborators Professor Robert Murphy and Dr Mamtha Raj were pivotal in the inclusion of the United States of America in the cognitive interviews (Chapter 6). My thanks also to the numerous international field-test sites and those involved in translation and cultural adaptation of the HAND-Q. All your efforts will help to ensure that the HAND-Q is an instrument that will serve the global community.

I would like to acknowledge the financial supports that have allowed me to pursue this research:

- 2016, 2017, 2018: Australian Government Research Training Program Scholarship
- 2016: The Flinders Clinician's Special Purpose Fund
- 2017: Flinders University Research Scholarship
- 2018: Flinders University Research Scholarship
- 2018: Royal Australasian College of Surgeons – Small Project Grant
- 2018: Australasian Foundation for Plastic Surgery – Plastic and Reconstructive Surgery Scholarship.

My sincere thanks to those who were kind enough to help edit and proofread this document. I am aware of the invaluable nature of time and appreciate you giving your time to help me produce a quality thesis.

Lastly, I would like to thank my family and friends for their patience and encouragement. Of special mention is my husband Ben, who has been amazing, thank you for support, belief in me always and facilitating my completion of this work. Thanks also to my PhD babies, Elliott and Frances, who are my greatest achievement during this candidature.

Contents

List of Tables

List of Figures

List of Abbreviations

ADLs	Activities of Daily Living
AUSCAN	Australian/Canadian Osteoarthritis Hand Index
BCTS	Boston Carpal Tunnel Syndrome Questionnaire
CAT	Computerised Adaptive Testing
CTS	Carpal Tunnel Syndrome
CTS-6	6-Item Carpal Tunnel Syndrome Scale
CTT	Classical Test Theory
DASH	Disabilities of the Arm, Shoulder and Hand
DIF	Differential Item Functioning
EFA	Exploratory Factor Analysis
FDA	Food and Drug Administration
GA	General Anaesthesia
HAT	Hand Assessment Tool
ICHOM	International Consortium for Health Outcomes Measurement
IRT	Item Response Theory
JHFT	Jebsen Hand Function Test

JIA	Juvenile Idiopathic Arthritis
M2DASH	Manchester-Modified DASH
MAM-16/32	Manual Ability Measure–16/32
MASS07	Modern Activity Subjective Survey of 2007
MIC	Minimal Important Change
MID	Minimal Important Difference
PEM	Patient Evaluation Measure
POS-Hand/Arm	Patient Outcomes of Surgery–Hand/Arm
PRO	Patient-Reported Outcome
PROM	Patient-Reported Outcome Measure
PROMIS-PF-UE	Patient Reported Outcome Measurement Information System–Physical Function–Upper Extremity
PRWE	Patient-Reported Wrist Evaluation
PRWHE	Patient-Reported Wrist/Hand Evaluation
PSI	Person Separation Index
QOL	Quality of Life
RA	Regional Anaesthesia
REDCap	Research Electronic Data Capture

RMT	Rasch Measurement Theory
ROM	Range of Motion
SDSS	Southampton Dupuytren's Scoring Scheme
TASD	Trapeziometacarpal Arthrosis Symptoms and Disability
TDLs	Tasks of Daily Living
TDX	Thumb Disability Examination
UEFI	Upper Extremity Functional Index
ULFI	Upper Limb Functional Index
US	United States
VAS	Visual Analogue Scale

Chapter 1: Literature Review

1.1 Introduction

> Measurement is the first step that leads to control and eventually to improvement. If you can't measure something, you can't understand it. If you can't understand it, you can't control it. If you can't control it, you can't improve it. (H. James Harrington)

Measurement within medicine is and always has been an integral part of diagnosis and treatment. From initial stages of training, clinicians are taught the utility of measurement, parameters such as height, weight, body mass index, haemoglobin, HbA1c and blood pressure (Hahn et al. 2007). They also learn about the invaluable nature of the patient's story, their symptomology and experience. This information is qualitative in nature and therefore more difficult to measure. With the use of patient-reported outcome measures (PROMs), the clinician can now transform qualitative data such as symptomology into quantitative data that can be used to track patient progress.

PROMs are a way of measuring outcome from the patient's perspective. This type of outcome measurement offers unique information about the issues that are of value to the patient. This information is especially valuable when the goal of treatment is to improve the patient's quality of life (QOL).

The aim of this chapter is to summarise the peer-reviewed literature on the measurement of outcomes in the field of hand surgery, PROMs that exist in hand surgery, the theoretical basis of PROMs and the characteristics needed for the clinical application of PROMs. This introduction will form the basis for the body of the thesis that will identify and critically appraise currently available PROMs relevant to hand conditions, explore the acceptability of PROMs to patients, examine the Disabilities of the Arm Shoulder and Hand questionnaire using

Rasch analysis and ultimately perform qualitative studies to develop a new PROM for hand conditions, the HAND-Q.

1.1.1 The Importance of Hands

Hands are complex physical structures with intricate anatomy that allows for the execution of delicate functions such as drawing, complex movements such as speed typing and strength such as exhibited in rock climbing. The opposing thumb is credited with being one of the major contributors to humankind's development beyond other animals within the mammalian realm (Alpenfels 1955). Hands are more than a physical tool; they are highly visible and an important communication asset (Alpenfels 1955).

Communication using touch is basic human requirement, from birth mammalian offspring require touch for comfort (Argyle 2010, 215). To touch and be touched are significant parts of the human emotional and sexual experience. Fisher and colleagues found that even fleeting interpersonal touch between strangers in a public setting lead to improvements in the individuals affect (Fisher et al. 1976). Hands are used to emphasize verbal communication and to replace it for those with difficulty hearing. Argyle discusses the importance of hand gestures in various cultures, many gestures are similar across languages and cultures (Argyle 2010, 53).The appearance of hands often tells a story about the person they are attached to, with visible indications of age, vocation and even socioeconomic status of the individual (Andersson 2011 and Staples 2003).

Hands hold important cultural significance in various communities around the world. In Thailand it is important to be able to place the palms together with straight fingers in greeting, in parts of Papua New Guinea fingers are amputated as a demonstration of grief (Warren 1972). In places where leprosy is the cause of much disability and social stigmatisation, the "claw hand" is claimed to be the hallmark sign associated with the condition (Warren 1972). Hand

difference refers to hands with structural anomaly from conventional anatomy and this can have significant effects of the psychological well-being of an individual. A study examining the degree of severity of hand difference and the association to self-concept found that the severity of deformity does not correspond to greater issues with self-concept; rather those with milder differences reported more psychological and social effects (Andersson 2011). The authors hypothesised this to be due those with milder hand differences attempting to be hide their disability or having expectations of being disability free, where as those with severe hand differences cannot hide their condition (Andersson 2011). Like the face, hands are highly visible, but unlike the face they are often in the field of view of the individual, providing a constant reminder of the perceived aesthetic issue (Johnson 2015).

1.1.2 The Specialty of Hand Surgery

The field of hand surgery as a specialty emerged after the Second World War from the necessity of managing limb-injured soldiers (Meals and Meals 2007). Because of the availability of rapid evacuation, major reconstructive facilities were developed in the United States for the treatment of hand and maxillofacial injuries (Mathes and Hentz 2006). Omer gives a historical account of the events leading to the development of hand surgery in the United States (US) in the wake of the Second World War. The volume of hand reconstruction required at this time in the US alone was estimated to be over 88 000 people (Omer 2000). The pioneers of Western hand surgery originated from various fields: general surgery, plastic and reconstructive surgery, and orthopaedic surgery (Glickel 2004). The father of hand surgery is widely regarded as Sterling Bunnell, who was a general surgeon who practised the principles of plastic surgery (Mathes and Hentz 2006).

In 1992, Sir Benjamin Rank published his recount of the early history of hand surgery in Australia (Rank 1992). He recalls the initial evolution of hand surgery from the traumatic war

injuries and the gradual shift post-war to hand injuries caused in the industrial setting. With the advent of the operating microscope, the field expanded to include the surgical management of congenital hand deformities and those secondary to chronic disease.

Current hand surgery services in Australia continue to be provided by surgeons from the previously listed three specialities, with the division of the workload varying depending on local conventions. In more recent times, hand surgery is increasingly being recognised as a specialty area in its own right, and the American Medical Association has included hand surgery on its list of designated specialties since 1975 (Omer 2000). The Australian Hand Surgery Society offers a Post-Fellowship Education and Training Programme to formally accredit Australian hand surgeons (McCombe 2018).

1.1.3 Evolution of the Practice of Hand Surgery

Hand surgery developed as a specialty in the US during the Second World War under the guidance of Sterling Bunnell (Burke, McGrouther and Smith 1990; Mathes and Hentz 2006). Technical improvements rapidly exploded as surgeons, working with instruments, sutures and precision operating microscope companies, pushed forward the boundaries of precise tissue repair (Jabaley 1981). The field was transformed by the availability of magnification, improved visual detail that allowed for a more precise surgical intervention (Burke, McGrouther and Smith 1990). The operating microscope allowed complex tissue reconstruction using a patient's tissue from remote areas and the replantation of divided parts (Mathes and Hentz 2006).

The field of surgery, in general, has evolved greatly from a time when patients were kept in hospital for prolonged periods of observation to the current day where patients are often discharged the same day that their surgery is performed. Upper limb surgery can often be performed under local or regional anaesthesia (RA) using nerve blocks with or without the use of sedation or general anaesthesia (GA). The benefits of nerve blocks are numerous and include

postoperative pain relief, reduced risks when compared with GA, reduced length of hospital admission, economic savings and environmental benefits (Hustedt et al. 2017; Merle and Dautel 2016). In addition, the analgesia provided by regional nerve blocks allows for early passive mobilisation, which can assist with the patient's return of function (Merle and Dautel 2016). In those patients who have GA in addition to a regional nerve block, the safety of the anaesthesia is improved with reduced risk of regurgitation and aspiration (Merle and Dautel 2016).

According to Merle and Dautel (2016), the majority of upper limb surgery can be completed with only RA. For these patients, there is the advantage of fewer side effects postoperatively, including drowsiness, respiratory distress, nausea and vomiting (Hustedt et al. 2017; McCartney et al. 2004). This leads to many patients being discharged home earlier, often only hours after the completion of surgery. The prolonged analgesic effect of RA delays the need for pain medication and helps to avoid the side effects of these medicines. Performing surgery without GA has economic savings due to decreased theatre time utilisation, reduced consumables and less invasive monitoring required postoperatively (Rhee et al. 2016).

The field of hand surgery is continually advancing techniques as a result of improved equipment and rehabilitation methods. For example, consider the repair of Zone II flexor tendon injuries (tendon damage within the region of the proximal finger); this anatomical region includes highly complex anatomy, which can result in poor postoperative outcomes (Gibson, Sobol and Ahmed 2016). Historically, this region of the flexor tendons was referred to as 'no-man's land', and the accepted management was to close the skin and later perform tendon grafting (Verdan 1960). In today's practice, it is standard to perform a primary repair. Improved repair techniques and suture material have allowed for controlled motion rehabilitation protocols that are improving outcomes.

There is no doubt that the practice of hand surgery will continue to evolve further in the future. As this service changes, it is important to continually assess the patient's perspective to ensure that we are improving overall patient outcomes.

1.2 Outcomes Measured in Hand Surgery

Operative procedures in the field of hand surgery target functional outcomes and improving the quality of an individual's personal and social life (Dubert 2014), but these concepts cannot be directly measured.

1.2.1 Clinical Measurements

Outcomes in the field of surgery have classically been based on clinical measures and the incidence of postoperative surgical complications, morbidity and mortality (Devlin and Appleby 2010). Clinical measures specific to hand function that have been used to assess outcomes include strength (grip, key pinch and tripod pinch), sensory testing (two-point discrimination and monofilament tests), mobility (passive and active range of motion [ROM] for each joint) and radiographic measurements (Macey et al. 1995; Schoneveld, Wittink and Takken 2009). These clinical measures are often poorly aligned with the outcome that is important to the patient. For example, evidence of bone healing on a radiograph does not mean that the patient is able to perform their activities of daily life (Giladi and Chung 2013).

1.2.2 Performance-Based Tests

The Jebsen Hand Function Test (JHFT) was developed in 1969 (Jebsen et al. 1969) to establish a functional assessment of the hand. The test involved observing a patient while they carried out seven standard common activities of daily living (ADLs). The time taken to complete each task was measured and then compared with standard timing measurements. McPhee (1987) questioned the appropriateness of a time measure as the primary measurement unit for hand

function. Following this, the Sollerman Hand Function Test was developed; it considers both the time that is taken to complete a standardised task and the grip quality (Sollerman and Ejeskär 1995). Both of these functional tests are designed to be carried out by an occupational therapist and take approximately 20 minutes per assessment (Hackel et al. 1992; Sollerman and Ejeskär 1995). The JHFT has been widely used as a measure to standardise measurement of hand function; it is appropriate for evaluating interventions and can be used in people with a large variety of conditions affecting hand function (Poole 2011). The Sollerman Hand Function Test has been used as a standardised method for assessment of hand function in quadriplegic patients. For this group of patients, it assists understanding their difficulties completing basic ADLs (Sollerman and Ejeskär 1995).

1.2.3 Clinician-Rated Scales

Several clinician-rated scoring scales have been developed that acknowledge the importance of patient satisfaction and symptoms experienced by the patient in the evaluation of functional outcome. Nakamura and Tamai designed a scoring scale for the evaluation of function following hand and digit replantation (Tamai 1982). This scale is a combination of traditional clinical measurements such as ROM, sensation and functional ADLs. It also has an item on subjective symptoms, cosmesis, patient satisfaction and current job status. The Ipsen scale is a modification of the Tamai classification but with criteria added to allow for evaluation of the wrist, elbow and shoulder (Ipsen et al. 1990; Kamburoğlu et al. 2011).

1.2.4 Measuring Impairment

The American Medical Association (AMA) *Guides to the Evaluation of Permanent Impairment* was developed to provide a standardised, objective approach of evaluating medical impairments. The original edition was published in 1958, with the most recent edition (sixth) published in 2008. Impairment is defined as ‘a loss, loss of use, or derangement of any body

part, organ system, or organ function' (Andersson and Cocchiarella 2001, 2). This document is used as a guide to quantify the impact of various diseases and states of injury on an individual. Its application to the upper extremity assessment considers only the anatomical impairment and does not take into consideration the functional or cosmetic evaluation because of an inability to measure these concepts in a precise and standardised fashion (Andersson and Cocchiarella 2001). This is suboptimal as the restoration of hand function is the primary purpose of performing hand surgery, which is not necessarily the restoration of normal anatomy. An assessment gives an overall impairment score for the hand, which can then be combined with other regional impairment scores to give an upper limb impairment score. Other medical and surgical issues can be considered and incorporated to reach a whole-person impairment score (Andersson and Cocchiarella 2001).

Although these assessment tools were a step forward, it was evident that the changes in these performance-based hand-scoring tools and the AMA guide may not be correlated with the patient's subjective scores of hand function. In the evaluation of surgical outcome, it has been established that there is a significant amount of subjectivity, which does not necessarily correlate with any objective clinical measurements (Pearl and Belcher 2013). During the 1980s, the World Health Organisation (WHO) developed the *International Classification of Impairments, Disability and Handicap* (WHO 1980, 27), which altered the way that outcome was measured. According to this manual, the terms impairment, disability and handicap were defined in the context of healthcare as follows:

- Functioning: is an umbrella term for body functions, body structures, activities and participation. It denotes the positive aspects of the interaction between an individual (with a health condition) and that individual's contextual factors (environmental and personal factors) (WHO 2001).

- Impairment: ‘Problems in body function and structure such as significant deviation or loss’ (WHO 2001).
- Disability: ‘is an umbrella term for impairments, activity limitations and participation restrictions. It denotes the negative aspects of the interaction between an individual (with a health condition) and that individual’s contextual factors (environmental and personal factors)’ (WHO 2001).

As a result of this holistic approach to outcome, it was no longer satisfactory to solely focus on physical impairments when evaluating outcomes (MacDermid 2005), but also on how that impairment causes activity limitation (disability) and affects the individual’s QOL. Badalamente et al. (2013) stated that for a given degree of measurable impairment, the degree of disability reported by the patient can vary widely.

1.2.5 Measuring Quality of Life

Fitzpatrick et al. (1999) pointed out that the objective of healthcare interventions is to improve the patient’s QOL. If this is the objective of the surgery, then it is only possible to measure the success of the intervention by gauging the change in the patient’s QOL. Measurement of QOL initially emerged in the 1970s as part of the paradigm shift towards holistic medicine and the theory of health being inclusive of physical, mental and social wellbeing (Cano, Klassen and Pusic 2009).

There is often a profound difference in the perception of the patient and that of the clinician (Cano, Klassen and Pusic 2009). This is perhaps most pronounced in the field of aesthetic surgery where improvement, as judged by the patient, is the sole purpose for the intervention (Malay and Chung 2013); however, this phenomenon is present in many fields. For example, a surgeon may be pleased with the aesthetic outcome following breast reconstructive surgery, but the patient might be quite dissatisfied with their post reconstructive appearance. The patient

and the clinician clearly have different perspective from which they consider the outcome. The patient may be disappointed with scarring, lack of sensation and ultimately that their reconstructed breast differs from their original breast (Dean and Crittenden 2016). The clinician often makes their assessment of the aesthetic outcome relative to the surgical difficulty that was encountered intra-operatively. Fitzpatrick et al. (1999) agree that proxy reports of patient's health status and wellbeing do not always agree with the opinion of the patient, and thus the need for the patient to directly report their assessment.

Patients are the best source of information on their own QOL (Hahn et al. 2007). Therefore, to assess an intervention, it is vital to ask the patient for their perspective; thus, the development of PROMs. Measuring an abstract concept such as QOL may initially seem to be unfamiliar territory for many health professionals (Hahna et al. 2007). The reality is that clinicians depend on measurements constantly, such as laboratory results, blood pressure and temperature—all of which are inherently associated with error (Hahn et al. 2007). It has been demonstrated by Hahn et al. (2007) that patient-reported outcome (PRO) data are as reliable as many clinical measurements that are accepted in routine clinical practice. Although once considered subjective and therefore unreliable, these measurement tools are now considered highly relevant to understanding the impact of clinical decisions and treatments (MacDermid 2014). PRO data are by necessity subjective; however, this does not mean it is not quantifiable or reproducible. Nowadays, most clinical trials include a PRO as the primary outcome, with other ratings of impairment or clinical measurements considered secondary outcomes (MacDermid 2014).

1.2.6 Qualitative Studies

Van der Giesen et al. (2010) explored hand function problems experienced by patients with rheumatoid arthritis resulting in swan neck deformity, and their splint preferences. They

identified seven hand function specific concepts: flexion initiation, painful joint hyperextension instability, appearance, small grip activities, big grip activities, application of pressure in activities and comprehensive hand function activities. The patient's preference for splint was based on the effectiveness of the splint, ease of use, appearance, comfort and associated side effects.

Gustafsson, Persson and Amilon (2002) reported on the stress factors and coping methods of patients with an acute traumatic hand injury. A variety of coping strategies were identified, including 'accepting the situation'.

The perspective of patients living with hand osteoarthritis was investigated with focus groups in five European countries by Stamm et al. (2009). This group then compared the concepts important to patients with hand osteoarthritis to the concepts measured by instruments used to measure functioning in this cohort. They found that a third of the concepts were measured in any way by existing instruments. They identified the following concepts as being important to patients and not measured by existing instruments: psychological consequences, different qualities of pain, aesthetic changes and leisure activities (Stamm et al. 2009).

The qualitative study performed for this thesis was informed by the methodology of previous qualitative studies performed for the development of the BREAST-Q, CLEFT-Q and BODY-Q (Klassen et al. 2009; Klassen et al. 2016; Wong Riff et al. 2018). These studies used purposeful sampling to recruit a heterogeneous cohort of patients who represented a broad range of patients in the field of interest. In-depth qualitative interviews were performed using the interpretive description approach (Thorne, Kirkham and MacDonald-Emes 1997). Verbatim transcription was performed to allow for extensive coding to identify common themes that are important to patients.

1.3 Patient-Reported Outcome Measures

The US Food and Drug Administration (FDA 2009, 2) defines PROs 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' (US Department of Health and Human Services FDA Center for Drug Evaluation and Research, US Department of Health and Human Services FDA Center for Biologics Evaluation and Research and US Department of Health and Human Services FDA Center for Devices and Radiological Health 2006). A PRO may be measured by self-report or by interview, provided that the patient's response is recorded without alteration by the interviewer. PROMs are questionnaires that directly capture the patient's perception of their functional status and general wellbeing (Dawson et al. 2010). Self-report is particularly well suited to evaluate social wellbeing, pain, satisfaction and QOL (Dubert 2014). The outcome measured may be in absolute terms (e.g., the severity of a symptom) or as the relative change from an earlier measurement.

The use of PROMs was initially limited to gauging treatment effectiveness in the setting of clinical trials (Fitzpatrick et al. 1999; Valderas, Alonso and Guyatt 2008; Basch 2014; MacDermid 2014). Increasingly, PROMs are being used more in routine clinical practice to establish the patient's perspective of their care and their resultant outcomes (MacDermid 2014). McGrail, Bryan and Davis (2010) believe that routine measurement of patient-reported outcomes should be instituted in Canadian healthcare. In the United Kingdom, the National PROMs Programme has been functioning since 2009. The programme includes patients who are undergoing four common surgical procedures: total hip replacement, total knee replacement, varicose vein surgery and groin hernia surgery (Black 2013).

Numerous studies have been performed that show PROM instruments are valid surrogates for more time consuming and costly professional clinical evaluations (Hahn et al. 2007;

MacDermid 2014). Stanger et al. (2016) performed a study of 1231 patients undergoing hip or knee joint replacement, comparing their PRO data with objective clinical assessments. They found the results of the different assessments were highly correlated and made recommendations that baseline PROM data should be obtained as an integral part of patient intake.

In the field of plastic surgery, patients are eager for information regarding the likely outcome for varied reconstructive procedures. Information relating to overall satisfaction, physical and social wellbeing, and aesthetic results at varied timepoints post operatively can assist the patient and clinician to make patient specific surgical decisions. PRO data are able to provide evidence on which to base these discussions (Malay and Chung 2013).

The information that PROMs provide is valuable, and health services and clinicians alike are increasingly recognising this worldwide. The US Department of Health and Human Services has provided funding through the National Institutes of Health to develop HealthMeasures, which is the official distribution centre for the Patient-Reported Outcomes Measurement Information System (PROMIS) (Cella et al. 2007). The International Consortium for Health Outcomes Measurement (ICHOM) is a not-for-profit organisation that is developing standard sets of outcomes for specific medical conditions to encourage healthcare providers to incorporate outcome measurement into their clinical practice. With this broad application of PROMs it is important to assess instrument quality to ensure those that are being implemented are appropriate to serve the purpose for which they are being applied.

1.3.1 Generic Patient-Reported Outcome Measures

PROMs are generally divided into those that measure general wellbeing or QOL and those that are targeted for specific regions or disease profiles (Szabo 2001; Black 2013). A commonly used general health status PROM is the Short Form 36 (SF-36), which is used to measure health

perception in a general population (Brazier et al. 1992; Ware and Sherbourne 1992), or the EuroQol (EQ-5D), which is the generic PROM chosen for the National PROMs Programme in England (Black 2013). One of the benefits of generic PROMs is that they allow comparison of results across different health conditions, which is of particular relevance when using PROM information to compare the value of procedures or interventions (Devlin and Appleby 2010). The disadvantage of generic instruments is that they often lack the sensitivity to measure the impact of many interventions.

1.3.2 Regional Anatomy Patient-Reported Outcome Measures

Numerous specialised PROMs exist for specific regions of the body, such as the Breast-Q (Pusic et al. 2009), Disabilities of the Arm, Shoulder and Hand (DASH) (Hudak et al. 1996) and Michigan Hand Questionnaire (MHQ) (K. Chung et al. 1998). According to Badalamente et al. (2013), the appeal of regional PROMs is due to the global nature of assessment that they provide and the ability to compare different conditions that affect the same region.

1.3.3 Condition-Specific Patient-Reported Outcome Measures

A variety of PROMs have also been designed for patients with specific conditions, such as the Boston Carpal Tunnel Syndrome Questionnaire (BCTQ) (Heybeli et al. 2002) and the Southampton Dupuytren's Scoring Scheme (SDSS) (Mohan et al. 2014). Because of their specific nature, these instruments are generally limited to use for patients who have the condition for which they were developed (Badalamente et al. 2013). Condition-specific PROMs are usually more sensitive to changes in the condition that they are designed to measure than generic tools (Badalamente et al. 2013). Mohan compared the sensitivity of the SDSS with that of the QuickDASH (a shortened version of the DASH) (Beaton et al. 2005) in a cohort of 61 patients pre- and post-surgical intervention for Dupuytren's contracture. It was found that the SDSS is a more sensitive instrument for this application (Mohan et al. 2014).

1.3.4 Patient-Reported Outcome Measures in Hand Surgery

In recognition that hand surgery procedures primarily aim to improve function and QOL, there have been many PROMs developed for use in this cohort. There are those that measure the whole upper limb as a single functional unit, such as the DASH (Hudak et al. 1996). Other PROMs are specific for the hand, such as the MHQ (K. Chung et al. 1998). There are also a variety of disease-specific PRO tools such as the previously mentioned BCTQ (Heybeli et al. 2002). A systematic review to identify and review both the development and validity of existing PROMs in the field of hand surgery will form the following chapter.

1.3.5 Work Compensation and Patient-Reported Outcome Measures

Spearing and Connelly (2011) performed a systematic meta-review examining the published evidence surrounding the question 'Is compensation "bad" for health?'. They found that in general the evidence published was of poor quality with the heterogeneity of compensation cover and outcomes measured making overall analysis difficult. In their review of 11 systematic reviews, only one study was identified as a high-quality study. This paper is a systematic review of the prognostic factors associated with delayed recovery post-whiplash injury; they found strong evidence of no association between litigation and worse health after whiplash injury (Scholten-Peeters et al. 2003).

Compensation status has been associated with poor outcome following surgery. In a meta-analysis performed by I. Harris et al. (2005), from 211 studies included in the analysis, a total of 175 (83%) concluded that the presence of compensation was associated with a worse outcome. Sub-analysis by geography suggested that this association was weaker in Australian studies compared with European studies. In the field of hand injuries, Wong (2008) published a study examining the factors that influenced the amount of time off work for people suffering

hand injuries in Hong Kong. She found that the severity of the injury, number of operations and the presence of a compensation claim were predictors for increased length of time off work.

A systematic review and meta-analysis by Fujihara et al. (2017) and colleagues examined the effect of worker's compensation on outcome measurement methodology in patients following upper extremity surgery. They found that patients without compensation were significantly more likely to show satisfactory improvement after surgery than those with insurance (OR 3.17, 95% CI 2.47–4.08). This trend was present regardless of the technique utilised to measure the outcome. However, its effect was not uniform between different outcome measurement methods. The disparity in outcome was greater between the insured and uninsured groups when the outcome was measured using PROMs compared with functional measures such as the arc of motion and grip strength (Fujihara et al. 2017). They suggested that results should always be measured both preoperatively and postoperatively to minimise the bias effect of workers compensation. The effect of people with workers compensation having worse postoperative outcomes was independent of the country from which the study originated. This suggests that the effect is not due to malingering or feigned impairment for financial incentive but a reflection of the psychological effects of a workplace injury (Fujihara et al. 2017).

1.3.6 Theoretical Underpinning of Patient-Reported Outcome Measures

DeVellis (2006, 2016) defines 'scales' as measurement instruments that are a series of questions (often termed items) that are intended to measure the level of a theoretical (or latent) variable that is not able to be directly observed. Examples of latent traits that are often measured this way include depression, anxiety and pain. He goes on to clarify that a 'scale' is distinct from an 'index' in that the items on a scale share a common cause (the latent trait) whereas an index is composed of various separate characteristics that do not have a common cause but may share an effect (DeVellis 2016).

Feinstein (1987) discusses several common indexes that are used frequently in clinical practice, such as the Apgar score (Figure 1.1), which is used to rate the clinical condition of a newborn baby and the TNM index (tumour, nodes, metastasis), which is used to stage cancer. The individual items that make up the scale are not necessarily from the same cause; for example, a person can have a metastatic disease without positive lymph nodes, but all contribute towards the ultimate outcome for the patient. Indexes are tools utilised by clinicians to determine prognosis or treatment. PROs are usually classified as 'scales' rather than indexes because they purport to measure a single underlying construct.

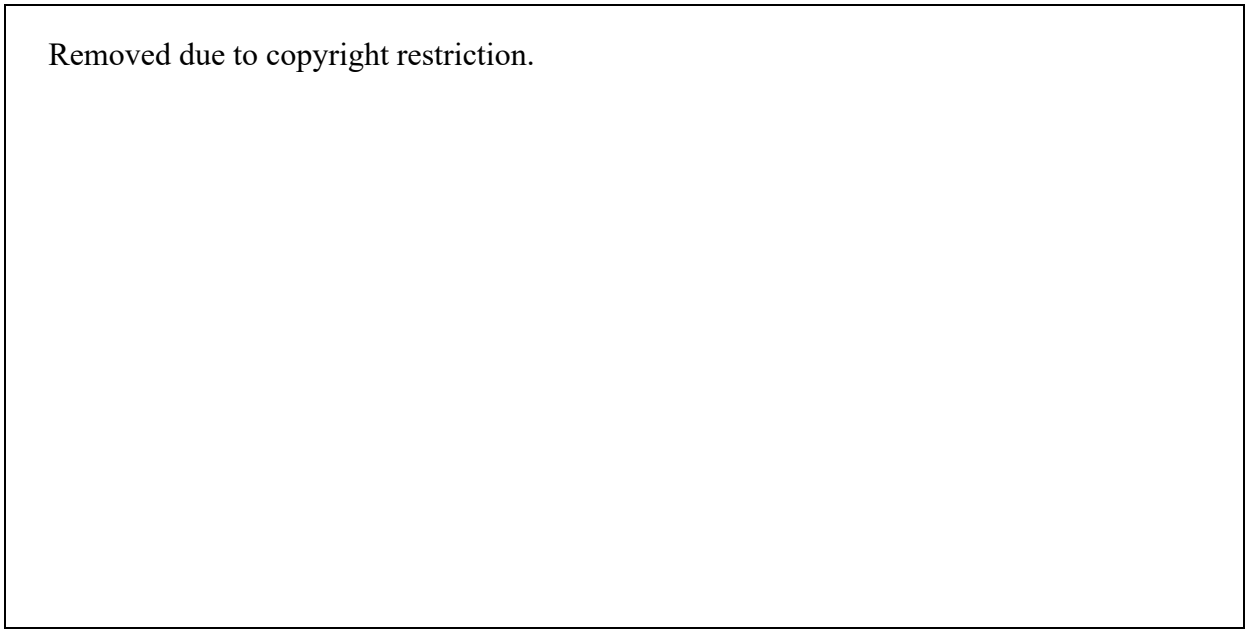

Figure 1.1: Apgar scoring system. A clinical index used to record the clinical condition of a newborn baby (Source: http://www.emsworld.com/article/10615556/apgar-scoring-of-newborns)rd the clinical condition of a newborn baby (Source: http://www.emsworld.com/article/10615556/apgar-scoring-of-newborns).

1.3.6.1 Classical Test Theory

The practice of making inferences about a construct that is not directly observable is an imperfect process, and therefore instruments that are used as proxies for the unobservable

variable will have a degree of error (DeVellis 2006). Spearman (1904) is credited for the early theories of what is now known as classical test theory (CTT) (also called classical measurement theory) (DeVellis 2006). This theory is the basis of many PROMs and is used as the reference point for instruments developed with other techniques (DeVellis 2006; Cano, Klassen and Pusic 2009). PROMs are questionnaires (often referred to as instruments) to measure constructs that cannot be directly measured; they act as a proxy for variables that cannot be directly observed. CTT is a set of concepts and techniques that determine how accurate the proxy indicator is in estimating the variable of interest (DeVellis 2006).

An assumption of CTT is that the observed score is determined by the true state of the variable of interest plus error introduced by the effect of other variables (Hays, Morales and Reise 2000; DeVellis 2006). The error is not differentiated into any subcategories but collated into a singular error entity (DeVellis 2016). According to CTT, the more items that are on a scale, the more reliable that scale is (DeVellis 2006). This results in lengthy scales with multiple items that are highly similar. This is a disadvantage of this methodology as long scales are undesirable because of the burden placed on the respondent (DeVellis 2006; Yafef et al. 2015).

There is a large amount of literature on the measurement theory of CTT, and this is an advantage for the technique as most researchers in the field will be familiar with CTT (DeVellis 2006, 2016). The majority of PROM scales that are available and the papers validating them are based on CTT principles (DeVellis 2006; Cano, Klassen and Pusic 2009).

Cano, Klassen and Pusic (2009) want clinicians to be aware of two key limitations of scales derived from CTT: first, that scales developed using traditional methodology produce ordinal data, and second, that these types of scales are only suitable for group comparison studies and are generally not precise enough for use in individual patient care.

1.3.6.1.1 Factor Analysis

Factor analysis is the primary CTT method for assessing the dimensionality of a scale; it may also be assessed using Rasch analysis (discussed later) (DeVellis 2006, 2016). A unidimensional scale is a scale where all items (questions) measure the same underlying variable; this enables the scale to give a true score of this unobservable variable (DeVellis 2006). The benefit of a unidimensional scale is that the construct itself that is be measured as directly as possible, it is not measuring a bi-product of a secondary influence on the construct. This feature results in more clearly interpretable scores than multidimensional scales. An example of the dimensionality of a scale would be a set of items asking about pain. At first observation, the scale may appear to be one-dimensional; however, if a proportion of the items are also asking about fatigue, then the scale is likely to be multidimensional. The characteristics of questionnaire items can mean that similarities to exist beyond that of the common latent variable that they are intended to measure. When scrutinised closely enough, there may be similarities found, but whether these are relevant similarities, requires expert judgement. An example is a unidimensional scale measuring upper limb function; some of the activities in the items may require large forces and other items fine motor skill. Variation is present between these subgroups of items. However, all the items are relevant to measuring the function of the upper limb. The objective of factor analysis is to identify those characteristics for which items differ substantially in a clinically relevant way to affect the dimensionality of a scale (Reeve et al. 2007; DeVellis 2016).

Factor analysis has been performed on existing scales to explore whether the scales are one-dimensional. Rodrigues et al. (2016) performed exploratory factor analysis on the DASH, with findings supporting the presence of two factors, which brings into question the validity of a single DASH score. Similar findings regarding the DASH being multidimensional were found by Lehman, Woodbury and Velozo (2011) and Franchignoni et al. (2010).

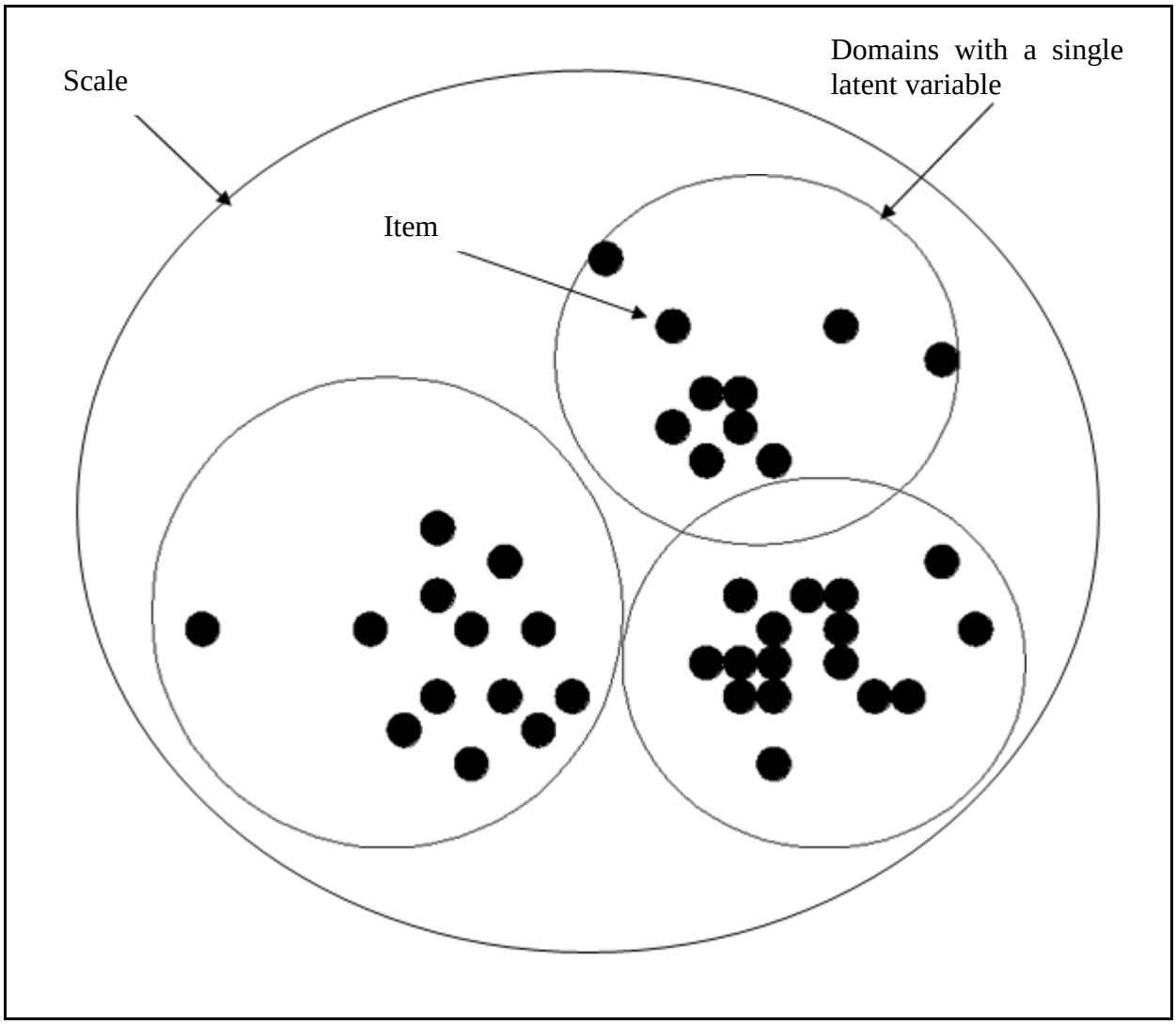

Figure 1.2: Exploratory factor analysis.

1.3.6.2 Item Response Theory and Rasch Analysis

Item response theory (IRT) is a collection of mathematical models and statistical techniques that describe, in probabilistic terms, the relationship between a respondent's answer to a survey question and their level of the latent trait (which is the term given to the unobservable phenomenon) that the PROM is designed to measure (Reeve et al. 2007). In IRT, a model is sought that fits the data being studied in an attempt to explain the data. Rasch analysis differs from this in that the aim is to determine the extent to which the data of interest fits the Rasch measurement model (B. Wright and Linacre 1989; Cano et al. 2011a), that is, whether it is

diagnostic. This subtle difference is not always appreciated, and the terms are often used interchangeably.

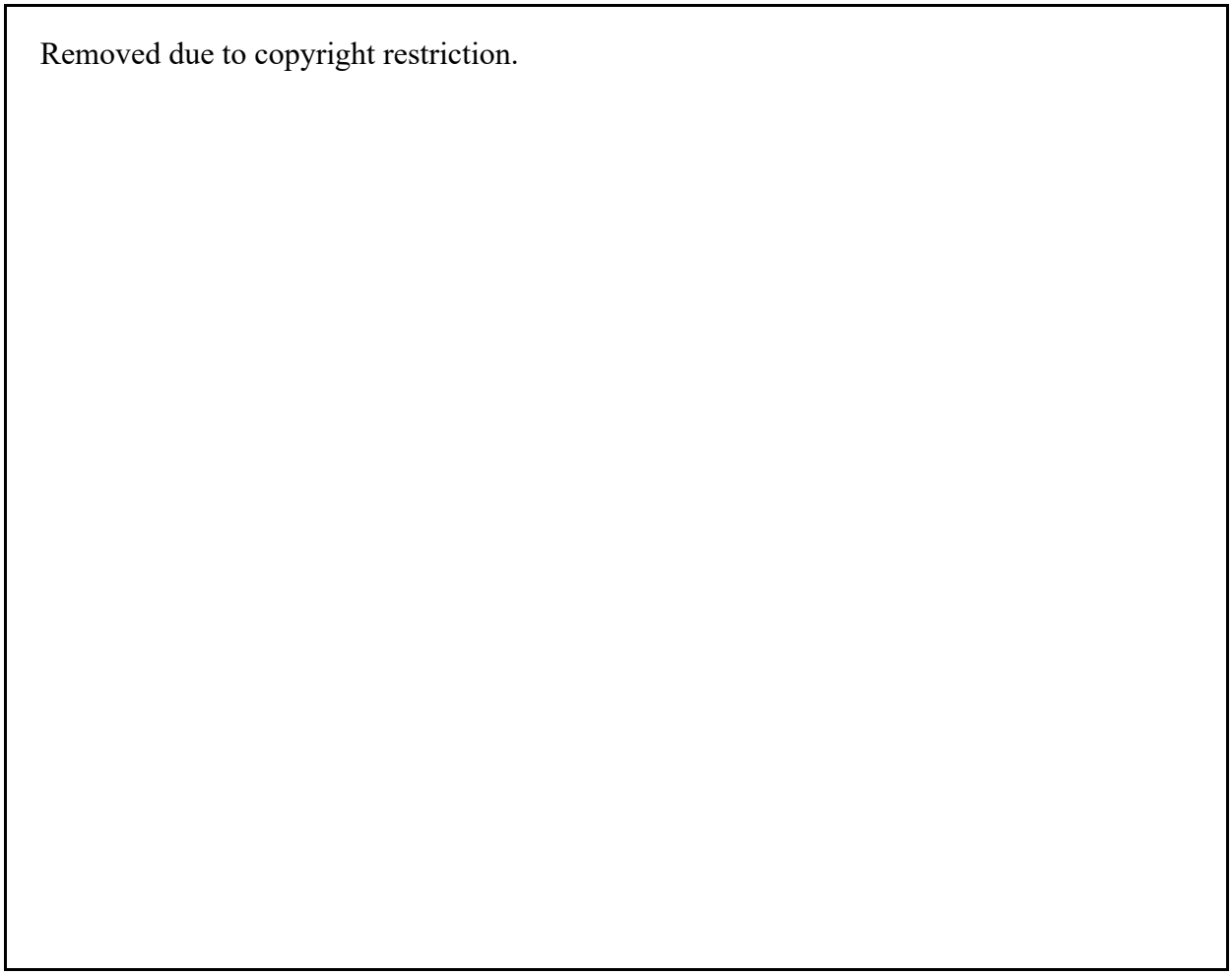

Figure 1.3: Georg Rasch.

Using the IRT approach to scale development involves the application of the Rasch measurement model, often referred to as Rasch analysis (Tennant and Conaghan 2007). The Danish mathematician Georg Rasch (Figure 1.3) was responsible for the development of a mathematical model that shows what should be expected in responses to items if accurate interval scale measurement of an underlying variable is to be achieved (Tennant and Conaghan 2007). Originally designed for use in the field of education testing, the Rasch model is used in many disciplines and increasingly in health sciences research (B. Wright and Linacre 1989; Tennant and Conaghan 2007; Hagquist, Bruce and Gustavsson 2009; Belvedere and de Morton

2010). It is useful for both the development of measurement questionnaires and the evaluation of existing instruments (Tennant and Conaghan 2007).

A fundamental requirement for implementation of the Rasch model is unidimensionality of the scale; if data fit the model then by definition the scale is unidimensional (Belvedere and de Morton 2010). The model constructs a hierarchical ordering of items from easy to hard (using an example of exam questions). If a respondent is able to answer a relatively hard item correctly, then it is highly probable that they can correctly answer the items rated easier. In clinical terms, if a patient has answered that he can walk 100 m, it is highly likely that they will answer that they can also walk 75 m. The model can be used for both dichotomous (yes/no) and polytomous (multiple response options) data. The model assumes that the likelihood of a respondent endorsing an item is a function of that person's level of latent trait and the level expressed by the item.

The measurement unit used in Rasch measurement is the logit (short for log-odds units) (Boone, Staver and Yale 2014). It is this unit that is used to express 'person measures' and 'item difficulties'. Figure 1.4 is an example of a plot of a respondent, Amy, who is placed at +0.25 logits, which is greater than Q8, which is −1.2 logits. This means that Amy will agree with Q8 the majority of the time (Boone, Staver and Yale 2014). Fitting data to the Rasch model with both the item and the person parameter estimates on the same logit scale gives a linear transformation of the raw score and provides interval data (Boone, Staver and Yale 2014).

Removed due to copyright restriction.

Figure 1.4: A plot of both 'person measures' and 'item difficulty' on the same scale.

(Boone, Staver and Yale 2014)

An advantage of the Rasch model is that it provides interval data that can be used for parametric statistical analysis. Tennant indicates that Rasch analysis should be used whenever a set of questionnaire items are intended to be summed together to provide a summed or total score (Tennant and Conaghan 2007; Boone, Staver and Yale 2014).

Overall fit statistics reflect how well the observed data fit the Rasch model. Individual person and item fit statistics are also generated, which allows refinement of the scale. Figure 1.5 illustrates this with an example of a Wright map. This graph shows a plot of the respondents ('person' on the left of the line) and questions ('items' on the right of the line. This plot demonstrates several possible improvements that can be made to this scale. There are redundant items as it seems that questions 8, 12, 14 and 15 are all measuring similar portions of the trait, as are questions 2, 5, 6, 11 and 13. This indicates that removing some of the items within these two clusters could shorten the scale and the scale would not lose any measurement precision (Boone, Staver and Yale 2014). The plot also shows that there would be a benefit in a question of difficulty between Q3 and Q4 to prevent clustering of the respondents at this level (Boone, Staver and Yale 2014).

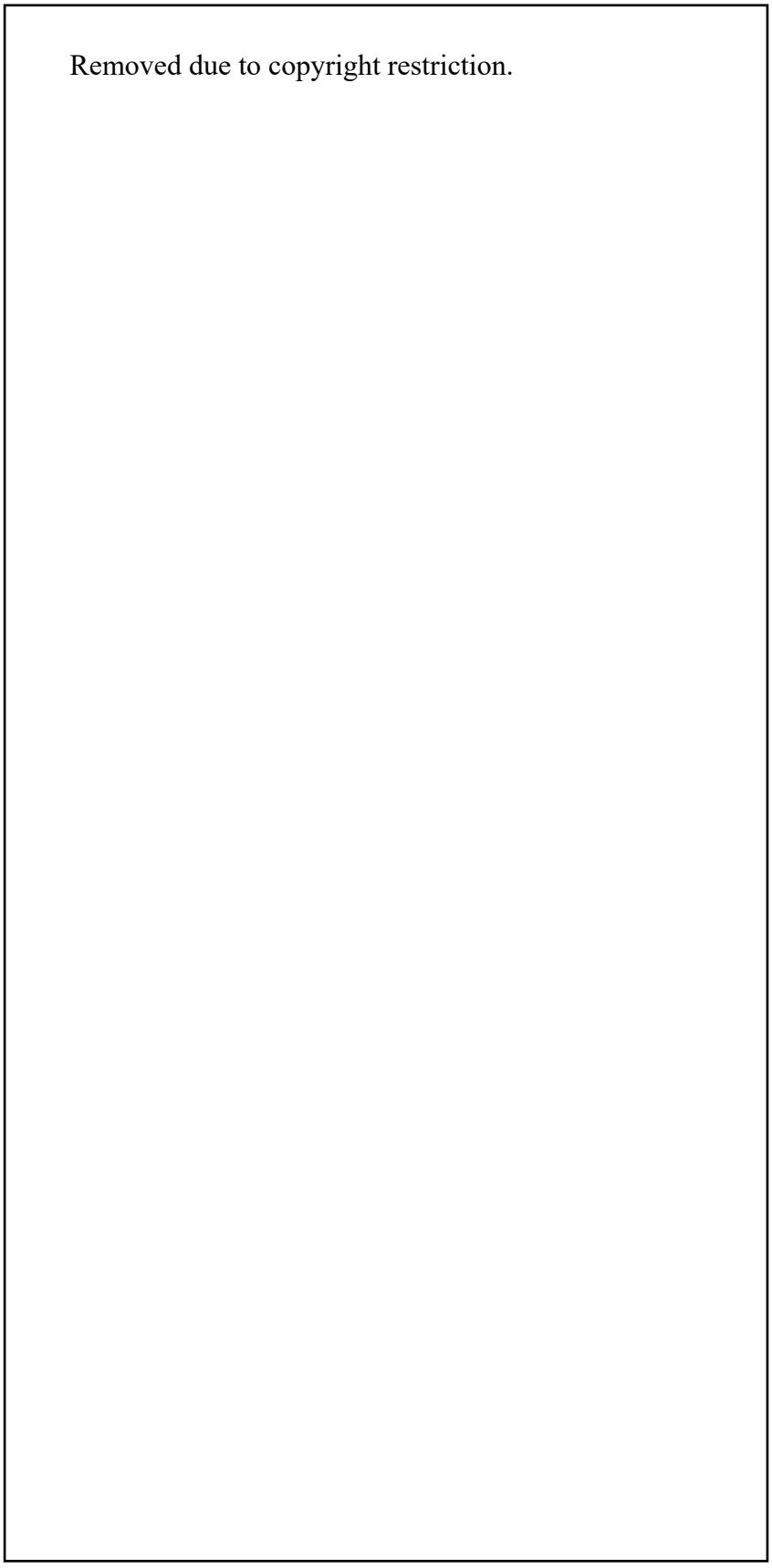

Figure 1.5: Example of a Wright map showing a plot of people (left) and items (right).

(Boone, Staver and Yale 2014)

Rasch models give the item and latent trait estimates that are stable between samples studied, with standard error conditional on the trait level that is linked to item content (Hays, Morales and Reise 2000). Unlike the problems with CTT-derived scales, Rasch-derived scales should function in the same way regardless of which group is being assessed (Pallant and Tennant 2007; Hagquist, Bruce and Gustavsson 2009). Pallant gives the example of an anxiety measurement scale; males and females should have the same probability of affirming an item at the same level of anxiety, that is, the probability of item endorsement is conditioned on the trait. If one gender had a different probability of endorsing an item, then that item would be displaying differential item functioning (DIF) and would defy the requirement of unidimensionality (Pallant and Tennant 2007).

Scales derived from the Rasch model are useful for monitoring change in the healthcare setting. As the trait level can be estimated from a subset of items, it is possible to establish an estimate of the trait at different points over time despite only some of the items being retested (Hays, Morales and Reise 2000). Hays expects that IRT estimates of health outcomes are more accurate than CTT estimates, and thus they should be more responsive to change in health over time (Hays, Morales and Reise 2000).

IRT is being used to complement and increasingly replace traditional methods. Tennant describes the Rasch measurement model as being the standard for modern psychometric evaluations of outcome scales (Tennant and Conaghan 2007). Cano and colleagues performed Rasch analysis on the DASH, the well-established PROM often used in upper limb research (Cano et al. 2011a). Traditional psychometric analysis of the DASH indicated that it was a reliable and valid measure of upper limb disability in patients with multiple sclerosis. Rasch analysis revealed that there were several problems with the DASH in this patient population; subsequently, the validity of its use in this patient cohort was in doubt (Cano et al. 2011a). The issues identified were item misfit, disordered thresholds and high residual correlations between

groups of items (this terminology will be explained further in Chapter 4). To illustrate their argument further, Cano et al. (2011a) used the example that traditional psychometric methods do not provide trustworthy scale evaluations and can result in misleading clinicians regarding the validity and reliability of PROMs. Therefore, to prevent misinformation, Rasch analysis should be used as an adjunct if not the primary method of analysis as it is based on strong measurement theory (Cano et al. 2011a). Only scales developed using IRT psychometric methodology are appropriate for use with individual patients for clinical care (Pusic et al. 2011).

Multiple software programs are available to analyse health outcomes data with IRT methods; however, their use is currently limited because of difficulties learning the programs. Widespread implementation of IRT in the health sector will be made possible by more intuitive software (Hays, Morales and Reise 2000). IRT is also ideal for the implementation of computerised adaptive testing (CAT) (discussed later).

1.3.6.2.1 Differential Item Functioning

An important factor in scales created with CTT is that the properties of the scale depend on the population included in the study. It follows that the comparison of data across different samples is not easily performed. A variety of characteristics of a sample population affect the properties of items and scales, and therefore data collected from different samples are not equivalent (DeVellis 2006). The items of the scale may perform differently according to the group of respondents; this is referred to as DIF. This effect may be uniform (predictable) or non-uniform (unpredictable). Uniform DIF is present where there is a consistent bias of a group's score in a certain direction; note that the predictable nature of uniform DIF allows for mathematical compensation. Non-uniform DIF is where there is variability of the direction and magnitude at various levels of the attribute, and because of the variability, this type of DIF cannot be effectively compensated for mathematically (Pallant and Tennant 2007). There are methods to

identify and allow for uniform DIF based on CTT. However, none as yet are able to solve the problem of comparing between groups (DeVellis 2006).

A further issue that is encountered with CTT-based instruments is that there is no consistent sensitivity across the whole of the score range (DeVellis 2006). This means that often the scales lose sensitivity at the upper and lower limits of the available range. Often it is those responses at either extreme end of the scale that are of most clinical concern; yet, these individuals have scores that are the most difficult to interpret accurately.

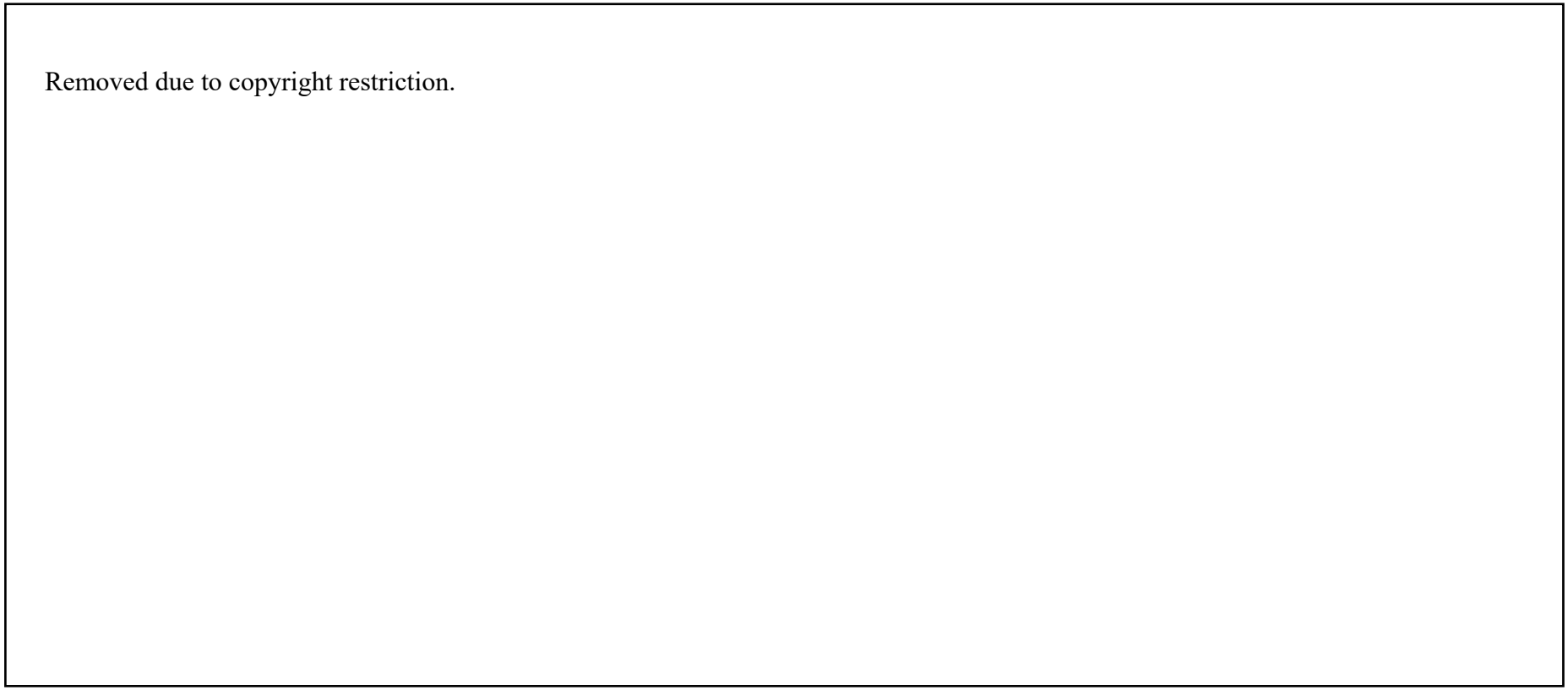

Figure 1.6: Item characteristic curves demonstrating uniform DIF (left) and non-uniform DIF (right).

1.3.6.3 Scale Scoring

Most PROMs result in an overall score of hand function or hand disability. This score is generated from the respondent's answers to each question. The producers of the scale determine the method of reaching the overall score of the scale. It may be a sum, average or another iteration of the item scores. Each PROM should come with specific instructions about how to calculate the overall score.

One of the issues when computing scale scores is the presence of reverse-scored items; these are items that are asking about the same construct as other items but approaching it from the opposite direction. An example of a pair of oppositely scored items might be an item 'I frequently struggle to sleep at night' and another item 'I sleep comfortably at night'. If the response options are 1–5, with 1 representing *strongly disagree* and 5 representing *strongly agree,* a person who has trouble sleeping might answer 5 to the first item and 2 to the second item. In the construction of such a questionnaire, it makes the most sense if a high score represents poor sleep for every item. Therefore, the second item's answer is reverse-scored so that the '2' is recoded to a '4'. This avoids a negatively worded item from cancelling out a positively worded item.

A further issue when creating a total score is whether all items should be treated equally or whether some items are worth more than others. Ideally, all items contribute equally to the measurement of the underlying concept; indeed, this is an assumption of both CTT and the Rasch model (DeVellis 2016). However, in reality, this is difficult to achieve, and there are often differences in item loadings on the relevant concept. Whether the scale developer decides to introduce weighting for items that show higher loading is a personal choice. According to Linacre and Wright (1995), the introduction of arbitrary emphasis of certain items degrades the methodological quality of a scale and biases the measurement in an unreproducible manner.

Many PROMs including the DASH (Hudak et al. 1996) produce a global score. For instruments that measure a single concept, this is a sound practice; however, if the instrument has multiple domains, it is a questionable practice (Dubert 2014). The appeal of a single score is understandable; it is simple, and on the surface, it would seem easy to compare with other global scores. The difficulty is that the summation of scores from different domains to create a global score is oversimplifying the information and in doing so sacrificing the depth of the information (Cano et al. 2011a). Belvedere and de Morton (2010) warn that summing of item scores of a multidimensional scale is misleading.

Many instruments produce ordinal measure scores, which are not suitable for parametric statistics, and nonparametric statistics should be utilised (Tennant, McKenna and Hagell 2004; Boone, Staver and Yale 2014). Despite this, many trials analyse ordinal data with parametric statistics, which is not sound practice (Boone, Staver and Yale 2014; MacDermid 2014). Some instruments have Rasch-based scoring algorithms that were released after the original tool development. Without this conversion from ordinal to interval data, clinicians should be cautious when applying a clinically important difference across the breadth of the scale as there are not necessarily equivalent levels of change between score intervals (Franchignoni et al. 2010; MacDermid 2014). Cano and colleagues agree that the use of ordinal data to measure change is meaningless (Cano, Klassen and Pusic 2009). Figure 1.7 illustrates a ruler with even spaces between each measurement point (representing interval measurement) and a ruler with varied distances between each measurement point (representing ordinal measurement). A measure of six points is consistent on the top ruler, regardless of where those six points fall on the ruler. A “measure” of six points on the lower ruler has a varied length depending on where those six points are measured along the ruler. This demonstrates why the use of ordinal data (the lower ruler) to measure any construct is not accurate when comparing results (either of

different respondents or the same respondent over time) because the distance between measurement points is inconsistent along the length the ruler.

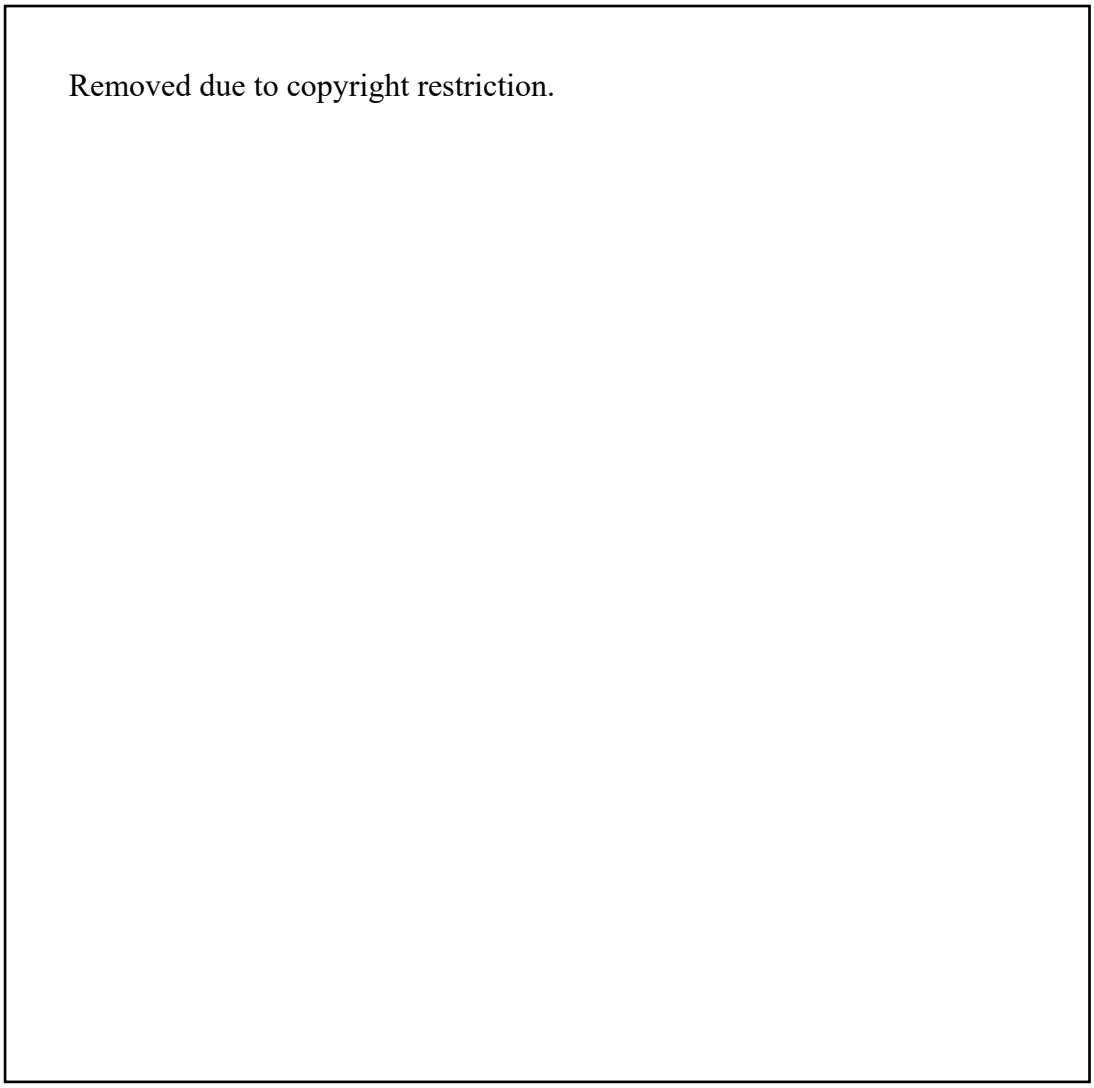

Figure 1.7: Interval vs ordinal data. (Boone, Staver and Yale 2014)

Normative data (also known as 'norms') is a valuable resource but is not readily available for many PROMs. Norms allow interpretation of individual scores and group averages in keeping with what is the normal score for someone in that age and gender group (Ware 1993).

1.3.7 Analysis of Existing Patient-Reported Outcome Measures

In the past, PROM scales were often produced by clinicians or experts within a field without the use of sound methodology (J. Wright and Feinsten 1992; DeVellis 2016). Cano states that

studies that implement ‘ad hoc’ questionnaires are of limited value as this type of questionnaire has not been developed and tested in a meaningful way (Cano, Klassen and Pusic 2009). They go on to advise that good quality PROMs should be developed with input from experts within the field, published literature and the patient cohort that the PROM is aimed to service (Cano, Klassen and Pusic 2009). To appreciate the full breadth of possible QOL issues faced by a patient population, it is important to conduct in-depth patient interviews. These may bring to light issues that experts and clinicians may not have thought were important to the outcome, but if the patient believes them to be influential to their outcome, then they are important to include. The FDA (2009) and Malay and Chung (2013) agree that patient input is vital, and warn that existing instruments cannot be assumed to be valid if patients have not been included in their development (see also US Department of Health and Human Services FDA Center for Drug Evaluation and Research, US Department of Health and Human Services FDA Center for Biologics Evaluation and Research and US Department of Health and Human Services FDA Center for Devices and Radiological Health 2006).

The subjective nature of the constructs, such as functional status and health-related QOL, that PROMs are designed to measure means that there is no way of *directly* measuring for the purpose of comparison (Mokkink, Terwee, Patrick et al. 2010). For example, it is not possible to directly measure someone’s health-related QOL to confirm the accuracy of their SF-36 score. Therefore, before implementing a PROM within the clinical setting, it is vital to ensure that the PROM is a useful tool without inherent bias. The International Society for Pharmacoeconomics and Outcomes Research (ISPOR), Good Research Practices Taskforce, has released a guideline on the methodology of PROM development, which includes qualitative phases to evaluate content validity (Patrick et al. 2011a).

To standardise terminology used to define the measurement properties of health-related PROs, the COSMIN (consensus-based standards for the selection of health measurement instruments)

group performed a Delphi study and published the consensus reached. The Delphi technique is a widely accepted method for gathering a consensus opinion from a panel of experts (Hsu and Sandford 2007). It involves a sequence of questionnaires which are anonymously completed, with multiple iterations and statistical analysis to arrive at a considered consensus opinion from a panel of experts from within the field.

The International Classification of Functioning, Disability and Health (ICF) has developed core sets, which provide a list of ICF categories to describe the functioning of individuals with specified conditions. Two core sets have been developed for individuals with hand conditions. The Comprehensive ICF core set has 117 categories, which covers all aspects of functioning relevant to individuals with hand conditions. The Brief ICF core set (Figure 1.8) has 23 categories and is representative of the minimum standard to describe a patient's functioning (Kus, Oberhauser and Cieza 2012). These ICF classifications assist in selecting an appropriate outcome evaluation tool as the instrument should ideally gather information from all relevant categories. The ICF core sets are not in themselves a measurement method (Dubert 2014).

ICF code	Title
b152	Emotional functions
b265	Touch function
b270	Sensory functions related to temperature and other stimuli
b280	Sensation of pain
b710	Mobility of joint functions
b715	Stability of joint functions
b730	Muscle power functions
b760	Control of voluntary movement functions
b810	Protective functions of the skin
s120	Spinal cord and related structures
s720	Structure of shoulder region
s730	Structure of upper extremity
d230	Carrying out daily routine
d430	Lifting and carrying objects
d440	Fine hand use
d445	Hand and arm use
d5	Self-care
d6	Domestic life
d7	Interpersonal interactions and relationships
d840 - d859	Work and employment
e1	Products and technology
e3	Support and relationships
e5	Services, systems and policies

Figure 1.8: The Brief ICF core set for hand conditions

1.3.7.1 Assessing the Properties of a Scale

1.3.7.1.1 Clinimetrics

The term ‘Clinimetrics’ was originally introduced by Alvan Feinstein to refer to the field of inquiry of indexes, rating scales and other measures used to describe or measure symptoms, physical signs and other clinically relevant occurrences (Fava, Tomba and Sonino 2012). Psychometrics is the field concerned with measuring psychological and social phenomena (DeVellis 2016). Although sometimes used interchangeably, there is varied opinion within the literature about whether this is appropriate. Feinstein’s commentary on ‘multi-item “instruments” vs Virginia Apgar’s Principles of Clinimetrics’ illustrates the differences between the two techniques (Feinstein 1999). In summary, the six principles of clinimetrics according to Feinstein are as follows:

1. variables are selected based on clinical experience and judgement
2. simple scoring (without weighting)
3. heterogeneity of variables
4. ease of clinical usage
5. face validity (common sense appraisal)
6. patients are the source of subjective components.

Clinimetric indexes ask direct, simple questions, with their development based on the clinical experience and judgement of the designing clinician (Feinstein 1999).

1.3.7.1.2 Psychometrics

It is important to appreciate that the validity of an instrument is not established by a single test but rather by a body of evidence (Smith et al. 2012). When PROM development was only in its infancy, the Medical Outcomes Trust created an independently functioning Scientific Advisory

Committee (SAC) to review health status and QOL instruments against rigorous criteria (Lohr 2002). In more recent times, the COSMIN group has developed a checklist of requirements to evaluate PROMs (Mokkink, Terwee, Knol et al. 2010). The following properties were determined to be of importance by that group

- reliability
    - reliability: the stability of a scale assessed by the correlations between repeat administrations of the scale on two occasions (test–retest reproducibility) (Cano et al. 2004)
    - internal consistency (also termed 'internal reliability'): the extent to which items on a scale measure the same construct (J. Wright and Feinsten 1992; Cano et al. 2004). This property is the average intercorrelation among all items; for items with ordinal responses, it is expressed by Cronbach's alpha (α), and for items with dichotomous responses, it is expressed by Kuder-Richardson 20 (KR-20) (J. Wright and Feinsten 1992). This property contributes to the confirmation that the scale is unidimensional with the minimum satisfactory level generally thought to be 0.8 (J. Wright and Feinsten 1992)
    - measurement error: the Standard Error of Measurement (SEM) is a measure of the standard deviation of errors, providing an measure of the accuracy of the scores (Harvill 1991).
- validity
    - content validity: the extent to which the items on the scale capture the concept of the domain that they are intended to cover, assessed by qualitative means during construction of the scale
    - construct validity: the degree to which a scale measures what it aims to measure

        - structural validity:
        - hypothesis testing: the ability of a scale to differentiate known groups, assessed by comparing scores between groups who are expected to score differently on a scale
        - cross-cultural validity
    - criterion validity: the correlation between the scale and a ‘gold standard’ that exhibits the same characteristics. Fitzpatrick states that the availability of a ‘gold-standard’ measure for comparison is rarely available and that if such a measure did exist, then it would negate the requirements for a new measurement tool (Fitzpatrick et al. 1999)

- responsiveness: the ability of the scale to detect clinically meaningful change, assessed by comparing scores before and after an intervention of known efficacy to calculate an effect size statistic (Cano et al. 2004)
- interpretability: the ability of the scale to reflect qualitative meaning in clinical practice, measured by minimal important change (MIC) or minimal important difference (MID), which is the smallest difference in the score that patients perceive as beneficial (Alrubaiy, Hutchings and Williams 2014)

In addition to the COSMIN assessment, the FDA also assess more pragmatic characteristics of PROMs, which include the concepts being measured, number of items, conceptual framework of the instrument, medical condition for intended use, population for intended use, data collection method, mode of administration, response options, recall period, scoring, weighting of items or domains, format, respondent burden, and translation or cultural adaption availability FDA (2009).

1.3.7.2 Readability

El-Daly and colleagues examined the readability of the most commonly used PROMs in orthopaedic research by using the Flesch Reading Ease Score (Kincaid et al. 1975). This scoring system is a validated readability tool that measures average sentence length and syllables per word (Kincaid et al. 1975). They found that the majority of PROMs analysed were incomprehensible to the average adult in the United Kingdom (El-Daly et al. 2016). This study included several instruments used in the field of hand surgery, namely, the MHQ, Patient-Rated Wrist/Hand Evaluation (PRWHE), DASH and QuickDASH. All of these instruments required a minimum education level of a 13 year old (El-Daly et al. 2016). If a patient is unable to understand a questionnaire, whether for reasons of cognitive difficulty or language difficulty, they are not able to give accurate answers and therefore should be excluded from any PROM-related research (Dawson et al. 2010). This would make the application of the instruments studied by El-Daly inappropriate for a large proportion of the population. For this reason, it is important to have PROM instruments that are appropriate for use in a heterogeneous population with varied language and cognitive abilities.

1.4 Clinical Implementation of Patient-Reported Outcome Measures

It has been established that PROs are valuable as they allow measurement of the variables that are most affected by medical and surgical intervention, that is, health-related QOL. Despite this, patients' reports on their functioning and wellbeing are not routinely collected in clinical practice. *The User's Guide for Implementing Patient-Reported Outcome Assessment in Clinical Practice* was created by the International Society for Quality of Life Research (ISOQOL) and summarised in a paper by Snyder et al. (Snyder et al. 2012). The purpose of this document is to assist clinicians who are considering implementing PRO measurement within their clinical setting. They advise that prior to any PRO implementation, it is important that clinicians

consider the goals of integrating PRO and assess what resources are available to assist with implementation (Snyder et al. 2012). Deciding on which patients are going to be included depends on the ultimate goal. It may be decided to survey all outpatients regardless of whether they have a specific need for monitoring or focus on a subgroup of patients with a specific condition. Inpatients may be a cohort of interest, but then one must consider whether they are likely to require assistance in completing the forms. The frequency at which patients are asked to complete the PROM is another consideration—do they complete a PROM at each appointment attended, or are they able to complete them at home prior to attending appointments?

It is important to use the appropriate PROM for the purpose that it is intended. This means that the PROM should be proved acceptable to patients, reliable, valid and responsive (Dawson et al. 2010). The evidence of these qualities should be from a patient population that is similar to the population in which it will be applied. The ISOQOL has also published recommendations on the minimum standards of quality required before a PROM is used in research (Reeve et al. 2013). These standards include the requirement for a conceptual model including the intended population for use, reliability, content validity, construct validity, responsiveness and interpretability of scores (Reeve et al. 2013).

When choosing which PROM to use for a research project, it is vital to consider the primary purpose of the study and to decide on the PROM most useful for that specific clinical question (Calfee and Adams 2012). When deciding on which PROM to integrate into clinical practice, the same well-defined clinical question cannot be used to guide the choice. It is a compromise—although having multiple PROMs in use within the same clinic is not practical, it is unlikely that any single instrument will be able to serve all patients effectively (MacDermid 2014).

Valderas et al. (2008) published a systematic review examining the impact of measuring PRO in clinical practice. There is little evidence that the routine collection of PRO data improves patient outcomes, but there is some evidence that it improves the process of care (Valderas et al. 2008; Snyder et al. 2009). Gonçalves Bradley and colleagues have published their protocol for determining the impact of the routine provision of information from PROMs in clinical practice (Gonçalves Bradley et al. 2015). The outcome of this Cochrane review is pending. As stated by Black, the impact of PROMs on routine clinical practice and the flow-on effects of improving health services are yet to be established (Black 2013).

Adoption of the routine use of PROMs by clinicians and hospitals is slowly increasing. However, the comprehensive implementation by health systems is limited to England, Sweden and sectors of the US (Black 2013). England's NHS has implemented PROMs on a national level to provide consumers with information regarding provider performance and to measure their production of 'health' rather than 'healthcare' (Devlin and Appleby 2010). Within Sweden and the US, the drive for PROMs has been led by the medical profession in an effort to improve care for patients. Bindra and colleagues expressed their hope that PROM completion will become part of the routine of a clinic visit, similar to filling in personal details currently (Bindra et al. 2003).

The National PROMs Programme of England's National Health Service (NHS) involves mandatory participation of all providers of four elective procedures (total hip replacement, total knee replacement, groin hernia surgery and varicose vein surgery) (Devlin and Appleby 2010). It is likely that this programme will extend beyond elective surgery to include a range of chronic medical conditions (Devlin and Appleby 2010). Providers in both the NHS and private hospitals are required to invite patients to complete the relevant questionnaires before they undergo surgery (Black 2013). This preoperative questionnaire collects patient demographic information, disease severity and other comorbidities. Along with this questionnaire are a

disease-specific PROM (Oxford Hip Score, Oxford Knee Score or Aberdeen Varicose Vein Score; there is currently no disease-specific PROM for inguinal hernia) and a generic PROM (EQ-5D index and EQ-Visual Analogue Scale). Patients who participate in the programme receive a postoperative questionnaire via postal mail after a predetermined period. The PROMs data are then linked with hospital episode statistics, and a summary for each provider is provided that takes into consideration patient characteristics and mean PROM change when comparing preoperative and postoperative data. This information allows for comparison of providers and identification of any practitioners that are outliers (Black 2013).

1.4.1 Clinical Uses of Patient-Reported Outcome Measures

1.4.1.1 Direct Patient Care

Routine patient information collected using PROs may assist to identify physical and/or psychological issues that might otherwise go unnoticed (Valderas et al. 2008). PROM data can be used to assist with predicting which patients will benefit from a particular intervention (Belvedere and de Morton 2010). This information can be used to base informed discussion during consent processes (Pusic et al. 2011). In England, it is now routine to administer a hip osteoarthritis PROM (Oxford Hip Score) every 3 months to help determine when the patient is likely to benefit most from hip replacement surgery (Black 2013). There is also the potential to develop threshold values that assist with the diagnosis of certain conditions (Van Vliet et al. 2013).

PRO instruments can be useful to monitor disease progression and provide feedback on treatment decisions (Valderas et al. 2008). PROM information assists clinicians to make clinical decisions in the same way as do other clinical investigations such as blood tests and radiographs (Black 2013). There are no definite criteria regarding when a patient with a certain PROM score must have ‘treatment X’, but rather a patient with this score is more likely to benefit from

'treatment X'. This guidance must then be considered along with other patient-specific factors before determining the best treatment option for each individual patient (Devlin and Appleby 2010).

Routine use of PRO in daily patient care may assist with establishing effective communication between the clinician and the patients (Valderas et al. 2008). PROM information can help to ensure that both parties have a common understanding of the patient's situation and expectations. Improving patient satisfaction with their healthcare providers is believed to improve their adherence to prescribed therapy (Valderas et al. 2008).

1.4.1.2 Health Service Monitoring and Development

PRO measurement data could be used to compare providers and/or healthcare organisations. PROs have a pivotal role in the reform of the NHS in England, which has transitioned to an outcomes-oriented performance model (Valderas, Fitzpatrick and Roland 2012; Gonçalves Bradley et al. 2015). Professor Black from the London School of Hygiene and Tropical Medicine is of the opinion that PROMs could help transform healthcare (Black 2013). He believes that the use of PROMs to compare healthcare providers will stimulate quality improvement by allowing patients to choose where they are treated on the basis of the reports of other patients. He also supports health providers reporting their outcomes publicly to their communities, which he believes will increase accountability and drive progress.

The English PROM survey aims to evaluate the relative performance of providers undertaking elective procedures such as hip replacement. An issue with using PROM data in this manner is that this type of data is prone to have substantial missing values, and this can potentially affect the judgments made from the summary information. Gomes et al. (2015) have suggested a mathematical strategy to address the missing data using multiple imputations.

In some areas of medicine, PROM data are used as the primary outcome measure in clinical trials. Therefore, it is conceivable that the data provided by PROs is the major influence on whether a procedure or treatment is given an endorsement, with subsequent influences on patients and clinical research. The appropriateness and adequacy of these decisions are rooted in the scientific quality of the instrument that is utilised.

1.4.2 Administration of Patient-Reported Outcome Measures

The mode of administration of PROs is another important consideration when integrating PRO into clinical practice. It is generally considered best practice to restrict the mode of administration to the method used in the development and validation of the specific PROM in question (DeVellis 2016). The primary choice is whether the patients will be asked to complete the PROM in person when attending the clinic or from home (Snyder et al. 2012).

1.4.2.1 Patient-Reported Outcome Measures at the Clinical Location

If completing the PROM while attending the clinic, there are still several modes of administration to consider; paper-based, computer-based or interview-based administration. Snyder is of the opinion that questionnaires completed in the waiting room may be influenced by the patient's anxiety prior to their review (Snyder et al. 2009). Dawson et al. state that collecting PROM data at follow-up clinical appointments may result in bias as patients are more likely to attend these review appointments if they have ongoing problems, thus resulting in negatively skewed data (Dawson et al. 2010).

1.4.2.2 Paper-Based Patient-Reported Outcome Measures

Paper-based questionnaires have low costs for implementation and are a straightforward method to start routine PROM collection (Snyder et al. 2012). However, this method has the disadvantage of requiring personnel to coordinate questionnaire completion and perform

manual data entry (Snyder et al. 2012). The labour costs involved in administrating and data entry are considerable. There is also the human errors that occur because of converting paper-based data into digital data (Rose and Bezjak 2009; Paulsen, Overgaard and Lauritsen 2012). There may also be issues with patients who have visual disabilities or low literacy (Rose and Bezjak 2009; Snyder et al. 2012). Yafef et al. (2015) found that respondents to paper questionnaires often provided more than one answer or left questions unanswered, which resulted in the inability to accurately score the questionnaire.

1.4.2.3 Interview-Based Patient-Reported Outcome Measures

Interview administration of PROMs is more personal, avoids literacy or vision issues, and allows for more in-depth questioning. However, there are concerns that people may not respond honestly, but instead answer with what they believe the interviewer would like to hear. This was demonstrated in a study by Weinberger et al. (1996) that compared respondent's answers to the SF-36 (a generic HRQOL questionnaire) when completed twice within a short timeframe using different modes of administration. They found that face-to-face administration resulted in a more optimistic outcome than did self-administration (Weinberger et al. 1996). There are also significant staffing costs with this model and possibly more inconvenience for patients, as they are required to attend scheduled interviews (Weinberger et al. 1996).

1.4.2.4 Computer-Based Patient-Reported Outcome Measures

Computer administration allows for efficient and accurate data collection, immediate scoring and recording of data into a database for later analysis (Yafef et al. 2015). This technique would potentially include the use of desktop, laptop or tablet devices used within the clinic setting. Van Den Kerkhof et al. (2005) explored the acceptability of computerised patient information collection in the setting of an anaesthetic preadmission clinic; patients were comfortable using the technology and expressed a preference for computerised methods over paper-based

methods. There is the benefit that a PRO report can be produced almost instantaneously with the summary score and graphical representations of progress to aid clinicians' interpretation of the PRO data at the consult that follows the PRO completion. This form of administration can also be integrated with an electronic medical record. Computer administration can also allow for the integration of CAT, which decreases the responder burden substantially without sacrificing meaningful data (Hays, Morales and Reise 2000; MacDermid 2014). It also minimises the floor and ceiling effects of a measurement tool, and decreases responder fatigue.

The drawbacks to this technique are the higher upfront costs involved in developing the PRO computer system, the requirement for specialised software and the personnel to manage the system. Ongoing costs for software updates, virus protection, system security and maintenance are also significant. There may also be a proportion of the respondents who are not comfortable using a computer. However, it is expected that this proportion will decrease with time (Rolfson et al. 2011).

1.4.2.5 Patient-Reported Outcome Measures at Home

If patients are completing the PROMs from home, then there are alternative methods to consider.

1.4.2.5.1 Mail-Based Patient-Reported Outcome Measures

Mailing of paper-based questionnaires is the classical technique; it is a low technology and relatively low-cost option. The Swedish Hip Arthroplasty Register and the National PROMs Programme in England both use this technique to collect their PRO data (Rolfson et al. 2011; Black 2013). However, there is the possibility of a low response rate and no way to ensure that patients are completing the instruments without the input of others, which can introduce respondent bias (Dawson et al. 2010). Dawson et al. suggest that a reminder letter should be

sent out if the questionnaire response has not arrived within 2–3 weeks (Dawson et al. 2010). This method still requires personnel to manage the mailing process and to score completed instruments.

1.4.2.5.2 Telephone-Based Patient-Reported Outcome Measures

Telephone-based interviews have been used in various studies and found to be a valid mode of data collection (Bot, Becker, et al. 2013). The telephone-based interview may be conducted with a live interviewer or via an automated system. This method avoids any issues with literacy and physical challenges in completing the questionnaire and therefore produces low levels of missing data (Weinberger et al. 1996). However, there are considerable costs in live phone interviews as well as the issue of patients responding in a socially desirable manner. Automated systems are costly and can be poorly tolerated by patients (Snyder et al. 2012).

1.4.2.5.3 Internet-Based Patient-Reported Outcome Measures

The most technologically advanced option for PRO administration is using the internet via web pages such as myClinicalOutcomes (Williams 2012) or PROM-specific apps. This allows immediate scoring, simultaneous data entry and real-time feedback to clinicians. CAT can be integrated into the software and patients can complete the PRO at a time and place of their convenience on their smartphone or other internet-enabled devices. Snyder et al. (2009) have developed a website to collect PRO in the setting of outpatient oncology, PatientViewpoint. This website automatically links the PRO data with the patient's electronic medical record. The website has been proved useful and acceptable to both patients and clinicians (Snyder et al. 2013), and the service continues to be improved. A study by Schamber et al. (2013) compared PROM completion rates when administered via postal mail and email. They found that electronic administration significantly improved completion rates.

The potential reasons that this technique is not yet widely adopted are the upfront costs with designing the system, costs of personnel to manage data collection, concerns regarding data security and privacy, patient preference and the possibility that not all patients have access to the internet.

1.4.2.6 Comparison of Different Modes of Administration

A Cochrane review comparing self-administered survey questionnaire responses from mobile apps versus other methods found no difference in overall scores from varied methods of collection, but there was a tendency for decreased missing data with collection via apps (Marcano Belisario et al. 2015). Others have also found that they gathered more complete data using a tablet or computer than paper-based surveys (Weinberger et al. 1996; Dy et al. 2012). With the prevalence of smart devices and increasing technology literacy within the community, this method is likely to be more commonly used in the future (Snyder et al. 2012).

As the majority of PROMs have been developed for administration on paper, there is legitimate concern about the equivalence of data when the administration is via electronic PROs (ePROs). The International Society for Pharmacoeconomics and Outcomes Research (ISPOR) has formed an ePRO Good Research Practices Task Force, which published recommendations on the evidence necessary to support measurement equivalence of ePROs to the paper-based PROMs from which they originated (Coons et al. 2009). They found that provided any modifications from the original version are minimal, evidence suggests that the psychometric properties of the original PRO will be applicable to the ePRO version (Coons et al. 2009). Alternatively, if substantial changes are made, then the full psychometric analysis should be repeated to establish the measurement equivalence of the ePRO and the original version (Coons et al. 2009).

It is important to be aware that the instructions given to patients about how to go about completing the PROM may alter the responses that are generated. PROMs should always be accompanied by instructions devised by the developers (US Department of Health and Human Services FDA Center for Drug Evaluation and Research, US Department of Health and Human Services FDA Center for Biologics Evaluation and Research and US Department of Health and Human Services FDA Center for Devices and Radiological Health 2006; FDA 2009). Depending on the patient population and their capabilities, the questionnaires may be self-reported, or some patients may require a family member or carer to complete the PROM for them (Dawson et al. 2010).

Ultimately, the mode of administration that is best for PROMs is dependent on the population on which you are focusing. Rolfson et al. (2011) performed a study comparing traditional pen and paper postal questionnaires and internet-based questionnaires on patients enrolled in the Swedish Hip Arthroplasty Register. They found that the response rate for internet-based questionnaires was only 34% compared with 92% in the paper-based method. The response rate to the internet questionnaire was found to decrease with increasing age of the participant, despite evidence that more than 90% of Swedes have access to the internet (Rolfson et al. 2011).

1.4.2.7 Computerised Adaptive Testing

Traditional scales derived from CTT tend to be lengthy as they are reliant on many items to decrease error and produce a measure with high internal consistency (also known as coefficient alpha, α) (Hays, Morales and Reise 2000; DeVellis 2016). Usually, a scale includes many repetitive items that disengage the respondent, and a range of items will be non-applicable to respondents at either end of the trait level. IRT methodology makes it possible to estimate the respondent's latent trait level from a small subset of items from a large item pool (Hays, Morales and Reise 2000). Computerised adaptive testing (CAT) takes advantage of this to

establish maximum information on a respondent with only minimal items, thus vastly reducing the respondent burden. CAT requires a large bank of highly discriminating items of varying levels of difficulty and a computerised algorithm. Items are then delivered to the survey respondent singularly, targeted to the trait level of the respondent. The first item is selected at random, usually of medium level difficulty; the following item is based on the respondent's answer to the first item. Subsequent items are based on the preceding responses.

1.5 Patient-Reported Outcome Measures to Assess Value

Devlin and Appleby (2010) published a report, 'Getting the Most Out of PROMs', where they discuss the stark current economic environment that currently faces the NHS of England. There is increasing pressure to justify services and treatments within the healthcare sector, and to justify spending with proof of output. Value is defined as 'health outcomes achieved per dollar spent' (Porter and Teisberg 2006). As discussed by Porter and colleagues, cost reduction without regard to the effect on outcomes achieved is potentially dangerous and may lead to greater expenditure in the long term (Porter 2010).

From first principles, it may seem logical that there is a linear relationship between the cost and value of a particular service or treatment. However, this relationship is not necessarily linear and, in some circumstances, high cost services have been found to have very low value. An example of this is percutaneous vertebroplasty, where cement is injected into a fractured vertebra to stabilise an osteoporotic vertebral fracture. The Medicare Benefits Schedule (the Australian Commonwealth funded healthcare body) no longer funds this procedure as in two randomised controlled trials the results following vertebroplasty were no better than placebo (Buchbinder et al. 2009; Kallmes et al. 2009). This expensive procedure was considered to have low value because the patient outcomes were not justifying the healthcare expenditure (Buchbinder, Osborne and Kallmes 2010).

Low-value care is use of an intervention where evidence suggests it confers no or very little benefit on patients, risk of harm exceeds likely benefit or, more broadly, the added costs of the intervention do not provide proportional added benefits. Choosing low-value care consumes resources that could have been expended on alternative forms of care conferring greater levels of benefit, either to the patient in question or to other patients (I. Scott and Duckett 2015)

Acknowledging this complex and non-linear relationship between cost and value is important for health services planning. PROMs can play a role in deriving more valid measures of the value of a service as they measure outcomes that matter to patients.

The NHS Outcomes Framework was announced in 2010 and included a shift from a process of care target to a performance-based model targeting patient outcomes. Integral to this framework is the systematic collection of preoperative and postoperative PROMs in selected elective surgical procedures. The large data set of PROM information from a real-world population gives an opportunity for facilitating research on the effectiveness of treatments (Black 2013).

Effectiveness is slightly but importantly different from efficacy (Johnson and Chung 2014). Efficacy-based studies are those that the medical field has depended on throughout the ages; epitomised by the randomised control trial, these studies show that a treatment ‘can’ work. Effectiveness studies go a step further and apply the treatment in real-world conditions, with a full variety of patients, and take into account pragmatic difficulties. Effectiveness studies define ‘what works for us’ rather than a treatment that is proved in clinical efficacy trials to ‘work on a select group of patients under ideal circumstances’.

Comparative effectiveness research (CER) is defined by the Institute of Medicine as:

> The generation and synthesis of evidence that compares the benefits and harms of alternative methods to prevent, diagnose, treat and monitor a clinical condition, or to improve the delivery of care. The purpose of CER is to assist consumers, clinicians, purchasers, and

policymakers to make informed decisions that will improve health care at both the individual and population levels.

CER is also known as ‘real-world research’. Roche et al. (2014) published *Quality Standards for Real-World Research*, which advises on methodological issues specifically related to comparative observational studies using large datasets of real patient data. As part of the *Affordable Care Act 2010*, US legislators established a national CER programme, the Patient-Centred Outcomes Research Institute (PCORI) (Johnson and Chung 2014).

Cost-effectiveness analysis (CEA) is a process whereby the cost of an intervention is compared with the consequence of this treatment measured in clinically relevant units, such as PROMs (Pusic et al. 2011; Malay and Chung 2013). Bindra and colleagues give an example in the hand surgery field as the cost per unit of DASH score improvement (Bindra et al. 2003). The inclusion of a generic PROM (the EQ-5D) in the NHS outcomes allows CEA to be performed to evaluate practices and policies quickly and cost effectively (Black 2013).

1.6 Barriers to the Use of Patient-Reported Outcome Measures in Clinical Practice

Black has identified a number of challenges that have become evident from England’s National PROMs Programme (Black 2013). The primary challenge in instigating routine PRO use is the time and cost of data collection, analysis and presentation. The implementation of techniques that incorporate more technological advances is anticipated to reduce this burden. Patient participation is another significant issue; it is vital to appeal to the whole spectrum of the population of interest to prevent underrepresentation of the patients from minority groups (Dawson et al. 2010). There is also the issue of appropriately and robustly accounting for differences in the case mix prior to making any meaningful comparisons between healthcare providers (at both the individual clinician level and the health sector level).

According to Valderas, there are both attitudinal and practical barriers to the implementation of PRO instruments into clinical practice (Valderas et al. 2008). There is the impression that clinicians are sceptical about the validity and potential utility of PRO instruments (Valderas, Alonso and Guyatt 2008). Hanh et al. (2007) state, 'To many clinicians, the assessment of health-related quality of life seems more art than science'. It is necessary to overcome this attitude to successfully integrate PROMs into patient care (Valderas, Alonso and Guyatt 2008).

Most PRO instruments are lengthy and may be perceived as burdensome by both patients and clinicians. The process of obtaining the PRO data in an easily comprehensible format in a timely fashion for the clinical use of practitioners is a resource-intensive process. Valderas et al. (2008) discuss the potential for PRO measurement in routine practice to potentially cause harm by encouraging patients to focus on problems that would otherwise go unnoticed. They also express concern about the potential for the clinician to lose control of the topic of discussion because of interference caused by the PROM. Their systematic review found that most studies had found positive effects of PROM integration on at least one aspect of process outcomes; however, effects of patient health status and outcome were less commonly examined and reported. They conceded that the heterogeneity of PROs used hindered the interpretation of the impact of PROMs.

1.7 Summary

Although PROMs have become well established in multiple surgical fields and have demonstrated usefulness in both research and clinical care, there remains a paucity of evidence on the use of PROMs in the routine care of hand surgery patients. It is necessary to identify and critically analyse the PROMs that are currently available in the field of hand conditions, this will be the focus of the following chapter.

Chapter 2: Measuring Quality of Life and Patient Satisfaction in Hand Surgery: A Systematic Review

This chapter has been published as an article in the *Australasian Journal of Plastic Surgery*, attached as Appendix A.1.

Sierakowski, Kyra L., Kathleen A. Evans Sanchez, Rachael A. Damarell, Nicola R. Dean, Philip A. Griffin and Gregory I. Bain. 2018. 'Measuring Quality of Life and Patient Satisfaction in Hand Conditions'. *Australasian Journal of Plastic Surgery* 1 (2): 85–98.

2.1 Introduction

This chapter follows on from the concepts introduced in Chapter 1 and focuses on appraising existing PROMs for patients with hand conditions. A systematic literature review was carried out to comprehensively identify all existing PROMs directly relevant to measuring outcome in patients with hand conditions. The hand can be considered anatomically with or without the wrist, for the purposes of this review both hand specific and hand/wrist instruments were included. Once the instruments were identified, the development methodology was reviewed and the domain content summarised for each instrument. This allowed for an understanding of the currently available instruments, learning from the considerable body of work in this area and identification of areas where improvements could be made.

The various approaches to the theoretical underpinning of PROMs were introduced in Chapter 1. There is a robust debate in the field of psychometrics concerning the best methodology for PRO development (Cano and Hobart 2011). CTT is being increasingly overtaken by IRT and Rasch analysis (Atroshi, Lyrén and Gummesson 2009). As discussed in greater detail in Chapter 1, there are numerous benefits of IRT and Rasch analysis; they allow for both person and item

parameters to be placed on the same scale and can be used to compare results between different populations. The scale can be analysed for DIF, which is the analysis of whether some items perform differently in subpopulations such as gender or age. IRT is being used to analyse and shorten existing scales by keeping only the items that provide the most useful information. The development methodology used for each PROM will be assessed according to the adherence to the international best practice guidelines (Lohr 2002).

The domain analysis will establish the breadth of concepts measured by each PROM. Interventions for patients with hand conditions often aim to reduce symptoms, improve functionality and potentially change the appearance of the hand. That is, they influence multiple domains. As hands are a highly functional body part, it is logical that measuring physical function is a priority. Hands are also integral in much of our day-to-day ability to perform in our social, professional and personal lives. Instruments limited to the measurement of only a single domain, such as physical function, do not capture the full spectrum of change produced by surgical or therapeutic interventions.

2.2 Methodology

2.2.1 Search Strategy

A systematic review of the English-language literature was performed to identify PROs that have been developed for use in patients with hand conditions. A broad range of relevant databases was included in the search: MEDLINE, CINAHL, EMBASE, Emcare, PsychINFO, HaPI, Scopus, Web of Science, ProQuest: Health & Medicine, and the Cochrane Systematic Review. The timeframe used for the search was from the database conception to June 2017. The purpose of the search was to find PROMs for the assessment of hand or upper limb conditions/surgery. The search strategy was developed with the aid of a medical research

librarian. Search terms included subject headings and text words for the following terms: ‘quality of life’, ‘health related quality of life’, ‘quality adjusted life years’, ‘health status’ or ‘functional status’ or ‘well being’ or ‘wellbeing’ or ‘patient reported outcome’ or ‘PROM’ or ‘PRO’ or ‘PROS’. Other search terms used included the specific names of hand PROMs known to the investigator: DASH, PRWHE, MHQ, PEM, POS-Hand/Arm and PROMIS. The terms ‘hand, metacarpus, finger/s, wrist, thumb and surg*’ were used to limit the results to the anatomical region of interest—the hand. An example of the full search strategy used for Medline is available for review (Appendix A.2).

Eligible papers were those that had been published in peer-reviewed literature; grey literature was excluded from analysis. Articles of interest were those discussing the development of, or psychometric analysis of, PRO tools used in hand conditions. Relevant review articles were included. Hand searching of references was performed to find any missing PRO questionnaires evaluating QOL, impairment and disability, or patient satisfaction after hand surgery.

Criteria for excluding papers from the study were that the paper described an instrument that was clinician reported (therefore, not a true PROM) or that the paper described an anatomic site other than the hand (for example, the elbow or shoulder) or that the paper described instruments that were developed for a specific subpopulation such as children, the elderly or those receiving workers compensation. Articles reporting on the translation of a non-English PROM or on ad hoc instruments were likewise excluded. The author and a colleague independently screened titles and abstracts in duplicate, discussed any discrepancies and established consensus.

2.2.2 Review Strategy

The review was conducted as per an a priori protocol that consisted of inclusion and exclusion criteria as well as data extraction templates. Information about the development process and psychometric evaluation of the PRO instruments was extracted from the articles. Instruments

that met the inclusion criteria were analysed by their domain content. Items and stems from each instrument were compiled in data extraction templates. In the case where more than one version of a tool was available, all versions of the instrument were included in the analysis.

Eligible instruments were those developed with a focus on the anatomical area of interest (hand or upper limb) or conditions specific to this region (carpal tunnel syndrome [CTS], Dupuytren's disease). Cano et al. (2004) described a framework for the assessment of PROMs based on the guidelines of the SAC of the Medical Outcomes Trust (Lohr 2002) and the US FDA (2009); this was used for the assessment of each instrument's development and validation methodology.

2.3 Results

Figure 2.1 summarises the results of the systematic search. The search yielded 3039 papers after the removal of duplicates. Following the screening of titles and abstracts to identify only relevant papers, 139 underwent full-text review. Exclusion of any articles at this stage was documented. A total of 87 articles met the inclusion criteria, identifying a total of 20 instruments relevant to hand surgery. Further reference searching found another four eligible instruments. A total of 24 instruments were identified: 10 regional upper extremity instruments, 6 regional hand/wrist and 8 condition-specific instruments.

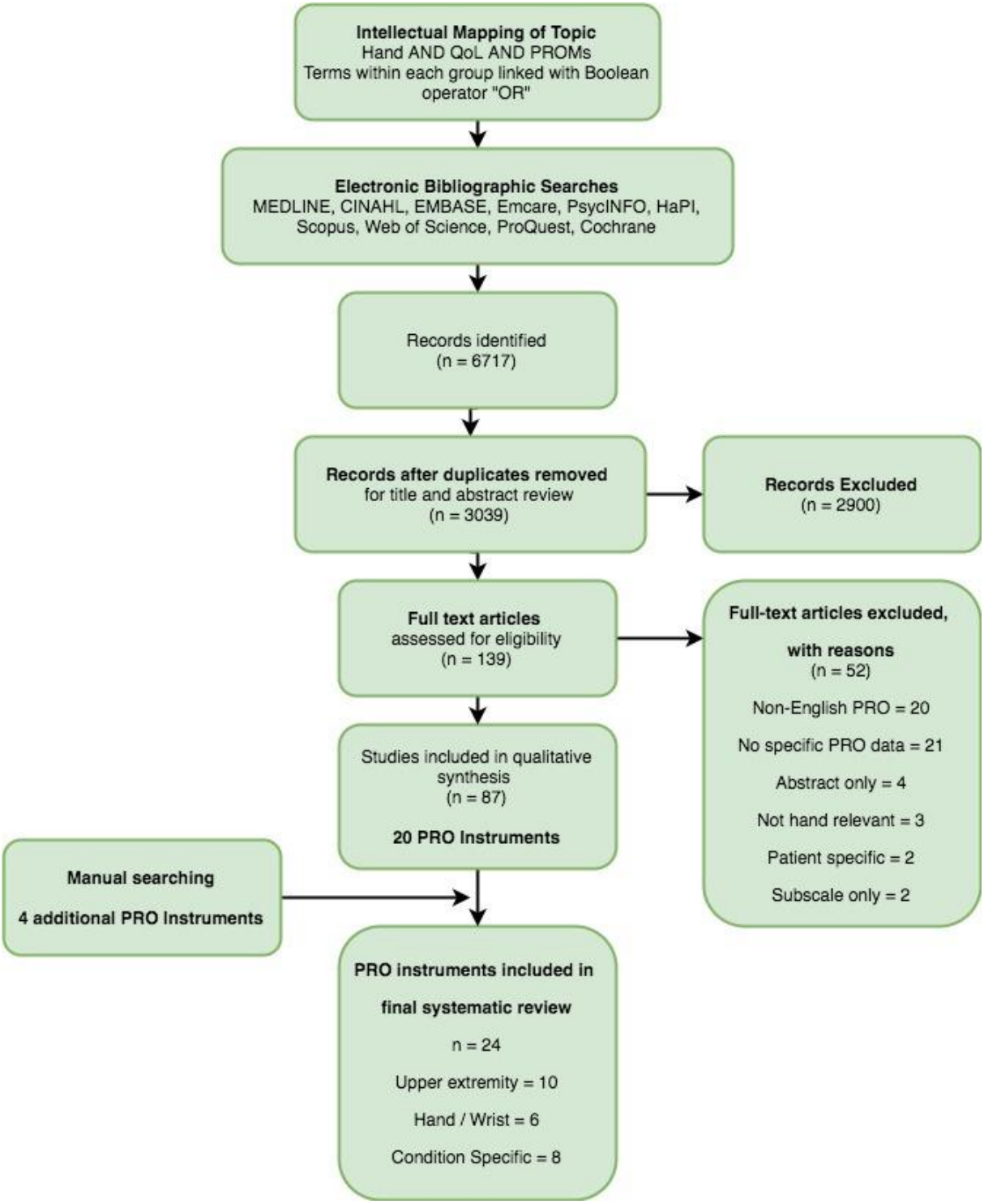

Figure 2.1: Flowchart of systematic search.

2.3.1 Regional Patient-Reported Outcome Measures—Upper Extremity

There were 10 PROMs found that focus on the upper extremity. The domain analysis, publication year and methodological technique are summarised in Table 2.1. The development and validation analysis for each PROM are summarised in Table 2.2. The in-depth domain comparison between PROMs is summarised in Table 2.3.

Table 2.1: Regional upper extremity PROMs

Instrument	Year	Author	Country of origin	Domains	Methodology
DASH	1996	Hudak	Canada, US	·Disability - functional status: physical, social, psychological - symptoms: pain, weakness, tingling/numbness, stiffness	CTT
UEFI	2001	Stratford	Canada	·Upper extremity function	CTT
POS-Hand/Arm	2004	Cano	UK	·Physical activities ·Symptoms ·Psychological functioning/cosmetic appearance ·Satisfaction (post-surgery)	CTT
MAM-16	2005	Chen	US	·Manual ability - hand tasks	Rasch
*QuickDASH	2005	Beaton	Canada, US	·Disability - functional status: physical, social - symptoms: pain, tingling, difficulty sleeping	CTT
ULFI	2006	Gabel	Australia	·Health-related QOL and upper extremity dysfunction (quantitative) ·Patient-specific index (qualitative) ·Overall status (VAS)	CTT/Factor analysis
*M2DASH	2008	Khan	UK	·Disability	CTT

				- functional status: physical, social, psychological - symptoms: tingling, weakness, stiffness, difficulty sleeping	
*MAM-36	2010	Chen	USA	·Manual ability - hand tasks	Rasch
*UEFI-15	2013	Hamilton	Canada	·Upper extremity function	CTT/Rasch
PROMIS-PF-UE	2013	Hays	US	·Disability of the upper extremity	CTT/IRT

* Developed from an existing PROM.

Table 2.2: Development and validation criteria of upper extremity PROMs

Criteria	Upper extremity PROMs									
	DASH	QuickDASH	M2DASH	POS-Hand/Arm	MAM-16	MAM-36	ULFI	UEFI	UEFI-15	PROMIS-PF-UE
Item generation										
Patient interviews				•						*
Literature	•	•	*	•	•	*	•	•	*	*
Expert opinion	•	•	*	•	•	*		•	*	*
Develop conceptual model	•	*	*	•						*
Item reduction										
Expert opinion	•	•		•	•		•			•
Item redundancy				•		•	•	•		
Endorsement frequencies	•	•	•	•		•		•		
Missing data		•		•				•		•
Factor analysis	•			•		•		•		•
Tests of scaling assumptions				•						
Item misfit (Rasch/IRT)					•	•			•	
Psychometric analysis										
Acceptability		•		•			•	•	•	•
Internal consistency reliability	•	•		•			•	•	•	
Item total correlations	•	•		•	•	•				
Interrater reliability										
Test–retest reliability	•	•		•			•	•	•	
Validity within scale	•	•		•		•	•		•	
Validity comparison with other measures	•	•	•	•	•	•	•	•		•
Validity hypothesis testing	•	•	•	•	•	•	•	•		
Responsiveness	•	•	•	•			•	•	•	

* Developed from an existing PROM.

Table 2.3: Domain analysis of upper extremity regional PROMs

	Upper extremity PROMs									
	DASH	QuickDASH	M2DASH	POS-Hand/Arm	MAM-16	MAM-36	ULFI	UEFI	UEFI-15	PROMIS-PF-UE
Physical Functioning										
Limitations of the whole upper limb	•	•	•	•	•	•	•	•	•	•
Limitations of hand/wrist										
Limitations of hand/digits										
Limitations of thumb										
Ability to perform ADLs	•	•	•	•	•	•	•	•	•	•
Ability to use smartphone/modern technology							•			
Ability to work	•	•	•				•	•	•	
Ability to travel	•						•	•	•	
Ability to participate socially	•	•	•				•	•		
Symptoms										
Pain issues	•	•		•			•			
Sensory changes (tingling/numbness)	•		•	•						
Stiffness	•		•	•			•			
Swelling	•		•	•			•			
Weakness	•		•	•			•			
Reduced ROM				•						
Insomnia	•	•	•	•			•	•		
Change in appetite							•			
Health-related QOL										
Satisfaction with treatment				•						
Satisfaction with outcome/overall assessment				•			•			
Inconvenience of medical/hospital										
Concerns re post-operative complications/recovery				•						
Expectations				•						

Psychological functioning						
Self-confidence/self-esteem	•		•	•		
Avoidance of uncomfortable situations						
Negative feelings about self				•		
Change in mood				•	•	
Body image						
Concerns regarding scarring				•		
Self-consciousness				•		
Satisfaction with hand appearance				•		
Sexual functioning	•		•			

2.3.1.1 Disabilities of the Arm, Shoulder and Hand

The DASH is a 30-item questionnaire that was developed as a joint initiative of the American Academy of Orthopaedic Surgeons, the Council of Musculoskeletal Specialty Societies and the Institute for Work and Health (Toronto, Ontario) (Hudak et al. 1996). The DASH is a brief self-administered instrument that measures symptoms and functional status for use in both research and daily clinical care. The DASH measures upper extremity disability at the person level (Hudak et al. 1996). The content was developed from a literature review, expert panels and existing measurement scales (Hudak et al. 1996). The items related to functional state (21 items) ask the patient to recall over the preceding week how their condition affected their ability to perform daily tasks, or their best estimate if they have not performed the task. Response options include *no difficulty* (1), *mild difficulty* (2), *moderate difficulty* (3), *severe difficulty* (4) and *unable* (5). The remaining nine items inquire about symptoms, interference with social interactions, sleeping and opinion of self. The DASH score is calculated using the following equation: (sum of the responses divided by the number of responses – 1) × 25. A score is only valid if there are less than 10% missing items (three or less). DASH scores are between 1 and 100, with a higher score indicating greater disability. There are also optional high-performance

modules for work or sports/performing arts; each has four items, scored 1–5. These modules are scored with the same technique but given a separate score.

The DASH is a widely adopted and well-established functional assessment tool for the upper extremity and used for many conditions affecting the upper limb. Construct validity of the DASH is supported by its moderate levels of correlation with a widely used generic instrument, the Medical Outcomes Study SF-36 (Baldwin and Wolf 2013; Sorensen et al. 2013). The minimally clinically important difference was found to be 10, using an anchor-based approach (Baldwin and Wolf 2013; Sorensen et al. 2013).

Reynolds and Thirkannad (2013) advocated the use of the 'recall DASH score', in which patients are asked to complete the DASH by remembering their pre-treatment condition. Although the 'recalled' DASH score was shown to be similar, it is not how the DASH was intended to be used and therefore of questionable use in practice.

Lehman et al. (2010) examined the factor structure of the DASH with exploratory factor analysis (EFA) and identified three potential factors: gross motor items, fine motor items and symptom items. Rasch-derived fit statistics were improved when the DASH was divided into three subscales (Lehman, Woodbury and Velozo 2011). They concluded that a multidimensional instrument would give an enhanced informative capacity of the tool, particularly when monitoring for status change as a result of targeted therapy. These findings were similar to those of Franchignoni et al. (2010), who also suggested that a three-factor solution was necessary.

Huisstede et al. (2009) suggested that the DASH is also useful for patients with neck complaints, which gives concern as to whether the DASH is specific enough for use in patients with upper limb pathology. The DASH was demonstrated to be unreliable in differentiating disability due to upper limb pathology versus lower limb pathology when being completed by

a person with both upper and lower limb issues (Dowrick et al. 2006). This resulted in the development of the Manchester-Modified DASH (M2DASH), which was developed to be a shorter questionnaire that is more specific to the upper limb (Khan et al. 2008).

2.3.1.1.1 Manchester-Modified DASH

The M2DASH was created by comparing the responses of a population of patients with lower extremity injury with those of a group of patients with upper limb injuries; the items most frequently endorsed by the people with lower extremity issues were removed. The authors also reformatted the questionnaire to fit on a single page for ease of use. The scores from the M2DASH were found to be highly correlated with the DASH (Khan et al. 2008). A further study by Khan compared the M2DASH to the Patient Evaluation Measure (PEM) and MHQ (K. Chung et al. 1998), and found the M2DASH to be highly correlated with both (Khan et al. 2009).

2.3.1.1.2 QuickDASH

The QuickDASH was created by the same developers as a more succinct version of the DASH suitable for day-to-day clinical use. Three different techniques were trialled for item reduction from the original DASH to form the QuickDASH. Conceptual methodology, equidiscriminative item total correlation and IRT (Rasch analysis) were separately implemented to produce three similar draft versions of the QuickDASH (Beaton et al. 2005). The content of the drafts differed slightly. The draft scales were evaluated and the final QuickDASH was chosen on the basis of the following parameters: (1) fewest number of items with greater than 40% in one response category (as this reflected the item had poor discrimination), (2) Cronbach alpha > 0.9 and (3) the highest correlation with the full DASH and most similar measurement properties to the DASH (Beaton et al. 2005). The group responsible for the development of the DASH selected the concept retention draft to form the finalised QuickDASH. This technique involved reducing

the original 16 domains to 11. Items were ranked firstly according to their importance and difficulty as viewed by patients and secondly according to the correlation with the item and the total DASH score.

Although it has been shown that scores from the DASH and the QuickDASH are highly correlated, work by Angst et al. (2009) showed that the two instruments are measuring slightly different concepts. The QuickDASH underestimates symptoms and overestimates disability compared with the DASH. It is for this reason that the original DASH is recommended for any clinical research studies (Angst et al. 2009).

London et al. (2014) have demonstrated that verbal administration of the QuickDASH replicates written results, with the benefit of minimising missing data. The minimal clinically important difference (or MID) was found to be 14, using an anchor-based approach. This approach to establishing responsiveness of a scale relies on respondents or clinicians reflecting whether the clinical condition has changed; this is discussed further in Chapter 9 (Sorensen et al. 2013).

2.3.1.2 Upper Extremity Functional Index)

The Upper Extremity Functional Index (UEFI) was developed as a self-report functional status measure for clinical practice and research applications. The scale has 20 items, each with a 5-point scale (0–4, *extreme difficulty/unable to perform activity* to *no difficulty*). Total scores are given between 0, the lowest functional status, and 80, which corresponds to the highest functional status (Stratford 2001). Item generation was based on items from existing measures. Item reduction was based on formal item analysis and factor analysis (Stratford 2001). Hamilton and Chesworth (2013) later performed Rasch analysis on the UEFI and found that the original 20-item UEFI did not fit the Rasch model. They went on to propose a revised UEFI-

15 as a valid and reliable interval-level measure of upper extremity function (Hamilton and Chesworth 2013).

2.3.1.3 Patient Outcomes of Surgery–Hand/Arm

The Patient Outcomes of Surgery–Hand/Arm (POS-Hand/Arm) was designed to evaluate the outcomes of surgery and include all clinically relevant domains as agreed by hand surgery experts (Cano et al. 2004). The development followed the international guidelines of the SAC; patient interviews were used to generate a pool of items, and field testing was used to show the best performing items and for evaluation of the instrument. The pre-surgery POS-Hand/Arm has three scales: physical activities (12 items), symptoms (12 items), and psychological functioning and cosmetic appearance (5 items). The post-surgery version includes the same three scales with the addition of a satisfaction scale (4 items). Summary scores are calculated by the addition of the items and transformation into a 0–100 scale where high scores indicate improved health. Satisfactory psychometric properties have been reported by the developer and independent authors; Cronbach alpha ranges between 0.73 – 0.93 for each of the scales, item-total correlations 0.38 -0.73 (Cano et al. 2004 and Scott et al. 2009).

2.3.1.4 Manual Ability Measure

The Manual Ability Measure–16 (MAM-16) was produced using an iterative technique of three rounds of data collection, item reduction and instrument refinement (Chen et al. 2005). The sample size of each round was 70, 30 and 15, respectively. Items were based on existing instruments and function tests. Initial item reduction was based on expert opinion of hand surgeons and therapists. A criticism of the MAM-16 is that the size of the population used for development was not the recommended minimum of 300 respondents as suggested by scale development expert Robert DeVellis (DeVellis 2006).

2.3.1.4.1 Manual Ability Measure–36

Chen and Bode (2010) went on to create the MAM-36 by using data from the same patient cohort as the MAM-16 along with a second cohort of patients with musculoskeletal or neurological conditions of the upper limb. Principle component analysis confirmed a unidimensional construct (Chen et al. 2005). The person separation reliability and item separation reliability were quoted to be 0.93 (person separation index 3.74) and 0.99 (item separation index 11.08). That is, the scales were able to measure a wide range of abilities and respondents could be stratified into five levels of ability.

2.3.1.5 Upper Limb Functional Index

The Upper Limb Functional Index (ULFI) was developed using the 'Guyatt model' of questionnaire development, which involves item generation from literature and existing PROMs. Item reduction was based on peer and patient feedback. The resultant instrument consists of 25 items relating to functional activities, a patient-specific index and an 11-point visual analogue scale (VAS) for rating the current 'overall status' (Gabel et al. 2006). Participants mark the boxes of the items that describe them (i.e., each item has a dichotomous response option). In the patient-specific index, respondents list five activities that are affected by their arm problem and rate these activities from 0 to 5, where 0 = *best* and 5 = *worst* (Gabel et al. 2006). The goal of the ULFI was to provide an instrument with improved clinical utility by the questionnaire having a short length, and ease of completion and scoring, while maintaining acceptable psychometric properties.

The developers of the ULFI later recognised the dichotomous response option led to suboptimal clinimetric properties, and altered the survey to have the response options 'yes', 'half' and 'no' with scores of 1, 0.5 and 0, respectively (Gabel et al. 2010). The ULFI is comprehensible to a level of education equivalent to seventh-grade schooling as opposed to the QuickDASH, which

is only understood by those with a 12th grade or higher level of education (Gabel et al. 2010). The ULFI has been found to be quicker to fill out and score, with decreased missing data compared with other PROMs (Gabel et al. 2010).

2.3.1.6 Patient Reported Outcome Measurement Information System–Physical Function–Upper Extremity

The PROMIS has been developed by the US National Institute of Health (NIH) to create standardised item banks to measure physical, mental and social health in adults and children (Cella et al. 2007). The PROMIS uses CAT (discussed in Chapter 1) that is based on IRT, which allows for the questions to be tailored to the participant according to their last response without sacrificing precision or content validity (Cella et al. 2007). Originally, the PROMIS created a Physical Function item bank that was for both upper and lower limb disability (Hays et al. 2013). From this item bank, the questions related to the Upper Extremity were used to create a 16-item subdomain item bank (Hays et al. 2013). This was further refined into what is now a 44-question PROMIS item bank (V2–Upper Extremity). The Patient Reported Outcome Measurement Information System–Physical Function–Upper Extremity (PROMIS-PF-UE) CAT presents each respondent with a maximum of 12 questions that are given in no specific order (Döring et al. 2014). The total score from administration of any PROMIS instrument ranges from 0 to 100, with higher scores indicating higher levels of disability. Conveniently, a score of 50 represents the mean score of the general population (in the US) (Döring et al. 2014). The psychometric properties of the upper extremity CAT have been found to compare favourably with the DASH while reducing the number of questions asked (6 vs 30) (Beckmann et al. 2016) and avoiding any floor or ceiling effects (these reflect the failure of an instrument to measure patients who are at the extremities (i.e., the floor or ceiling of the concept being measured) (Döring et al. 2014).

2.3.2 Regional Patient-Reported Outcome Measures—Hand/Wrist

There were six PROMs identified that focus on the hand (with or without the wrist). The domain analysis, publication date and methodological technique are summarised in Table 2.4. The development and validation analysis are summarised in Table 2.5. The in-depth domain comparison between instruments is summarised in Table 2.6.

Table 2.4: Regional hand/wrist PROMs

Instrument	Year	Author	Country of origin	Domains	Methodology
PEM	1995	Macey	UK	·Treatment ·Hand health profile ·Overall assessment	CTT
PRWHE	1996	MacDermid	Canada	·Pain ·Specific function ·Usual Function	CTT
MHQ	1998	Chung	US	·Overall hand function ·ADLs ·Pain ·Work performance ·Patient satisfaction with hand function ·Aesthetics	CTT
MASS07	2008	Alexander	USA	·Functional limitation during modern high-frequency activities	CTT
HAT	2009	Naidu	USA	·Activity limitation (without aid) ·Functional tasks ·Pain ·Aesthetics	EFA
*Brief MHQ	2011	Waljee	USA	·Overall hand function ·ADLs ·Pain ·Work performance · Patient satisfaction with hand function ·Aesthetics	CTT Concept retention

* Developed from an existing PROM.

Table 2.5: Development and validation criteria of hand/wrist PROMs

Criteria	Hand/Wrist PROMs					
	MHQ	BMHQ	PRWHE	HAT	PEM	MASS07
Item generation						
Patient interviews	•	*	•			
Literature	•	*	•	•		•
Expert opinion	•	*	•	•		
Develop conceptual model						
Item reduction						
Expert opinion	•	•	•			
Item redundancy	•					
Endorsement frequencies		•	•			
Missing data						
Factor analysis	•			•		
Tests of scaling assumptions						
Item misfit (Rasch/IRT)						
Psychometric analysis						
Acceptability						
Internal consistency reliability	•	•	•	•		•
Item total correlations						
Interrater reliability						
Test–retest reliability	•	•	•	•		•
Validity within scale	•	•				
Validity comparison with other measures	•	•	•	•		•
Validity hypothesis testing	•	•	•			•
Responsiveness		•				

* Developed from an existing PROM.

Table 2.6: Domain analysis of hand/wrist regional PROMs

	Hand/Wrist PROMs					
	MHQ	BMHQ	PRWHE	HAT	PEM	MASS07
Physical functioning						
Limitations of the whole upper limb						
Limitations of hand/wrist	•	•	•	•	•	•
Limitations of hand/digits	•	•		•	•	•
Limitations of thumb	•	•		•	•	•
Ability to perform ADLs	•	•	•	•	•	•
Ability to use smartphone/modern technology						•
Ability to work	•	•	•		•	
Ability to travel						
Ability to participate socially			•	•		
Symptoms						
Pain issues	•	•	•	•	•	
Sensory changes (tingling/numbness)	•	•		•	•	
Stiffness	•				•	
Swelling	•				•	
Weakness	•				•	
Reduced ROM	•	•				
Insomnia	•			•		
Change in appetite						
Health-related QOL						
Satisfaction with treatment					•	
Satisfaction with outcome/overall assessment	•				•	
Inconvenience of medical/hospital						
Concerns re postop complications/recovery						
Expectations					•	
Psychological functioning						
Self-confidence/self-esteem						
Avoidance of uncomfortable situations	•					
Negative feelings about self	•					

Change in mood	•				
Body image					
Concerns regarding scarring					
Self-consciousness	•	•			•
Satisfaction with hand appearance	•	•	•	•	•
Sexual functioning					

2.3.2.1 Patient Evaluation Measure

The PEM originated from an international consensus meeting of multidisciplinary hand surgery experts in Derby, United Kingdom (Macey et al. 1995). It is composed of 18 questions, which when summed together given an overall hand health score. The evaluation is composed of three parts: Part One—Treatment (5 items), Part Two—How the hand is now (hand health profile, 11 items) and Part Three—Overall assessment (3 items) (Macey et al. 1995). Response options are rated on a 7-point Likert scale. The overall score is represented as a percentage of the maximum score possible, with higher numbers indicating worse outcome (Waljee et al. 2011). No detail is given on the methods used to develop this PROM. The psychometric properties of the PEM were later validated in patients with a fractured scaphoid and CTS by Dias et al. (2001) and Hobby, Watts and Elliot (2005), respectively. The tool was found to be reliable, valid and responsive in these patient populations. Both Hobby, Watts and Elliot (2005), and Forward, Sithole and Davis (2007), found that the PEM was more sensitive than the DASH score; however, it was also noted that the PEM measures symptoms whereas the DASH measures disability and that this factor could account for the difference in sensitivity.

2.3.2.2 Patient-Rated Wrist/Hand Evaluation

The original Patient-Rated Wrist Evaluation (PRWE) was designed to measure the outcome following distal radius fracture. It was developed from a survey of International Wrist

Investigators (members of which pursue academic, clinical and research focuses) to determine the structure and content of the scale. This group determined that the key issues for patients recovering from a distal radius or scaphoid fracture were the pain, functional ability and patient satisfaction (MacDermid 1996). Items were based on interviews with wrist-injured patients and hand surgeons, literature, and existing PROMs. The items were reduced by expert consensus and statistical analysis of pilot data to select the 'best' items for the subscales (MacDermid 2007). The scales consist of 15 items separated into two domains: pain (5 items) and function (10 items). For each item, a numerical scale from 0 to 10 is used to gain a response. Scoring entails the addition of the pain score to the halved function score to give a total score of 0–100, where lower scores indicate better function and less pain. The PRWE was modified to the PRWHE by changing the word 'wrist' to 'wrist/hand' throughout the questionnaire; an aesthetics item was also added at this stage, which is not included in the overall score. This altered version, the PRWHE, was shown to correlate with the DASH (MacDermid and Tottenham 2004).

The PRWHE is preferred by clinicians because of its short format and ease of scoring, improving the scale's practicality for routine use (MacDermid and Tottenham 2004). The MID was found to be 14, using anchor-based approach (Sorensen et al. 2013).

The developers of the PRWHE have identified the benefit of modern IRT and have attempted to retrospectively fit the PRWHE to the Rasch model (T. Packham and MacDermid 2013). This analysis suggested several changes were necessary to the current scale and the scoring to make it an acceptable Rasch model fit. The required changes include dividing the PROM into three subscales: pain, specific activities and usual activities. The authors recommend that clinicians exercise caution when considering the changes in any single items and summed PRWHE scores as they remain ordinal in nature (Taylor and Kersten 2014).

The fact that this scale was not developed with hand conditions as the primary focus leads to questionable content validity and applicability for the hand patient cohort. The 11-point response scale for each item is not in keeping with recommendations for survey design (Khadka et al. 2012). The content validity is not fully established as some of the activities referred to in the survey might not have been performed by patients; estimates of their experience and non-response may both introduce inaccuracies (Taylor and Kersten 2014).

2.3.2.3 Michigan Hand Outcomes Questionnaire

The MHQ was developed with input from a panel of 20 (B. Chung and Morris 2014) patients with hand conditions, hand therapists and hand surgeons (K. Chung et al. 1999). It is designed to measure a wide range of hand conditions(B. Chung and Morris 2014). The MHQ is a hand-specific outcomes instrument composed of 25 questions to be answered for each hand and a further 12 questions relating to both hands, totalling 62 questions (B. Chung and Morris 2014; Wehrli et al. 2016). These items are separated into six distinct scales: overall hand function, ADLs, pain, work performance, aesthetics and patient satisfaction with hand function. Each hand is evaluated separately (B. Chung and Morris 2014). Scoring of the MHQ results in a number between 0 and 100, with 100 being the 'best' possible outcome (Shauver and Chung 2009) with the exception of the pain scale (Waljee, Kim et al. 2011). Each of the six domains results in separate scores (B. Chung and Morris 2014). A summary score can then be calculated by averaging the scores from each domain, after reversing the pain score. Each scale requires a minimum of 50% response rate to be used. Depending on the patient's condition, a score can be generated for either the left or the right hand, or an average between the two can be generated for a combined score for relevant conditions that have bilateral effects. The MHQ is the only questionnaire to adjust for hand dominance and distinguish disability between hands (Waljee et al. 2011).

K. Chung et al. (1999) established the responsiveness of the MHQ by correlating patient's self-assessment of their progress with repeated MHQ scores over time. MID represents the value by which a subscale score would need to change over time to indicate a relevant change in the clinical status of the patient's hand (B. Chung and Morris 2014). In patients with rheumatoid arthritis, CTS and distal radius fracture, this value may vary from 3 to 23 points (Shauver and Chung 2009).

The MHQ is long and repetitive compared with other PROMs. As a result, the completion rates vary between studies. In a study of 194 participants, only 116 (59%) completed the survey while in the waiting room of the clinic. In an effort to establish reliability, subjects were asked to complete the survey a second time in the following week. Only those with complete subscale responses at both administrations were eligible, this resulted in reliability being calculated based on 53 responses (B. Chung and Morris 2014). Poor completion rates are attributed to the length of the survey, taking at least 15 minutes to complete.

Confirmatory factor analysis performed by an independent group found that the original factor structure failed to meet the criteria for model retention (B. Chung and Morris 2015). They proposed a shortened version of the MHQ that maintains the separation between affected and non-affected hands (B. Chung and Morris 2015); however, this version has not undergone psychometric testing.

2.3.2.3.1 Brief MHQ

Original authors of the MHQ recognised the benefit of a shortened version of the MHQ. Studies of reliability measures indicated item redundancy with Cronbach alpha values greater than 0.9 (B. Chung and Morris 2014). It is accepted that longer surveys yield lower response rates and problems with data quality. The development of the Brief MHQ aimed to keep the validity and reliability of the original MHQ, with improved implementation in multicentre and nationwide

studies (Waljee et al. 2011). Unlike its predecessor, the Brief MHQ does not distinguish between laterality of hand symptoms. Item reduction was performed using a concept-retention technique, which allows the retention of items that are clinically relevant regardless of their statistical value. Two items were kept from each scale, resulting in 12 items in the Brief MHQ. The choice of the items was based on their correlation with the original MHQ score. A score can only be calculated if all 12 items are completed (Waljee et al. 2011).

This technique of item reduction has the disadvantage that items with high clinical importance overall may be removed if they are ranked lower than second within the subscale ranking (B. Chung and Morris 2015). This methodology has been criticised because of its subjective nature compared with other psychometric approaches such as the Rasch or equidiscriminative techniques (Waljee et al. 2011).

2.3.2.4 Modern Activity Subjective Survey of 2007

The Modern Activity Subjective Survey of 2007 (MASS07) was designed as a short subjective functional assessment of the wrist and hand to give clinicians the ability to quantify patients' functional limitations during modern high-frequency activities (Alexander, Franko et al. 2008). The developers specifically aimed to capture the use of modern technology such as computers and handheld devices (Alexander et al. 2008). The developers generated a list of activities and the resultant draft measure was administered to a group of 50 patients. Feedback resulted in the instrument consisting of 10 questions, each graded from 0 to 10, with the total score ranging from 0 to 100 and higher scores reflecting worse ability to function.

2.3.2.5 Hand Assessment Tool

The Hand Assessment Tool (HAT) is a region-specific activity limitation measure for the hand, wrist and forearm axis. It includes domains of functional tasks, pain symptomatology and hand

aesthetics. EFA was performed during instrument development and identified a seven-factor model (Naidu, Panchik and Chinchilli 2009); despite this, the HAT is summed to give a single summary score. Designed to measure activity limitation as a direct result of the individual's health state without assistance from others or devices, as per the ICF definition 'difficulties an individual may have in executing activities' (Naidu, Panchik and Chinchilli 2009), this instrument deliberately avoids items relating to work activities to avoid possible self-report bias, which is distinct from the DASH and MHQ.

2.3.3 Condition-Specific Patient-Reported Outcome Measures

There were eight PROMs identified that focus on various conditions of the hand. The domain analysis, publication date and methodological technique are summarised in Table 2.7. The development and validation analysis are summarised in Table 2.8. The in-depth domain comparison between instruments is summarised in Table 2.9.

Table 2.7: Condition-specific PROMs

Instrument	Year	Author	Country of origin	Domains	Methods
Carpal tunnel syndrome					
BCTS	1993	Levine	US	·Symptom severity ·Functional status	CTT
*CTS-6	2009	Atroshi	Sweden	·Symptom severity	IRT
Rheumatoid arthritis					
Hand-VAS	2002	Massy-Westropp	Australia	·Handicap level	
Osteoarthritis					
AUSCAN	2002	Bellamy	Canada / Australia	·Pain ·Stiffness ·Physical disability	CTT
Base of thumb arthritis					
Nelson Score	2007	Citron	UK	·Pain ·Disability	CTT
TDX	2017	Noback	USA	·Arthritic thumb pain ·Thumb pain ·Satisfaction thumb	CTT
TASD	2016	Becker	USA	·Symptoms ·Disability	CTT/Factor analysis
Dupuytren's contracture					
SDSS	2014	Mohan	UK	·Disability ·Discomfort ·Activities - personal - domestic - work/social - hobbies	CTT

* Developed from an existing PROM.

Table 2.8: Development and validation criteria of condition-specific PROMs

	Condition-specific PROMs							
	SDSS	BCTQ	CTS-6	Hand-VAS	AUSCAN	TDX	TASD	Nelson Score
Item generation								
Patient interviews	•	•	*		•	•	•	•
Literature					•	•	•	•
Expert opinion		•	*		•		•	
Develop conceptual model								
Item reduction								
Expert opinion						•		
Item redundancy					•	•	•	•
Endorsement frequencies	•	•			•			
Missing data		•				•		
Factor analysis			•			•	•	
Tests of scaling assumptions								
Item misfit (Rasch/IRT)								
Psychometric analysis								
Acceptability	•					•		
Internal consistency reliability	•	•	•		•	•	•	•
Item total correlations			•					
Interrater reliability								
Test–retest reliability	•	•	•	•	•	•		•
Validity within scale						•	•	

Validity comparison with other measures	•	•	•	•	•	•	•	•
Validity hypothesis testing		•			•	•	•	
Responsiveness	•	•	•		•	•		•

* Developed from an existing PROM.

Table 2.9: Domain analysis of condition-specific PROMs

	Condition-specific PROMs							
	SDSS	BCTQ	CTS-6	Hand-VAS	AUSCAN	TDX	TASD	Nelson Score
Physical functioning								
Limitations of the whole upper limb								
Limitations of hand/wrist	•				•			
Limitations of hand/digits	•			•	•	•	•	•
Limitations of thumb				•		•	•	•
Ability to perform ADLs	•	•			•	•	•	•
Ability to use smartphone/modern technology								
Ability to work	•							
Ability to travel	•							
Ability to participate socially	•					•		•
Symptoms								
Pain issues	•	•	•		•	•	•	•
Sensory changes (tingling/numbness)		•	•					
Stiffness					•		•	
Swelling							•	
Weakness		•				•	•	•
Reduced ROM						•	•	•
Insomnia		•	•			•		•
Change in appetite		•	•			•		
Health-related QOL								
Satisfaction with treatment								
Satisfaction with outcome/overall assessment						•		•
Inconvenience of medical/hospital								
Concerns re postop complications/recovery								
Expectations								
Psychological functioning								
Self-confidence/self-esteem								

Avoidance of uncomfortable situations				
Negative feelings about self				
Change in mood			•	•
Body image				
Concerns regarding scarring				
Self-consciousness				
Satisfaction with hand appearance	•			
Sexual functioning				

* Developed from an existing PROM.

2.3.3.1 Carpal Tunnel Syndrome

2.3.3.1.1 Boston Carpal Tunnel Questionnaire

The BCTQ is the original condition-specific PROM in the field of hand surgery. It was developed with input from hand surgeons, rheumatologists and patients, who identified six domains important in the evaluation of CTS: pain, paraesthesia, numbness, weakness, nocturnal symptoms and overall function (Levine et al. 1993). The resultant Symptom Severity Scale is composed of 11 items, each scored from 1 (*mildest*) to 5 (*most severe*). The panel also named 12 activities that are affected by CTS; during pilot testing, the number of items was reduced to eight, and these items make up the Functional Status Scale. Each item is scored from 1 (*no difficulty*) to 5 (*cannot perform the activity at all*). The mean of each scale is calculated separately to establish the overall symptom severity and overall functional status scores. The recall period requires the respondent to answer in relation to a 'typical twenty-four-hour period during the past two weeks'.

The developers of the BCTQ correlated the scores from their scales to a variety of physical measurements, such as grip strength, two-point discrimination and monofilament testing, and found only moderate correlations at best (Levine et al. 1993). The clinically significant change in scores was 0.47 in a carpal tunnel surgery population (Amirfeyz et al. 2009). De Carvalho

Leite, Jerosch-Herold and Song (2006) performed a systematic review of the psychometric properties of the BCTQ and found it to be a valid, reliable, responsive and acceptable instrument for outcomes of CTS.

2.3.3.1.2 Six-Item Carpal Tunnel Syndrome Scale

Atroshi, Lyrén and Gummesson (2009) and Atroshi et al. (2011) performed EFA and IRT-based analysis on the BCTQ, which resulted in the production of the Six-Item CTS Scale (CTS-6), which maintained similar measurement properties to the original BCTQ.

2.3.3.2 Rheumatoid Arthritis

2.3.3.2.1 Visual Analogue Scale–Hand

That Hand-VAS is a single-item measure developed for self-report of handicap caused by rheumatoid arthritis of the hands. It poses the question 'Considering your needs for everyday life, what is your handicap level due to rheumatoid arthritis in your hands?' (Massy-Westropp, Ahern and Krishnan 2005). With responses recorded on a 10 cm horizontal VAS, from 0 (*no handicap*) to 10. The Hand-VAS is to be used in conjunction with the VAS-RA, a global measure of handicap due to rheumatoid arthritis.

The developers of this single-item PROM did not publish how the question was developed but did perform validation of the measure with patients. No correlation was found between the VAS-Hand and other measures of impairment, pain and disability as measured with the MHQ or SF-36 scales or objective measures such as grip strength, pinch strength and active ROM (Massy-Westropp, Ahern and Krishnan 2005).

2.3.3.3 Osteoarthritis

2.3.3.3.1 Australian/Canadian Osteoarthritis Hand Index

The Australian/Canadian Osteoarthritis Hand Index (AUSCAN) is designed to provide a reliable, valid and responsive self-administered questionnaire to probe pain, stiffness and physical disability in patients with osteoarthritis of the hand (Bellamy, Campbell, Haraoui, et al. 2002b). Item generation was based on existing instruments and the expert opinion of rheumatologists, orthopaedic surgeons and physiotherapists. Patient interviews were used to rate items according to their prevalence, perceived importance and attribution of their hands. Item rationalisation was based on the frequency of patient experience, and the relative importance of the item. The analysis includes excluded items to examine differences in item characteristics in early versus late item exclusion. The resulting instrument is composed of 27 items, divided among three subscales: pain (5), stiffness (1) and physical function (9). The scores may be given for each subscale or added together for an overall score. The developers of the AUSCAN published a subsequent paper that establishes the clinimetric properties of the AUSCAN (Bellamy, Campbell, Haraoui, et al. 2002a). EFA performed by Allen et al. (2006) confirmed the internal consistency, factor structure and construct validity of this instrument.

2.3.3.4 Base of Thumb Arthritis

2.3.3.4.1 Trapeziometacarpal Arthrosis Symptoms and Disability Questionnaire

The Trapeziometacarpal Arthrosis Symptoms and Disability Questionnaire (TASD) aims to evaluate the symptom severity and disability associated with trapeziometacarpal arthritis and allow for comparison of scores pre- and post-intervention (Becker et al. 2016). The initial item bank was developed from literature, expert opinion and patient input (Becker et al. 2016), although the details of the patient input are not specified. Item reduction was performed on the

basis of expert judgement to remove redundant, gender-specific, dominant hand or ambiguously worded items (Becker et al. 2016). The instrument consists of 12 items on two subscales (symptoms and disability), both measured with a 5-point Likert scale. Scoring is performed as per the QuickDASH, with the total score given out of 100, with higher scores representing more symptoms and greater disability. Principle component analysis confirmed the questionnaire's structure of two scales. Internal consistency of each of the subscales was good, and the instrument showed good convergent validity, with significant correlations with the DASH with moderate to high coefficients (Becker et al. 2016).

2.3.3.4.2 Thumb Disability Examination

The Thumb Disability Examination (TDX) is a PROM designed specifically for basal joint osteoarthritis (Noback et al. 2017). Item generation was performed by collating items from existing scales, and this was enhanced by patient interviews and finally expert input. A panel of experts and patients carried out item reduction with particular emphasis on removing items that were only relevant to the dominant hand, items that did not require force to be put through the thumb and otherwise irrelevant items. A preliminary questionnaire was then developed and underwent review by an experienced epidemiologist. The pilot testing of the TDX was conducted on 50 patients with basal joint arthritis. Factor analysis was performed to remove any items that displayed cross-loading between domains. The final instrument is composed of 20 items, divided into three domains; pain, function and satisfaction with the affected thumb. The overall score is given between 0 and 100, with higher scores correlating to greater disability. The TDX has a higher level of responsiveness to treatment and is less burdensome on patients than the DASH (Noback et al. 2017).

2.3.3.4.3 Nelson Hospital Score

The Nelson Hospital Score (Nelson Score) was developed from surgeon–patient interviews during the consultation to establish common symptoms both and after surgery (Citron, Hulme and Wardle 2007). The questions posed during these interviews were based on existing instruments. Items that were specific to the dominant hand or respondent's gender were removed to shorten the questionnaire. This PROM is criticised for employing poor methodological quality and lacks factor analysis, and as a result is rarely utilised in clinical practice (Becker et al. 2016).

2.3.3.5 Dupuytren's Disease

2.3.3.5.1 Southampton Dupuytren's Scoring Scheme

The SDSS aims to provide a way to quantify the disability caused by Dupuytren's disease, which allows for the prioritisation of treatment and audit of outcomes (Mohan et al. 2014). Developed in keeping with the recommendations of the Derby Outcomes Conference, item generation was based on patient survey. Item reduction was performed according to the frequency that problems were mentioned. This resulted in five items, each rated between 0 (*no problem*) and 4 (*severe problem*), with a total score of 0–20. The SDSS has been shown to be more sensitive to change than the QuickDASH when measuring outcome at 6 months post-surgery. Interestingly, there was no correlation found between either the preoperative SDSS or QuickDASH scores and preoperative deformity measured with a goniometer and total extension deficit (Mohan et al. 2014). Thus, those with very severe deformities did not necessarily have worse preoperative scores, reflecting the ability of individuals with a prolonged hand deformity to use compensatory mechanisms to participate in their roles.

2.4 Discussion

The ideal PROM for the field of hand surgery is an instrument that engages with patients by asking meaningful questions, and is easy to complete and score to aid clinical implementation. The PROM must be a valid and reliable instrument that allows for accurate measurement at a single point in time. To enable the measurement of intervention it is important that the PROM can measure accurately at the individual level over time. This means that the score needs to be in interval format so that there is consistency in the measurement between integers (as discussed in Chapter 1). This ensures that a change in score from 20 to 40 is of similar size as a change in score from 40 to 60. Only when we have a scientifically and mathematically sound basis to our measurement can we hope to have meaningful clinical inclusion of PROMs.

The instruments discussed above aim to measure the patient's experiences of their hand surgery or hand condition. Despite the multitude of instruments that have been identified, it remains the case that 'no gold standard, objective criterion measurement tools for patient-rated hand outcomes exist' (B. Chung and Morris 2014). Only the POS-Hand/Arm has been developed in accordance with the criteria of the SAC (Cano et al. 2004) and FDA (2009), and has been proved to be psychometrically sound by an independent author (A. Scott et al. 2009); however, its use within the surgical literature has been minimal and it does not employ modern development methodology.

The most widely utilised PROM for the upper limb is commonly accepted as the DASH (Naidu, Panchik and Chinchilli 2009). This instrument measures disability without reference to the side of hand dominance or to the side of injury. This disregard for these important influencing factors leads to potential misjudgement of outcomes. Further, the DASH has been proved to fail unidimensionality testing (Noback et al. 2017); consequently, the meaning of the overall

DASH score is unclear and potentially misleading to clinicians as it does not reflect a single underlying concept (Cano et al. 2011a).

The attempt to assess the upper limbs as a functional unit is attractive; however, there are disadvantages in this approach. The effects of an intervention may be confused because of other injuries or conditions. Because of the frequency of coexisting upper limb pathologies, the DASH cannot discriminate the changes specific to the intervention of interest (Noback et al. 2017). Therefore, use of an upper limb regional instrument has the risk of minimising the effects of hand-specific interventions (Kamal and Hand Surgery Quality Consortium 2016). Further, it has been discussed in the literature that the DASH is not specific for the upper limb, measuring disability for patients with lower limb pathology and cervical spine pathology (Khan et al. 2008; Huisstede et al. 2009). This lack of specificity is what led to the Manchester group revising the DASH to create the M2DASH in an effort to make the instrument more specific to the upper limb (Khan et al. 2008).

Since the design of the most commonly utilised hand PROMs, there has been much progress in the field of questionnaire science and design. CTT is based on the assumptions that (a) the more items on a scale the less it will be affected by random error, (b) reliability and validity estimates are only applicable to the sample studied or a population well represented by the sample and (c) any changes to the scale would require re-evaluation of the psychometrics of the scale (T. Packham and MacDermid 2013). The techniques used in the field have transitioned from the CTT techniques used to construct many of the instruments identified in this study; to the more modern IRT and Rasch analysis techniques. IRT was used in the development of the PROMIS-PF-UE. The use of this improved method results in instruments that are more scientifically sound and allow for valid application between populations. One of the main applications of PROs in hand surgery is their use to measure change over time, often pre- and post-intervention. Instruments that are currently widely utilised in the field, such as the DASH, MHQ and

PRWHE, all give a score that is in ordinal format, which is not theoretically suitable for measuring change over time. Ordinal data do not have consistent spacing between digits and therefore measuring change cannot be performed accurately. An advantage of these modern psychometric techniques is that they produce an interval scoring system, which means that it is mathematically sound to compare measurement outcomes over time.

Condition-specific instruments are more sensitive to relevant surgical interventions. As the items are more specific to the individual's clinical condition, there is improved face validity and potentially fewer missing data. There has been an increase in the PROMs available to measure the outcome of specific hand conditions such as the BCTQ (Levine et al. 1993), SDSS (Mohan et al. 2014) and TDX (Noback et al. 2017). The improved sensitivity and specificity of condition-specific instruments allow for a more appropriate measurement of change resulting from an intervention. However, none of the condition-specific instruments in this review gives scores in interval format, so measuring change is still not mathematically sound. The difficulty with the clinical implementation of condition-specific instruments is that in any given hand clinic you would require a whole suite of different questionnaires, each with their own instructions, format and scoring systems. This would not result in a practical option for implementation in the majority of hand services.

What is lacking in this field is an instrument constructed using the gold standard of PRO development according to the SAC and US FDA, and that allows for measurement of specific hand conditions by using a series of independently functioning scales. Such an instrument would comprise a set of scales developed using IRT or Rasch analysis to give interval format scores that can be used to measure change over time and to compare scores across patient populations. This would ensure that the most sensitive and specific scales are being utilised but with the convenience and clinical utility of a single PRO system.

Hand aesthetics is an important issue for many hand surgery patients, particularly those with osteoarthropathies such as rheumatoid arthritis and Dupuytren's contracture (Bogoch, Escott and Ronald 2011; Johnson et al. 2015). Hands are highly visible within daily practices, and anomalies of the hands often draw unwanted attention. Despite this acknowledged importance, there is no accepted measure for hand aesthetics (Johnson et al. 2015). Hand aesthetics is a scale in the MHQ and a single question in the PRWHE. Hand aesthetics has not been suitably explored as a concept. This may be due to the desire of scale developers and clinicians to have a PROM that produces a single overall score. Hand aesthetics often will not align with other measures of outcome and thus is not suitable to be combined with other domains such as function and satisfaction.

Patient satisfaction and the perioperative experience is an area that is not adequately explored by currently available PROMs. There are several areas within hand surgery where the functional outcomes between different surgical approaches are considered equivalent. The preference of the patient should be a consideration in assessing which technique is employed. Patient satisfaction is a complex concept that is not suitably measured by a single item, as has been the common practice (Graham 2016). As written by Graham (2016, 931), 'it is crucial that we approach the measurement of satisfaction with the same methodological rigour and insight that we consider all of our clinical outcomes'. The current PROs that attempt to measure patient satisfaction do not tend to address the complexity of the concept. The POS-Hand/Arm, MHQ and PEM have a subscale dedicated to measuring patient satisfaction. The POS-Hand/Arm has items relating to scar lumpiness, results compared with expectations, speed of recovery compared with patient expectation and whether the patient would recommend the operation to a friend with a similar issue. The MHQ has a cluster of six questions exploring patient satisfaction with their hand function, motion, strength pain and sensation. The PEM (Macey et al. 1995) has three items that explore patient satisfaction, being patient satisfaction with the

hospital, patient satisfaction with their hand and an attempt to gauge whether the patient's expectations have been met given their original injury. This is an area that the field of hand surgery needs to establish a better understanding of so that we can improve the experience of hand surgery for our patients.

2.5 Summary

This systematic review of PROMs relevant to patients with hand conditions has shown that there are few instruments that satisfy international guidelines for PROM development. There is still much work to be done in this field to achieve accurate clinically integrated outcome measurement. There is a need for an instrument developed according to international guidelines, and to allow for measurement of all concepts that are relevant to patients with hand conditions (Lohr 2002). The use of IRT or Rasch analysis to develop these scales will result in interval-level scores that can be used to measure change over time and to compare scores across patient samples with accuracy. This would ensure that the most sensitive and specific scales are being employed, with the convenience and clinical utility of a single PROM system. Later chapters of this manuscript will discuss the development of such an instrument, the HAND-Q.

Chapter 3: Prospective Cohort Study to Explore the Acceptability of Patient-Reported Outcome Measures to Patients of Hand Clinics

3.1 Introduction

Before embarking on the development of a PROM, it is vital to learn as much as possible from the currently available instruments. As discussed in Chapter 2, there are many PROMs available in the field of hand conditions. The most commonly cited PROMs in the relevant literature are the DASH (Hudak et al. 1996), PRWHE (MacDermid 1996) and MHQ (K. Chung et al. 1998). The intention of a newly developed PROM would be to produce an instrument that is suitable for integration into routine clinical care. Before pursuing this, it is important to assess whether patients of hand clinics are willing and able to complete PROMs while waiting for their clinical review appointment. Researchers in the field of oncology have found that incorporation of PROMs into the clinical setting has lead to improvements in communication between the patient and clinician, and overall improved patient satisfaction (Detmar 2002). There is however considerable difference in the environment, patient cohort and clinical interaction when comparing hand clinics to oncology clinics.

The primary aim of this study was to assess the acceptability of PROM completion in the hand clinic. A secondary aim of this study was to examine patient responses to the selected existing PROMs to learn about which items are unanswered and how this missing data affect the ability of the instrument to produce a score.

Acceptability of a PROM was assessed by the participation, retention and withdrawal rates of participants, and data quality (Fayers and Machin 2013). The quality of the data was established by examining the proportion of missing data and the number of questionnaires that were unable to be scored (McHorney et al. 1994; Cano, Klassen and Pusic 2009).

3.2 Methodology

This project was a prospective randomised cohort study. The study received ethics and governance approval from the Southern Adelaide Clinical Human Research Ethics Committee (SAC HREC OFR#404.16; Appendix B.1). Inclusion criteria were any hand clinic patient aged 14 years or older who could read English and understand the questionnaires.

3.2.1 Participant Recruitment

The study was conducted at the Plastic and Reconstructive Surgery Hand Clinic of Flinders Medical Centre between February and June of 2017. Participants were recruited by a member of the research team after they had checked in with clerical staff for their scheduled appointment. The study was explained to the participant and they were provided with the Participant Information and Consent form and were given the opportunity to ask questions about the research. Participants were advised that the study was to compare the three instruments to decide which was most suitable for future use in the clinic. Participants were not specifically told that the completeness of their responses would be a variable of interest in the study. An online random number generator performed randomisation, with the participant enrolment number deciding each participant's PROM allocation. Participants received a paper version of one of the included PROMs: the DASH (Hudak et al. 1996), MHQ (K. Chung et al. 1998) or PRWHE (MacDermid 1996). In addition, each participant received the written question 'How long did it take you to complete this form?' and was asked to respond by

selecting from the following options: ‘less than 5 minutes’, ‘5–10 minutes’, ‘10–15 minutes’ or ‘greater than 15 minutes’.

Once a participant was assigned to receive a PROM, they would then receive the same questionnaire on each subsequent clinic visit, with a minimum time period of 2 weeks between administration. This longitudinal aspect was to assess dropout rates over repeated administration. For the purposes of the study, the maximum number of questionnaires administered to any single participant was capped at five.

Participants were left to independently read the directions on the questionnaire and complete the form to the best of their ability. In some cases, participants had family members aid them by acting as a scribe. Questionnaires were not reviewed for completeness at the time of collection. Participant demographics including age, gender and hand pathology were noted by a member of the research team. Information relating to the affected side was collected retrospectively from the participant’s case notes. Data were collated in a database (Microsoft Access 2010).

3.2.2 Patient-Reported Outcome Measures

The PROMs selected for this study were introduced and discussed in greater detail in Chapter 2. In brief, the DASH (Appendix B.2) is a 30-question PROM that measures disability of the upper limbs as a single functional unit (Hudak et al. 1996). The PRWHE (Appendix B.3) is a 17-question PROM that measures pain and function. (MacDermid 1996). The MHQ (Appendix B.4) has 62 questions and is the only included PROM that can give a score for the left and right hand separately; the mean of these scores results in the total score, which is only relevant for patients with bilateral hand problems. For unilateral hand problems, the MHQ has 37 core questions. Specification of the side/s affected is required to calculate the relevant MHQ score (K. Chung et al. 1998).

3.2.3 Statistical Methods

Sample size calculation was calculated a priori on the basis of the null hypothesis that the proportion of scorable questionnaires was not dependent on the questionnaire group. A sample size of 100 (per group, total n = 300) produces a two-sided 95% confidence interval with a width of 0.178 when the sample proportion is 0.25 (Newcombe 1998). Descriptive statistics were used to describe the study sample, the completeness of the data and the time taken to complete the questionnaires. Data normality was determined by review of the histogram, skewness and kurtosis (Kim 2013). The independent samples two-tailed *t*-test was used to compare the means between groups where the data estimated a normal distribution. The relationship between categorical variables was explored with chi-square tests of independence using Yates's correction for continuity (when correlation between greater than two dichotomous variables was performed) (Pallant 2016). Bonferroni adjustment of the alpha level (0.05) was used for post hoc analyses (Bland and Altman 1995).

3.3 Results

3.3.1 Sample Population

The characteristics of the study participants and non-participants are summarised in Table 3.1. Age data approximated a normal distribution with skewness / kurtosis of -0.150 /-.0974 and 0.176/-0.890 for non-participants and participating groups respectively. In terms of age and gender distribution, there was no significant difference between the sample who participated and those who did not. The two-tailed *t*-test showed no significant difference in the age of participants and non-participants ($t(575) = 1.76$, $p = 0.08$). The chi-squared test for independence indicated no significant association between gender and whether or not a subject chose to participate ($\chi^2(1, n = 577) = 0.37$, $p = 0.54$).

3.3.2 Participant Rate, Retention Rate and Withdrawal Rate

Overall, 85% of those approached were willing to participate in the study; both genders were similar in their participation rate: males (84%) and females (86%). Of those who declined to participate, many gave practical reasons such as they did not have their reading glasses or they were in too much discomfort to complete the forms. Recruitment, enrolment and randomisation are summarised in Figure 3.1. The demographics of participants assigned to each PROM group are shown in Table 3.2. The hand pathology was categorised into traumatic and elective conditions, which were similar between groups. Further detail on elective conditions is listed in Table 3.3.

A total of 137 participants attended the clinic for a second visit during the study period, 129 of whom agreed to continue in the study, which gives a retention rate of 94%. There were nine participants who withdrew from the study, four males and five females. Five of those who withdrew were from the DASH cohort, the remaining four were from the MHQ cohort. There were no withdrawals from the PRWHE cohort. Participants who withdrew gave reasons such as lack of interest, physical discomfort and the length of the questionnaire.

Table 3.1: Demographics of participants and non-participants

		Participants	Non-participants
Sample population (n = 577)		491	86
Mean age (range)		49 (14–91)	52 (18–89)
Gender n (%)	Male	282 (57)	53 (62)
	Female	209 (43)	33 (38)

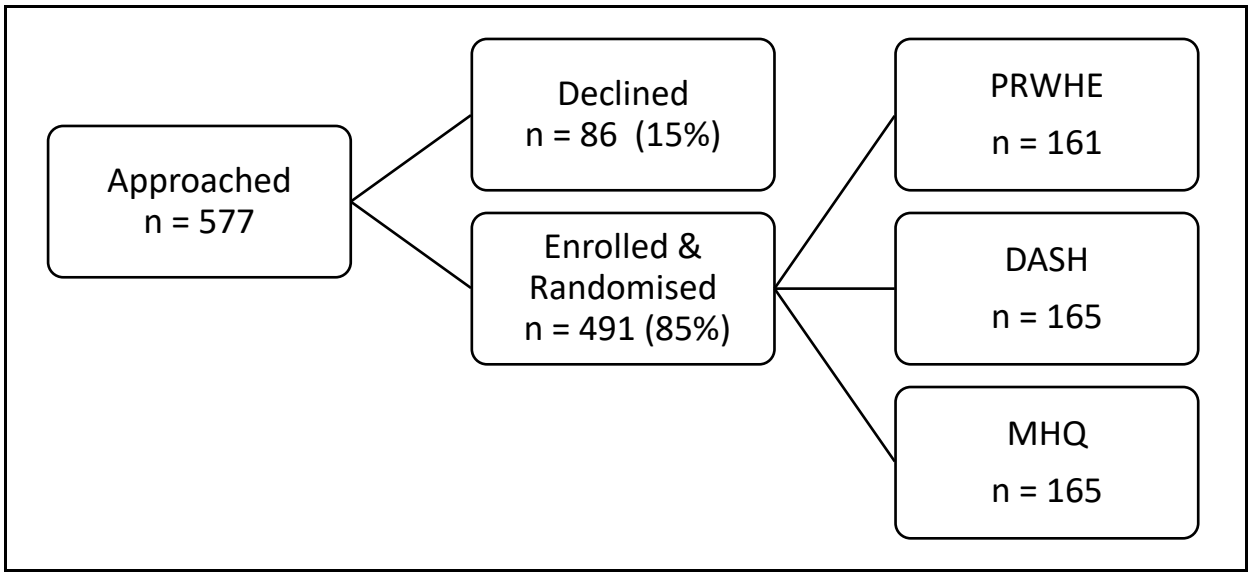

Figure 3.1: Flow diagram of study participant enrolment and randomisation into groups.

Table 3.2: Characteristics of participants in each questionnaire group

Participants		**Questionnaire group**		
		PRWHE	DASH	MHQ
Number (n = 491)		161	165	165
Age in years mean (range)		47 (14–88)	46 (14–90)	49 (14–90)
Gender, n (%)	Males	86 (53)	98 (59)	97 (59)
	Females	75 (47)	67 (41)	68 (41)
Pathology, n (%)	Trauma	82 (51)	96 (58)	94 (57)
	Elective[#]	79 (49)	69 (42)	71 (43)

[#] Further detail of the elective conditions is shown in Table 3.3.

Table 3.3: Elective conditions in each questionnaire group

Elective pathology	Questionnaire group		
	PRWHE	DASH	MHQ
Arthropathy, n (%)	8 (5)	6 (4)	7 (4)
Trigger finger, n (%)	7 (4)	8 (5)	11 (7)
Nerve compression, n (%)	30 (19)	25 (15)	23 (14)
Tumour, n (%)	4 (2)	4 (2)	1 (1)
Dupuytren's contracture, n (%)	8 (5)	11 (7)	11 (7)
Ganglion, n (%)	0 (0)	1 (1)	2 (1)
Undefined, n (%)	22 (14)	14 (8)	16 (10)

3.3.3 Incomplete Questionnaires

A total of 673 questionnaires were submitted, 328 (48.7%) of which were incomplete (i.e., missing at least one response). The percentage of incomplete responses was 35.0% of the PRWHE group, 29.9% of the DASH group and 81.3% of the MHQ group. These differences were significant (χ^2 (2, n = 673) = 144.92, $p < 0.000$). Table 3.4 shows the summary data for the three PROMs.

Table 3.4: Questionnaires completed

Questionnaire		PRWHE	DASH	MHQ	Total
Items per PROM		15	30	62	107
PROMs returned		214	234	225	673
Items administered		3210	7020	13950	24180
Items missing a response n (%)		135 (4.2)	271 (3.9)	1644 (11.8)	2050 (8.5)
Number of questionnaires with (x) items left unanswered	x				
	0	139	164	42	345
	1	37	38	16	91
	2–3	29	11	5	45
	≥4	9	21	162	192
Incomplete questionnaires n (%)		75 (35.0)	70 (29.9)	183 (81.3)	329 (49)
Questionnaires that could not generate a score n (%)		0 (0)	22 (9)	25 (11)	47 (7)

3.3.3.1 Sub-analysis of Michigan Hand Questionnaire

The MHQ has questions that specifically ask about the left and right side independently, regardless of the side(s) of interest. When including only the data relating to the side/s of interest, number of incomplete MHQs decreased from 81% to 33%, bringing the completion rate to a similar level as the other PROMs (Table 3.5).

Table 3.5: Sub-analysis of MHQ responses

	All MHQ	Side of interest only	
		Unilateral condition	Bilateral condition
Items per PROM	62	37	62
PROMs completed	225^	199	8
Items administered	13950	7363	496
Items missing a response, n (%)	1644 (11.8)	442 (6.0) 497 (6.3)*	55 (11.0)
Incomplete questionnaires, n (%)	183 (81.3)	65 (32.7) 69 (33.3)*	4 (50.0)
Questionnaires that could not generate a score, n (%)	25 (11.1)	8 (4.0) 9 (4.3)*	1 (12.5)

***** Calculated from the MHQ core questions relevant to the affected side in participants with a unilateral condition and all questions for those with a bilateral condition, n = 207.

^ There were 18 of the 225 MHQ PROMs where the side of interest was not known; these cases were included in the overall analysis.

3.3.3.2 Analysis by Item

The frequency of missing data for each item of the PROMs was calculated (see Table 3.6, Table 3.7 and Table 3.8). The most frequently unanswered question in the DASH was 'Q21. Sexual activities', with 15.4% of participants not giving an answer. The PRWHE had two questions that were most unanswered by 8.9% of participants: 'Q3. Rate your pain … When lifting a heavy object' and 'Q14. Rate the amount of difficulty … Work (your job or usual everyday work)'. The most frequently unanswered question in the MHQ was 'QVI6. Satisfaction with … Sensation (feeling) of your _ hand', with 12% of participants not answering this item.

Table 3.6: DASH item analysis

Items		Frequency of missing data	%
1	Open a tight or new jar	1	0.4
2	Write	7	3
3	Turn a key	2	0.9
4	Prepare a meal	4	1.7
5	Push open a heavy door	3	1.3
6	Place an object on a shelf above your head	5	2.1
7	Do heavy household chores (e.g., wash walls, wash floors)	3	1.3
8	Garden or do yard work	3	1.3
9	Make a bed	3	1.3
10	Carry a shopping bag or briefcase	5	2.1
11	Carry a heavy object (over 10 lbs)	5	2.1
12	Change a lightbulb overhead	14	6
13	Wash or blow dry your hair	4	1.7
14	Wash your back	4	1.7
15	Put on a pullover sweater	3	1.3
16	Use a knife to cut food	4	1.7
17	Recreational activities which require little effort (e.g., card playing, knitting).	7	3
18	Recreational activities in which you take some force or impact through your arm, shoulder or hand (e.g., golf, hammering, tennis)	10	4.3

19	Recreational activities in which you move your arm freely (e.g., playing frisbee, badminton)	16	6.8
20	Manage transportation needs (getting from one place to another)	8	3.4
21	Sexual activities	36	15.4
22	During the past week, to what extent has your arm, shoulder or hand problem interfered with your normal social activities with family, friends, neighbours or groups?	12	5.1
23	During the past week, were you limited in your work or other regular daily activities as a result of your arm, shoulder or hand problem?	12	5.1
24	Arm, shoulder or hand pain	13	5.6
25	Arm, shoulder or hand pain when you performed any specific activity	15	6.4
26	Tingling (pins and needles) in your arm, shoulder or hand	16	6.8
27	Weakness in your arm, shoulder or hand	15	6.4
28	Stiffness in your arm, shoulder or hand	18	7.7
29	During the past week, how much difficulty have you had sleeping because of the pain in your arm, shoulder or hand? (circle number)	12	5.1
30	I feel less capable, less confident or less useful because of my arm, shoulder or hand problem. (circle number)	11	4.7
Total		271	3.9

Table 3.7: PRWHE item analysis

Stem	Items	Frequency of missing data	%
PAIN Rate the average amount of pain in your wrist/hand over the past week:			
Rate your pain:			
1	At rest	4	1.9
2	When doing a task with a repeated wrist/hand movement	7	3.3
3	When lifting a heavy object	19	8.9
4	When it is at its worst	8	3.7
5	How often do you have pain?	7	3.3
FUNCTION			
A. Specific Activities: Rate the amount of difficulty you experienced performing each of the items listed below over the past week:			
6	Turn a door knob using my affected hand	5	2.3
7	Cut meat using a knife in my affected hand	9	4.2
8	Fasten buttons on my shirt	7	3.3
9	Use my affected hand to push up from a chair	7	3.3
10	Carry a 10 lb object in my affected hand	8	3.7
11	Use bathroom tissue with my affected hand	10	4.7
B. Usual Activities: Rate the amount of difficulty you experienced performing your usual activities in each of the areas listed below, over the past week:			
12	Personal care activities (dressing, washing)	4	1.9
13	Household work (cleaning, maintenance)	5	2.3

14	Work (your job or usual everyday work)	19	8.9
15	Recreational activities	16	7.5
APPEARANCE – OPTIONAL*			
16	How important is the appearance of your hand?	14	6.5
17	Rate how dissatisfied you were with the appearance of your wrist/hand during the past week?	11	5.1
Total		135	4.2

* Appearance questions are not included in the calculation of missing data as they are optional.

Table 3.8: MHQ item analysis

Stem	Items	Frequency of missing data (L + R)	%	Frequency of missing data related to affected hand*	%
QI. The following questions refer to the function of your hand(s)/wrist(s) during the past week.					
1	Overall, how well did your _ hand work?	24	5.3	6	2.8
2	How well did your _ fingers move?	23	5.1	5	2.3
3	How well did your _ wrist move?	27	6	7	3.3
4	How was the strength in your _ hand?	25	5.6	6	2.8
5	How was the sensation (feeling) in your _ hand?	28	6.3	8	3.7
QII. How difficult was it for you to perform the following activities using your _ hand?					
1	Turn a door knob	29	6.5	8	3.7
2	Pick up a coin	29	6.5	8	3.7
3	Hold a glass of water	29	6.5	8	3.7
4	Turn a key in a lock	29	6.5	8	3.7
5	Hold a frying pan	31	6.9	8	3.7
QIIC. How difficult was it for you to perform the following activities using both of your hands?					
1	Open a jar	5	2.2	5	2.4
2	Button a shirt/blouse	7	3.1	7	3.4
3	Eat with a knife/fork	6	2.7	6	2.9
4	Carry a grocery bag	5	2.2	5	2.4

5	Wash dishes	8	3.6	8	3.9
6	Wash your hair	6	2.7	6	2.9
7	Tie shoelaces/knots	6	2.7	6	2.9
QIII. The following questions refer to how you did in your normal work (including both housework and school work) during the past four weeks.					
1	How often were you unable to do your work because of problems with your hand(s)/wrist(s)?	14	6.2	14	6.3
2	How often did you have to shorten your workday because of problems with your hand(s)/wrist(s)?	17	7.6	17	7.7
3	How often did you have to take it easy at your work because of your hand(s)/wrist(s)?	16	7.1	16	7.2
4	How often did you accomplish less in your work because of problems with your hand(s)/wrist(s)?	16	7.1	16	7.2
5	How often did you take longer to do the tasks in your work because of problems with your hand(s)/wrist(s)?	16	7.1	16	7.2
QIV. The following questions refer to how much pain you had in your hand(s)/wrist(s) during the past week.					
1	How often did you have pain in your _ hand(s)/wrist(s)?	39	8.7	13	6.0
2#	Please describe the pain you had in your _ hand(s)/wrist(s)?	186	41.3	27	14.9
3#	How often did the pain in your _ hand(s)/wrist(s) interfere with your sleep?	169	37.6	26	12.6
4#	How often did the pain in your _ hand(s)/wrist(s) interfere with your daily activities (such as eating or bathing)?	168	37.4	26	12.1
5#	How often did the pain in your_ hand(s)/wrist(s) make you unhappy?	168	37.4	26	12.1
QV. The following questions refer to the appearance (look) of your _ hand during the past week.					

1	I am satisfied with the appearance (look) of my _ hand.	53	11.8	19	8.8
2	The appearance (look) of my _ hand sometimes made me uncomfortable in public.	51	11.4	18	8.4
3	The appearance (look) of my _ hand made me depressed.	51	11.4	18	8.4
4	The appearance (look) of my _ hand interfered with my normal activities.	51	11.4	18	8.4
QVI. The following questions refer to your satisfaction with your _ hand/wrist during the past week.					
1	Overall function of your _ hand	51	11.4	17	7.9
2	Motion of the fingers in your _ hand	52	11.6	18	8.4
3	Motion of your _ wrist	52	11.6	19	8.8
4	Strength of your _ hand	52	11.6	18	8.4
5	Pain level of your _ hand	51	11.4	18	8.4
6	Sensation (feeling) of your _ hand	54	12	21	9.8
	Total	953	7.7	391	5.7

* Only missing data pertaining to the side of interest is counted.

[#] Questions QIV2, QIV3, QIV4 and QIV5 are excluded from the calculation of total missing data (in this table) as these questions are not always applicable; whether they should be answered is dependent on the respondent's answer to QIV1.

3.3.4 Analysis by the Ability to Generate a Score

Questionnaires were scored according to the published guidelines relevant to each instrument. Whether a questionnaire could be scored was dependent on how the scoring algorithm for that specific instrument managed missing data, as summarised in Table 3.9.

Table 3.9: Scoring rules for missing data for each PROM

PROM	Scoring missing data rules	Source
DASH	'A DASH score may not be calculated if there are greater than 3 missing items.'	Disabilities of the Arm, Shoulder and Hand (Hudak et al. 1996)
PRWHE	'If there is an item missing, you can replace the item with the mean score of the subscale.'	Patient-Rated Wrist Evaluation (PRWE) User Manual (MacDermid 2007)
MHQ	'Missing values in each scale may affect the validity of the scores. If 50% or more of the items in a scale are missing, then that particular scale cannot be scored. For scales with less than 50% missing, the average of the existing scale items may be imputed for the missing items.'	Michigan Hand Questionnaire website: http://mhq.lab.medicine.umich.edu/scoring-the-mhq

The PRWHE questionnaires all resulted in a score as any missing values can be substituted with the mean value of that subscale (MacDermid 2007). There is no maximum number of missing items allowable; thus, a subscale score can potentially be calculated even with only one item answered. An example of a participant's answers for the function subscale of the PRWHE is given below (Table 3.10); they only answered half of the items on the Function subscale, but a score could still be calculated.

Table 3.10: Example of responses for PRWHE Function subscale

Function subscale questions	Responses	With mean score substitution for missing responses
Q6	—	7.8
Q7	—	7.8
Q8	5	5
Q9	—	7.8
Q10	—	7.8
Q11	—	7.8
Q12	7	7
Q13	7	7
Q14	10	10
Q15	10	10

According to the scoring rules, the mean value of the given responses in the scale can be substituted for the missing data. Therefore, the mean value of the given responses in Table 3.10 is calculated as follows: (5 + 7 + 7 + 10 + 10)/5 = 7.8. Substituting the mean for the missing values allows for calculation of the Function score as follows: (7.8 + 7.8 + 5 + 7.8 + 7.8 + 7.8 + 7 + 7 + 10 + 10)/2 = 39

The DASH had 22 (9.4%) questionnaires that could not generate a score as they had more than three items unanswered. A total of 25 (11%) MHQ questionnaires were unable to produce a score for either one or both hands. When the side of interest was considered, 9 (4%) of the questionnaires could not be scored. Because of these varied approaches to managing missing data, the PRWHE was significantly more likely to result in a scorable questionnaire ($\chi^2(1, n = 673) = 4.30$, $p < 0.001$), and the DASH was significantly more likely to result in an un-scorable questionnaire ($\chi^2(1, n = 673) = 15.21$, $p < 0.001$).

3.3.5 Time Taken to Complete the Questionnaires

The participant-reported time taken to complete each questionnaire is shown graphically in Figure 3.3. In total, 82% of participants reported their questionnaire completion took less than 10 minutes, regardless of which questionnaire they received. The PRWHE was the quickest to complete, with 60% of respondents reporting being able to complete the questionnaire within 5 minutes. The MHQ was the slowest to complete, with only 21% reporting completion in less than 5 minutes; this difference was significant ($\chi^2(8, n = 673) = 83.91$, $p < 0.001$).

3.3.6 Participant Characteristics Relationship with the Ability to Generate a Score

There was no relationship between a participant's age or gender and whether they returned a score-generating questionnaire ($\chi^2(3, n = 491) = 6.788$, $p > 0.05$; $\chi^2(1, n = 491) = 0.06$, $p > 0.05$, respectively). There was also no association with the nature of a participant's condition (i.e., elective or traumatic) and the likelihood of them returning a questionnaire that could generate a score ($\chi^2(1, n = 491) = 0.789$, $p > 0.05$).

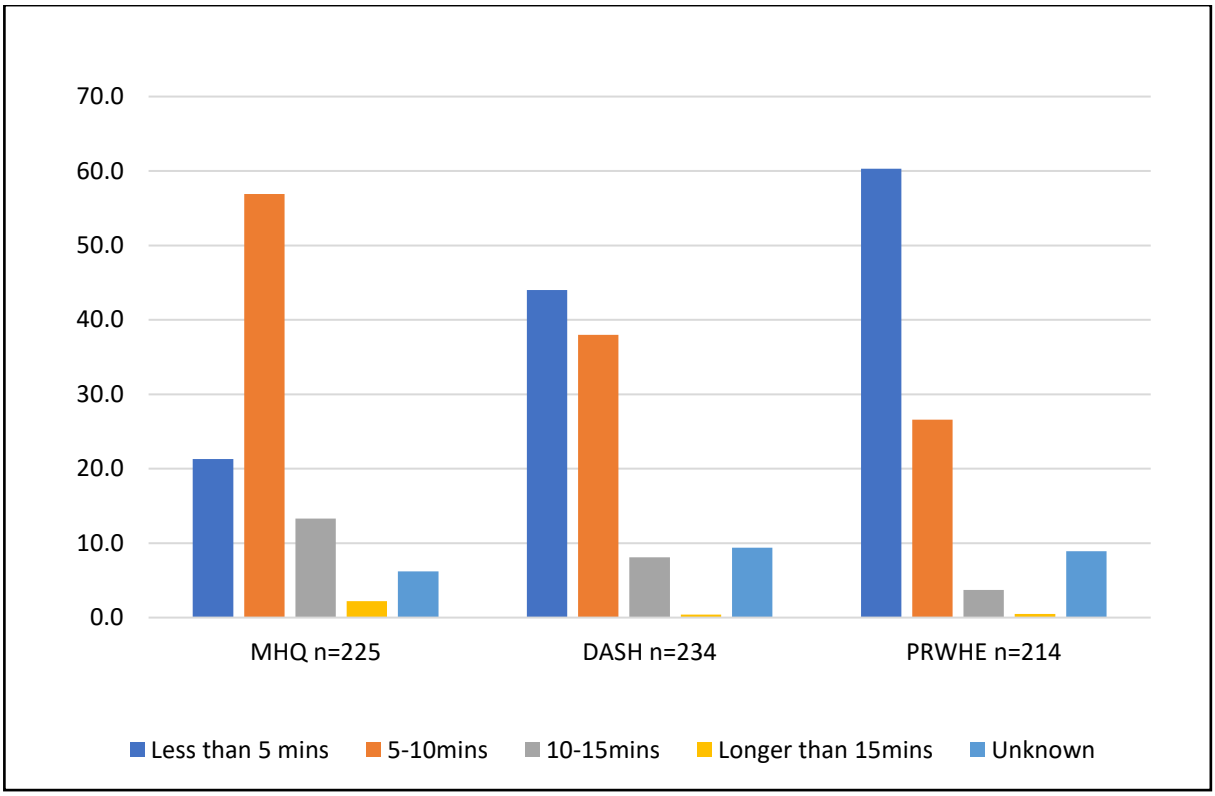

Figure 3.3: Patient-reported time taken to complete questionnaires.

3.4 Discussion

3.4.1 Participation Rate, Retention Rate and Withdrawal Rate

Administration of PROMs in the clinic waiting room resulted in a participation rate of 85%, which is similar to other studies with hand surgery patients (Nota, Strooker and Ring 2014). Nota, Strooker and Ring (2014) performed a study comparing follow-up response rates with various forms of PROM administration. The participation rate in this study, where PROMs were collected in the waiting room, was considerably higher than when PROM data were collected by post (34%) or email (24%), and even slightly higher than when participants were asked to complete the PROM via phone interview (80%). The high participation rate in a study sample with minimal exclusion criteria implies that this result is applicable to the whole hand clinic population. This proposition is further supported by the finding that age and gender did not influence the likelihood of participation. The ability to collect longitudinal PROM data is

helpful in tracking the progress of individuals and when trying to gauge the success of any intervention. This study showed high retention rates of participants (94%), which may show that if patients are initially willing to complete a PROM, they will continue to do so throughout the course of their treatment.

Although the number of participants with a hand immobilised in a cast or splint was not directly measured in this study, it is reasonable to infer that a substantial proportion of those attending clinic would have had their hand(s) immobilised. The high participation rate despite the physical challenge associated with a hand injury or immobilisation suggests that these physical barriers are less of an obstacle to PROM administration and completion than might have been expected. Participants were willing to use their time waiting in clinics to complete the PROMs.

3.4.2 Incomplete Questionnaires

The proportion of incomplete MHQs was analysed in two ways: first, where only the core questions of the MHQ related to the 'side of interest' were considered and, second, where the entirety of the questionnaire was considered. The analysis using the first method yielded results that all three instruments had similar rates of incomplete responses (PRWHE, 35%; DASH, 30%; and MHQ, 33%) despite the different length of each PROM. When the MHQ in its entirety was considered, the number of incomplete questionnaires increased to 81%. It is reasonable to infer that participants did not engage with the questions that related to their unaffected limb, despite the instructions of the questionnaire asking that they complete all questions.

3.4.3 Frequently Unanswered Items

The item that was most frequently left unanswered (15.4%) in the DASH was Q21, which enquires about sexual activities. In the PRWHE, the problematic items related to pain with

lifting a heavy object and difficulty at work. It is possible that these items are often unanswered as the content may not seem relevant to the participant or may be too personal for the respondent to disclose. The MHQ had a more complex pattern to the missing answers, with a tendency for increased missing data in the later part of the questionnaire, as illustrated in Figure 3.2.

3.4.4 Ability to Generate a Score

The tolerance for missing data within the scoring rules for each instrument had an impact on the number of questionnaires that could be scored. The MHQ had the greatest amount of missing data among the groups. However, when only the questions related to the *affected* hand were considered, the number of MHQ questionnaires unable to generate a score was lower than that of the DASH (4% and 9%, respectively). The PRWHE scoring rules allow for any number of missing items to be substituted by a mean score (MacDermid 2007). As a result, only those who did not answer an entire section were unable to be scored. In contrast, the DASH and MHQ scoring rules are more stringent in their management of missing data. However, the unlimited substitution of the mean for missing answers may compromise the overall reliability of the PRWHE as the noise-to-signal ratio is increased by computing mean values for missing data points (McHorney et al. 1994).

3.4.5 Time Taken to Complete the Questionnaires

A total of 82% of participants reported the time to complete as less than 10 minutes. In the clinic setting that this study was conducted, patients are often waiting for more than 10 minutes before their consultation, and therefore the completion of PROMs did not cause any delays. Of course, other hand clinics may be more efficient in their appointment management. The time that is considered to be practical would be dependent on the clinical setting of implementation.

3.4.6 Participant Characteristics Relationship with Participation and Ability to Generate a Score

This study showed that gender and age did not influence the likelihood of participation or whether participation would result in a scorable questionnaire. This is in contrast with the findings of Bot, Anderson, et al. (2013), who found that non-responders to postal DASH questionnaires were more likely to be young and male. The lack of gender and age bias in this study supports PROM administration in the clinic waiting room as opposed to mail-based PROM collection.

3.4.7 Limitations

There were some limitations in this study. There was no detail collected on the characteristics of the traumatic hand pathology cohort. Hand trauma is a heterogeneous group including fractures and soft tissue injuries, and further detail on this subgroup might have been beneficial. Further information on the current status of the hand pathology and the severity of the condition would have also aided analysis to ensure that these factors were not influencing participation or completion rates. The information regarding the affected side was collected retrospectively through case note review. Unfortunately, this resulted in an inability to be sure of the relevant affected side for 18 participants. Time taken to complete the questionnaires was not measured objectively with a stopwatch but was self-reported in categories. Therefore, the time data might not have been an exact representation of the actual time taken by respondents. This element of the design of the study was chosen to minimise the labour involved in the administration of the study, as timing each individual would have significantly limited the number of total participants approached to enrol in the study. Although the time taken for questionnaire completion was not objectively measured, the perceived length of time taken from the patient's perspective is relevant.

This study measured the acceptability of a PROM-based research study to patients—participants were approached, informed about the purpose of the study, consented and enrolled in the research project. Therefore, the acceptance rates could be a reflection of the patient's willingness to participate in the study, rather than to complete a PROM for clinical care. However, we have proved that the model of collecting PROM outcome data in the clinic waiting room is superior in terms of participation rates to PROM data collection via email or postal methods. If PROM administration were to become part of routine practice in the setting of hand clinics then participants may be even more willing to participate, especially if patients could see that clinicians were reviewing their PROM data during the clinical consultation. These theories will be the focus of further research following routine PROM implementation.

According to this study, PROM administration in the hand clinic waiting room results in a high participation rate and good quality data collection; however, this mode of administration may be prone to bias. Dawson et al (2010) suggest that collecting follow-up PROM data at outpatient appointments is not ideal because of the inconsistent follow-up time schedules and the selection bias that patients with ongoing problems are more likely to return to clinic appointments. Any studies reporting on the PROM data collected in the waiting room should consider this potential bias. The participating sample and non-participating sample were shown to be no different in terms of their gender and age. Ideally, it would be valuable to know if these cohorts were also the same in terms of their pathologies, hand immobilisation, pre- vs post-surgical status and complications.

The practical implementation of PROMs into routine clinical care remains a challenge. The time, labour and resources needed to design and implement PROM collection can be a barrier in a high-volume clinical practice setting. In this study, two members of the research team were present at the hand clinics to approach patients and collate the data, which is seldom economically feasible in the long term. When considering routine administration of PROMs in

hand clinics, rigorous exploration of acceptability, efficacy and practicality is necessary to ensure optimal practice adoption.

3.5 Summary

The primary aim of this study was to assess the acceptability of PROM completion in the hand clinic. The high participation and retention rate show that completing a PROM while waiting for a clinical consultation in a busy hand clinic is acceptable for patients. The participation rates in the clinic waiting room were higher than other methods of PROM data collection reported in the literature (Bot, Anderson, et al. 2013; Nota, Strooker and Ring 2014). The secondary aim of this study was to examine the items that often resulted in missing data and how these missing data were managed according to the PROM scoring rules. Items that were often unanswered were identified and consideration given as to why this was the case. The content of these items can be re-assessed when developing new items to improve response rates. The relationship between the management of missing data by the PROM's scoring rules and whether the instrument can provide a score was clearly demonstrated. The amount of missing data that is tolerated by the scoring rules has flow-on effects on whether the PROM can generate a score and indeed the reliability of this score. This study has supplied useful insight that will be considered when developing a new PROM for hand conditions, the HAND-Q.

Chapter 4: Rasch Analysis of the Disabilities of the Arm, Shoulder and Hand Questionnaire

4.1 Introduction

The most widely reported instrument in the field of hand conditions is quoted as the DASH (Hudak et al. 1996), as well as its abbreviated version; the QuickDASH (Beaton et al. 2005). As introduced earlier in Chapter 2, the DASH was developed in 1996 by a collaboration of the American Academy of Orthopaedic Surgeons, the Council of Musculoskeletal Specialty Societies and the Institute for Work and Health. It purports to measure disability of the upper limb, viewing the whole upper limb as a single functional unit.

Concepts that form the basis of the DASH are symptoms and functional status, which includes physical, social and psychological function. However, when examining the legitimacy of summing responses together to form a total score, the logic of adding together scores from contrasting concepts becomes problematic (Cano et al. 2011a). Interpreting change in the total score is particularly difficult because an improvement in the total score may reflect a true improvement in the patient's function but could also represent a change in the individual's circumstances. For example, an improvement in score may be observed with an improvement in transportation methods but in fact a worsening of the patient's symptoms (Cano et al. 2011a). These factors are easily distinguished on clinical consultation and could be detected in instruments that have individual domain scores. However, they are not discernible from a single summary score, such as in the results of the DASH.

The aim of this study was to provide clinicians in the field of hand surgery a broad psychometric evaluation of the DASH to illustrate that analysis with Rasch methodology can uncover

measurement weaknesses where traditional psychometric analysis shows sound measurement characteristics. Responses to the DASH were collected from a heterogeneous patient group with varied hand pathology (i.e., a cohort analogous to that used in the original development of the DASH) (Hudak et al. 1996). The data provided by this group then underwent three analyses. First, the DASH was analysed using traditional psychometrics. Second, psychometric evaluation of the DASH based on Rasch measurement theory (RMT) was completed. Last, the DASH was conceptually restructured into two separate scales: ADLs and symptoms; properties of each were then assessed with RMT.

4.2 Methods

4.2.1 Participants, Recruitment and Data Collection

Patients attending a busy publicly funded tertiary hospital hand clinic were invited to take part in the study. Participants completed the DASH independently in the clinic waiting room while they waited for their clinical review. Written informed consent was obtained for each participant. Inclusion criteria required study participants to be aged over 14 years, and able to read, write and speak English. The Southern Adelaide Clinical Human Research Ethics Committee approved the study (Appendix B.1). Anticipated sample size was greater than 300 questionnaire responses as this is the recommended figure for item analysis (Cano et al. 2011a).

4.2.2 Structure and Scoring of the Disabilities of the Arm, Shoulder and Hand Questionnaire

The DASH has 30 questions that are grouped together in a single scale to measure ‘symptoms and functional status’ in individuals who have an upper limb disorder (Hudak et al. 1996). All questions are scored from 1 (*no difficulty*/*not at all*/*not limited at all*/*none*/*no difficulty*/*strongly disagree*) to 5 (*unable*/*extremely*/*extreme*/*so much difficulty that I can’t sleep*/*strongly agree*).

A total score is given by the addition of each item score and calculation of the mean item score. This is then transformed into a score between 0 and 100 by summing the responses, dividing by the number of responses, and then subtracting 1 and multiplying by 25. When interpreting the DASH summary score, a higher number represents worse 'symptoms and functional status'. A copy of the DASH is located in Appendix B.2.

4.2.3 Traditional Psychometric Methods

Traditional psychometric theory, or CTT as it is also known, is based on descriptive analyses and correlational statistics to assess scaling assumptions, reliability and validity of a scale. The mathematical details of this technique are beyond the requirements of clinicians and are described in detail elsewhere (DeVellis 2006; Hobart and Cano 2009; DeVellis 2016). In this study, the DASH data were examined for data quality (i.e., the percentage of missing data per question), scaling assumptions (the variance of item means and corrected item-total correlations), scale-to-scale targeting (score mean, standard deviation, floor and ceiling effects) and internal consistency reliability (Cronbach's alpha) (Cronbach 1951).

4.2.4 Modern Psychometric Methods

Modern psychometric methods encompass IRT and RMT. RMT analysis of the DASH response data was conducted using RUMM2030 software (RUMM Laboratory, Perth, Australia). The analysis uses a variety of statistical and graphical tests to examine the function of each item of the scale, and from this information, the quality of the overall scale can be assessed. When the data 'fit' the Rasch model, then the scales are considered robust. The following tests were performed.

4.2.4.1 Thresholds for Item Response Categories

Response options scored consecutively from 1 '*No difficulty*' to 5 '*Unable*', which indicates a continuum of difficulty performing each given task (Figure 4.1). The term threshold refers to the point on the scale where a respondent is equally likely to choose the option either side of the threshold. A person located at the threshold point between '*mild difficulty*' and '*moderate difficulty*' would have a 50% choice of choosing either response. If the response options are constructed in a way that allow for logical used of the scale then each of the threshold points between consecutive response options should be in order. Ideally, each response category represents a similar breadth of the construct being measured and is the most likely response of a group of respondents within a range of person locations (Pallant 2007). If a response option is never the most likely to be selected, then it is not providing any helpful information and the response categories are not functioning well for that item, this is called disordered thresholds (Linacre 2002). Disordered thresholds are commonly caused by respondents not understanding the response options due to confusing labelling, or too many response options so that the respondent cannot differentiate between options (Pallant 2007). This concept is represented graphically for each item by the category probability curve; an example of an item with ordered thresholds is given in Figure 4.2.

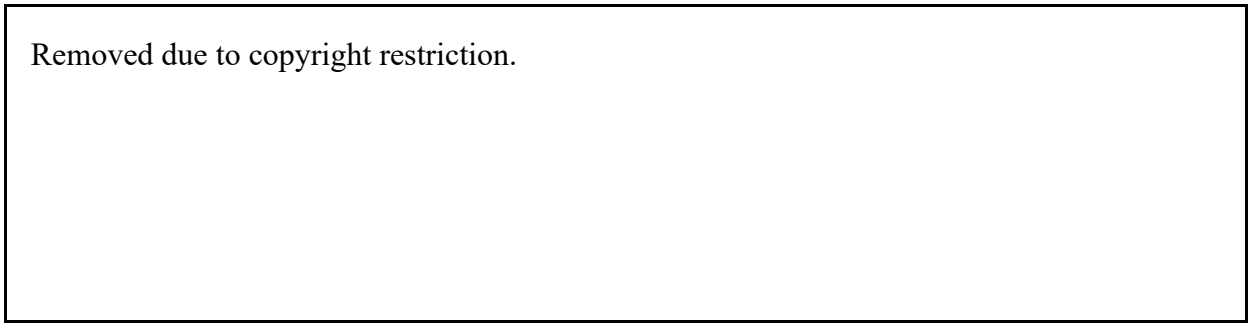

Figure 4.1: Example of categories of response options used in the DASH.

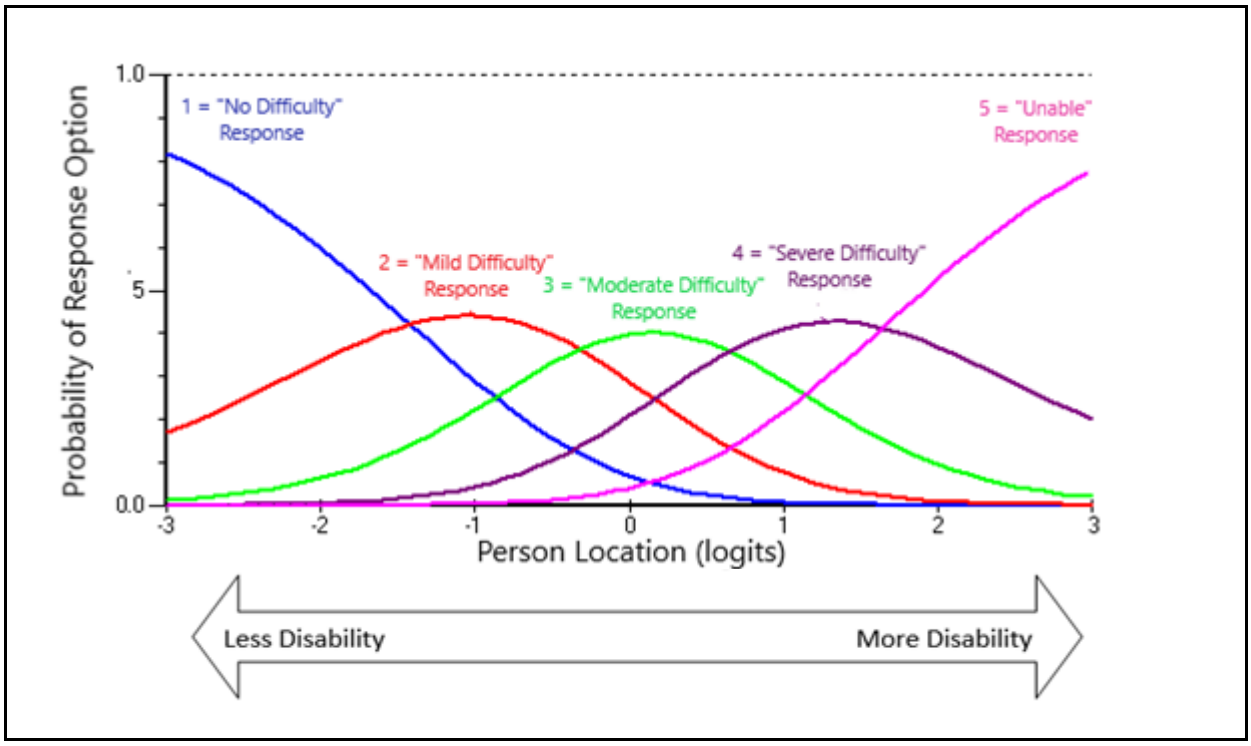

Figure 4.2: Example of a category probability curve with ordered thresholds. This figure shows the category probability curve for question 4 of the DASH 'Rate your ability to … Prepare a meal'.

In Figure 4.2, the y-axis represents the probability of the response category being selected (0–1). The x-axis represents respondents with varying amounts of the underlying concept, which in the case of the DASH is disability. Those people with a logit location of −3 have almost no disability, and those people with a logit location of 3 have severe disability. Each of the coloured lines indicates the probability of a different response category labelled 1–5, (1 (blue line) = *no difficulty*, 2 (red line) = *mild difficulty*, 3 (green line) = *moderate difficulty*, 4 (purple line) = *severe difficulty* and 5 (pink line) = *unable*). From this graph, it can be seen that those at either extremity of the concept of disability are most likely to answer with the corresponding extreme response option. For example, those with severe disability (logit > 2) are most likely to select response option 5, which corresponds with '*unable*', and those with minimal disability (logit < −2) are most likely to select response option 1, which corresponds with '*no difficulty*'.

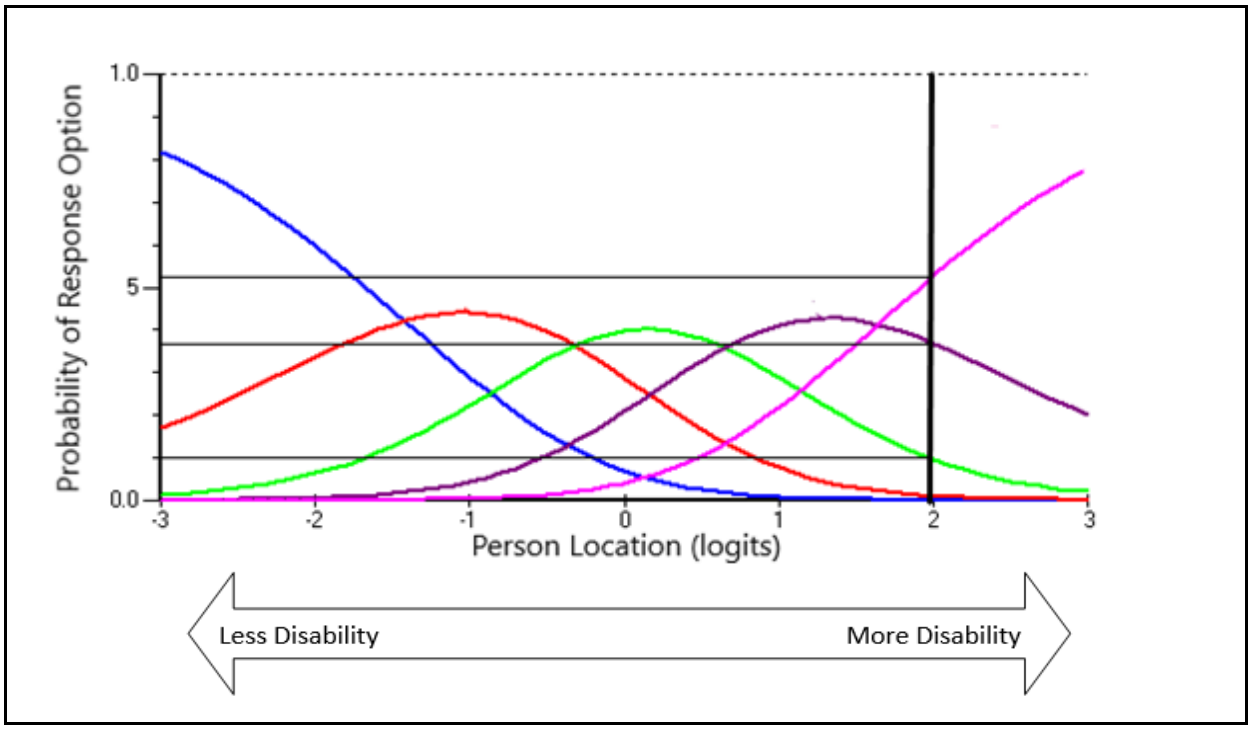

Figure 4.3: Response Option Probabilities. The figure shows the same category probability curve as shown in Figure 4.2 with a mark at the person location of 2, demonstrating that the response option probabilities at any person location add up to 1. For example, a person with a disability level of 2 logits will have approximately a 50% chance of responding to this question with '*unable*'(pink line), a 40% chance of responding with '*severe difficulty*' (purple line) and a 10% chance of responding with '*moderate difficulty*'(green line).

It is also important to assess that each of the response options is the most likely option at a particular person location, and that the order of the thresholds corresponds with the continuum of presented options. That is, the peaks of the probability curves for each of the response options for this item are in order, *no difficulty*, *mild difficulty*, *moderate difficulty*, *severe difficulty* and *unable* (1, 2, 3, 4 and 5), and not disordered, such as *no difficulty*, ***moderate difficulty***, ***mild difficulty***, *severe difficulty* and *unable* (1, 3, 2, 4 and 5).

4.2.4.2 Item Fit Statistics

To examine if the DASH measures a unidimensional construct, that, is a clinical hierarchy, three indicators of fit were examined.

- Log residuals (also known as fit residuals) represent the interaction between the item and all of the people who responded to that particular item. It is a summary statistic of the differences between the observed and expected responses from all responses. Assuming the data fit the Rasch model, the difference between the observed and expected responses would be only due to random error. On the basis of this assumption, the fit residuals for each item should be between −2.5 and +2.5, with 99.5% of values being in this range (Hobart and Cano 2009).
- Chi-square values represent the interaction between an item and the trait being measured by the scale. To calculate this statistic, the sample of respondents is divided into similarly sized groups according to their level of the trait of interest to create 'class intervals'. For each item, the mean observed score of each class interval is compared with the expected scores of the items predicted to be at the mean of the class interval. The chi-square values are the sum of the chi-square values calculated for each interval scale. The associated chi-square probabilities are the likelihood that the difference between observed and expected is due to chance (Hobart and Cano 2009). Therefore, for data that fit the Rasch model, the chi-square probabilities would be non-significant after Bonferroni adjustment.
- Item characteristic curves are a graphical illustration of fit that is helpful to understand the above statistics. An item characteristic curve can be generated for each item. It displays the expected response as predicted by the Rasch model at each level of the measurement continuum (solid line). Black dots represent the observed mean scores for

each of the class intervals (defined above). The chi-square values (discussed above) correspond to the coherence or lack thereof between the observed and expected scores (Hobart and Cano 2009).

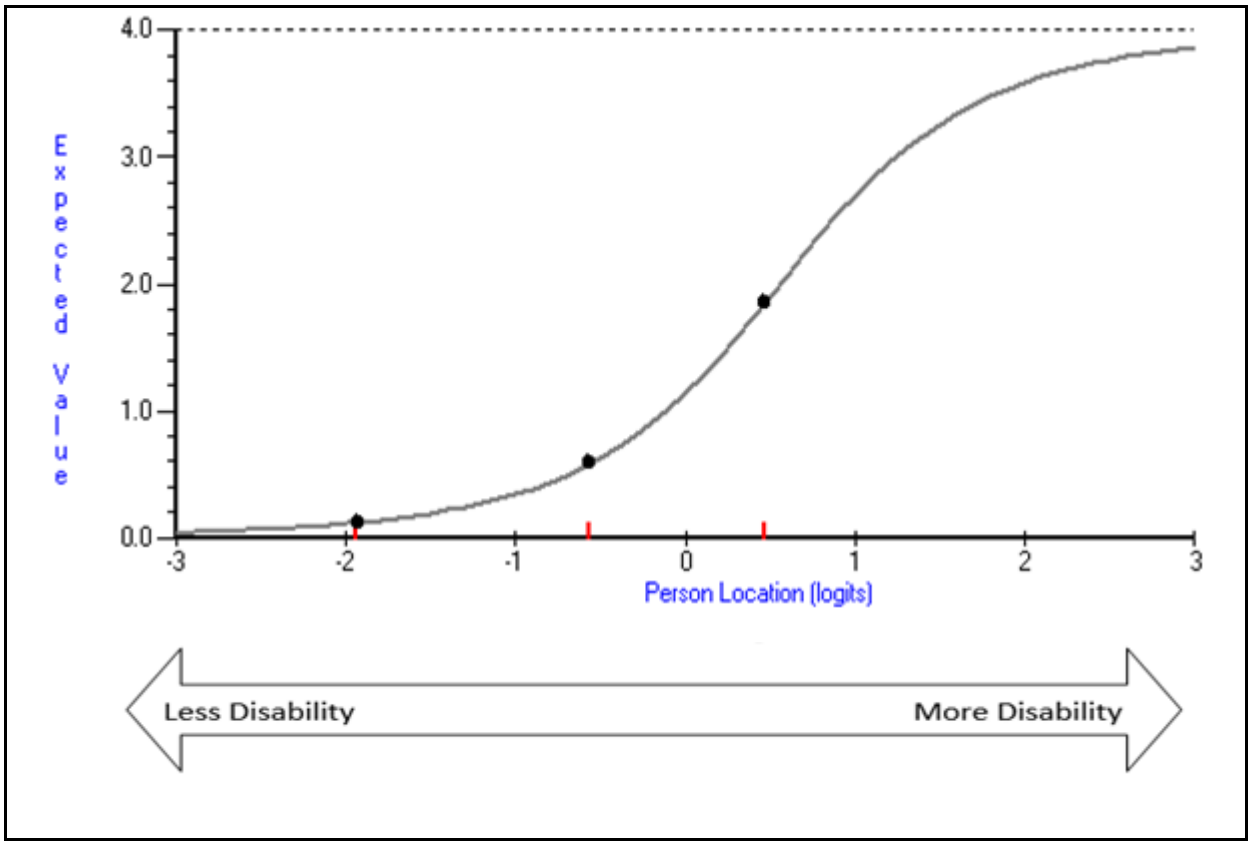

Figure 4.4: Example of an item characteristic curve for Q21 from the DASH, 'Rate your ability … Sexual activities'. The solid line represents the response as predicted by the Rasch model; the black dots stand for the observed mean scores for each of the class intervals. This figure illustrates good fit between the predicted and actual scores as the dots are located along the solid line.

4.2.4.3 Item Locations and Targeting

According to RMT, the items of a scale should define a continuum of a single construct (Hobart and Cano 2009). Visualising where items are positioned along a continuum of the construct being measured graphically illustrates how well the items of an instrument map out the

construct being measured. The items of a scale should be evenly distributed over a clinically meaningful range with person locations populating the same range; this means the scale is well targeted to the population for which it is measuring. This is further elaborated on in Results 4.3.3.3.

4.2.4.4 Dependency

The Rasch model needs each item to be independent, that is, the response to one item does not influence the response to another item. An example of two items that show dependency are 'Rate your ability to tie your shoelaces' and 'Rate your ability to tie a bow in a ribbon'. Both items require similar ability in terms of hand function. Therefore, the answer to both questions is likely to be the same and the items are 'dependent'. To test for item dependency, the residual correlation between items is calculated. Residual correlations greater than 0.30 represent 10% shared variance (Hobart and Cano 2009). Cohen's guidelines are used for interpretation of the correlation *r* value (small, 0.1–0.29; moderate, 0.3–0.49; and strong, >0.5) (Andrich 1988). Items with residual correlations greater than 0.3 are identified as they can artificially inflate the reliability of a scale (B. Wright and Masters 1982). That is, if the same question is asked multiple times in slightly different ways a scale is likely to be highly reliable as items are highly dependent, but they would not measure a spectrum of the construct being measured.

4.2.4.5 Person Separation Index

The Person Separation Index (PSI) is a measure of the reliability of a scale in a given sample population (Hobart and Cano 2009). It indicates the error associated with the measurement of individuals within a given sample. It is analogous with the Cronbach alpha used in traditional psychometrics. Higher values indicate greater reliability.

4.3 Results

4.3.1 Cohort Demographics

The participants in this study represent the population attending hand clinics. Overall, the study included 165 participants, who submitted a total of 234 DASH questionnaires. The demographics of this study cohort are summarised in Table 4.1.

Table 4.1: Cohort demographics

Characteristic		(n)	%
Total participants		165	
Gender			
	Female	66	40
	Male	99	60
Age (years)			
	Mean (SD)	47 (20)	
	Range	14–90	
Condition			
	Trauma	96	58
	Nerve compression	25	15
	Dupuytren’s contracture	11	7
	Trigger finger	8	5
	Other	19	11
	Arthropathies	6	4
Questionnaires completed			
	One	165	
	Two	48	
	Three	14	
	Four	4	
	Five	3	

4.3.2 Traditional Psychometric Analysis

The traditional psychometric analysis supports the DASH as a quality PROM to measure upper limb disability. The data quality showed items missing responses from up to 16% of participants. Overall scale scores were able to be calculated for 90% of respondents; the remaining 10% had three or more items unanswered. The mean item scores were similar, and the corrected item-total correlation range was small; this satisfies the scaling assumptions. The scale targeting was good in that the scores spanned the range of the instrument, with minimal floor or ceiling effects. The internal consistency reliability was high (Cronbach's alpha = 0.96).

Table 4.2: Traditional psychometric analysis of the DASH

Psychometric Property		Total
Questionnaires, n		234
Data quality		
	Item missing data (%)	0.9–15.8
	Computable scale scores (%)	90
Scaling assumptions		
	Item mean scores, range	1.6–3.2
	Item SD, range	1.1–1.6
	Item variance, range	1.2–2.7
	Corrected item-total correlations, range	0.41–0.78
Targeting		
	Mean score (SD)	36.4 (21.6)
	Possible score range	0–100
	Observed score range	0–91
	Floor/ceiling effect (%)	0/0.9
	Skewness	0.356
Reliability		
	Cronbach's alpha	0.96
	Mean inter-item correlation	0.43

4.3.3 Modern Psychometric Analysis

4.3.3.1 Thresholds for Item Response Categories

The item response category thresholds were disordered in 10 of the 30 items, as shown in Figure 4.4. In this figure, the items are represented by horizontal bars, and the items are ordered according to the item location (from the most difficult item at the top to the least difficult item at the bottom). The colours on the horizontal bars represent the response options for that item. For example, items 1–21 have the response options: *no difficulty* (blue), *mild difficulty* (red), *moderate difficulty* (green), *severe difficulty* (dark purple) and *unable* (pink). Items that are missing a coloured bar have disordered thresholds, which means that the response options do not function properly for those items.

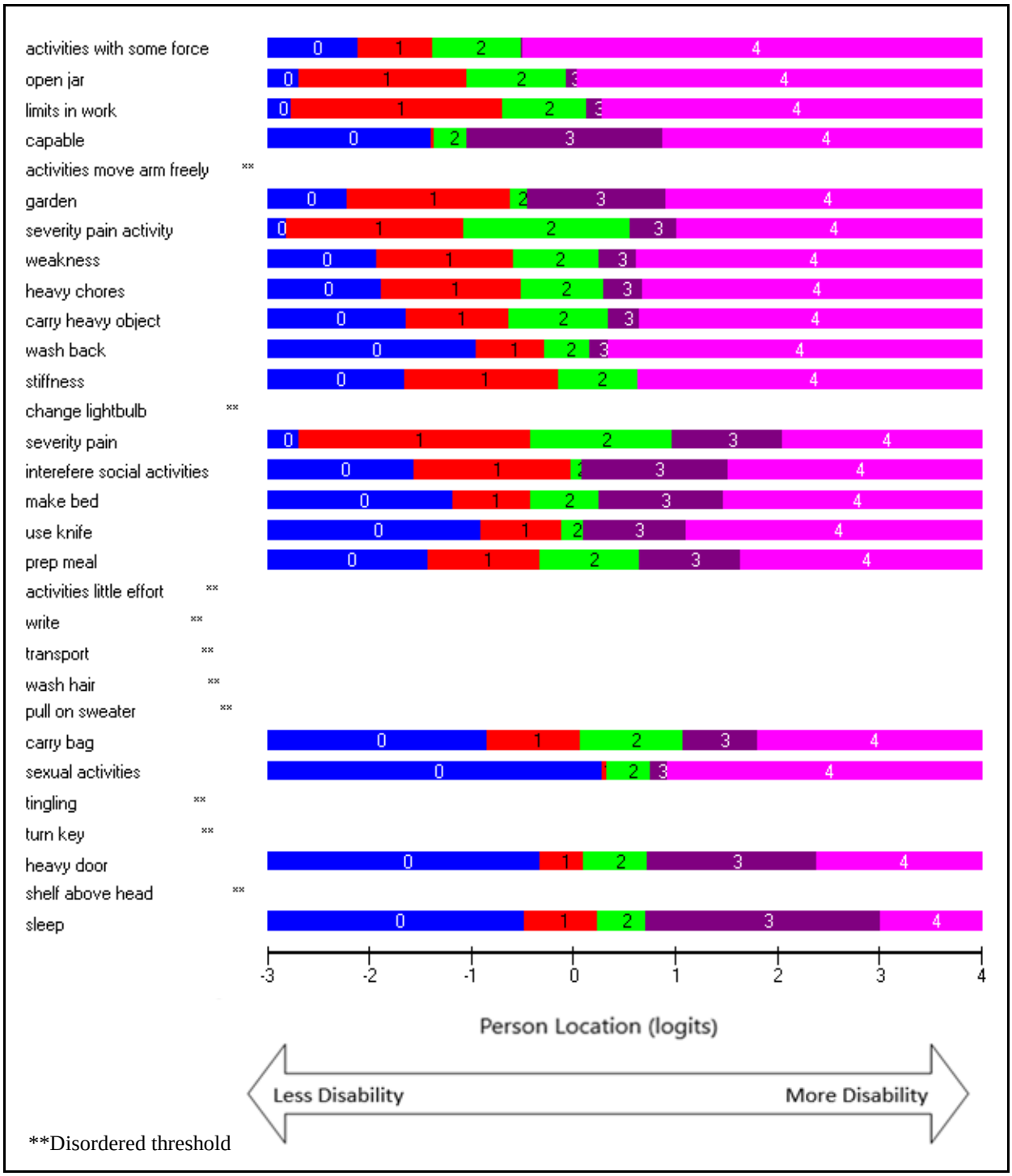

Figure 4.5: DASH item threshold map.

An example of an item with disordered thresholds is item 12, 'Change a lightbulb overhead'. Figure 4.6 shows the category probability curve for this item. In this figure, the x-axis represents the location of respondents along the construct being measured (disability) (i.e., the person location). People located on the left side of the graph have less disability, and people on the

right of the graph have more disability. The y-axis shows the probability of respondents endorsing the response categories (0–1). Each curve represents one of the response options for this item: blue line, labelled 0, represents the response '*no difficulty*'; red line, labelled 1, represents the response '*mild difficulty*'; green line, labelled 2, represents the response '*moderate difficulty*'; purple line, labelled 3, represents the response '*severe difficulty*'; and pink line, labelled 4, represents the response 'unable'.

For this item, the response option of '*severe difficulty*' (dark purple line, labelled 3 in the figure) was never the most likely to be chosen, indicating the response options are not suitable for this item. This means at no point along the whole spectrum of disability is the respondent more likely to answer this item with '*severe difficulty*', the respondent can complete the task with *'no difficulty', 'mild difficulty'*, '*moderated difficulty*' or they are '*unable*'.

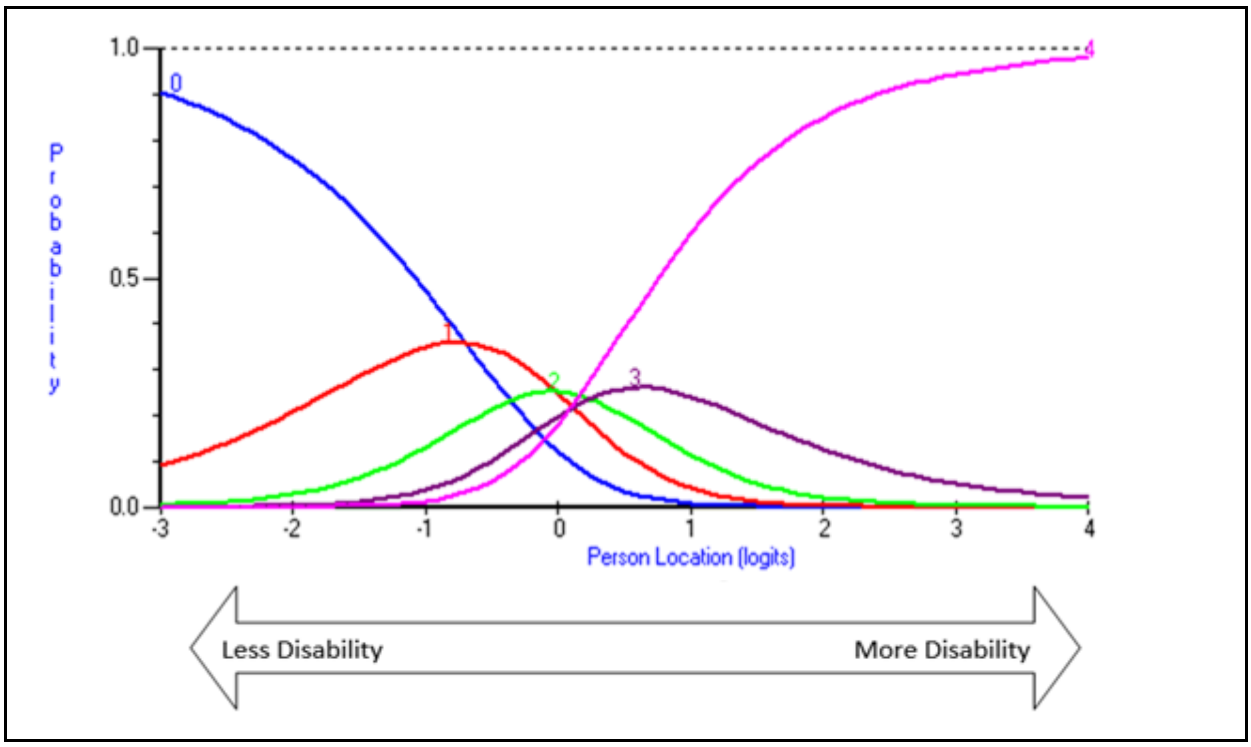

Figure 4.6: Category probability curves for item 12, 'Change a lightbulb overhead'.

4.3.3.2 Item Fit Statistics

Fit statistics are summarised in Table 4.3. Log fit residuals were within the recommended criteria of −2.5 to +2.5 for 23 of 30 items. Of the remaining seven items, four had notable misfit: items 8, ‘Garden …’; 26, ‘Tingling …’; 28, ‘Stiffness …’; and 30, ‘I feel less capable …’. The Bonferroni adjusted chi-square *p*-values were significant for the latter three listed items.

Item characteristic curves of these items reflected some deviation between the observed scores and the Rasch-expected scores, suggesting that items often under-discriminated disability. Figure 4.7 shows the item characteristic curves graph for item 28. This plot shows the response to a specific item predicted by the model (solid line) throughout the measurement continuum. The dots on the graph represent the scores of respondents divided into three groups with varying levels of ‘disability’ (or class intervals). The central and left dots are above the line, and the right dot is below the line. This means that among those respondents with a lesser disability (represented by the left dot), there was more reported stiffness than predicted, and among those with elevated levels of disability (represented by the right dot), there was less reported stiffness than predicted.

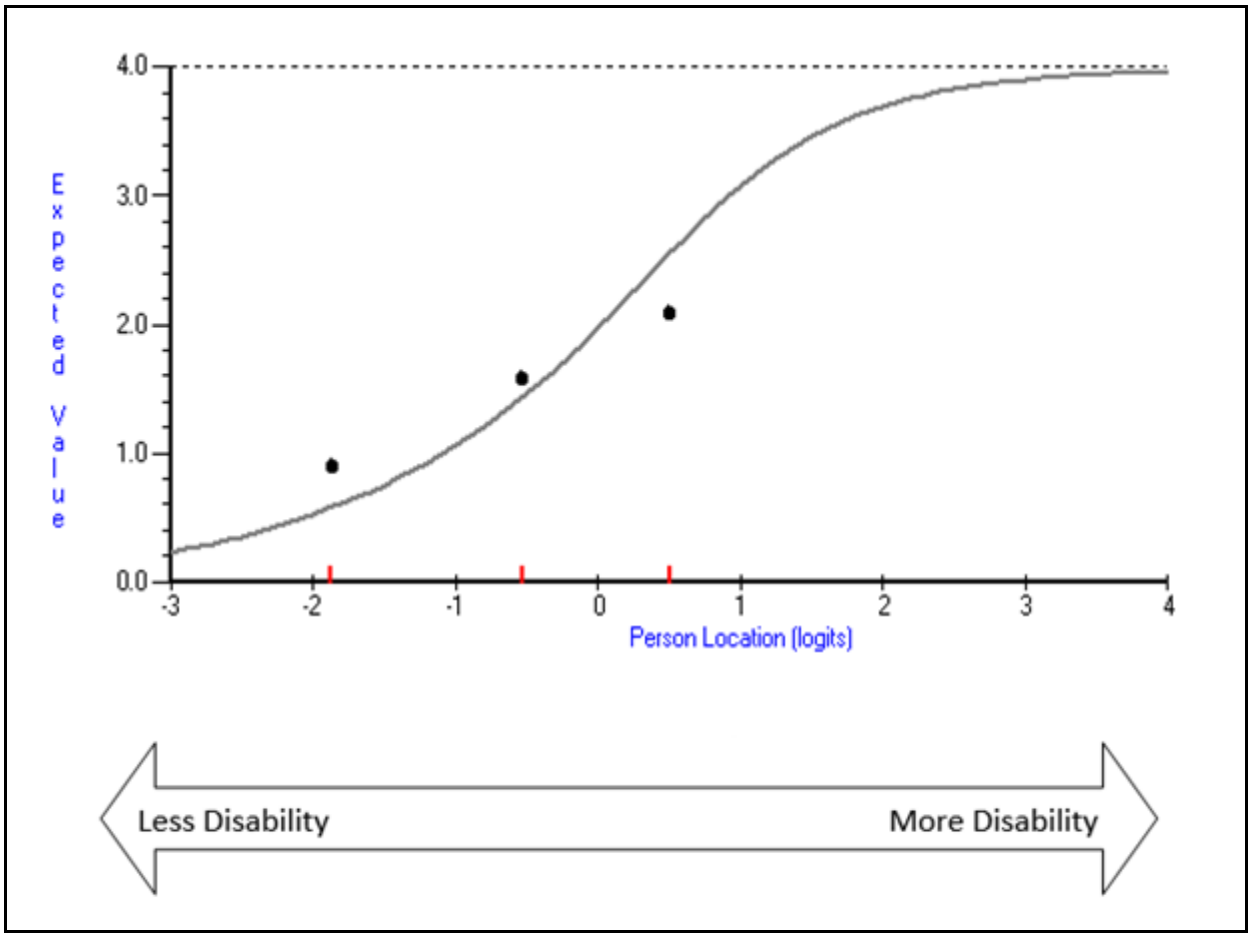

Figure 4.7: Item characteristic curve for item 28,

'Stiffness in your arm, shoulder or hand'.

4.3.3.3 Item Locations and Targeting

The DASH demonstrated satisfactory targeting as the range of disability measured by the scale items covered 99% of those measured in the sample. The item locations distributed between −1.12 and +0.87, which indicates that the DASH items define a reasonable continuum (shown in Table 4.3). Figure 4.8 shows the person-item threshold distribution. This figure shows the targeting of the questionnaire by comparison of the person and item locations and their distributions on the virtual ruler measuring 'disability' (x-axis). The measurement unit of this x-axis is the Rasch software derived figure called the 'Logit'. This axis has no discrete minimum or maximum value as it is a product of probability (which approaches but never

reaches 0 or 1), as a result the range of the x-axis spans from -infinity to +infinity (Hobart and Cano 2009, 22-23). The mathematics behind this theorem are beyond the scope of this thesis and are detailed elsewhere (Hobart and Cano 2009, 22-32 and Andrich 1988). To further explain Figure 4.8; the top histogram shows where the participants are located along the ruler; the person location. The bottom histogram represents the location of the items of the DASH along the same metric; the item location. A ceiling effect is shown by the lack of items (represented in the lower histogram) located at a position less than −3.0 logits, which means that the DASH is unable to accurately score those participants located on this part of the ruler, which is those with the least disability. Clinically this means that participants with minimal disability are not measured appropriately by the DASH. This property of the DASH is likely the reason why the developers created further optional modules (Work and Sports / Performing Arts) that were not included in this study.

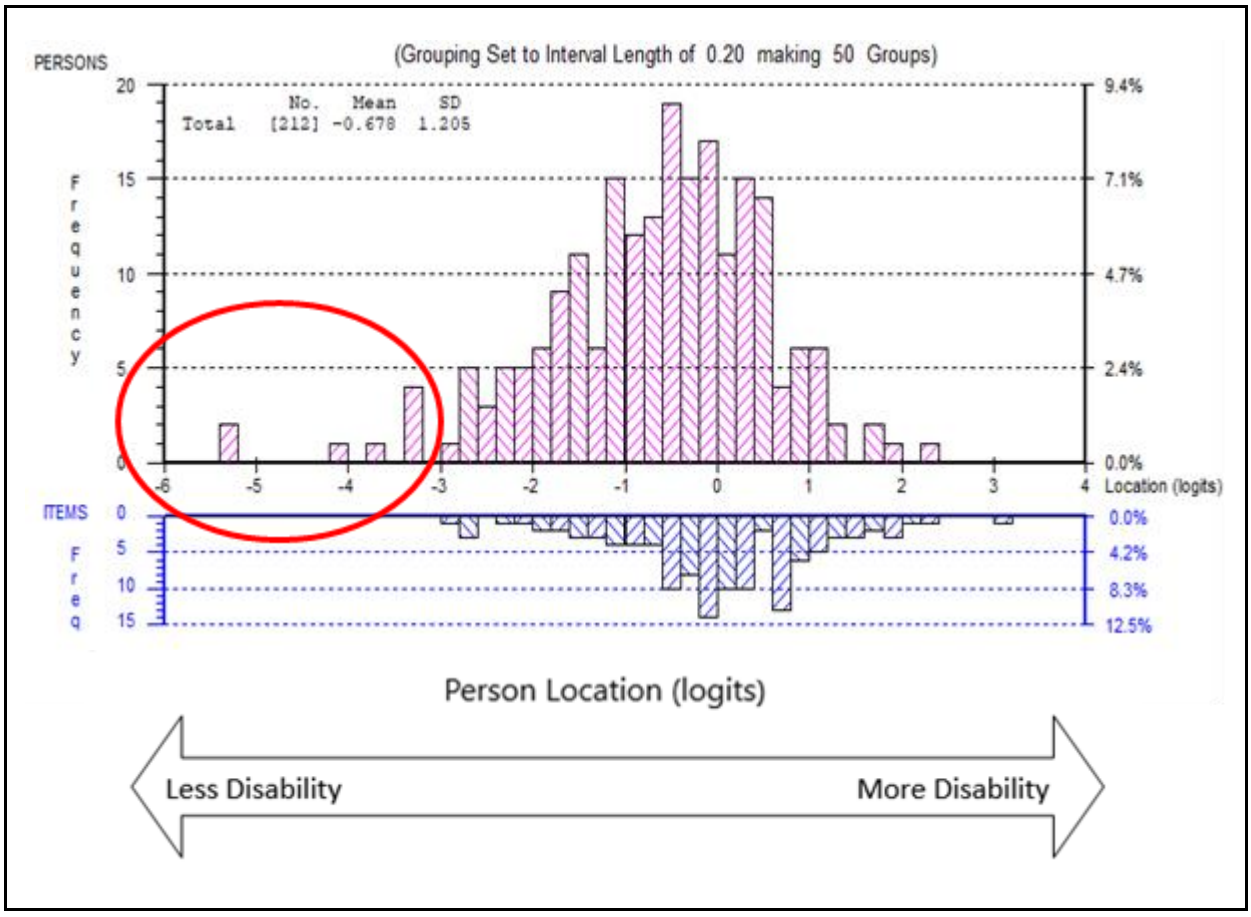

Figure 4.8: DASH person-item thresholds dependency.

Medium residual correlations were found in 13 pairs of items; one pair were highly correlated indicating that the response to item 24 (‘Arm, shoulder or hand pain’) influenced the response to item 25 (‘Arm, shoulder or hand pain when you perform any specific activity’). High correlations may lead to inaccurate results as respondents are likely to answer highly correlated questions in the same way with these items.

4.3.3.4 Person Separation Index

The PSI without extremes was 0.95.

4.3.3.5 Score Transformation

Figure 4.9 shows the relationship between the DASH score (y-axis) and the construct of 'disability', which the DASH purports to measure (x-axis). The left side of the x-axis corresponds to low amounts of disability and thus low DASH scores. A linear relationship between 'disability' and DASH score would indicate a consistent change in disability for each change in DASH score (i.e., interval format data). The relationship between these two variables then follows an S-shaped rather than a linear pattern (i.e., ordinal format data). This shows that a one-point change in the DASH score is associated with a variable amount of change in the construct of disability depending on the position of the score. Thus, the amount of 'disability' between integers of the DASH score metric is inconsistent throughout the scoring metric as DASH scores are ordinal and not interval level data. This inconsistency results in a limited value in comparing the change of scores, as the same value in score change can stand for an entirely different amount of 'disability' depending on where the scores fall in the score range. This means that there is no consistency between integers of the DASH score, as the raw scores are in ordinal format and not interval format, which is required to have consistent measurement over the breadth of possible scores.

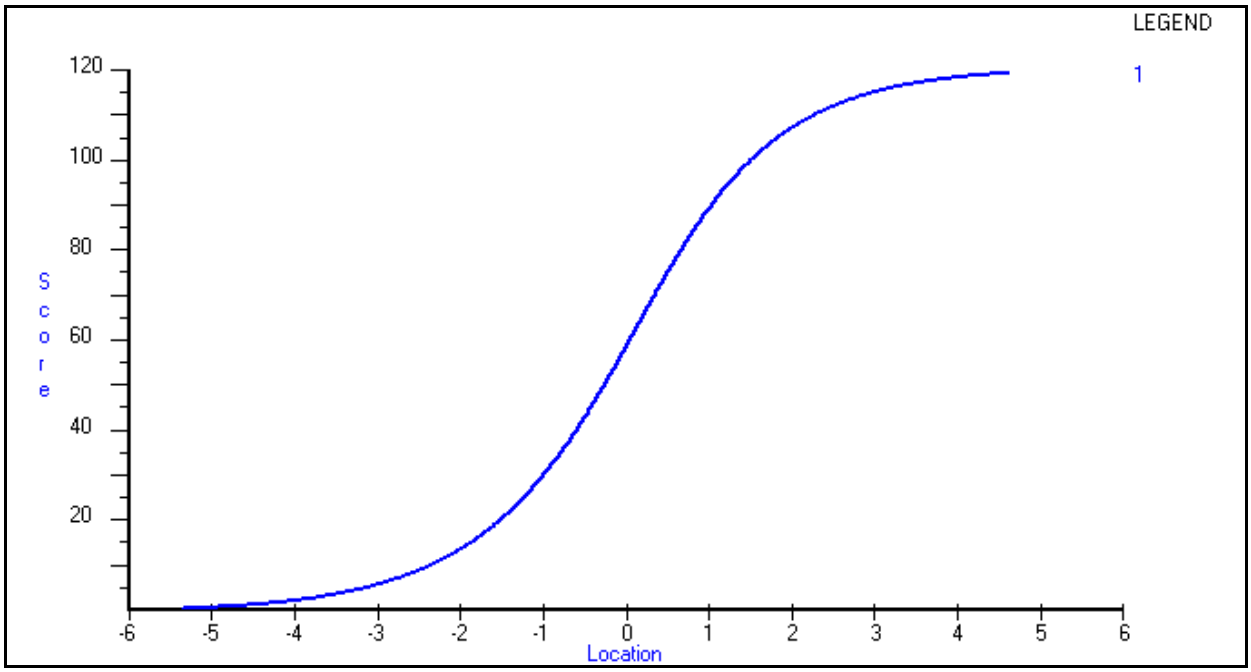

Figure 4.9: Raw score to interval metric transformation.

Table 4.3: RMT statistical indicators of fit

Stem	Items	Original DASH					Revised DASH scales—ADLs and Symptoms				
		Item location	Standard error	Fit residual	χ^2	*p*-value	Item location	Standard error	Fit residual	χ^2	*p*-value
1	Open a tight or new jar	−0.94	0.08	1.03	2.40	0.302	−1.05	0.08	2.53*	12.88	0.002
2	Write	0.36	0.08	2.60*	4.29	0.117	0.39	0.08	3.27*	10.92	0.004
3	Turn a key	0.68	0.08	0.15	1.79	0.410	0.75	0.09	1.03	3.01	0.222
4	Prepare a meal.	0.14	0.08	−1.63	10.07	0.007	0.14	0.09	−1.05	6.38	0.041
5	Push open a heavy door	0.73	0.08	0.26	1.82	0.403	0.79	0.09	0.73	4.40	0.111
6	Place an object on a shelf above your head	0.76	0.08	0.14	0.67	0.716	0.85	0.09	−0.12	2.81	0.246
7	Do heavy household chores (e.g., wash walls, wash floors)	−0.36	0.08	−2.88*	11.20	0.004	−0.42	0.08	−2.27	6.17	0.046
8	Garden or do yard work	−0.60	0.08	−3.44*	14.16	0.001	−0.69	0.08	−2.73*	8.31	0.016
9	Make a bed	0.03	0.08	−2.62*	13.42	0.001	0.02	0.08	−2.14	9.34	0.009
10	Carry a shopping bag or briefcase	0.53	0.08	0.72	2.71	0.258	0.59	0.09	1.27	6.47	0.039
11	Carry a heavy object (over 10 lbs)	−0.32	0.08	−0.60	0.38	0.827	−0.37	0.08	0.16	1.20	0.550
12	Change a lightbulb overhead	−0.10	0.07	−0.36	0.61	0.739	−0.12	0.08	−0.06	0.25	0.882
13	Wash or blow dry your hair	0.43	0.08	−1.02	3.54	0.170	0.48	0.08	−0.86	3.89	0.143
14	Wash your back	−0.18	0.07	−1.87	6.71	0.035	−0.22	0.08	−1.32	4.64	0.098

15	Put on a pullover sweater	0.50	0.08	−2.32	8.40	0.015	0.56	0.09	−1.71	4.95	0.084
16	Use a knife to cut food	0.05	0.08	−1.54	6.17	0.046	0.05	0.08	−0.42	3.60	0.165
17	Recreational activities which require little effort (e.g., card playing, knitting)	0.18	0.08	0.29	0.87	0.646	0.18	0.08	1.44	0.49	0.783
18	Recreational activities in which you take some force or impact through your arm, shoulder or hand (e.g., golf, hammering, tennis)	−1.12	0.08	−1.65	10.15	0.006	−1.26	0.08	−1.28	4.63	0.099
19	Recreational activities in which you move your arm freely (e.g., playing frisbee, badminton)	−0.60	0.07	0.41	4.64	0.098	−0.69	0.08	1.44	0.23	0.891
20	Manage transportation needs (getting from one place to another)	0.41	0.08	0.58	6.13	0.047	—	—	—	—	—
21	Sexual activities	0.57	0.08	0.53	0.26	0.877	—	—	—	—	—
22	During the past week, to what extent has your arm, shoulder or hand problem interfered with your normal social activities with family, friends, neighbors or groups?	0.01	0.08	−0.37	1.77	0.412	0.154	0.08	0.885	0.45	0.797
23	During the past week, were you limited in your work or other regular daily activities as a result of your arm, shoulder or hand problem?	−0.76	0.08	−0.35	1.92	0.383	−0.659	0.08	0.926	3.20	0.202
24	Arm, shoulder or hand pain	−0.02	0.09	0.58	6.00	0.050	0.174	0.099	−1.413	7.31	0.026
25	Arm, shoulder or hand pain when you performed any specific activity	−0.58	0.09	0.28	4.70	0.095	−0.522	0.091	−1.437	7.77	0.021

26	Tingling (pins and needles) in your arm, shoulder or hand	0.63	0.08	4.48*	16.60	0.000*	0.821	0.084	2.804*	3.40	0.183
27	Weakness in your arm, shoulder or hand	−0.41	0.08	0.88	5.09	0.078	−0.316	0.081	−1.331	7.46	0.024
28	Stiffness in your arm, shoulder or hand	−0.14	0.08	3.76*	23.92	0.000*	0	0.083	1.432	2.85	0.240
29	During the past week, how much difficulty have you had sleeping because of the pain in your arm, shoulder or hand?	0.87	0.09	1.08	3.23	0.198	1.09	0.089	−0.724	2.07	0.356
30	I feel less capable, less confident or less useful because of my arm, shoulder or hand problem	−0.74	0.08	3.51*	29.05	0.000*	−0.742	0.081	0.876	6.81	0.033

Revised DASH scales: ADLs (items 1–19) and Symptoms (items 22–30). *Indicates extreme fit residuals and statistically significant p values.

Table 4.4: Reliability statistics

Scale	Included sample	Rasch coverage	PSI without extremes	χ^2	*p*-value	Cronbach α	Items with disordered thresholds (n)	Item pairs with residual correlations >0 (n)	Items with significant χ^2, *p*-value (n)	Items with fit residual outside ±2.5 (n)
DASH	210	99%	0.95	202.67	0.000	0.958	10	13	3	7
Revised DASH scales										
ADLs	208	98%	0.94	94.58	0.000	0.958	4	3	0	3
Symptoms	208	98%	0.87	41.33	0.001	0.890	1	1	0	1

4.3.3.6 Rasch Analysis of Revised Disabilities of the Arm, Shoulder and Hand Questionnaire—Activities of Daily Living and Symptom Scales

The DASH item content was reviewed and divided into two conceptually separate scales: 'ADLs' and 'Symptoms' scales. The ADLs scale consisted of the first 19 items of the DASH. The Symptoms scale contained items 22–30 of the DASH. Item 20, 'Manage transportation needs' and item 21, 'Sexual activities', were not considered to fit conceptually with either scale and were therefore excluded from further analysis.

The revised scales retained good targeting, with both scales measuring 98% of the sample; however, there were persisting ceiling effects for both scales. There was a total of five disordered thresholds between both scales; the ADL scale had four disordered items (2, 3, 17 and 19) and the Symptoms scale hand only one disordered item (26). Fit statistics for each item are shown in Table 4.3. Log residuals were within the recommended criteria of −2.5 to +2.5 for 24 of the 28 items. The remaining items had log fit residuals less than ±3.3. The Bonferroni adjusted chi-square *p*-values for all items were non-significant, showing good item fit. There were three pairs of items in the ADLs scale and a single pair of items in the Symptoms scale that showed residual correlations >0.3, ranging from 0.309 to 0.463. Table 4.4 lists the reliability statistics for both revised scales. The PSI was 0.94 for the ADLs scale and 0.87 for the Symptoms scale, suggesting adequate separation of individuals by the items.

4.4 Discussion

This study has demonstrated a psychometric analysis of the DASH using both traditional and modern techniques. The results indicate that although the DASH performs well when examined with CTT, there are inadequacies revealed by RMT analysis. Traditional psychometric analyses both in this study and others indicate that the DASH is a reliable and valid instrument (Kennedy

et al. 2011). However, since the DASH was published (Hudak et al. 1996), there have been improvements in the field of measurement science psychometrics, which provide a more sophisticated measurement of the patient's perspective (Cano and Hobart 2011). The RMT findings in this study support others that have found that the DASH does not comply with the Rasch measurement model and therefore its use should be re-evaluated.

The DASH has been subjected to Rasch analysis in Canadian patients with Dupuytren's contracture (Forget et al. 2014) and in Italian patients with general musculoskeletal disorders of the upper limb (Franchignoni et al. 2010). Both of these studies found that the DASH failed to comply with the Rasch measurement model. Braitmayer et al. (2017) performed Rasch analysis on the DASH in a German population with hand injuries and diseases. Their analysis also did not support the validity of the DASH (i.e., they could not confirm that the DASH measures what it purports to measure—upper limb disability). This study examined the DASH in an Australian English-speaking cohort with varied hand injuries and hand conditions.

Perhaps the most fundamental obstacle to meaningful measurement with the DASH is the lack of a unidimensional construct (Cano et al. 2011a). The developers of the DASH described the scale as a measure of 'upper-limb disability and symptoms', but in trying to measure both concepts, it does not measure either concept well. The items of the DASH make sense when considered from a clinical perspective; the items reflect many of the questions asked of patients during a consultation, such as symptoms, function and psychological wellbeing (Fayers and Machin 2013). Factor analysis studies have found that the DASH is composed of at least two or three separate constructs (Lehman, Woodbury and Velozo 2011; Rodrigues et al. 2016). The lack of a unifying conceptual basis for the items of the DASH means that there is no sensible hierarchy created by the items, making measurement illogical.

Our remodelling of the DASH into two separate scales with coherent conceptual bases led to improved measurement characteristics. The number of disordered thresholds was halved, the number of residual correlations dropped by 70%, and other item fit statistics also improved as listed in Table 4.4. This simplistic revision serves to illustrate how the DASH could be improved.

There are limitations to this study. Our sample size was 210, which is considered adequate for Rasch analysis of dichotomous scales (Linacre 1994); however, as the DASH has five response options, a larger sample would have been preferable (Linacre 1999). A larger sample size is unlikely to have improved the fit of the Rasch model as smaller sample sizes are more likely to show falsely high indices of fit (Hobart and Cano 2009). The sample also included some participants who completed the DASH more than once. However, multiple measurements at different points in time are not considered repeated measurements. They are best considered independent assessments as the participant's status is expected to be different at various stages of clinical recovery.

This study has illustrated the limitations of traditional psychometric analysis and the risks of hand clinicians being misled by the results of this analytic approach. Rasch analysis has identified issues regarding the validity of the DASH and the appropriateness of its use in research and clinical practice. Mathematical analysis aside, the logic of adding together items that ask about symptoms with those that ask about function and transportation is questionable; symptoms lead to functional impairment in a cause-and-effect pattern (Fayers et al. 1997). Clearly interpretable information for clinical use requires separate concepts are measured with individual scales. The summing together of varied concepts in a total summary score may be appealing, but this practice makes interpretation incredibly difficult and often inaccurate. PROs are an important part of clinical work and research in our field; selecting instruments that satisfy

the requirements of both traditional and modern psychometrics is key to having valid, responsive and interpretable data.

4.5 Summary

This study shows that the DASH does not perform well when assessed with modern psychometrics. The use of the DASH in its current form, as the PROM of choice for hand surgery clinics, should be reconsidered. We propose that important concepts such as function and symptoms should be measured with separate scales that satisfy modern psychometric analysis. This will ensure that clinicians are using scientifically sound instruments that are clearly interpretable and allow for meaningful application.

Chapter 5: Developing the Hand-Q Phase 1: Qualitative Study

A methods protocol of this chapter has been published as an article in the in the *British Medical Journal Global Open*, attached as Appendix C.1.

Sierakowski, Kyra, Nicola R. Dean, Andrea L. Pusic, Stefan J. Cano, Philip A. Griffin, Gregory I. Bain, Anne Klassen, and Donald Lalonde. 'International multiphase mixed methods study protocol to develop a cross-cultural patient-reported outcome and experience measure for hand conditions (HAND-Q).' *BMJ Open* 9, no. 3 (2019): e025822.

5.1 Introduction

Based on the findings of the systematic review in Chapter 2, the acceptability study in Chapter 3 and the in-depth analysis of the DASH in Chapter 4, it is evident that there is a requirement for a methodologically robust PROM in the field of hand conditions.

Developing a new PROM begins with a preliminary conceptual framework to guide a qualitative study. This framework includes a description of the concepts of interest, and the relationship that exists between these concepts within the population of interest. Concepts of interest were generated from existing PROMs and from content identified as lacking in existing instruments (Chapter 2). This preliminary conceptual framework was used as the basis for an interview guideline (Table 5.1). Qualitative interviews with patients who have lived experience of hand conditions and the associated treatments were used to establish the concepts that matter to this patient cohort.

The aim of this study was to use qualitative techniques to identify the concepts that are important to patients with a hand condition in terms of their outcome and their experience. This information from the perspective of the patient was gathered in a series of in-depth interviews and then used to build a comprehensive framework, which will form the structural foundation of the HAND-Q scales.

5.2 Methodology

5.2.1 Qualitative Theory

The design typology used for the development of the HAND-Q is in keeping with that described by Creswell and Plano Clark's (2011) definition of exploratory sequential design—instrument development variant. In this approach, the findings from a qualitative study with a small cohort (Chapter 5) informs the development of a questionnaire (Chapter 6). The generalisability of the questionnaire is then tested by its application in a large cohort using a quantitative study (Chapters 7 and 8).

The qualitative interview technique employed was an approach from the applied health sciences field known as interpretive description (Thorne, Kirkham and MacDonald-Emes 1997). This approach acknowledges that there is pre-existing theoretical and clinical knowledge informing the study, which is appropriate given that much is known already about this field. In addition, this approach aims to produce knowledge relevant to the clinical context with the proviso that the individual's understanding of a concept is of the greatest importance, regardless of the clinical or theoretical explanation (Welford, Murphy and Casey 2011). The consolidated criteria for reporting qualitative studies (COREQ) checklist is reported on where applicable (Tong, Sainsbury and Craig 2007).

5.2.2 Ethics

Ethics approval was obtained from the Southern Adelaide Clinical Human Research Ethics Committee (HREC/17/SAC/5), (Appendix C.2) and the Horizon Health Network Research Ethics Board (2017-2499) (Appendix C.3). These approvals were to perform in-depth qualitative interviews and cognitive interviews (discussed in Chapter 6).

5.2.3 Participant Recruitment

Australian participants were recruited from a publicly funded tertiary hospital (Flinders Medical Centre) and from the private practices of two hand surgeons (Professor Gregory Bain and Dr Philip Griffin). Canadian participants were recruited from the clinic of a hand surgeon (Dr Don Lalonde) with an attachment to a tertiary hospital. Potential participants were invited to participate in a qualitative interview by a member of the healthcare team at each site. Invitation to participate was extended either in person or over the telephone. Patients were eligible if they were currently receiving treatment for a hand condition and had experienced hand surgery in the preceding 5 years (ideally within the last 12 months), with a minimum of 4 weeks since their last surgery. Shorter periods post-surgery were acceptable if the participant had previous past experience of hand surgery. Participants were excluded if they were unable to converse fluently in English or if a cognitive problem prevented their ability to contribute. Participants were purposively sampled to include a heterogeneous population with respect to age; gender; hand condition; setting of surgery (hospital operating theatre vs private rooms); funding for surgery (public vs private); and whether surgery was performed with GA, sedation or local anaesthetic.

5.2.4 Data Collection

Interviews were scheduled at a time of mutual convenience and held either in person (at Flinders Medical Centre or at the private practice of the referring surgeon) or on the telephone depending on the preference of the participant. Written consent was obtained from all participants. Participants could have an accompanying person attend for their comfort; however, it was emphasised that it was the participant's experience that was of interest. Interviews were conducted in a semi-structured manner, which allowed for topics to be introduced by the interviewer and then elaborated on by the participant (Shauver and Chung 2010). Audio recording of the interviews allowed the interviewer to focus their attention on the discussion without the need to produce comprehensive field notes (although some notes were taken), while still producing a record for transcription and analysis.

An interview guide (Table 5.1) was constructed on the basis of the preliminary conceptual framework. Open-ended questions were used to encourage discussion. The interviewer actively probed for new concepts throughout the interviews. The interview guide was adapted to reflect any updated content. Demographic data were collected; this included participant age, gender, hand condition and date of most recent hand surgery. For consistency, a single female interviewer (author KS) performed all of the qualitative interviews.

The interviewer was a full-time PhD student at the time of the study, with a background as a medical doctor. Although she had no prior experience with qualitative work, she received informal training in qualitative methods by experts in the field prior to the commencement of the study. The interviewer had no significant prior relationship with any of the participants; however, participants were aware that the interviewer was a medical officer, studying a higher degree, with an interest in hand surgery. As a medical officer with ambitions to train in the field of plastic and reconstructive surgery, the interviewer would have entered this study with prior

knowledge and possible biases that may have influenced this study. To mitigate potential bias several strategies were used to ensure robust qualitative analysis. Examples include: coding analysis of the data was carried out by two researchers, with multiple sources consulted (patients and clinicians, therefore performing triangulation), member checking was utilised to ensure that interpretation of data was consistent and peer debriefing was used throughout the study.

Table 5.1: Preliminary interview guide

Experience of care
1. What treatments have you had for your condition?
2. What was good or bad about the treatment?
3. If the participant has had surgery:
a. What was your experience of the anaesthetic used? (probe: general anaesthetic, block, local)
b. Would you have considered having treatment under local anaesthetic? (probe: why, why not)
4. Who do you see when you come to the hospital clinic? (probe: receptionist, nurse, doctor, occupational therapist)
5. What are the people like who care for you? (probe: friendly, made you feel comfortable, easy to talk to, listened to you)
6. What kind of verbal and written information did they give you? (probe: gave enough information, let you ask questions, answered your questions, provided information about recovery)
Physical function
7. Does your condition create any functional problems? (probe: work, personal care, hobbies)
8. What specific things do you have difficulty with due to your hand problem? (probe: getting dressed, cooking, typing, sport)
9. Do you experience any symptoms related to your functional problem? (probe: pain, discomfort, embarrassment, mood disturbance)
Psychological wellbeing
10. How does your hand problem make you feel? (probe: frustrated, angry, upset, worried, stressed)
11. How does your hand problem make you feel about yourself? (probe: self-esteem, body image, confidence, self-conscious, different from others)
Appearance
12. How would you describe the appearance of your hand/s? (probe: from close up, from far away, symmetry, texture, attractiveness)
13. How has your hand appearance changed since your treatment? (probe: scarring, descriptive detail)
14. What do you like or dislike about your hand appearance?
15. Is there anything about your hand appearance that you would like to change? (probe: for details)
16. Do you ever hide your hands? How do you do this?
17. How important is the appearance of your hands to you?
Other
18. Is there anything I have not asked you that you think it is important for me to know?
19. Would you like to receive a copy of the transcript from today's discussion?
20. Would you be interested in participating in cognitive interviews?

5.2.5 Data Analysis

Interviews were transcribed verbatim with all identifiers removed. Transcripts of the interviews were provided to participants if they expressed interest in reviewing their transcript.

A 'line-by-line' approach was used to code the data. An iterative technique was used, whereby data collection and analysis took place concurrently to explore the relevance and importance of any newly identified concepts. Quotations from the participants about any aspect of outcome or experience of care were copied into an Excel spreadsheet along with demographic information. Using a process of constant comparison to identify common concepts of interest, these data were categorised into conceptual domains, themes and subthemes (Pope, Ziebland and Mays 2000). Initial coding was performed by a team member (KES) and then confirmed by the author (KS). A third team member (AK) oversaw this process and resolved any conflicts. The members of the study team discussed the data analysis throughout the study and thus used peer-debriefing to provide consistency (Cohen and Crabtree 2008).

5.3 Results

5.3.1 Demographics of Participants

A total of 62 interviews were performed with participants from Australia and Canada; summarised demographic information is shown in Table 5.2, and participants' individual demographic information is presented in Appendix C.4. Interviews averaged 34 minutes in duration (range, 13–61 minutes). Participants were on average 6 months post their most recent hand surgery (range, 1 week to 4 years). No participants withdrew from the study.

Table 5.2: Participant characteristics

Characteristics	Australian	Canadian	Total
Number of participants	40	22	62
Age, mean (range)	63 (38–78)	67 (27–85)	65 (28–86)
Gender, male/female	18/22	10/12	28/34
Condition			
Trigger finger	4	4	8
Osteoarthritis	8	0	8
Rheumatoid arthritis	1	0	1
CTS	8	12	20
Trauma	6	1	7
Dupuytren's contracture	11	3	14
Other	2	2	4

5.3.2 Coding

The line-by-line coding technique that was used to analyse the interview transcriptions is illustrated below. A participant, male aged 66 years with Dupuytren's contracture (a condition in which one or more fingers become permanently bent in a flexed position), said the following:

> '… oh yes the most annoying thing for me was shaking someone's hand because every time I did that they'd want to talk about it and I didn't want to talk about it.' (Ref 28)

This quotation resulted in three codes as follows: 'Psychological/Annoy' for feeling annoyed, 'Physical/Function' for difficulty shaking hands and 'Social/Isolation' for not wanting to talk about their condition.

5.3.3 Concept Elicitation

Coding of interview transcripts resulted in the identification of five wellbeing domains and five satisfaction domains.

5.3.3.1 Wellbeing Domains

5.3.3.1.1 Hand Function

A major domain that emerged from the interviews was hand function. Subthemes were identified as tasks of daily living (TDLs), accommodation, activities and mobility actions. The frequency of the subtheme codes is displayed graphically in Figure 5.1. An example of how a participant's quotation would be coded is shown in Table 5.3.

Table 5.3: Example of coding for hand function

Ref	Country	Gender	Age	Condition	Quotation	Coding
40	Australia	Male	76	Rheumatoid arthritis	'Well in my hands it causes me disability, being **unable to clothe myself**, I have to have **special knives and forks**. I find that I find it **difficult to do up buttons**. I like writing and **I find it difficult to write**, but I do write, I wrote all that there.'	- TDL: dressing - Accommodation: equipment—cutlery - TDL: eating—cutlery - TDL: dressing—buttons - TDL: writing

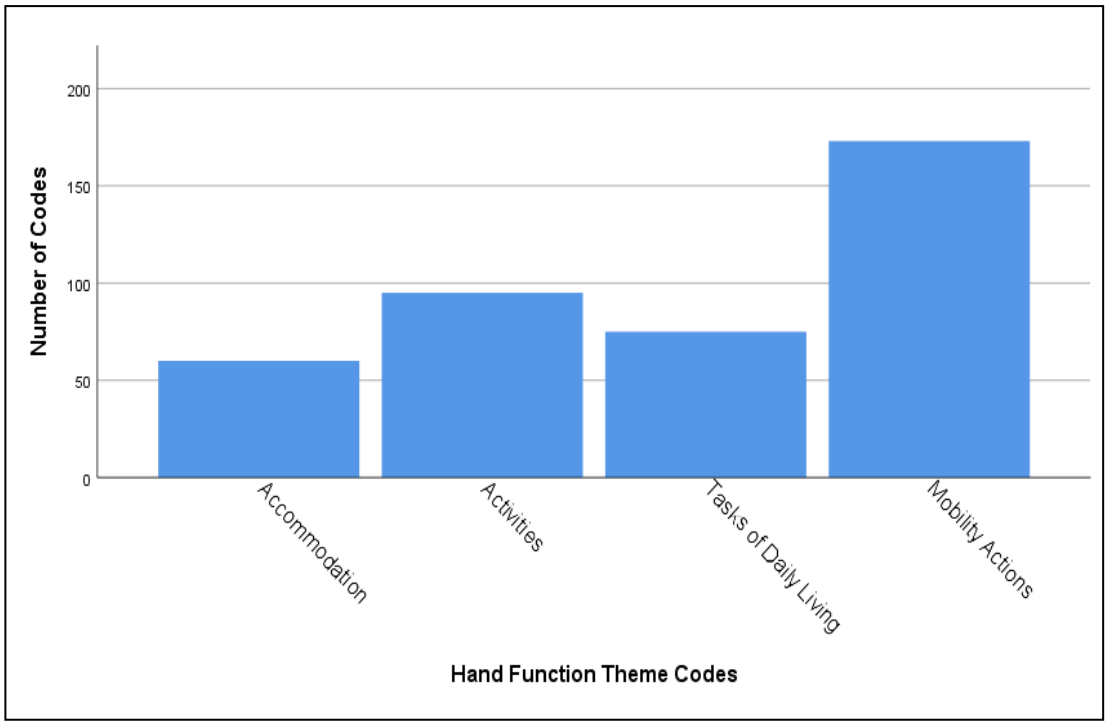

Figure 5.1: Hand function codes.

TDLs encompassed functional difficulties that participants experienced with their hand(s) while carrying out everyday tasks. Most of these tasks apply widely to most people. Tasks such as dressing, cooking, eating, self-grooming and personal hygiene were common issues:

> ‘So many um ADLs that I just couldn’t do, I couldn’t cut my own toenails, I couldn’t pluck my eyebrows, I couldn’t use scissors, so many things that put pressure on that thumb … I couldn’t tie my shoelaces up.’ (Ref 24)

> ‘But it was more about cooking! It was so annoying, you known you might be flattening something down and you’ve got the knife in your hand, the way the finger was I couldn’t get right down on something.’ (Ref 2)

> ‘I just sort of can’t take weight, like a heavy plate—last night passing a plate of food around the table … I felt as though my hands couldn’t hold it, because it was too heavy.’ (Ref 4)

> ‘Oh I think that the biggest problem is that I can’t do finer things with my hands. Like I can’t unscrew … umm we use long life milk and I can’t unscrew the top on that.’ (Ref 40)

> ‘I was getting to the point where I was having lots of trouble holding a pen and writing even the fatter ones I was having trouble with.’ (Ref 26)

The **accommodation** subtheme included how participants had adapted to their hand condition, at times by using their unaffected hand; others used equipment or changed behaviours:

> ‘So things that I would normally do with my left hand I do with my right hand.’ (Ref 6)

> ‘I ended up having to get a lot of ergonomic things in so like kettle tippers, knives with the handles that stick up rather than, and forks, spoons that tilt.’ (Ref 27)

> ‘If we go out to dinner or to lunch or something … usually lunch. I have to order something that can easily be cut up into little bits … so that I can eat it.’ (Ref 36)

The **activities** subtheme was used to code for any activities that the participant named as being affected because of their hand condition. Many participants described having to give up on activities that they previously enjoyed, such as playing instruments, gardening, crafts and sports. Some had also experienced the ability to return to these activities post treatment:

> ‘I couldn’t knit or crotchet do all the things I love to do I love to sew, I like to knit, I like to do a lot of things … I’m back in to my gardening and sewing and knitting and. In fact I just finished doing my fifth pair of socks for Christmas.’ (Ref 58)

> ‘Riding my bike was probably the worst. Um, I got to a point where I couldn’t actually pull the brake.’ (Ref 8)

Mobility actions were specific hand movements or actions that participants identified as being problematic for them, not specific to activities per se. Many participants described difficulty gripping things, dropping things, and difficulty picking up objects or holding objects. Often mentioned problems were using pockets and shaking hands:

> ‘To carry something that doesn’t have a handle I would use my right hand because I don’t feel like I have a good grip … No, it’s more gripping things I think. Picking things up and holding things.’ (Ref 6)

> ‘I can’t hold a book either that really bothers me. I can’t hold my book anymore I’m an avid reader so to hold a book that’s hard.’ (Ref 50)

‘I’m a big broad guy you want to be able to shake hands like a man. You know, and he is the sort of guy you know, you want to give him a good squeeze and you know.’ (Ref 8)

5.3.3.1.2 Symptoms

Hand symptoms were another major domain that was identified with a total of 535 codes; subthemes were classified as pain, sleep disturbance, numbness, stiffness and weakness. Table 5.4 shows an example of the coding for this domain. Figure 5.2 shows the frequency of the subtheme codes; the most common codes related to pain (269) and numbness (92).

Table 5.4: Example of coding for hand symptoms

Ref	Country	Gender	Age	Condition	Quotation	Coding
22	Australia	Female	75	Osteoarthritis	‘This one has been a real **pain in the proverbial here, it’s still aching**, it has not stopped aching, it drives me insane at night because it, during the day I don't seem to notice it **but at night it aches** and it’s still aching, it’s aching up here today.’	- Symptom: sleep—disturbance - Symptom: pain—aching

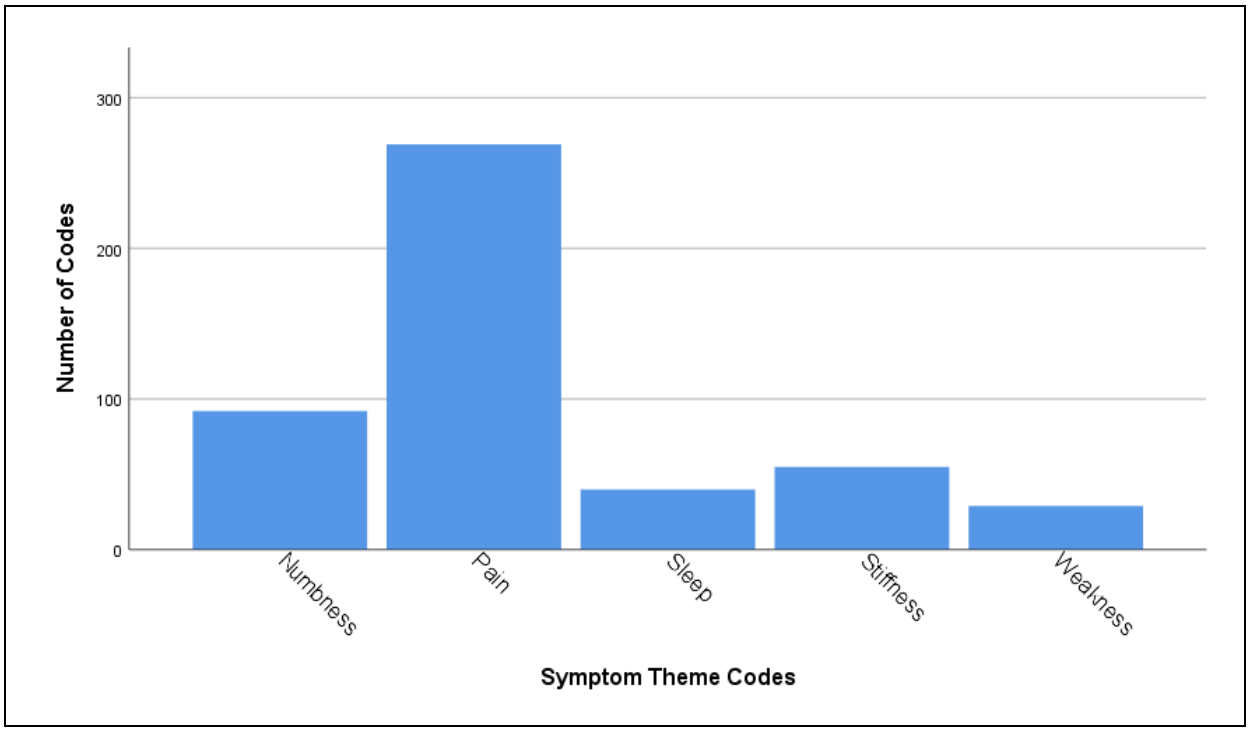

Figure 5.2: Symptom codes.

Pain was described in terms of its character and severity:

> 'I think I am now more comfortable than preop. Cos there was a lot of grating and grinding within that joint.' (Ref 20)

> 'I got into the car to turn the ignition on and I got this, felt like a knife going in to the joint.' (Ref 24)

> 'The pain was horrendous that I'd just be in tears because it hurt so much.' (Ref 11)

Numbness and tingling were common and often described as different symptoms:

> 'And I had a lot of numbness and tingling and a lot of pain with the discomfort in my arm it was just like a, I would say like an ache, and then the numbness feeling for, at night I'd wake up I would be completely pins and needles, no feeling.' (Ref 58)

Sleep disturbance was another common theme, whether it was due to pain or numbness disturbing sleep or the inability to become comfortable because of the hand condition:

> 'I had to get in the right position to try and get off to sleep you know where it was less painful.' (Ref 20)

‘It started waking me. It started waking me up from a dead sleep and that, you know having to run your hands under cold water and try to stop the pain, it was horrible.’ (Ref 23)

‘Oh, it used to be painful. Wake up night time and the pain I’d have like 2 or 3 Tylenol and then it’d ease off and I’d go back to bed. Same thing every night.’ (Ref 46)

Many participants also reported **stiffness** and **weakness**:

‘Anytime I move them they’re stiff, they’re sore, they feel like they’re, they’re not swollen but they feel swollen.’ (Ref 50)

‘It just feels so weak. It just feels so weak.’ (Ref 11)

5.3.3.1.3 Hand Appearance

Hand appearance was identified as a major domain with a total of 385 codes generated on this topic. Table 5.5 shows an example of the coding for this domain. Subdomains were used to describe the anatomical location: hand, finger, knuckle, nail, palm, skin, thumb and scar. The majority of codes described the appearance of the hand as a whole (n = 158) or the fingers (n = 102).

Table 5.5: Example of coding for hand appearance

Ref	Country	Gender	Age	Condition	Quotation	Coding
40	Australia	Male	76	Rheumatoid arthritis	‘I think in those 10 years, coz my fingers went off very quickly. I think that was one of the reasons that **I thought that a girl would never have me.** Because my fingers **looked like claws.** Just a bit better than a birds claw.’	- Appearance: qualitative - claw - Psych: other - confidence

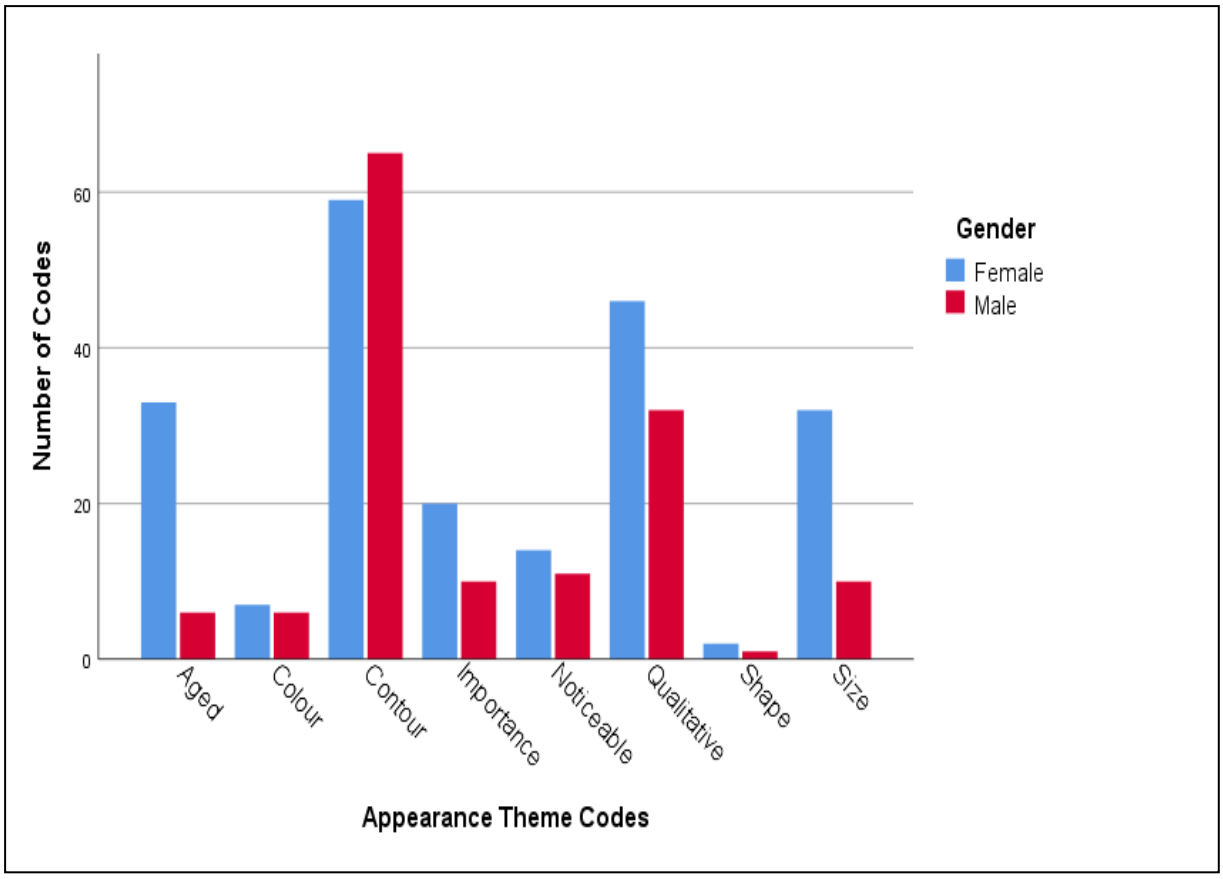

Figure 5.3: Hand appearance codes by gender of participant.

Subthemes were used to describe the content: aged, contour, qualitative, size, shape and importance. The most common code was to do with the contour (n = 124) or the qualitative description (n = 78) of the hand. Figure 5.3 shows that both male and female participants produced codes; more codes were generated from female participants in all appearance subthemes except contour:

> ‘Um, so my hands look older like you know when I look at my hands, you know they strike me as a 54 year old women’s hands and I really don't think of myself as older so when I look at my hands it reminds me of my ages and I don't necessarily like that.’ (Ref 61)

> ‘I mean I try to do a fist and I have fingers going in all sorts of funny directions that you put them out straight and you don’t notice it so much.’ (Ref 26)

> ‘I used to say they were like talons.’ (Ref 13)

> ‘But I don’t think that it looks particularly good. And whilst I was told that there would be a small hole in the little finger, I don’t know … I think that is a bit more than small.’ (Ref 31)

There was a lack of consensus among participants about the importance of hand appearance, with the cohort divided between those who thought it was of great importance and those who were indifferent:

> ‘I pretty much wouldn’t leave the house unless I had the splint on, cause I was that self-conscious of how my hand looked like.’ (Ref 53)

> ‘They’re not the prettiest hands anymore but I don’t really care so long as they work.’ (Ref 18)

5.3.3.1.4 Psychological Impact

Participants described the impact of their hand condition(s) on their emotional state; this resulted in 324 codes. Most statements were about the negative psychological impact of their condition (n = 294). There were several participants who reflected on their condition with acceptance (n = 30), which presented a different subtheme to the other psychological constructs. Table 5.6 shows examples of coding for negative and positive psychological effects.

Table 5.6: Examples of coding for psychological impact

Ref	Country	Gender	Age	Condition	Quotation	Coding
50	Canada	Female	68	CTS	'It's just I'm aware that it's going to hurt and I don't like it. I uh, it **saddens me, it depresses** me, i**t's frustrating and** it can be **embarrassing**.'	- Psych: depressed mood - Psych: frustrating - Psych: embarrassment
33	Australia	Male	66	Trauma	'I mean **I get annoyed that I can't do things that I used to be able to do**, but you just **have to find a way around it. Or do nothing. And doing nothing for the rest of my life is not really an option.**'	- Psych: acceptance - Psych: annoyance

Subthemes identified within this domain were annoyance, anxiety, concern, frustration, depressed mood, self-consciousness and embarrassment. The frequency of these subthemes and the conditions that affected participants that contributed these codes are shown in Figure 5.4. A variety of conditions contributed to each subtheme, showing that the psychological effects of hand conditions span a variety of pathologies:

> 'Obviously you know there is going to be days that you can't be you can't surpass the sadness and frustration.' (Ref 53)

> 'I suppose I did start to feel sorry for myself.' (Ref 8)

> 'That was quite devastating for me, because I was looking to … ah … no I will stop there. It was quite devastating, and I didn't handle it very well.' (Ref 40)

> 'When it is like this I do get self-conscious about it and would hide my hands under the table.' (Ref 2)

There was also a separate construct, that of acceptance and the participant embracing their condition and living as best they can with their hand problem.

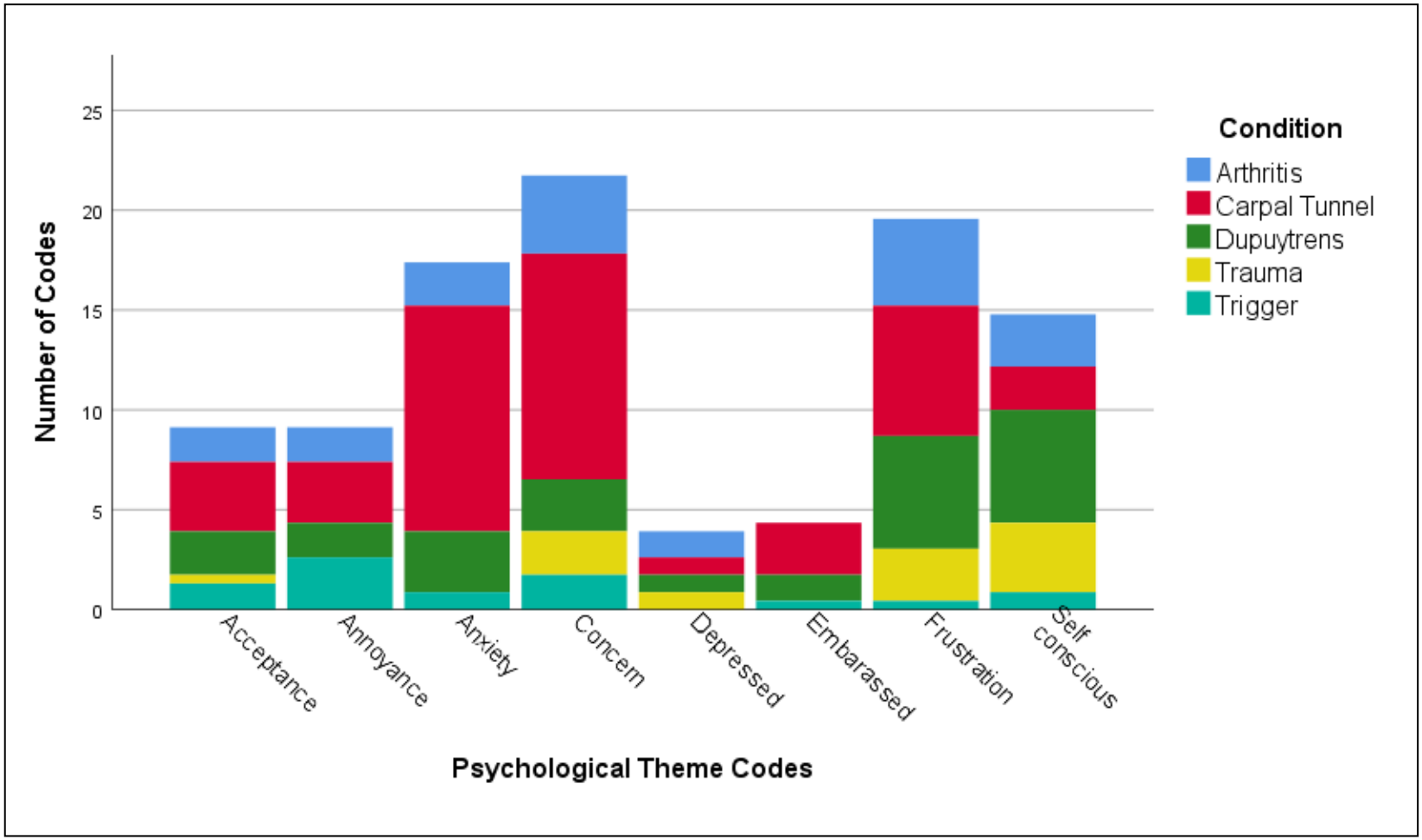

Figure 5.4: Psychological impact codes by participants' hand condition.

5.3.3.1.5 Social Impact

Another domain identified was the impact of the hand condition on the participant's social life (n = 231); subdomains were social function, social isolation and relationships. **Social function** included both work- and leisure-related issues; however, most codes were related to work (116 of 120):

> 'I'm at a point where I'm actually thinking should I be doing this work; do I want to be doing this work?' (Ref 8)

> 'So I'd be counting money or signing documents or something and all of a sudden I'd have to really stop what I was doing because I had a lot of fatigue in my wrist and I'd have sharp pains coming up my hands.' (Ref 48)

> 'Well I have practically been out of work now for 18 months, with both hands because it has just been ongoing.' (Ref 5)

Social isolation included codes relating to comments made by others about their hands or their inability to participate in social greetings:

> 'Or I remember someone made a comment about "oh you've got crooked fingers" and I went yeah you know and explained the situation and that but that does make you feel like you know.' (Ref 12)

> 'Got to a point where it was hard to shake people's hands, because your fingers had started to curl right over.' (Ref 31)

> 'Well I couldn't drive ... I was stuck at home virtually … apart from taxis.' (Ref 33)

Relationships were described as being important by those living with and recovering from a hand condition or surgery:

> 'You gotta have someone at home that can do that, like, hubby cooked for me for 4 weeks. Which was great. You gotta have some help when you do this stuff, yep.' (Ref 24)

> 'And my partner would massage my hands for me, and that would make me feel relieved, it would relieve the hands.' (Ref 12)

5.3.3.2 Satisfaction Domains

5.3.3.2.1 Anaesthesia

Participants described their perioperative experience, which was influenced by the type of anaesthesia that they had for their surgery. Subthemes were GA and local anaesthesia. Experiences varied between those who had either technique. Patients who had GA reported:

> 'I felt like absolute rubbish when I came out of the anaesthetic.' (Ref 5)

> 'I've had lots of anaesthetic and no dramas whatsoever just dozed off to sleep and woke up later thinking it's all done, good.' (Ref 20)

Meanwhile, patients who had local anaesthesia, an awake procedure, reported:

> 'I asked what he was doing, he showed me every step of the way, showed me what the nerve was supposed to look like and what it looked like when they opened me he showed me how purple it was, he said it's supposed to be the colour as a pencil and I was like wow so purple and that's how badly it was pinched off.' (Ref 41)

> 'Ah monumental because I just walk out of the surgery, I can drive home and ah no lethargic, or anything, once it wears off there is a bit of pain there but that is minimal. And ah if I could have all surgery like that then I would go with that every time.' (Ref 40)

5.3.3.2.2 Satisfaction with Staff

Satisfaction with the interaction with various members of staff was a minor domain (n = 197); subthemes were quality care and respectful care. There were positive (n = 153) and negative codes (n = 44):

> 'They make you feel like you are the only person for that period of time that you are there.' (Ref 8)

> 'They were just a little bit—the girls were talking about their relationships and stuff like that, instead of actually remembering the patient that's laying in bed.' (Ref 12)

> ‘Obviously the doctor didn’t seem to know what he was doing. Or didn’t seem to recognise what it was.’ (Ref 59)

5.3.3.2.3 Information

Information about surgery or treatment was identified as a minor domain (n = 118); subthemes were written, questions, enough, more needed, explained and google.

> ‘I really didn’t get a lot of the information, I had assumed that I would have maybe a week of work I came out of surgery and they offered a sick certificate for a month and I’m going “what”. I’ve only been at my job for a year, I don’t have sick leave like that.’ (Ref 16)

> ‘The doctors umm that came to see me. Yeah, they were pretty good. They explained the risks; they umm explained exactly what they were going to do, about the anaesthetic. That this will happen, and that will happen. Umm. Yeah I felt quite comfortable.’ (Ref 6)

> ‘Umm he could not, not could not but would not tell me with a tendon transfer what tendon I would lose … and therefore what function I would lose.’ (Ref 33)

5.3.3.2.4 Splint

If participants had to wear a splint or brace as part of the treatment of their hand condition, this was an experience that they wished to reflect on and share. Splints were generally thought of as a necessary inconvenience:

> ‘Had to try and keep it on overnight as well and it was just so uncomfortable it drove me crazy.’ (Ref 22)

> ‘The splint was just it was so cumbersome.’ (Ref 53)

> ‘And that was huge I mean I couldn’t ... there were very few clothes that I could wear. I couldn’t wear a jacket.’ (Ref 33)

5.3.3.2.5 Overall Outcome

When asked how they felt about their satisfaction with their treatment overall, participants varied in their response. Some felt very happy with their outcome, others were very dissatisfied and some were still unsure:

> ‘I’ve certainly gone the other way. It’s worse than when I went in for surgery.’ (Ref 7)

> ‘All I know is it doesn’t do what it used to do, which is good. So ultimate result is awesome.’ (Ref 8)

> ‘Um, if I still had the pain, I would be saying it wasn’t worth it, but I don’t have the pain so it was worth having it done actually.’ (Ref 57)

5.4 Discussion

There are concepts of interest found in this qualitative study that are known to be important constructs when measuring outcomes in patients with hand conditions. Existing PROMs in the field focus on hand function and symptoms as the main outcomes of interest. Unsurprisingly, this study confirmed that hand function and symptoms are important to patients with a hand condition.

Many hand functions identified have previously been reported on in other qualitative studies, such as picking up small objects, using zippers, writing, dressing and opening jars (van der Giesen et al. 2010). These functions are the focus of items in many existing instruments such as the DASH (Hudak et al. 1996) and MHQ (K. Chung et al. 1998), as well as more recent instruments such as the UEFI-15 (Hamilton and Chesworth 2013).

Although these concepts were not previously unrecognised, this study did show that these functional hand problems do span across varied hand conditions in the heterogeneous population interviewed in this study.

Another concept included in many existing instruments is hand symptoms; again, this study confirmed this as a key concept that is important to those with hand conditions. The primary symptom experienced by participants was pain, in varied severity and nature depending on their hand condition and their stage of disease or recovery. The second most common symptom was numbness. Other symptoms identified were weakness, stiffness and sleep disturbance. These symptoms are covered by currently available instruments with varied levels of comprehensiveness.

This study also revealed other concepts that are considered important by this cohort (although perhaps not as important as function and symptoms), concepts such as hand appearance, psychological impact, social impact and QOL. These concepts are not measured adequately by currently available instruments (see domain analysis in Chapter 2).

Hand appearance was particularly important to many participants, as hands are highly visible and an important aspect of our human interaction:

> 'They have a male appearance which for me is important because of the kind of work I do I try to engender a feeling of confidence in people. And I find that the use of hands and the appearance of hands is important in that context.' (Ref 45)

Bogoch and Judd describe the hands as a second face due to their importance in body image (Bogoch and Judd 2002). Although appearance was usually a secondary motive for surgery, it was still an important one for many participants, and the source of dissatisfaction with outcome if they were not happy with their hand appearance postoperatively. This finding agrees with that of others who have reported on motivation for surgery in the rheumatoid arthritis population (Alderman et al. 2006; K. Chung et al. 2006; Bogoch, Escott and Ronald 2011). One participant in this study with rheumatoid arthritis said:

> ‘They don’t look very pretty. I am aware of that … think that was one of the reasons that I thought that a girl would never have me. Because my fingers looked like claws. Just a bit better than a bird’s claw.’ (Ref 40)

The importance of hand appearance was also acknowledged by Stamm et al. (2009, 1455) in a study in mainly female participants with osteoarthritis of the hand, describing that they ‘stopped speaking with hands’ because of aesthetic changes. Within our study, a subgroup including both woman and men believed hand appearance to be important, although women did generate more codes in this domain. Other participants had the sentiment that they did not care about what it looked like so long as it worked:

> ‘It doesn’t bother me at all. Like I say, I’m never going to be a hand model. I never was.’ (Ref. 8)

The psychological effects of living with a hand condition are complex and varied depending on nature of the condition and the mechanism or injury:

> ‘It’s just I’m aware that it’s going to hurt, and I don’t like it. I uh, it saddens me, it depresses me, it’s frustrating and it can be embarrassing.’ (Ref. 50)

There is previous qualitative work analysing how patients with an acute traumatic hand injury manage stress factors (Gustafsson, Persson and Amilon 2000) and develop coping mechanisms (Gustafsson, Persson and Amilon 2002). Stamm et al. (2009) showed that psychological problems were a consistent theme identified in all five European countries involved in their study examining hand osteoarthritis. This study had a mixed cohort between elective and traumatic conditions, and although there would have been profound differences in the early traumatic experience, many of the psychological experiences on a day-to-day basis of living with a hand condition were shared. There was an overwhelming sense of frustration and concern at their inability to use their hand in the way they used to or in the way they would have liked.

These feelings of frustration developed into feelings of sadness or depression for some, with some describing feeling overwhelmed and useless:

> 'There is lots of days that there is some depression that sets in.' (Ref 53)

Some also had feelings of embarrassment or being self-conscious of their hand, not wanting people to notice their condition. These feelings led to concealing behaviours by some participants. They reported holding their hands in certain ways that would make their hand difference less visible, putting their hands behind their backs for photographs:

> 'So now I am bit a more conscious, if I'm having my photo taken—not that I like having my photo taken, I hate it. I do tend to hide them.' (Ref. 3)

The social impact of living with a hand condition was considerable. Shaking hands was a concept that repeatedly came up as a time of difficulty for people living with a hand condition. This was due to both the physical difficulty of shaking hands (due to discomfort when their hand is squeezed) and the unwanted attention that this social practice brings to their hand condition. Social isolation because of not being able to drive, and not being able to care for oneself, was also raised. The dependency on others to help care for them was an area of difficulty for many, the loss of independence causing frustration and grief for many participants.

The concept of acceptance or adjustment was present in many interviews. Many participants had the approach to just carry on with things, and cope as best they can with their hand condition. This was more evident as a psychological mindset but was also demonstrated by the many adjustments and workarounds that these participants made to enable them to perform physical tasks with their hand condition.

Another major concept explored by this study was that of patient satisfaction with the processes of care. We explored participants' relationship with their treating surgeon and hand therapist (if applicable), interactions with others as part of their journey, perioperative experience and

overall feelings of satisfaction with their treatment. Measuring patient satisfaction is not a simple concept. As discussed in Chapter 2, there is a tendency to oversimplify the measurement of patient satisfaction by use of a single item. In this study, we explored with participants what made them feel satisfied or dissatisfied with various aspects of their treatment and outcome. For many participants, it was their relationship with their surgeon that shaped their satisfaction with treatment, regardless of their satisfaction with their outcome:

> 'I thought he is a nice, young, with it, obviously a very skilful person, lovely personality.' (Ref 36)

Most participants described positive relations with their surgeon; however, this might have been due to recruitment bias discussed later. There were some participants who had negative relationships with their surgeon, describing not feeling respected or listened to.

Information exchange about the surgery and treatment was another domain that was important to participants. Many described not feeling like they were included in the decision-making process. Obviously, this is a different circumstance in a trauma versus elective surgery setting. Many participants did not realise the length of recovery post-surgery and the effect this would have on their ability to work and function independently:

> 'So I mean his personality, I couldn't ask for a nicer guy to explain everything and to go through everything, so he's very much a gentleman. So yeah no I have no complaints, I really don't.' (Ref 58)

The perioperative experience was discussed with a focus on the participants' experience of their anaesthesia and the awake procedure. Themes emerged about the side effects of anaesthesia, with some participants feeling unwell after general anaesthesia. Those that had their surgery awake reported varied experiences; many enjoyed being involved in the procedure, discussing the surgery as it was taking place and even seeing the surgical reasons for their hand issue.

Others described the awake procedure as being unpleasant and expressed preference for being under GA with any later treatments.

5.4.1 Potential Weaknesses of Study

The main limitation of this study is the sample of participants did not include the full breadth of hand conditions that are seen in clinical practice. While common conditions such as carpal tunnel and trigger finger were well represented the sample did not include many patients with hand deformity or amputation and none with a congenital hand difference. Absence of these cohorts was not deliberate but due to an inability to enrol participants who have these conditions. Further qualitative work can be performed to review the completed HAND-Q scales with individuals with conditions that were not included to ensure that they are relevant and applicable to these patient cohorts.

This study is not without weaknesses. The interviews were conducted at clinical locations that may have introduced some bias with participants feeling less inclined to report dissatisfaction while at the place of their treatment. There is also the possibility of selection bias in that people who were dissatisfied with their treatment or outcome might have been less likely to have been asked to participate and potentially less likely to agree to take part. The interviewer was known to be a medical doctor training in the field of plastic surgery, and this might have influenced the participants' responses.

Although international, the interviews were performed in two English speaking countries with similar economic and cultural environments. This study would have been strengthened by the inclusion of participants from varied cultures and countries with varied economic status. This weakness is addressed by the cultural adaptation and field testing of the HAND-Q in varied languages and cultures in later chapters.

5.5 Summary

This qualitative study has explored what concepts are important to people with a hand condition in terms of their outcome and their experience. Some of these concepts were confirmed as relevant as they had been already identified and are measured with currently available PROMs (hand function, hand symptoms). Other concepts were more unique and are not measured well by currently available PROMs (hand appearance, psychological impact, social impact, quality of life, satisfaction with process of care, satisfaction with splint). The conceptual map forming the basis of the HAND-Q development has been refined and further detailed by this study. This rich qualitative data set will form the basis of the HAND-Q scales in the following chapter.

Chapter 6: Developing the HAND-Q Phase I: Item Generation and Content Validation

6.1 Introduction

The qualitative study described in Chapter 5 resulted in the identification of key wellbeing and satisfaction domains. The purpose of the HAND-Q is to provide a suite of scales that are able to legitimately measure the concepts that are important to patients whom experience a hand condition. Each scale will be scored separately, as each measures a separate unidimensional construct. There would be no unifying or global score, but rather independent scale scores to ensure ease in clinical interpretation and measurement of change. The target population for the HAND-Q scales includes all patients with a hand condition that are able to read and comprehend the scales, for this reason the readability of the scales was aimed to be simple and easily interpretable. A subset of the scales are only applicable to the post-surgery or post intervention patient cohort. The scales are designed for use in clinical practice by clinicians (surgeons, hand therapists etc.), researchers and also service providers performing quality assurance assessments (clinic administrators, service evaluations).

The aims of this chapter are twofold. First, a draft HAND-Q instrument was developed using the same methodology used to create other Q-PROMs (Wong Riff et al. 2017 and Pusic et al. 2009) . The rich qualitative data from the previous study formed the basis for item generation. A comprehensive item bank was developed, with questions that capture the issues that matter to patients. Items were then organised into individual scales, each exploring a separate concept. Accompanying instructions, response options and recall periods were developed alongside the scales.

Second, these scales were validated through processes with both patients and experts in the field. A diverse sample of the target population participated in cognitive interviews. The purpose of these interviews was to assess comprehension of the questionnaire and evaluate comprehensiveness (Patrick et al. 2011b). Experts were asked if the scales explored clinically relevant concepts and whether there was any additional content that was necessary. The recommendations of the COSMIN methodology for assessing the content validity of PROMs was used as a guide throughout this content validation study (Terwee et al. 2018).

6.2 Methodology

6.2.1 Ethics

Ethical approval to perform cognitive interviews with the Australian and Canadian participants was included in the cited applications in Chapter 5. To include a participant cohort from the US, ethical approval was obtained from the Office of Research and Innovation, Lehigh Valley Health Network, Allentown, Pennsylvania, United States (STUDY00000046) (Appendix D.1).

6.2.2 Item Generation

From the qualitative dataset produced in Chapter 5, participant quotations were used to develop a set of items for each domain identified. As stated by Streiner, Normal and Cairney "patients and potential research subjects are an excellent source of items" (2015).

As an example; this quotation was used as the basis for several items on different scales:

> '… oh yes the most annoying thing for me was shaking someone's hand because every time I did that they'd want to talk about it and I didn't want to talk about it.' (Ref 28)

Table 6.1: Example of coding used for item generation

Coding		Associated stem/Item	Response options
Domain	Subtheme		
Function	Mobility actions	Stem: How difficult is it to use your hands? With your hand problem in mind, how difficult would these tasks have been: Item: Shaking someone's hand?	Not at all difficult A little difficult Moderately difficult Extremely difficult
Social	Isolation	Stem: Does your hand problem affect your social life? With your hand problem in mind, how much do you disagree or agree with each statement: Item: I avoided greetings (e.g., waving or shaking hands).	Definitely disagree Somewhat disagree Somewhat agree Definitely agree
Psychological	Annoyance	Stem: Does your hand problem affect how you feel? With your hand problem in mind, how often have you felt: Item: Annoyed?	Never Sometimes Often Always

When developing items, the wording of original participant quotations was kept intact as much as possible. This was to ensure that the items were easy to understand and the concepts resonated with patients. The item pool was used to form scales with the psychometric approach of RMT (B. Wright and Masters 1982). This approach requires that items map out a concept of interest by way of a clinical hierarchy (measuring from a little to a lot of a concept). Therefore, each item was designed to measure the concept of interest in varying amounts. For example, in

the Psychological scale, the items range from those that would be easy to endorse for most people with a hand problem (e.g., frustration) to more difficult to endorse (e.g., stressed) to the most difficult to endorse (e.g., sorry for self or overwhelmed). Similarly, the items for each scale were designed to measure a different quantity of the concept of interest.

For each scale, the response options were limited to four labelled options for simplicity and to abide by the recommended guidelines (Khadka et al. 2012). The response options deliberately avoided a neutral response option as the amount of the construct measured by a neutral option is unclear and does not fit the mathematical model of RMT. The response options were chosen according to what was logical to measure the construct of interest. For example, the Symptoms scale response options are *none*, *mild*, *moderate* and *severe*. The Hand Appearance scale response options allow for positive or negative response options: *very dissatisfied*, *somewhat dissatisfied*, *somewhat satisfied* and *very satisfied*. Response options for all scales are listed in Tables 6.2 and 6.5.

Table 6.2: HAND-Q response options

Response format	Response options			
Satisfaction	Very dissatisfied	Somewhat dissatisfied	Somewhat satisfied	Very satisfied
Difficulty	Not at all difficult	A little difficult	Moderately difficult	Extremely difficult
Severity	None	Mild	Moderate	Severe
Amount	Not at all	A little bit	Quite a bit	Very much
Frequency	Never	Sometimes	Often	Always
Agree/Disagree	Definitely disagree	Somewhat disagree	Somewhat agree	Definitely agree
Bothered	Not at all bothered	A little bothered	Moderately bothered	Extremely bothered

Instructions were developed to orientate the respondent to the task required by each scale. The format of the scale instructions, items and response options was based on existing Q-PROMs such as the CLEFT-Q (Tsangaris et al. 2017). Wording was kept as brief and simple as possible to aid patient interpretation and translation.

The timeframe that respondents are asked to reflect on when answering the items is termed the recall period. The recall period is an important consideration when designing a scale as it dictates the context of the concept that is being measured. For example, if you ask the how much pain the respondent has at the time of completing the questionnaire this would not capture that they have had severe pain for the last several days but are currently experiencing a reprieve (Patrick et al 2011b). The selection of a recall period is a compromise between being too long (with the risk that participants are unable to give an accurate reflection of their condition) and being too short (where participants may not have attempted to perform each of the items). For most of the HAND-Q scales, the recall period was set as the preceding week. This recall period is used in existing instruments including the other Q-PROMs (Pusic et al. 2009; Klassen et al. 2016; Wong Riff et al. 2017). Several scales do not specify a recall period as the concepts they are measuring are not necessarily appropriate to reflect on a set time period; an example of this is the Hand Acceptance scale. The recall periods for each of the HAND-Q scales are listed in Table 6.5.

6.2.3 Cognitive Interviews

The preliminary scales were shown to participants with hand conditions in semi-structured cognitive interviews. These interviews were conducted by telephone, audio recorded, transcribed and analysed using the same techniques as described in Chapter 5. The purpose of the cognitive interviews was to ensure that the target population understood the instructions, items and response options. The ‘think aloud’ technique was used; this involves the participant

talking through the questionnaire and flagging any content that is unclear (Van Someren, Barnard and Sandberg 1994; Collins 2003). The interviewer also sought feedback on the relevance of the items to the individual. If a phrase was unclear to a participant, it was discussed how the wording could be improved to facilitate understanding. The interviewer probed for any further content that the participant thought would be a valuable addition.

A total of 20 cognitive interviews were performed to confirm understanding of the HAND-Q scales. These interviews were performed in three consecutive rounds to allow for changes to be made to the scales between rounds. There were 9 participants who contributed in round 1, 7 in round 2 and 4 in the final third round. An item tracking matrix was used to document the changes made to the items between rounds of interviews (Patrick et al. 2011b). At the completion of the third round of interviews, there were only very minor changes made to the HAND-Q and it was determined that saturation had been achieved for the cognitive evaluation.

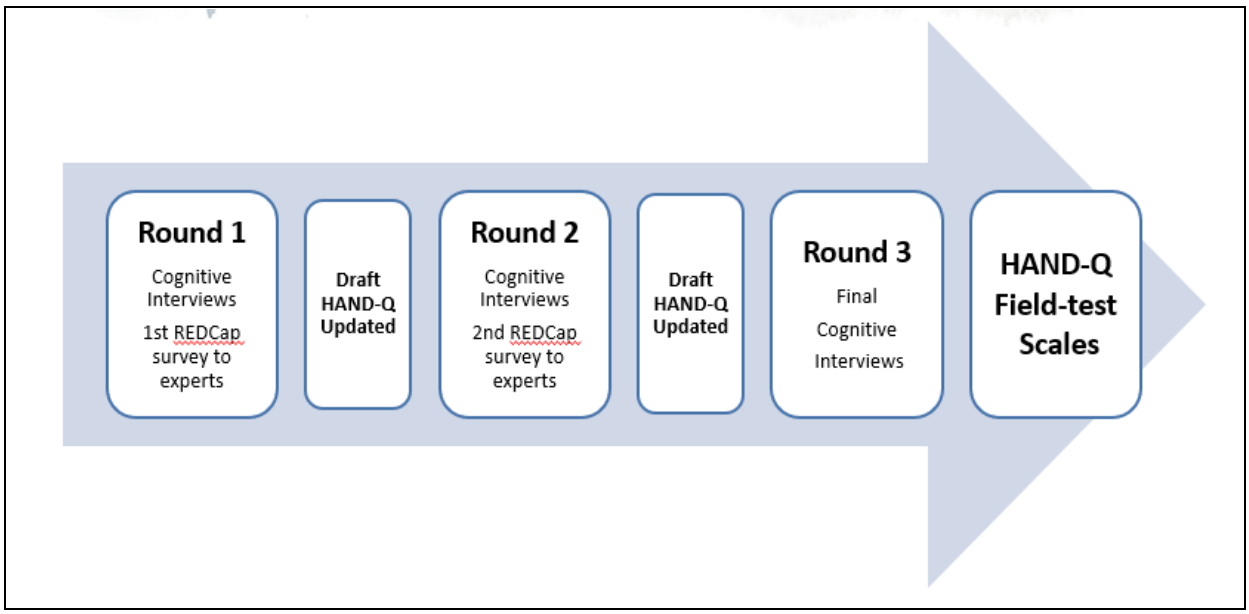

Figure 6.1: The process of cognitive interviews and professional feedback.

6.2.3.1 Participant Recruitment

Cognitive interview participants from Australia and Canada were a sub-cohort of those who participated in the qualitative interviews (details of which are in Chapter 5). During the in-depth qualitative interview, participants were asked if they would be willing to take part in a further interview to review the initial scales of the HAND-Q. A cohort from the US was included to ensure content validity of the HAND-Q in this country. American participants were invited to participate in a cognitive interview by members of the healthcare team (Professor Robert Murphy and Dr Mamtha Raj).

6.2.4 Professional Input

Professional feedback was sought on the preliminary scales to ensure that they were comprehensive and relevant to the full spectrum of hand conditions. An online REDCap (Research Electronic Data Capture) survey was created to collate this input (P. Harris et al. 2009). Professionals were identified through networks and sent the survey link via email in January and February 2018. Reminder emails were sent prior to the closing date of the survey. The survey was anonymous to encourage honest and open feedback. Professional feedback was collected in two rounds and incorporated into the updated version of the HAND-Q, as illustrated in Figure 6.1.

6.3 Results

6.3.1 Cognitive Interviews

Table 6.3 and Appendix D.2 show the participant characteristics in the cognitive interviews. The HAND-Q scales were discussed with 20 participants, purposively sampled from three different countries. Sixteen of these participants took part in the initial qualitative study (Chapter 5). Interviews were on average 70 minutes in duration (range, 14–114 minutes).

Table 6.3: Demographic information of cognitive interview participants

Characteristics	Australian	Canadian	American	Total
Number of participants	9	7	4	20
Age, mean (range)	61 (47–76)	64 (55–76)	56 (32–76)	60 (32–76)
Gender, male/female	3/6	2/5	2/2	7/13
Condition				
Trigger finger	1	1	2	3
Osteoarthritis	4	2	3	7
Rheumatoid arthritis	1	0	0	1
CTS	2	5	7	9
Trauma	2	2	3	5
Dupuytren's contracture	2	1	1	3
Other	1	1	2	3

6.3.1.1 Instructions

The following feedback was obtained from participants regarding the instructions:

- Function scale instructions: Is the note about 'gadgets' understood by respondents?

'Okay, so this is opening a jar I would put 4, I do have a device which I use, but without it I cannot open a jar.' (Ref 27)

- QOL scale instructions: For the question 'How much has your hand problem interfered with the following?', are patients able to think of QOL in terms of only their hand problem?

'Yeah, so this is a tricky one … cause it's my brain and not my hand always! But I answered this according to my hand.' (Ref 24)

- Recall period: Are patients able to reflect on just the past week?

> ‘So the next one I found a bit hard, because you’ve written “in the past week” and I know it’s post-surgery, I just think it’s really different because some things I couldn’t manage at all by the time I had surgery, and some I couldn’t manage then and I still find them very difficult. And it’s changed the way I do things.’ (Ref 24)

6.3.1.2 Response Options

The format of the response options was improved as a result of one participant who suggested that the response options be ‘floating titles’ or at least be repeated on every page so that respondents do not lose track.

6.3.1.3 Modifying Items

There were several phrases that did not retain their intended meaning when interpreted by participants from the US. The phrase ‘hand condition’ is an ideal example; it was used throughout the early drafts of the HAND-Q to describe the participant’s hand pathology (including both traumatic and non-traumatic issues). Several participants from the US and one from Canada misunderstood the term ‘hand condition’:

> ‘Wouldn’t it be more like “what side is your hand like issue?” … No, because “condition” to us is like what type of condition is it in, is it workable? Is it non-workable?’ (Ref 65)

> ‘The “hand condition” is confusing for us, for me anyway and I think most Americans would be … but I figured it was for what’s bothering you or whatever.’ (Ref 63)

When asked whether replacing the word ‘condition’ with ‘problem’ would work, participants responded favourably.

Examples of other changes that were made to improve clarity to the US cohort were the addition of ‘symptoms’ in the instructions for the Symptoms scale (‘How do your hands feel (symptoms)?’) and ‘faucet’ to the Function scale item (‘Turning a tap (i.e., faucet)’).

Other modifications were made to the wording of instructions and items to enhance participant understanding. Many of the item revisions were minor, such as changes in spelling, punctuation or simplification of wording.

6.3.1.4 Excluding Items

Some items were unclear to multiple participants and therefore were excluded from the field-test scales. Examples are given below:

- From the Satisfaction with Information scale, the item 'What other patients like you experience after hand surgery?' caused confusion with many participants:

> 'Um, what other patients with my hand condition experienced after surgery? What do you mean by that?' (Ref 50)

- From the Symptoms scale, in the item 'Hands feeling tender?', participants were unsure about the term 'tender':

> 'As far as tender, I'm not sure what you meant by tender?' (Ref 51)

6.3.1.5 Confirmation of Concepts

Many of the concepts explored by the HAND-Q are not explored by other hand relevant PROMs. To confirm these issues were considered important to patients, the cognitive interviewer probed for feedback. There were also specific items that were confirmed whether they were acceptable to patients. Representative quotations are given to support each scale as follows.

6.3.1.5.1 Appearance

> 'It's nice to actually read that and realise I am not really vain or dramatic, because a lot of people would just brush things off … it is nice to actually look at it and realise that things I

think about, to see it written down, obviously someone else has thought about it as well.' (Ref 27)

'Well I've never liked the look of my hands, I've got farmers hands from my father.' (Ref 39)

6.3.1.5.2 Function

The following quotations were received specifically regarding the item 'Cleaning (e.g., wiping) yourself after a bowel movement?':

'That doesn't offend me at all. I mean we all do go to the toilet, don't we?' (Ref 6)

'It is one of those everyday parts of life and totally appropriate to ask that.' (Ref 27)

6.3.1.5.3 Symptoms

The following quotations pertain specifically to whether there was redundancy in the items 'Tingling in your hands' and 'Hands feeling numb (i.e., less feeling)'.

'That part of my hand is numb constantly, while I get tingling in the fingers, and they are definitely different.' (Ref 27)

'Yeah, well tingling I usually get on the tip of my fingers, and along my baby finger, and along my palm there. I get tingling right there. Numbness is sometimes I feel like I go to grab something, and it's almost like I've got nothing in my hand even though I've got my hand on it.' (Ref 49)

6.3.1.5.4 Sleep

'Falling asleep not so much but staying asleep and comfort were huge … yeah numbness would affect my sleep, and of course some days I would feel super tired because I was constantly struggling with my hand at night.' (Ref 48)

6.3.1.5.5 Sex

'Let's face it, sex is a part of life. It is an important part and I think these questions are quite relevant.' (Ref 11)

‘Yeah, and it did become a problem with my hands, gripping and being intimate and stuff. And my hands would go numb … I went through it and thought they are all necessary and not offensive.’ (Ref 48)

‘Yeah they’re fine. I don’t think they pry into your life or anything like that, but they are still very good questions to ask for sure … your sex life is something that nobody’s gonna bring up, and it’s a big part of everyone’s life.’ (Ref 27)

6.3.1.5.6 Psychological

‘One thing I have learned, if I can control this, these questions here, it makes a difference to the actual arthritis. If you let this run over you, you’re on the downward go!’ (Ref 40)

‘Now I’m normally a very positive person … my answers are all based on the fact that 99% of the time I talk myself into being positive. It still had a huge impact on me. You know, not being able to use my hands.’ (Ref 24)

6.3.1.5.7 Social

‘I did feel like that people didn’t understand what I was going through. They didn’t have it so they don’t know.’ (Ref 63)

‘My wife has to do so many things for me ... yeah that’s a pretty big loss. I think you’ve done a good job here, able to empathize with most of your patients I reckon.’ (Ref 40)

6.3.1.5.8 Work

‘This is very important. In my particular field for sure, if I can’t get stuff ready to go in the moment, it makes me feel incompetent and its crucial for the job. Very pertinent questionnaire.’ (Ref 50)

6.3.1.5.9 Hand Acceptance

‘Yeah, and some of these things it’s taken me 20 or 30 years to deal with! A long time. Yeah, so, I have learned to live with my problem … yes, exactly, I have learned. I’ll put 4 for that, but 30 years ago I would have put 1 … I think you’ve hit it right on the head, some people even my age would still strongly disagree that they’ve learned to live with them.’ (Ref 40)

6.3.1.5.10 Overall Outcome

‘I’m not back to having a hand that works perfectly, but I’m thrilled with how it works compared to how it was before.’ (Ref 24)

‘… having surgery changed my life for the better. Well, I have to agree with that one, because there are some conditions that were bothering me which are no longer there.’
(Ref 49)

6.3.1.5.11 Quality of Life

‘And I think this is great because it goes from thinking about yourself, to how you interact with other people.’ (Ref 2)

‘They’re good questions though to ask because with Dupuytren’s they most definitely would make a difference. You know, they are good questions to ask of somebody if you’re dealing with Dupuytren’s because they would make a difference, those questions as to how you would feel.’ (Ref 51)

6.3.1.5.12 Anaesthesia

‘A lot of people think “I don’t want to be awake” but I didn’t watch it, and you know, before I turned around it was finished. And not having to go through all the preparations and having a [general] anaesthetic was great.’ (Ref 27)

‘Having my hand numbed was the best thing I ever done. It was awesome. I had no effects from the anaesthesia. I didn’t get vomiting dopey feeling, you know, sleepy. I just left a happy camper. I didn’t have to stay at the surgical unit long. I got out, dressed and left.’ (Ref 63)

6.3.1.5.13 Post-Anaesthesia Symptoms

‘Yes, they’re quite clear. I don’t know why anybody would not understand them.’ (Ref 59)

6.3.1.5.14 Awake Procedure

‘I did find when you go to the dentist with a toothache and they give you that injection, it does sort of work, and then when they give you a second and third jab you don’t really feel them. In this one though, I felt every jab, and there was like 12 of them.’ (Ref 2)

‘How it felt while it was taking place, um, I was a little anxious during that, because I could hear some plucking sounds and feeling pressure, so it made me a bit uncomfortable.’ (Ref 48)

6.3.1.5.15 Information

‘I mean, they let me know what I should expect, and how long it takes me to heal up, but you know, the first hand I had done, I had an issue with that. He thinks I had a reaction to the material used in my surgery, and uh, I went into the ER and he did do some work, he opened it up and I went on an antibiotic just in case it was an infection, and he didn't think it was, and he explained all that to me real good!’ (Ref 64)

6.3.1.5.16 Surgeon

‘Yeah for some you just feel like a transaction and they discard your input, and I think that's the biggest point to be addressed. Their level of knowing what's going on and getting what needs to be done. Not just being treated as a...as a person on the list. That's the hardest part I found with a lot of doctors. The interaction.’ (Ref 50)

‘Well the first thing he made me very comfortable. Very professional. Like I said, he made me feel comfortable. He was really a perfect gentleman I can't say enough about him.’ (Ref 58)

6.3.1.5.17 Hand Therapist

‘But the most important thing is that they listen to what you say. Now I’ll say “this has happened to this hand since I last saw you”, and they actually listen to what you have to say, and feel your hand, and do the appropriate treatment. Because I’ve been going for a full 14 years, they don’t need to explain what each treatment is for. But they’re great at listening.’ (Ref 27)

‘I think the staff were brilliant and all that was great, it did a good job.’ (Ref 2)

6.3.1.5.18 Office Staff

‘The staff were wonderful and very well organized.’ (Ref 58)

‘I don’t really think there’s anything there that should cause anybody not to understand.’ (Ref 59)

6.3.1.5.19 Hand Clinic

‘There are some people that expect everything to be done right away, right on time, and in our health system the way it is right now, that’s not possible.’ (Ref 59)

‘Like, you know, the clinic’s just a disaster which is why, you know, the foundation that I work for, we’re actually campaigning to do the renovation of that spot for the next two years because it’s awful.’ (Ref 61)

6.3.1.5.20 Splint

‘I looked like Edward Scissorhands.’ (Ref 6)

‘I think that they were all good, but some of them were a bit uncomfortable. They looked fantastic because they had the colour on them I liked.’ (Ref 39)

‘I had to take the splint off my right hand to get dressed, to put a shirt or anything on, because I couldn’t get my arm in through the sleeve.’ (Ref 51)

6.3.1.6 HAND-Q Overall

‘Obviously you are wanting to understand how people feel about all this and those questions are really going to nail it.’ (Ref 61)

‘They were all very easy to answer; very easy. No, they were well worded and easy to answer. There was nothing I had to think about on those at all.’ (Ref 51)

‘I think it is quite interesting because it actually made me stop and think about it myself … to a bit of a more in-depth degree that I probably had, so I think that is probably even good for me.’ (Ref 26)

‘It is going to be hard to decide which questions you want to, how you are going to sort of filter it because they are all pretty relevant.’ (Ref 6)

‘I found it very easy to understand, and you know, the ones regarding sexual intimacy were respectful, not too graphic or anything. I don’t think you’d insult anyone. I was surprised at first, but I think that makes it more thorough and professional.’ (Ref 48)

> ‘My husband was also sitting next to much while I was doing it, which was actually very valuable, because I don’t think he realised quite how much of an impact this has had on my life. Um, yea, so that was actually huge …’ (Ref 24)

> ‘There could be a lot more help out there, and I think you’re doing your best to capture that.’ (Ref 24)

> ‘I think good on ya for what you’re doing, and if it's going to make people’s lives easier with answers and things, all the better. But yeah, good on you for your perseverance with this stuff.’ (Ref 8)

6.3.1.7 Problematic Content

6.3.1.7.1 Length of Scales

Several participants expressed concern about the length of the HAND-Q and doubts as to whether participants would complete the scales when left to do so independently. This issue will be addressed by shortening the scales significantly during the process of item reduction at the completion of the field test.

6.3.1.7.2 Function Scale

Items relating to hand function are divided into single-handed activities (e.g., using a remote control) and activities that include both hands (e.g., food preparation). An effort was made when constructing the items to focus on tasks that are mainly hand dependent, as opposed to activities that require the entire upper limb (e.g., washing your hair) or even the whole body (e.g., gardening or yard work). The other consideration was whether activities are performed by an individual’s dominant hand (e.g., writing) versus a task that can be completed by either the dominant or the non-dominant hand (e.g., picking up a coin). The measurement of hand function is therefore unavoidably complex, and it was important to include patients with both dominant and non-dominant hand conditions as well as those with bilateral conditions to ensure that the scale was applicable across varied clinical scenarios.

6.3.1.7.3 Work Scale

The underlying concept of a Work scale was directed by those people who were still taking part in the paid workforce rather than those who were unpaid, volunteers or retired. There were differing views on whether focusing on paid work was appropriate. One expert was concerned that it would be potentially discriminatory to not include homemakers and volunteers in the work scale. A participant who was a stay-at-home mother said:

> 'If you consider me being a mom, I do housework … So, I'm a single mom, I do everything! I have a home, I live at home with two boys, I could answer in that way.'
> (Ref 66)

There were concerns raised by one participant who was a farmer:

> 'Paid employment is an interesting concept when you are self-employed.' (Ref 26)

Other participants felt strongly that the concepts measured by the Work scale should be for those who are working in paid roles. A participant who worked as a theatre nurse said:

> 'Yes you have to … [keep scales for only paid work] This is very important. In my particular field for sure, if I can't get stuff ready to go in the moment, it makes me feel incompetent and its crucial for the job. Very pertinent questionnaire.' (Ref 50)

A retired participant agreed that the Work scale did not apply to his situation:

> 'Plenty of work at home, but that's different.' (Ref 39)

A retired participant who also acted as a carer for a relative responded with:

> 'I think that's a different thought process when you're in a job where you're working to live and your hands affecting that.' (Ref 24)

It was decided that the Work scale would be only for those in the paid workforce as the same issues regarding loss of income, loss of employment and financial security do not apply to those who are retired, carers for family members or homemakers. Although the tasks people in these

roles perform could certainly be considered ‘work’, it was felt that changes in their QOL could be measured with the other HAND-Q scales. The terminology of the stem was revised to include only those who work in a paid position.

6.3.1.7.4 Anaesthesia Scale

The concept behind the Anaesthesia scale was to capture the patient experience of their hand surgery anaesthesia. This was a difficult concept for which to create pertinent items as all items had to be relevant to all participants regardless of whether they had experienced a local anaesthetic, sedation or a GA. During the cognitive interviews, it was clear that some participants did not comprehend the relevance of some of the items to their experience. Some participants that had local anaesthesia did not consider that they had any ‘anaesthesia’ and therefore did not realise that this scale was applicable to them. Instructions were clarified to aid participants’ interpretation of whether the scale applied to them.

6.3.2 Professional Input

Overall 25 professionals provided input in the development of the HAND-Q. Round one had 14 responses, and round two collected a further 11 responses. Professionals were from a variety of fields and geographical locations; the details are displayed in Table 6.4. The sample included 12 hand surgeons and 10 hand therapists. In total, 23 of the 25 professionals reported that patients with hand conditions were the focus of their practice.

Table 6.4: Profession and nationality of contributing professionals

Profession	**Nationality**							**Total**
	Australian	Canadian	British	Dutch	American	Indian	French	
Plastic surgeon	1	0	1	2	2	1	0	7
Orthopaedic surgeon	1	0	1	0	1	0	2	5
Hand therapist	5	2	2	1	0	0	0	10
Researcher	0	0	0	1	0	0	0	1
Physiotherapist	1	0	0	1	0	0	0	2
Total	8	2	4	5	3	1	2	25

Professional feedback was helpful to refine some items and add some new items. An example of an item that was modified as a result of professional feedback was an item on the Function scale, ‘Wiping after you go to the toilet?’. A plastic surgeon from the United Kingdom, commented, ‘Wiping after going to the toilet may have ethnic and cultural variation with washing used in some parts of the world, and in ethnic minority communities in the West’. As a result, the item was changed to be more broadly culturally appropriate. The final item was ‘Cleaning yourself (e.g., wiping) after a bowel movement?’. This terminology was acceptable to participants (see Section 6.3.1.5.2).

Another helpful insight provided by a hand therapist from the Netherlands relates to the Overall Outcome scale, where initially the items asked how the respondents feel about their most recent ‘hand surgery’. It was suggested that the wording be changed to ‘hand treatment’, which would broaden the application of the scale to include non-surgical treatments. Any changes resulting

from professional input were discussed with participants at later cognitive interviews to ensure that the changes were acceptable to patients.

Much of the feedback from the professionals consulted related to the overall length of the field-test HAND-Q and that in some scales there was redundancy in some items. These issues will be addressed in the item reduction phase during Rasch analysis at the end of the international field test (Chapter 7).

A conceptual issue that was brought up by the professionals was whether the scales should ask the respondent to answer the Wellbeing scales in respect to the hand that bothered them the most or the hand that was most recently treated. A hand therapist from the Netherlands, who was concerned about the instructions for the Function scale, stated, 'Occupational therapists use devices in order to improve a patient's hand function in daily life; this scale cannot measure that effect if this instruction is used. I would prefer to use the function with gadgets/devices included'. It was decided to keep the scales as they are for the field test with the future option of modifying the instructions to suit varied applications.

6.3.3 Summary of Changes

Over the course of the three rounds of cognitive interviews and two rounds of feedback from professionals, there were many changes made to the items of the HAND-Q. These changes were documented in an item tracking matrix. A summary of the changes per scale are listed in Table 6.5.

Table 6.5: HAND-Q scales

	Response options	Recall period	Initial items	Items added	Items revised	Items dropped	Items for field test
Wellbeing scales							
1. Appearance	Satisfaction	Now	29	1	10	0	30
2. Function	Difficulty	Past week	34	3	14	2	35
3. Symptoms	Severity	Past week	18	6	17	2	22
4. Psychological	Frequency	Past week	16	3	0	0	19
5. QOL	Severity	Past week	9	2	1	0	11
6. Sleep	Frequency	Past week	8	1	3	1	8
7. Social	Agree/Disagree	Past week	13	0	4	0	13
8. Sexual	Bothered	None	9	0	0	0	9
9. Work	Agree/Disagree	None	9	2	3	0	11
10. Acceptance	Agree/Disagree	None	7	0	6	0	7
Satisfaction scales							
11. Anaesthesia	Bothered	None	17	0	5	3	14
12. Post-anaesthesia symptoms	Severity	None	12	2	0	1	13
13. Awake procedure	Satisfaction	None	17	1	8	1	17
14. Information	Satisfaction	None	21	1	7	2	20
15. Surgeon	Agree/Disagree	Recent appointments	25	1	8	1	25
16. Hand therapist	Agree/Disagree	Recent appointments	20	1	4	2	19
17. Hand clinic	Agree/Disagree	Recent appointments	14	0	3	1	13

18. Overall outcome	Agree/Disagree	Most recent treatment	10	0	13	1	9
19. Office staff	Agree/Disagree	Recent appointments	13	1	2	0	14
20. Splint	Satisfaction	Most recent splint	11	2	1	1	12
Total				28	109	15	319

Numerical values refer to number of items.

6.4 Discussion

The domains elicited in Chapter 5 were developed into a set of 20 independently functioning HAND-Q scales, each measuring a separate concept. The scales are divided into those that measure the wellbeing of the participant, and those that measure their satisfaction with their experience. These scales were refined through cognitive interviews and feedback from professionals within the field. The foundations of the HAND-Q scales have strong content validity as they are based on in depth qualitative studies performed with a sample of the target population, with further confirmation of validity by professionals in the field.

The HAND-Q scales have been developed in keeping with the methods used for other Q-PROMs such as the BREAST-Q, CLEFT-Q and BODY-Q (Pusic et al. 2009; Klassen et al. 2016; Wong Riff et al. 2017). The conceptual framework from the qualitative study (Chapter 5) had similarities with the frameworks of these instruments. The concept of a wellbeing domain and satisfaction domain is similar between the HAND-Q and the BREAST-Q (Pusic et al. 2009). Within the wellbeing domain, there are scales that measure hand appearance, function, symptoms, work, psychological impact, sleep, sex, social impact, QOL and acceptance of hand condition. Within the satisfaction domain, scales measure the patient's satisfaction with aspects of their anaesthesia (anaesthesia, awake procedure and post-anaesthesia symptoms), their care givers (surgeons, hand therapist, hand clinic and associated administrative staff), their splint and their overall treatment outcome.

Many of the concepts measured by these scales are not included or only superficially covered in existing instruments. The process of cognitive interviewing was used to confirm the relevance of these concepts to the target population. An example is the psychological impact of having a hand condition. It is recognised that injuries and conditions affecting the hand and

upper limb can have profound effects on psychological wellbeing. Bailey et al. (2009) studied patients with upper extremity nerve damage and found that 39% had signs of clinical depression. In currently available PROMs, such as the DASH (Hudak et al. 1996), PRWHE (MacDermid 1996) and MHQ (K. Chung et al. 1998), there are single items enquiring about psychological wellbeing. These items are mixed with other concepts to produce an overall score, which compromises the measurement of any psychological impact as it is combined with other questions on function or symptoms. In the literature, there already exists multiple scales that measure depression and anxiety, such as the Depression Anxiety Stress Scales (Crawford and Henry 2003; Pilkonis et al. 2011). However, in the qualitative interviews, it was evident that the psychological impact of a hand condition is overwhelmingly one of frustration and irritation, which are unique to depression and anxiety. In this study, the psychological impact scale was verified as relevant and important with both participants and professionals in the field.

The scale on sexual QOL was developed from only a small set of qualitative data (having only been raised by one participant), but there is evidence in the literature that sexual QOL is a problem in those with musculoskeletal problems. Shauver and Chung (2010), and Shauver, Aravind and Chung (2011), described an unexpected topic of difficulty with sexual activity in their qualitative study with patients recovering following severe tibial fracture. In addition, a qualitative study that explored sexuality, emotions and relationships of those with a traumatic brachial plexus injury found that this population reports detrimental effects on their sexual activity (Wellington 2010). The effects of juvenile idiopathic arthritis (JIA) on sexual QOL were reported by Hill, Herstein and Walters (1976); they found that those with severe disease had limitations on their sexual activity due to pain, position or fatigue. J. Packham and Hall (2002) also reported on sexual activity in this population (adults with JIA), finding that patients who were not sexually active had higher levels of physical disability, as measured with the Health Assessment Questionnaire (Kirwan and Reeback 1986), and poor body image. Of those

that were not sexually active, 30% felt this was a result of their JIA, with poor body image being a more common issue (66.6%) than physical disability (8.3%). The psychological effects of hand difference are acknowledged in the paediatric population (Gupta, Kay and Scheker 2000), and it is logical that these effects would persist in adulthood and may influence patient's sexual QOL.

In existing PROMs, the DASH has a single item asking about 'Sexual activities'; this item was identified as being frequently unanswered in Chapter 4. Having an item that is so different from other items on the scale is not ideal from multiple measurement principles. The concept of having a scale dedicated to the concept means that if participants are not comfortable to answer the questions then they will skip the whole scale, rather than missing an item. Missing items result in reduced reliability for that scale, as is the case with the question regarding sexual activity in the DASH. This scale was tentatively discussed at the cognitive interviews to discover what participants thought of it. Overwhelmingly participants were positive about the scale; none of the participants reported offense by any of the items. Some participants did express concern that potentially the older demographic may find the scale inappropriate; however, the average age of those interviewed was 60 years and the range of 32–76 years was inclusive of older individuals. Many participants responded that the scale did indeed resonate with them, that they had noticed their hand condition affecting their sexual QOL. While not anticipated to be widely implemented in all hand clinics, the sexual QOL scale does seem to be measuring a concept that is important to people with a hand condition. One participant responded to the Sexual QOL scale:

> 'I think you did a really good job. I was sitting outside with my husband, and said, "oh this one's interesting" and I think there were a few moments of daylight for him!' (Ref 24)

The Anaesthesia scale, Post-Anaesthesia Symptoms scale and Awake Procedure scales are all unique to the HAND-Q. These concepts were theoretically thought to be important in the field

of hand surgery, where many procedures can and often are performed on awake patients with the assistance of local anaesthesia. There are benefits to this technique, detailed in Chapter 1, but there did not exist a meaningful way to measure the patient experience and therefore establish patient preference. These concepts were validated with participants during the qualitative interviews, and thus scales were developed to measure these concepts. Cognitive interviews revealed some confusion about who the anaesthesia scale is targeting, and this led to revision of the instructions for this scale.

Using cognitive debriefing interviews to confirm the content and structure of the preliminary HAND-Q scales resulted in many minor and some more substantial changes. Patients from three different English-speaking countries have confirmed that they understand the content and relate to the concepts that are being measured.

6.4.1 Potential Weaknesses of Study

This study has some potential weaknesses. The interviews were all performed by a single interviewer who is a developer of the HAND-Q. The interviewer had limited training in conducting cognitive interviews. The potential for bias due to the medical background of the interviewer is discussed in Chapter 5. Due to the length of the instrument, it was not possible to perform an in-depth review and discussion on every item or every scale. As a result, some scales were not discussed with every participant to ensure that the scales found later in the booklet were also discussed. Efforts to ensure that the included sample population was varied in terms of age, gender and condition were made. However, given the broad heterogeneity of hand conditions, it was impossible to include participants with every hand condition. The readability of the scales has not been quantitatively assessed using the Flesch–Kincaid statistic.

6.5 Summary

The content validity of the HAND-Q scales has been established with patient cohorts from Australia, Canada and the US. Relevant professionals from around the world and from varied professions have given feedback on the scales. The HAND-Q scales have been finalised for the international field test (see Appendix D.3).

Chapter 7: Developing the HAND-Q Phase II: International Field Test

7.1 Introduction

Field testing involves collecting scale responses from a broad range of participants from the target population. To ensure that the HAND-Q is optimised for use in multiple countries of application, the field test includes international sites with a variety of cultures, languages and economic statuses. This approach has been previously used for the development of the CLEFT-Q which included thirty hospitals in twelve countries (Klassen et al. 2018). Field testing in multiple languages requires translation and cultural adaptation of the draft HAND-Q, this process will be carried out following the guidelines of the International Society for Pharmacoeconomics and Outcomes Research (Wild et al. 2005). The aim of this study is to collect responses to the HAND-Q from field test sites in Europe, Asia pacific and America which will provide an international data set on which the final Rasch analysis can be carried out and the HAND-Q scales produced. This study is currently underway and the final results will be published separately and not included in this thesis.

7.2 Methodology

7.2.1 Site Recruitment

To acknowledge the importance of including multiple cultures, languages and countries with varied economic statuses in the HAND-Q field test, international collaboration was vital. Researchers sought out collaborators in the Indian subcontinent and Europe by emailing professional contacts. Interested collaborators received written information and the opportunity to discuss the project further via telephone or teleconference. Prerequisites for any field test site

were to have a hand clinic that could provide a minimum of 200 participants over a timeframe of approximately 6 months. Field-test collaborators were responsible for attaining local ethics approval with the support of the HAND-Q team. Translation and cultural adaptation were the responsibility of the collaborating site, with oversight by Dr Sierakowski at Flinders University.

7.2.2 Participant Recruitment

A member of the healthcare team at each site will perform recruitment. Responses to the HAND-Q will be collected using paper booklets or electronically using a REDCap survey (P. Harris et al. 2009). Participants' language and cognitive abilities will be satisfied by their ability to read the questionnaire in the language that it is presented to them; no external judgement of literacy or cognitive ability will be made.

7.2.3 Participating Sites

Recruitment for the HAND-Q field test will be undertaken at the centres listed in Table 7.1 and represented on a map in Figure 7.1. The goal is to recruit a minimum of 200 participants from each country, as this sample size will allow for a minimum of 50 participants in four class intervals to allow for DIF by country or field-test site. This sample will also result in item calibrations that are stable within 0.5 logits (person location estimates) with a 99% confidence interval (Linacre 1994).

Table 7.1: International HAND-Q field test

Country	Hospital	Hand clinic specialty	Language
Australia	Southern Adelaide Local Health Network	PRS	English
	Royal Adelaide Hospital	Orthopaedic	English
Canada	New Brunswick	PRS	English
	Hamilton	PRS	English
United Kingdom	Oxford University	PRS	English
US	Dartmouth-Hitchcock, New Hampshire	PRS	English
	SUNY Downstate, Brooklyn, New York	Orthopaedic	English
	Michigan Medicine, Ann Arbor, Michigan	PRS	English
India	Ganga Hospital, Coimbatore	PRS	Tamil
	Christian Medical College Vellore, Vellore	Orthopaedic	Bengali, Hindi
Pakistan	Rawalpindi Medical University, Rawalpindi	Orthopaedic	Urdu
Netherlands	The Hand Clinic, Amsterdam	PRS	Dutch
France	Hand Surgery Centre, Caen	Orthopaedic	French
	Nimes University Hospital, Nimes	Orthopaedic	French
Finland^	Helsinki University Hospital, Helsinki	Hand	Finnish

PRS, plastic and reconstructive surgery.

^Finland has hand surgery as a separate surgical speciality.

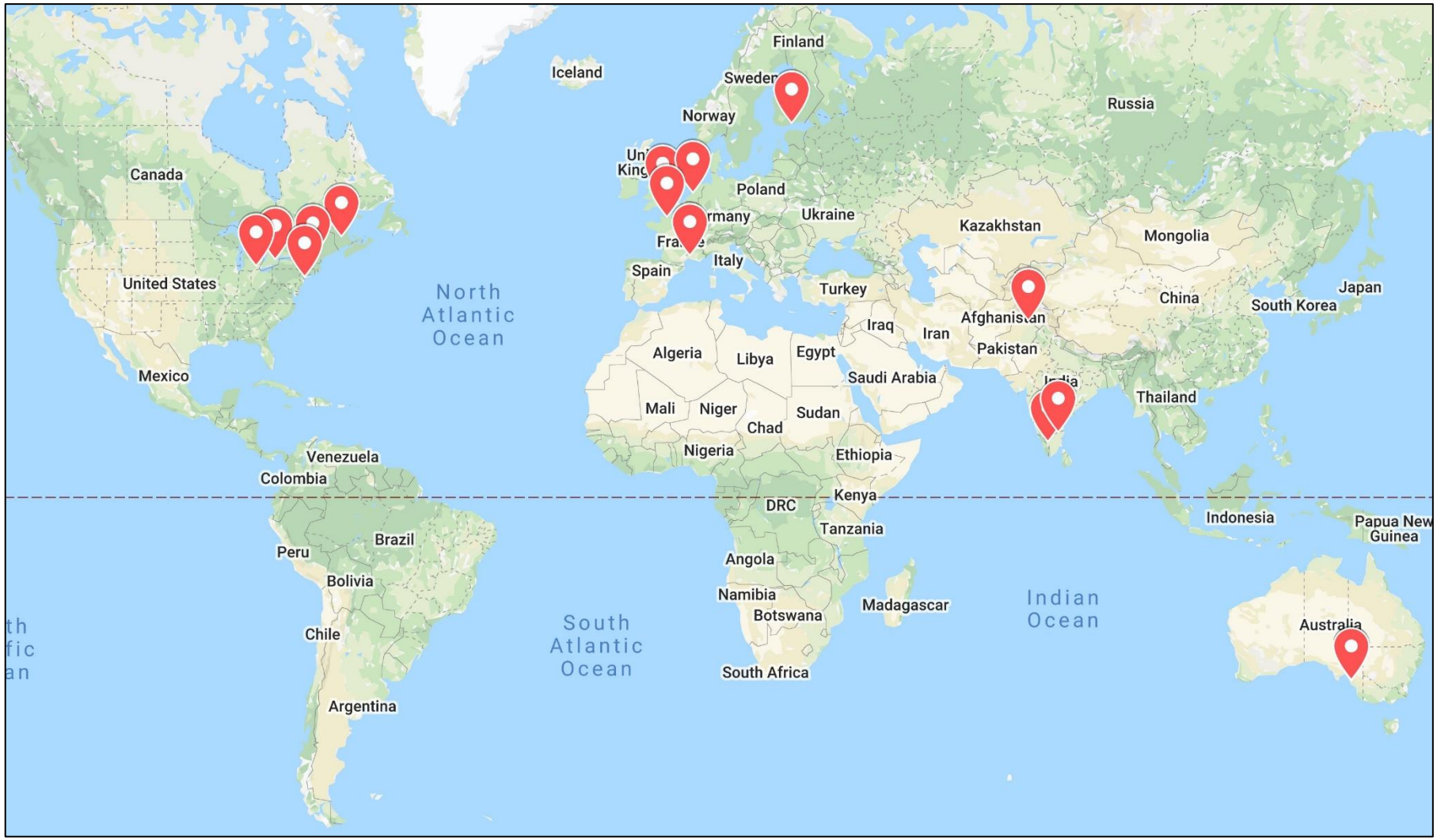

Figure 7.1: World map showing distribution of HAND-Q field-test sites.

7.2.4 Data Collection

Data will be collected with either paper booklets or electronic survey REDCap platform (P. Harris et al. 2009). Figure 7.2 shows a screen shot of the REDCap HAND-Q field-test questionnaire. Data entry from paper-based responses to the REDCap database will be performed by trained members of the research team.

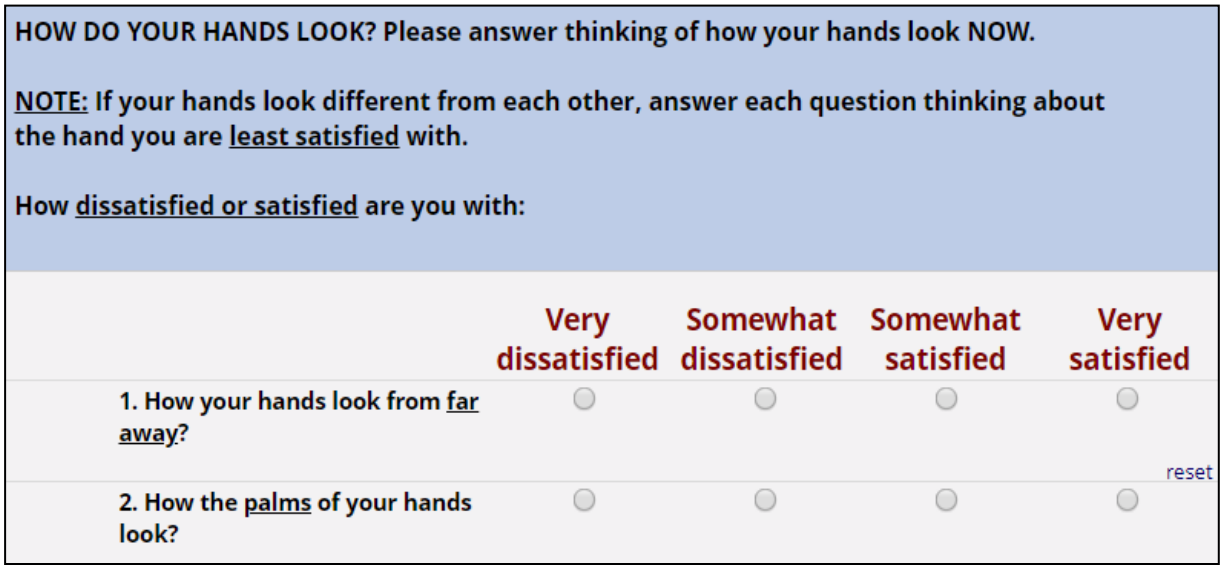

Figure 7.2: Screen shot of the REDCap HAND-Q field-test questionnaire.

7.2.5 Data Analyses

7.2.5.1 Rasch Analysis

RMT (Rasch 1960) will be used to refine the items that make up each scale. This analysis will be conducted with RUMM2030 software (RUMM Laboratory, Perth, Australia) (Andrich and Sheridan 1997–2011). This analysis involves making judgments about the suitability of an item in a scale on the basis of the evidence presented by several statistical and graphic tests. Statistical details from each item are combined to consider the scale as a whole. When data are found to fit the Rasch model, RMT is supported by the data and therefore the estimations from the model are considered accurate. The following tests will be performed as part of the analysis; further details are available in Chapter 4:

1. Thresholds for item response options will be examined for each scale to ensure that response options have appropriately ordered thresholds. If any scales display disordered thresholds, the scores will be recoded.
2. Item fit statistics will be examined, first, using the item–person interaction (log residuals, ideally between −2.5 and +2.5); second, using the item–trait interaction (chi-square values, ideally non-significant after Bonferroni adjustment); and last, using the item characteristic curves. This statistical information will be considered along with the clinical usefulness of the item before any decision is made regarding whether the item should be retained.
3. Targeting and item location will be examined using the item map. How well the items cover the spectrum of a construct will be established. Redundant items will be removed (items measuring the same amount of the underlying construct).
4. PSI will be calculated.
5. Dependency will be established by calculating residual correlations between items. Correlations greater than 0.3 will be identified and subjected to further subtest analysis to ensure the effect on the reliability of the scale is minimal.
6. DIF will be examined to show if items function differently for participants with different hand conditions, or from varied countries or languages.

Further analysis will be performed with SPSS. Parametric analyses (independent *t*-tests and one-way ANOVAs) will be used when data distribution approximates a normal distribution, an alpha level of 0.05 will be used. Normality will be assessed by the shape of the data distribution on a histogram and the degree of skewness and kurtosis. Traditional psychometric tests include the following: Cronbach alpha, the proportion of participants with scale-level missing data, and scores at the floor and ceiling. Rasch logit scores will be transformed into scores ranging from

0 (worse) to 100 (best). These scores will be used in the following hypothesis tests of construct validity:

1. Correlation between scales to show the extent that each scale measures a separate but related construct. *It is hypothesised that these intercorrelations will range between r = 0.30 and r =0.70 as these scales were developed to measure distinct concepts but with clear clinical relations*. Hand Appearance is expected to have lower correlations than the other scales as it is a more separate concept and not necessarily related to Symptoms and Function
2. Correlation between scales and patient characteristics of gender and age. This is to determine the extent that a scale may be vulnerable to bias based on these variables. *It is predicted that these correlations will be low (<0.30)*
3. Correlation between appearance scores of patients with hand conditions that have a pronounced effect on appearance (Dupuytren's contracture and rheumatoid arthritis) and those where the aesthetic impact is minimal (CTS). *It is hypothesised that the former group will have lower appearance scores than the latter group*
4. The relationship between patients who describe their hand condition as 'severe', and those who describe their condition as 'moderate' or 'mild'. This will be established with one-way between-groups ANOVA. *It is hypothesised that the former group will have lower scores on the QOL, Function and Symptoms scales than the other groups*
5. Comparison of mean scores of those who believe that they require further surgery for their hand condition with those of patients who do not believe they require further surgery. This will be performed using an independent samples *t*-test. *It is anticipated that the group reporting that they require further surgery will have lower scores than those who report not needing further surgery*.

7.2.6 Translation and Cultural Adaption

Cultural and linguistic validation of the HAND-Q field test is currently underway in the languages listed in Table 7.1. The translation process being followed is the best practice guidelines of the International Society for Pharmacoeconomics and Outcomes Research (Wild et al. 2005) and the Mapi Research Trust (Acquadro et al. 2012). In short, these guidelines recommend two independent forward translations performed by a native speaker of the target language. The two versions are then combined into a single version, which is then back-translated into English to ensure that no meaning has been lost in the process. The resultant translated HAND-Qs will then be discussed with patients in the target country to ensure the accuracy of the translation. This process is currently underway to allow for recruitment to commence in 2019.

7.3 Results

This study is currently in progress. Sites are at various stages of navigating translation, ethics and recruitment, as summarised in Table 7.2. Field-test sites that initially showed interest in taking part but withdrew from the study before commencement are listed in Table 7.3.

It is anticipated that data collection will continue until September 2019. After this time, the responses will be analysed as detailed in the Methods section. This analysis will produce the finished HAND-Q scales, which will then undergo further psychometric testing in Phase III (detailed in Chapter 9).

Table 7.2: Status of current international field-test sites

Country	Hospital	Status as of December 2018	Participants recruited
Australia	Southern Adelaide Local Health Network Royal Adelaide Hospital	On hold	449
Canada	New Brunswick, Canada	Active	42
	Hamilton	Ethics	—
United Kingdom	Oxford University	Ready to commence Jan 2019	—
US	Dartmouth-Hitchcock, New Hampshire	Ready to commence Jan 2019	—
	SUNY Downstate, Brooklyn, New York	Active	88
	Michigan Medicine, Ann Arbor, Michigan	Active	15
India	Ganga Hospital, Coimbatore	Translation	—
	Christian Medical College Vellore, Vellore	Translation	—
Pakistan	Rawalpindi Medical University, Rawalpindi	Translation	—
Netherlands	The Hand Clinic, Amsterdam	Translation	—
France	Hand Surgery Centre, Caen	Translation	—
	Nimes University Hospital, Nimes	Translation	—
Finland	Helsinki University Hospital, Helsinki	Translation	—
Total			594

Table 7.3: Withdrawn field-test sites

Country	Hospital	Reason for withdrawal
US	Mayo Clinic, Rochester	New medical records system with other PROMs incorporated, too burdensome to add HAND-Q field test
	UT Southwestern Medical Centre, Texas	Support staff not available to conduct the field test
	UW Medicine, Harborview Medical Centre, Seattle	Contact lost, reason for withdrawal unknown
India	Medanta Hospital, Gurgaon	Logistical issues, concerns about cultural appropriateness of some of the HAND-Q content
Netherlands	Xpert Clinic, various locations	Potential conflict of interest as staff already involved with the ICHOM standard set for hand conditions
	Eramus University Medical Center, Rotterdam	

7.4 Discussion

Creating a PROM with a large international sample of participants who speak multiple languages is unique in the field of hand conditions. Because of this, the HAND-Q has the potential to be the most globally applicable and robust PROM available for the clinical care and research of those with hand conditions.

This method of PROM development has been previously implemented with the other Q-PROMs; perhaps the best example of this is the CLEFT-Q (Klassen et al. 2018). For the development of the CLEFT-Q, there were 12 countries that took part in the field test, with 30 different sites. In total, the CLEFT-Q field test recruited 2434 children or young adults with a cleft lip and/or palate. The resulting CLEFT-Q scales were examined for DIF according to age, gender and language to ensure that the scales performed consistently between groups. The findings of this analysis supported the use of the scales with the same scoring algorithm.

CLEFT-Q scales have subsequently been incorporated into the ICHOM standard set for cleft lip and palate (Arora and Haj 2016).

7.5 Summary

This study is currently in progress. The results of this study will be published in a separate article following the completion of the study.

Chapter 8: Developing the HAND-Q Phase II: Preliminary Rasch Analysis

8.1 Introduction

Previous chapters described the qualitative study and the process of refining the HAND-Q scales with cognitive interviews and feedback from professionals. The next step of development of a PROM is field testing. This requires a large number of patients from the target population (in this case, people suffering from a hand condition) completing the field-test version of the HAND-Q. The responses provided by participants provide information on the appropriateness of each of the items and how the scales function overall. The international field test is currently underway as described in Chapter 7. This chapter describes the preliminary analysis conducted on data collected prior to 1 October 2018. The analysis was performed to refine the Wellbeing scales that had a minimum of 150 participants responses after exclusion of outliers.

8.2 Methodology

8.2.1 Participant Recruitment

8.2.1.1 Australia

The study was granted ethics approval by the Southern Adelaide Clinical Human Research Ethics Committee (HREC/13/SAC/223) (Appendix E.1). Participants were recruited in person at South Adelaide Local Health Network (Flinders Medical Centre and Noarlunga Hospital) and the Royal Adelaide Hospital from both the Plastic and Reconstructive Surgery Hand Clinic and the Orthopaedic Hand Clinic. Eligible participants were patients aged over the age of 14 who were being treated for a hand condition at the clinic. Excluded from the study were those

whose English language skills made participation difficult and those aged under 14 years of age. Participants gave their written consent before completing the HAND-Q questionnaire.

A postal pack including the HAND-Q field-test questionnaire was sent to 56 patients who had been scheduled for surgery on their hand within South Adelaide Local Health Network. Included in the package was a letter of invitation to participate in the study, information about the study, reply paid envelope and the HAND-Q field-test questionnaire. Returning the questionnaire was considered implied consent to take part. Twenty-two patients returned a completed HAND-Q survey, giving a postal response rate of 39.2%.

8.2.1.2 Canada

The study was approved by the Horizon Health Network Research Ethics Board (2017-2574) (Appendix E.2). Participants were recruited in person at the offices of Dr Donald Lalonde. Eligible participants were any patients over the age of 18, able to read English and who were being treated for a hand condition at the clinic. Exclusion criteria were cognitive disability or language difficulty that prevented participation.

8.2.2 Data Collection

Data were collected with either paper booklets or electronic survey using the REDCap platform (P. Harris et al. 2009). Data entry from paper-based responses to the REDCap database was performed by trained members of the research team (KS, KES, KB, NF, ST and SA, see Acknowledgements section for further detail).

8.2.3 Data Analyses

The analyses are described in Chapter 7, Section 7.2.5.

8.3 Results

8.3.1 Demographics

Participant characteristics are shown in Table 8.1. In total, 491 patients were recruited in person and by post. Plastic surgery participants composed 87% of the sample with the remaining 13% participants from Orthopaedic surgery clinics.

Table 8.1: Participant demographics

Characteristics		
Age in years, mean (SD), range		48 (19), 15–86
Gender, n (%)		
Female		225 (46)
Male		248 (51)
Missing		18
Nationality n (%)		
Australia		445 (90.6)
Canada		42 (8.6)
Missing		4 (0.8)
Specialist clinic, n (%)		
Plastic and reconstructive		428 (87.2)
Orthopaedic		50 (10.2)
Missing		13 (2.6)
Education level, n (%)		
Primary		15 (3.1)
High		254 (51.7)
Further		183 (37.3)
Other		21 (4.3)
Missing		18
Method of recruitment, n (%)		
Face to face		469 (95.5)
Post		22 (4.5)
Hand condition, n (%)		
Elective	CTS	86 (17.5)
	Dupuytren’s contracture	34 (6.9)
	Trigger finger	32 (6.5)
	Osteoarthritis	20 (4.1)
	Rheumatoid arthritis	13 (2.6)
Traumatic	Soft tissue injury	125 (25.5)
	Fracture	120 (24.4)
Unknown	Other	47 (9.6)
	Missing	14 (2.8)

8.3.2 Rasch Analysis

8.3.2.1 Item-Level Analysis

8.3.2.1.1 Item Response Option Thresholds and Item Fit Statistics

Analysis based on RMT showed all items on the proposed scales displayed ordered thresholds. Figure 8.1 shows the threshold map for the Function scale. The stem of this scale asks about the difficulty experienced by the respondent when performing a series of everyday tasks (See Appendix D.3 for further detail). The response options for this scale are '*Not at all difficult'* represented by the purple bar (3), '*A little difficult*' represented by the green bar (2), '*Moderately difficult*' represented by the red bar (1) and '*Extremely difficult*' represented by the blue bar (0).

Table 8.2 lists the fit statistics for each item. Log fit residuals were within the recommended criteria of −2.5 to +2.5 for 78 of 84 items (Hobart and Cano 2009). The remaining items had log fit residuals less than ±4. These items were retained as they were acceptable based on other indices. The Bonferroni adjusted chi-square *p*-values for all items were non-significant, showing good item fit. The item characteristic curves and the clinical usefulness of items were considered as part of the decision process to determine whether each item should be retained or excluded.

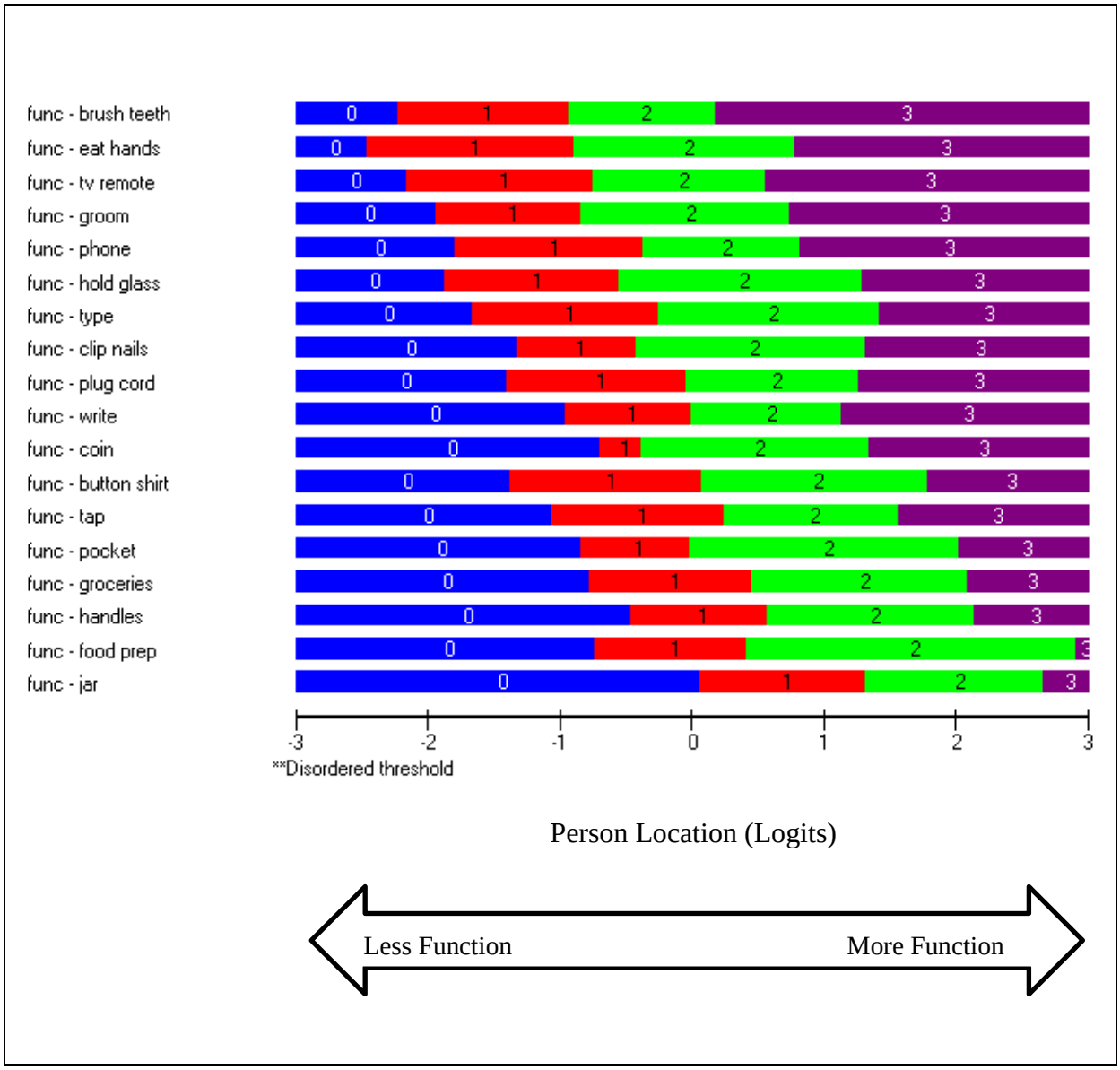

Figure 8.1: Threshold map for HAND-Q Function scale.

8.3.2.1.2 Targeting, Item Locations and Person Separation Index

In Table 8.2, the items are listed by location serial order for each of the scales. For example, in the hand Function scale, the easiest item for participants to respond '*Not at all difficult'* was 'Brushing your teeth?', which had an item location of −.99, and the hardest item for participants to respond '*Not at all difficult'* was 'Opening a jar?', which had an item location of +1.35. Figure 8.2 shows an example of the item map from the Function scale. It graphically displays the items of the scale and their relationship to patient responses.

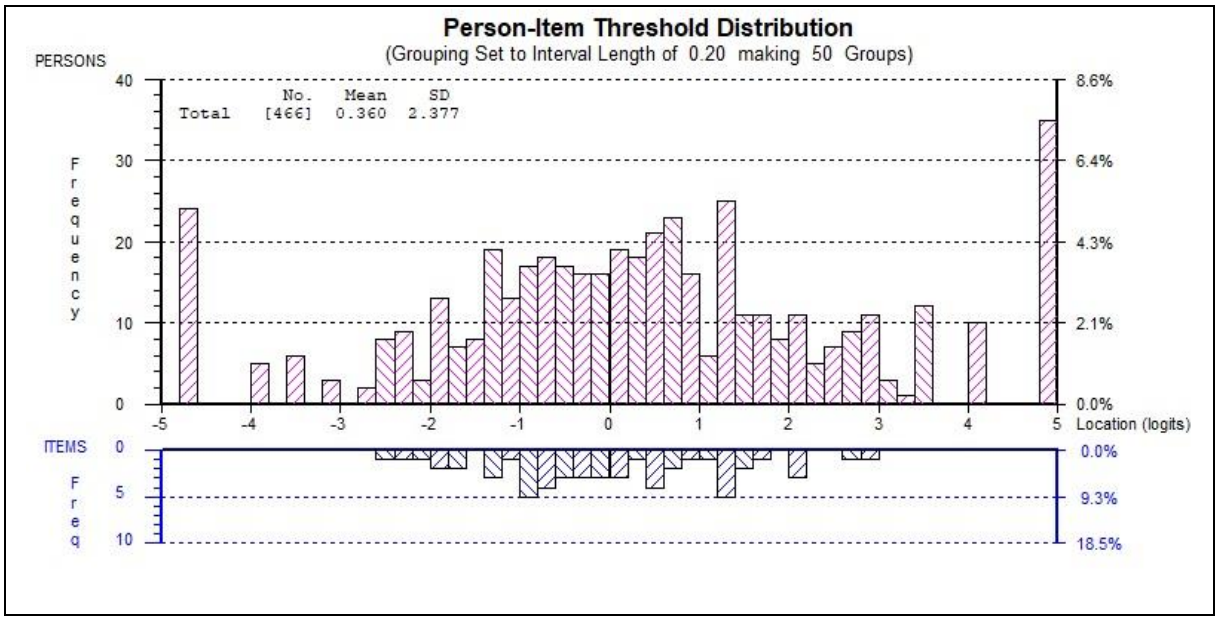

Figure 8.2: Person–item map for HAND-Q Function scale.

8.3.2.1.3 Dependency

There were two pairs of items (each in different scales) identified that had residual correlations >0.3. Subtest analysis on these pairs of items revealed minimal effect on the reliability of the two scales (Appearance and Symptoms; <0.01 difference in PSI).

8.3.2.2 Scale-Level Analysis

Table 8.2 lists the scale reliability statistics. The Rasch coverage varied from 65% to 95%. A non-significant chi-square value was achieved for four of the seven scales. The PSI ranged from 0.84 to 0.94, which shows good reliability.

Table 8.2: RMT statistical indicators of fit

Scale and items^	Item location	SE	Fit residual	Chi-square	DIF	Prob
Appearance						
Wave	−0.45	0.10	−1.61	8.22	3	0.04
Proportioned	−0.39	0.10	0.53	5.17	3	0.16
Shape fingers	−0.25	0.10	−0.09	9.35	3	0.03
Line up	−0.24	0.10	−1.39	2.21	3	0.53
Knuckles	−0.23	0.10	1.44	7.74	3	0.05
Palm on table	0.14	0.10	−2.45	4.36	3	0.23
Hold glass	0.20	0.10	0.37	4.24	3	0.24
Normal	0.24	0.10	0.30	7.24	3	0.06
Straight fingers	0.31	0.10	0.89	4.68	3	0.20
Compared to others	0.66	0.10	−1.87	11.29	3	0.01
Function						
Brush teeth	−0.99	0.08	0.66	3.88	7	0.79
Eat with hands	−0.86	0.08	−2.35	8.89	7	0.26
TV remote	−0.78	0.08	−1.06	6.24	7	0.51
Grooming	−0.68	0.08	0.68	11.36	7	0.12
Phone to ear	−0.45	0.08	0.44	4.00	7	0.78
Hold glass*	−0.38	0.08	−3.64	20.37	7	0.00
Type	−0.16	0.08	1.92	11.39	7	0.12
Clip fingernails	−0.15	0.08	−0.52	9.83	7	0.20
Plug in cord	−0.06	0.07	−0.71	4.86	7	0.68
Write	0.06	0.07	2.08	10.99	7	0.14
Pick up coin*	0.09	0.07	3.35	5.46	7	0.60
Button shirt	0.16	0.08	−2.35	11.91	7	0.10
Turn tap	0.25	0.07	−1.96	18.97	7	0.01
Use pocket	0.39	0.08	−0.29	6.93	7	0.44
Hold groceries	0.59	0.08	1.76	5.06	7	0.65
Hold handles	0.75	0.08	−0.35	15.40	7	0.03
Prepare food	0.86	0.08	−1.82	13.70	7	0.06
Open jar	1.35	0.08	0.99	9.49	7	0.22

Scale and items^	Item location	SE	Fit residual	Chi-square	DIF	Prob
Symptoms						
Pain at rest	−0.79	0.07	−2.49	16.55	6	0.01
Tingling	−0.23	0.07	1.29	6.85	6	0.33
Throbbing	−0.16	0.06	−1.21	8.00	6	0.24
Sleep disturbance	−0.15	0.06	−2.27	8.82	6	0.18
Numbness	−0.09	0.06	1.45	2.64	6	0.85
Clumsiness	−0.07	0.06	0.77	3.16	6	0.79
Swelling*	−0.04	0.07	3.19	15.17	6	0.02
Stiff	0.41	0.07	1.21	6.25	6	0.40
Pain with use	0.50	0.07	−1.57	12.61	6	0.05
Weakness	0.64	0.06	−1.05	6.60	6	0.36
Psych						
Sorry for self	−0.88	0.09	1.01	10.37	7	0.17
Overwhelmed	−0.86	0.08	−1.19	5.20	7	0.64
Depressed	−0.71	0.08	−1.53	9.35	7	0.23
Anxious	−0.69	0.08	0.68	6.39	7	0.49
Stressed	−0.27	0.08	−1.04	13.49	7	0.06
Upset	−0.01	0.08	−2.09	13.39	7	0.06
Irritated	0.47	0.08	−0.50	9.75	7	0.20
Fed-up	0.55	0.07	−0.77	12.24	7	0.09
Annoyed	0.59	0.08	0.20	7.09	7	0.42
Frustrated	1.79	0.09	0.08	7.16	7	0.41
QOL						
Close relationships	−1.08	0.08	0.06	3.63	5	0.60
Social life*	−0.72	0.08	−2.84	11.41	5	0.04
Relax	−0.57	0.08	0.98	2.12	5	0.83
Mood	−0.33	0.08	−1.67	10.15	5	0.07
Bathe	0.15	0.08	2.06	4.13	5	0.53
Independent	0.26	0.07	−1.81	10.19	5	0.07
Physically active	0.92	0.07	0.99	4.11	5	0.53
Activities enjoy	1.38	0.07	0.09	1.43	5	0.92

Scale and items^	Item location	SE	Fit residual	Chi-square	DIF	Prob
Sleep						
Fall asleep	−0.70	0.10	0.91	6.87	4	0.14
Not enough	−0.10	0.10	−1.58	8.85	4	0.07
Disturb	0.13	0.10	2.49	4.38	4	0.36
Woke up*	0.21	0.10	−3.66	18.05	4	0.00
Comfortable	0.46	0.10	0.10	5.66	4	0.23
Sex						
Enjoyment	−0.59	0.11	−1.56	12.45	7	0.09
Distraction	−0.45	0.11	−1.68	9.82	7	0.20
Symptoms interfere	−0.26	0.11	2.46	13.99	6	0.03
Give pleasure	−0.14	0.11	−1.32	7.94	7	0.34
Awareness	−0.09	0.11	−0.47	7.96	7	0.34
Tender*	0.63	0.11	2.70	12.41	7	0.09
Function interfere	0.89	0.11	−1.74	9.40	7	0.23
Splint						
Replace	−0.69	0.12	2.40	3.39	2	0.18
Look	−0.63	0.11	0.08	4.04	2	0.13
Socialise	−0.62	0.11	−1.91	6.33	2	0.04
Enjoy life	−0.09	0.10	−0.93	1.35	2	0.51
Care for hand	0.23	0.11	−0.86	3.18	2	0.20
Comfortable	0.26	0.11	1.18	0.94	2	0.63
Sleep	0.32	0.10	0.07	0.67	2	0.72
Be active	0.51	0.11	1.04	1.80	2	0.41
Dress self	0.72	0.10	0.52	3.85	2	0.15

^ Items are in location order for each scale.

* Indicates item with fit residual ±2.5.

8.3.3 Traditional Psychometric Analysis

Table 8.3 lists the traditional psychometric results for each of the scales. Scale reliability was supported by high Cronbach alpha coefficients (>0.90). The scales demonstrated minimal floor effects, but ceiling effects were as high as 42% for the Sex scale. Missing data were less than 10% for four of the seven scales, and less than 20% for the remaining scales.

Table 8.3: Reliability statistics of HAND-Q scales

Scale	Included sample	Rasch coverage	PSI without extremes	Chi-square	*p*-value	Cronbach α	Floor n, %	Ceiling n, %	Missing data n, %
Appearance	265	95%	0.89	64.49	0.000	0.95	1, 0	85, 32	43, 16
Function	401	86%	0.94	155.02	0.001	0.98	23, 6	34, 8	71, 18
Symptoms	429	93%	0.84	86.66	0.014	0.90	4, 1	24, 6	44, 10
Psychological	416	90%	0.88	94.43	0.028	0.94	4, 1	42, 10	17, 4
QOL	398	88%	0.85	47.17	0.203	0.92	7, 2	46, 12	16, 4
Sleep	322	75%	0.85	43.81	0.002	0.94	20, 6	83, 26	9, 3
Sex	213	65%	0.87	37.59	0.014	0.95	7, 3	101, 47	19, 9

8.3.3.1 Construct Validity Testing

Scale validity was supported by interscale correlations as shown in Table 8.4. Higher scores were moderately to highly correlated between all scales except the Appearance scale, which showed only a small correlation.

The correlation between scale scores and demographic variables such as gender and age were low (<0.3); thus, the scales were not affected by bias due to these respondent characteristics.

Mean Appearance scale scores were compared between those with CTS and those with either Dupuytren's contracture or rheumatoid arthritis (using an independent samples *t*-test). It was found that those in the former group have significantly higher scores (i.e., they are more satisfied with their hand appearance; mean [SD], 70 [24]) than those in the latter group (56 [27]; $p = 0.003$).

One-way between-groups ANOVA to explore the relationship between QOL, Function and Symptom scores with self-described condition severity is summarised below:

- QOL scores were significantly different depending on self-rated condition severity ($F(2,438) = 7.734$, $p = 0.001$). Post hoc comparisons using the Tukey HSD test indicated that the mean scores for the 'mild' (69.4 [20.4]) and 'moderate' (63.4 [20]) groups were significantly different from the 'severe' (57.4 [23.3]) group. There was no significant difference in the QOL scores when comparing those who described their condition as 'mild' and 'moderate'.
- Function scores were also significantly different depending on self-rated condition severity ($F(2,454) = 7.414$, $p = 0.001$). Post hoc comparisons using the Tukey HSD test showed that the mean scores for the 'mild' (62.4 [25.7]) and 'moderate' (55.7 [24]) group were significantly different from the 'severe' (49.5 [23.3]) group. Again, there

was no significant difference in Function scores when comparing those who described their condition as ‘mild’ and ‘moderate’.

- Symptoms scores were significantly different between the three severity groups ($F(2,446) = 15.254$, $p = 0.000$). Post hoc comparisons using the Tukey HSD test showed that the mean scores for the ‘mild’ (66.6 [17.3]), ‘moderate’ (59.8 [16.4]) and ‘severe’ (52.8 [19.3]) groups were all significantly different.

Comparison of means using an independent samples *t*-test showed that those who believed they needed further surgery (58 [21]) had statistically significant lower QOL scores than those who did not (69 [21], $p = 0.006$).

Table 8.4: Convergent and discriminant construct validity of the HAND-Q scales with patient characteristics

	Appearance		**Function**		**Symptoms**		**Psych**		**QOL**		**Sleep**		**Sex**	
	r	*p*	*r*	*p*	*r*	*p*	*r*	*p*	*r*	*p*	*r*	*p*	*r*	*p*
Appearance	1													
Function	.297	<0.01	1											
Symptoms	.384	<0.01	.615	<0.01	1									
Psychological	.340	<0.01	.501	<0.01	.615	<0.01	1							
QOL	.251	<0.01	.583	<0.01	.600	<0.01	.695	<0.01	1					
Sleep	.211	<0.01	.390	<0.01	.642	<0.01	.572	<0.01	.618	<0.01	1			
Sex	.216	<0.01	.413	<0.01	.466	<0.01	.533	<0.01	.627	<0.01	.482	<0.01	1	
Age	−0.068	0.16	0.012	0.80	0.013	0.79	.113*	0.02	.182**	<0.01	.108*	0.03	.210**	<0.01
Gender	.145**	<0.01	0.077	0.10	.144**	<0.01	.096*	0.04	−0.018	0.71	.109*	0.03	−0.054	0.34

<0.3	Low
>0.3 to <0.5	Moderate
>0.5	High

Interpretation of correlation co-efficient based on recommendations of Cohen (1998, 79-81).

8.4 Discussion

The HAND-Q is being developed using a mixed-methods approach to provide a robust outcome measurement system suitable for both clinical and research applications. The psychometric analysis described in this chapter supplies evidence that the proposed seven scales are both reliable and valid. The scales analysed in this study all displayed ordered thresholds, which means that respondents can interpret the response options and answer in a meaningful manner (Pallant 2007). Individual item fit was supported by low log fit residuals and non-significant Bonferroni adjusted chi-square *p*-values (Pallant 2007). Residual correlations were examined for their impact on overall scale reliability (PSI) and were found to be minor (<0.01).

Person–item maps for each scale showed good targeting of the items to the population of interest; however, there were considerable ceiling effects for the Appearance, Sleep and Sex scales. This is also reflected in the Rasch coverage (indication of the proportion of respondents that were captured by the scale), which was found to be lower than 80% in the Sleep and Sex scales. These concepts are relevant to only a subset of the hand clinic population, and therefore the scales should be completed by those who report their hand condition causing problems with their appearance, sleep or sexual life. Hand appearance is a concept that is highly relevant to a part of the population. Indeed, it has been reported to be a major motivator for surgery in the rheumatoid arthritis population (K. Chung et al. 2006). If this is the case, then it is appropriate that hand appearance be measured as an outcome in this patient cohort. Hand appearance is not a primary concern for all hand clinic patients as many conditions have minimal effect on appearance and for many it is not a priority of their treatment (Bogoch, Escott and Ronald 2011). For this reason, the Hand Appearance scale should only be used in the cohort who report or are expected to report hand appearance as an issue.

As expected, all scales were found to correlate by varying degrees. Scales (except for the Appearance scale) were all correlated moderately to highly (r = 0.39–0.695). That is, those participants with a low Symptoms score (i.e., they suffer many symptoms) were also likely to have low Function, Psychological, Sleep, Sex and QOL scores. This corresponds with what would be expected clinically. The Appearance scale correlated only weakly to other HAND-Q scales (r = 0.211–0.384). Hand appearance is a more independent concept (Johnson 2015), which is possibly why hand appearance is not measured with most currently available PROMs (Sierakowski et al. 2018), or excluded in the scoring of the PROM, as is the case with the PRWHE (MacDermid 1996).

The scales were not biased on the basis of the respondent's gender or age as the correlation between scale scores and these demographic variables was low. This means that the scales are equally sound to evaluate the concepts that they measure regardless of the age or gender of the respondent. Further analysis will be performed on the final international data set will include Differential Item Functioning (DIF). This technique examines whether the scales behave the same when comparing different cohorts, for example to establish if the scales have sound measurement qualities for patients from different countries, varied languages or different hand conditions (Hays, Morales and Reise 2000). Items that show significant DIF will be considered for removal from the final HAND-Q scales.

The Appearance scale performed in the expected manner when comparing the scores between respondents with a deforming hand condition such as Dupuytren's contracture or Rheumatoid Arthritis and those with a condition that has minimal impact on hand aesthetics such as Carpal Tunnel Syndrome. The QOL, Function and Symptoms scales also performed as expected when tested in a clinical scenario; the scale scores corresponded to the respondent's self-rated severity and the QOL scale score was significantly lower for those who believed they need further hand

surgery. The performance of these scales in these a priori clinical hypotheses is reassuring in that they are valid and meaningful in their application.

The HAND-Q study design has limitations that are important to consider. First, recruitment staff at participating hand clinics invited all patients who were cognitively able and could understand English to complete the HAND-Q scales. The benefit of this recruitment strategy is that the final version of the HAND-Q scales is well targeted to the general population of hand clinics, rather than only a subgroup of patients. The disadvantage of this approach is that the dataset collected is not specific enough to answer questions related to a specific patient or clinical characteristic and contributes to the large ceiling effects observed in the dataset. Second, the sample was composed of more participants with traumatic than elective conditions, as this is the case mix routinely reviewed in publicly funded hand clinics. Third, the majority of participants in the sample were under the care of the plastic and reconstructive surgery service as opposed to the orthopaedic surgery service. This was a result of the recruitment clinics active during this initial field-test period. The data collected by the completion of the international field test will be from a more equal distribution of plastic and orthopaedic surgery hand clinics. Four, the response rate to the mailed survey portion was lower than ideal; however, this group formed only a small portion of the overall sample. Five, recruitment staff might have introduced some bias at the time of individual enrolment and overlooked a subgroup of patients. Future studies will include a consecutive sample of patients to avoid any possible recruitment bias. Last, data entry processes were not subjected to any quality control analyses and this analysis would improve confidence with the data quality. Because of the format of the scales and the data entry platform used, any data that might have been skipped during data entry were bought to the attention of the data entry personnel at the end of each scale. This allowed for the data for that scale to be reviewed to ensure accuracy.

8.5 Summary

This chapter details the preliminary results from the field test of the HAND-Q. Seven scales were refined on the basis of RMT. These proposed scales were shown to be psychometrically sound. Although the remaining scales had insufficient data for analysis at the time of writing, projected data collection from international sites is likely to allow for analysis of all scales at the completion of the field test. The preliminary scales from this study are available in Appendix E.3.

Chapter 9: Developing the HAND-Q Phase III: Psychometric Evaluation

9.1 Introduction

The HAND-Q scales will be finalised at the completion of the international field test, which is detailed in Chapter 7. Phase III will involve additional psychometric evaluation of the HAND-Q. The aim of this Psychometric Evaluation is to further evaluate the reliability, validity and responsiveness of the HAND-Q scales. This chapter outlines the methodology that will be used for the psychometric evaluation of the HAND-Q scales which will be performed outside the scope of this thesis.

9.2 Methodology

The concepts of traditional psychometric evaluation have been discussed in Chapter 4 but are reiterated here with specific reference to how they will be used in the context of the final HAND-Q. Sample size calculation of 100 was based on 95% confidence level, allowing for 10% margin of error, an alpha level of 0.05 will be used.

9.2.1 Reliability

Reliability is not an inherent property of a scale but rather a result of interaction between the instrument, the sample cohort and the context in which the test occurs (Streiner, Norman and Cairney 2015). In Phase II of HAND-Q development the reliability of the scales will be explored with Cronbach alpha, Person separation index (PSI) these measures reflect the internal consistency of the scales and the amount of error associated with measurement of an individual respectively (See Chapter 7, section 7.2.5). In this psychometric analysis the test–retest

reliability of the scales will also be established. Test-retest reliability is the ability of the scales to reproduce the same score when completed by a stable participant at different time points. This characteristic is relevant to the clinical application of the HAND-Q scales, as a change of score should reflect a change in the clinical status of the participant, not error within the measurement instrument. To assess the test–rest reliability of the scales, a cohort of 100 participants with hand conditions will be asked to complete the HAND-Q scales at two time points, with a week interceding. The time interval used between administration is controversial, it is a compromise between recollection bias (where the respondent remembers how they answered during the previous sitting and proceeds to give the same answers based on memory) and unwanted clinical change (as only patients with a currently stable condition will be asked to participate in this study) (Marx 2003). The appropriate time interval is also dependent on the construct that is being measured, a scale designed to measure current mood state is not likely to remain stable over a period of a few weeks, where as a scale measuring personality traits should remain stable over a period a several years (Pallant, 2016). A week was chosen as the time interval in this study as it was thought to be long enough to avoid recollection bias and short enough to avoid clinical progression of a stable clinical condition.

9.2.2 Validity

Evaluation of the validity of a measurement instrument is not a dichotomous variable but rather a constellation of information to support the validity of the instrument in the sample and context for which it is being tested (Streiner, Norman and Cairney 2015). For the purposes of this study the validity of a PROM can be simplistically divided into three separate properties: content validity, construct validity and criterion validity (Mokkink, Terwee, Knol et al. 2010). Content validity was established by the qualitative study in Chapter 5. Construct validity was confirmed during cognitive interviews in Chapter 6.

Further construct validation will be performed by comparison of the HAND-Q with existing PROMs such as the MHQ (K. Chung et al. 1998) and DASH (Hudak et al. 1996). This form of construct validation is routinely used in instrument validation studies, where a new instrument is compared with existing instruments to demonstrate either convergent or divergent relationship depending on the constructs are that are being compared (Frost, 2007).

No single existing PROM examines the full breadth of concepts explored by the HAND-Q. We hypothesise that the HAND-Q scales for similar constructs (e.g., Function and Symptom scales) will show moderate correlation with the scores of these existing instruments as they are measuring similar constructs.

Criterion validity is how well an instrument reflects the findings of the 'gold standard', unfortunately, no such standard exists for PROMs (Mokkink, Terwee, Knol et al. 2010). Convergent and discriminant validity is often incorrectly used as a proxy for criterion validity; instead, this comparison provides evidence of construct validity as discussed above (Mokkink, Terwee, Knol et al. 2010).

9.2.3 Responsiveness

Responsiveness establishes the ability of an instrument to detect clinically meaningful change over time. There are varied approaches to establishing the responsiveness of an instrument; most cited are the anchor-based approach or distribution-based approach. In the anchor-based approach, a separate variable is used to establish whether clinically relevant change has occurred; this variable may be clinician or patient rated (Revicki et al. 2008). The distribution-based method estimates the MID from the distribution of scores from a population based on a statistical calculation (Guyatt et al. 2007; Guyatt et al. 2002). RMT analysis has also been demonstrated to be highly sensitive to detect responsiveness (Hobart, Cano and Thompson

2010). As the best way to measure responsiveness is debated in the literature, we plan to employ all three methods mentioned above.

To establish whether the HAND-Q scales are responsive to change, a cohort of 100 participants will be asked to complete the questionnaire prior to undergoing treatment and at a point 12 weeks later.

9.2.4 Participant Recruitment

Participant recruitment for the test–retest reliability and content validity testing will occur simultaneously. The inclusion criteria will be the same as used for the field test (Chapter 7); however, as the sample size for this Phase III study is smaller, recruitment will only go ahead at a limited number of sites (participating sites to be confirmed).

To enable measurement of reliability, participants who are currently stable will be asked to complete the HAND-Q scales. Clinical stability will be decided by the treating clinician.

To enable measurement of responsiveness, participants who are undergoing carpal tunnel release and trigger finger release and primary release of Dupuytren's contracture will be recruited prior to their surgery. These conditions have been selected as they are common in hand clinics and the treatment effectiveness of the surgical interventions is well established (Watchmaker and Watchmaker 2018, Everding et al. 2015 and Dias and Aziz 2018), that is the pre and post-operative scores would be expected to be improved in almost all participants.

9.2.5 Data Collection

Participants will be asked to complete the HAND-Q scales in addition to the MHQ and DASH on an electronic device using REDCap (P. Harris et al. 2009). Contact information for participants will be collected so that participants can be sent a link to the scales for completion

1 week later. The same categories of demographic data collected in the field test will be recorded.

For the responsiveness study, participants will be asked to complete the HAND-Q scales preoperatively. Contact information will be collected, and participants will be sent a link to complete the HAND-Q again at 12 weeks post-surgery.

9.2.6 Data Analysis

9.2.6.1 Test–Retest Reliability

HAND-Q scores will be calculated for the two completions by each participant. The test–retest reliability will then be calculated in SPSS using intraclass correlation coefficient methodology.

9.2.6.2 Construct Validity

The HAND-Q, MHQ and DASH scores will be calculated for each participant. Scores on each of the scales will then be compared using a Pearson's r correlation in SPSS.

9.2.6.3 Responsiveness

Anchor-based techniques will be used to calculate the MID from the HAND-Q transformed Rasch scores. Distribution-based calculation of MID will also be performed. The transformed Rasch HAND-Q scores will be compared using paired *t*-tests; effect size and standardised response means will also be calculated. These analyses provide group-level comparisons to establish responsiveness. RMT analysis will be used to analyse individual person-level comparisons (Hobart, Cano and Thompson 2010). Application of the variety of comparisons listed here will ensure that the responsiveness of the HAND-Q is thoroughly established.

9.3 Summary

This study to evaluate the psychometric properties of the HAND-Q instrument is planned to commence after completion of the international field test. The results of this study will be published separately after the completion of Phase II.

Chapter 10: Conclusion

10.1 Summary of Thesis

Chapter 1 introduced the concepts of outcome measurement in hand conditions and relevant literature. Chapter 2 presented a systematic review of the available PROMs relevant to hand conditions. From this review, it was evident that there was scope to develop a new PROM that explored the patient experience and their outcome in a more comprehensive and methodologically sound manner. Before embarking on such a project, it was important to acknowledge the work already achieved in this field and to learn from existing PROMs. To this end; Chapter 3 compared the data quality and acceptability of three commonly used PROMs in hand treatment: the DASH, MHQ and PRWHE. Chapter 4 explored the psychometric properties of the DASH using both traditional and Rasch analyses. This study demonstrated the weaknesses of the DASH when it was assessed with modern psychometrics and reiterated the importance of considering both traditional and modern psychometric analyses.

The international qualitative study (Phase I) that informed the development of the HAND-Q was described in Chapter 5. From this qualitative dataset, the items of the HAND-Q were generated and grouped into twenty separate scales each designed to measure a unique construct. Professionals in the field and participants with hand conditions reviewed initial drafts of the HAND-Q scales as documented in Chapter 6. The following chapter described the international field test (Phase II), which has commenced but is not yet completed. The field test involved centres located in nine countries. This international collaboration has required the HAND-Q to be translated into seven languages (Dutch, French, Finnish, Urdu, Hindi, Bengali and Tamil), which is work in progress. Preliminary Rasch analyses were performed on the data collected prior to October 2018. Chapter 8 details these analyses which resulted in seven preliminary

HAND-Q scales (Appendix E.3). Finally, Chapter 9 details the planned Phase III study, which will further define the psychometric properties of the HAND-Q scales.

10.2 Significant Original Contribution to Knowledge

This body of work has provided a significant original contribution to knowledge by exploring the lived experience of those with hand conditions and establishing what outcomes concepts are meaningful to them. Many of the concepts that were identified are not addressed by existing PROMs. Examples of areas of deficiency include the patient experience of their treatment and their satisfaction with their care providers.

The mixed methods approach that had proved successful in the development of other PROMs such as the CLEFT-Q, BREAST-Q, BODY-Q and FACE-Q was used as a framework for the development of the HAND-Q (Pusic et al. 2009; Klassen et al. 2015; Klassen et al. 2016; Klassen et al. 2018). The protocol utilised for the HAND-Q development has been published in peer reviewed literature (Sierakowski et al. 2019). Qualitative studies were conducted with participants from Australia, Canada and the United States of America. From these in-depth qualitative studies, the HAND-Q scales were each developed to reflect an issue important to people with a hand condition. The draft scales are being translated and culturally adapted for testing in a further seven languages. No currently available PROM relevant to hand conditions has such a broad inclusive testing population, most are developed within a single country using only local language (usually English) and translated only after completion of the scales. The benefit of including multiple language and cultural groups in the development stages of an instrument are that the scales can be designed to suit the varied participant cohort, not just adapted to different languages after scale items have been finalised. This consideration is unique in the field of hand related PROMs. The HAND-Q will be the product of international

multilanguage and culturally diverse collaboration, and as a result, it will be suitable for international application, allowing and encouraging international collaborative research.

10.3 Future Research

Chapter 8 describes the HAND-Q field test (Phase II) that is currently underway at multiple sites internationally. This field test will allow completion of the HAND-Q scales by identifying the items that resonate with participants from all languages tested and are psychometrically sound. Chapter 9 describes further psychometric testing of the HAND-Q (Phase III) that is planned to be conducted on the completed scales. Following this study, the HAND-Q will be made freely available for use in both clinical and research applications. As discussed throughout this thesis the validity of an instrument is never proven by a particular test but supported by a body of work. It is intended that this body of work will continue to grow with the application of the HAND-Q broadly internationally. For example, potential weaknesses in the spectrum of the conditions represented by the sample in the qualitative studies could be addressed with further cognitive interview testing to include those with a congenital hand difference or a severe hand deformity.

The completed HAND-Q will be continually evaluated and updated as necessary. Adoption of computerised adaptive technology will be explored. Future studies to establish the relationship between scoring of existing instruments and the HAND-Q will allow for use of retrospective data using crosstalk scoring tables (Hobart and Cano 2009). This would allow existing research data to be used and allow for meta-analysis of outcome data using different outcome measurement instruments.

10.4 Concluding Comments

The HAND-Q has been developed to provide a methodologically robust PROM tool for clinicians to use with individual patients and at the cohort level. It will provide interval level data that allows for mathematically sound comparison of outcomes between groups and measurement at an individual level. The HAND-Q has been patient-centred throughout its development as this is the essence of patient-reported outcome measurement. It is anticipated that the HAND-Q scales will be useful tool for clinicians, researchers and patients alike.

References

Acquadro, C., K. Conway, C. Giroudet and I. Mear. 2012. *Linguistic Validation Manual for Health Outcome Assessments*. Lyon: Mapi Institute.

Alderman, Amy K., Anjali S. Arora, Latoya Kuhn, Yongliang Wei and Kevin C. Chung. 2006. 'An Analysis of Women's and Men's Surgical Priorities and Willingness to Have Rheumatoid Hand Surgery'. *The Journal of Hand Surgery* 31 (9): 1447–53.

Alexander, M., O. I. Franko, E. C. Makhni, D. Zurakowski and C. S. Day. 2008. 'Validation of a Modern Activity Hand Survey With Respect to Reliability, Construct and Criterion Validity'. *Journal of Hand Surgery (European Volume)* 33 (5): 653–60.

Allen, Kelli D., Joanne M. Jordan, Jordan B. Renner and Virginia B. Kraus. 2006. 'Validity, Factor Structure, and Clinical Relevance of the AUSCAN Osteoarthritis Hand Index'. *Arthritis & Rheumatology* 54 (2): 551–6.

Alpenfels, E. J. The Anthropology and Social Significance of the Human Hand. *Artificial Limbs*. 2 (2): 4 -21

Alrubaiy, Laith, Hayley A. Hutchings and John G. Williams. 2014. 'Assessing Patient Reported Outcome Measures: A Practical Guide for Gastroenterologists'. *United European Gastroenterology Journal* 2 (6): 463–70.

Amirfeyz, Rouin, Alanna Pentlow, Julian Foote and Ian Leslie. 2009. 'Assessing the Clinical Significance of Change Scores Following Carpal Tunnel Surgery'. *International Orthopaedics* 33 (1): 181–5.

Andersson, G. B. J. and L. Cocchiarella. 2001. *AMA Guides to the Evaluation of Permanent Impairment*. American Medical Association.

Andersson, G. B., C. Gillberg, E.Fernell, M. Johansson and A. Nachemson. 2011. 'Children with surgically corrected hand deformities and upper limb deficiencies: self-concept and psychological well-being.' *Journal of Hand Surgery (European Volume)* 36 (9): 795-801.

Andrich, David. 1988. *Rasch Models for Measurement.* Sage.

Andrich, David and Barry Sheridan. 1997–2011. *RUMM2030*. Software program. RUMM Laboratory, Perth (Australia).

Angst, Felix, Jörg Goldhahn, Susann Drerup, Matthias Flury, Hans-Kaspar Schwyzer and Beat R. Simmen. 2009. 'How Sharp is the Short QuickDASH? A Refined Content and Validity Analysis of the Short Form of the Disabilities of the Shoulder, Arm and Hand

Questionnaire in the Strata of Symptoms and Function and Specific Joint Conditions'. *Quality of Life Research* 18 (8): 1043–51.

Argyle, Michael. 2010. Bodily Communication. Routledge

Arora, Jason and Mona Haj. 2016. *Implementing ICHOM's Standard Sets of Outcomes: Cleft Lip and Palate at Erasmus University Medical Centre in the Netherlands*. London: International Consortium for Health Outcomes Measurement (ICHOM).

Atroshi, Isam, Per-Erik Lyrén and Christina Gummesson. 2009. 'The 6-Item CTS Symptoms Scale: A Brief Outcomes Measure for Carpal Tunnel Syndrome'. *Quality of Life Research* 18 (3): 347–58.

Atroshi, Isam, Per-Erik Lyrén, Ewald Ornstein and Christina Gummesson. 2011. 'The Six-Item CTS Symptoms Scale and Palmar Pain Scale in Carpal Tunnel Syndrome'. *The Journal of Hand Surgery* 36 (5): 788–94.

Aviram M. Giladi and Kevin C. Chung. 2013. 'Measuring Outcomes in Hand Surgery'. *Clinics in Plastic Surgery* 40 (2): 313–22.

Badalamente, Marie, Laureen Coffelt, John Elfar, Glenn Gaston, Warren Hammert, Jerry Huang, Lisa Lattanza, et al. 2013. 'Measurement Scales in Clinical Research of the Upper Extremity, Part 1: General Principles, Measures of General Health, Pain, and Patient Satisfaction'. *The Journal of Hand Surgery* 38 (2): 401–6.

Bailey, Ryan, Vicki Kaskutas, Ida Fox, Carolyn M. Baum, and Susan E. Mackinnon. 2009 'Effect of upper extremity nerve damage on activity participation, pain, depression, and quality of life.' *The Journal of Hand Surgery* 34 (9): 1682-1688.

Baldwin, Paul C. and Jennifer Moriatis Wolf. 2013. 'Outcomes of Hand Fracture Treatments'. *Hand Clinics* 29 (4): 621–30.

Basch, Ethan. 2014. 'New Frontiers in Patient-Reported Outcomes: Adverse Event Reporting, Comparative Effectiveness, and Quality Assessment'. *Annual Review of Medicine* 65: 307–17.

Beaton, Dorcas E., James G. Wright, Jeffrey N. Katz, P. Amadio, C. Bombardier, D. Cole, A. Davis, et al. 2005. 'Development of the QuickDASH: Comparison of Three Item-Reduction Approaches'. *Journal of Bone & Joint Surgery* 87 (5): 1038–46.

Becker, Stephanie J. E., Teun Teunis, David Ring and Ana-Maria Vranceanu. 2016. 'The Trapeziometacarpal Arthrosis Symptoms and Disability Questionnaire: Development and Preliminary Validation'. *Hand* 11 (2): 197–205.

Beckmann, James T., Man Hung, Maren W. Voss, Anthony B. Crum, Jerry Bounsanga and Andrew R. Tyser. 2016. 'Evaluation of the Patient-Reported Outcomes Measurement

Information System Upper Extremity Computer Adaptive Test'. *The Journal of Hand Surgery* 41 (7): 739–44.e4.

Bellamy, N., J. Campbell, B. Haraoui, E. Gerecz-Simon, R. Buchbinder, K. Hobby and J. C. MacDermid. 2002a. 'Clinimetric Properties of the AUSCAN Osteoarthritis Hand Index: An Evaluation of Reliability, Validity and Responsiveness'. *Osteoarthritis and Cartilage* 10 (11): 863–9.

Bellamy, N., J. Campbell, B. Haraoui, R. Buchbinder, K. Hobby, J. H. Roth and J. C. MacDermid. 2002b. 'Dimensionality and Clinical Importance of Pain and Disability in Hand Osteoarthritis: Development of the Australian/Canadian (AUSCAN) Osteoarthritis Hand Index'. *Osteoarthritis and Cartilage* 10 (11): 855–62.

Belvedere, Shane L. and Natalie A. de Morton. 2010. 'Application of Rasch Analysis in Health Care is Increasing and is Applied for Variable Reasons in Mobility Instruments'. *Journal of Clinical Epidemiology* 63 (12): 1287–97.

Bindra, R. R., J. J. Dias, C. Heras-Palau, P. C. Amadio, K. C. Chung and F. D. Burke. 2003. 'Assessing Outcome After Hand Surgery: The Current State'. *The Journal of Hand Surgery: British & European Volume* 28 (4): 289–94.

Black, Nick. 2013. 'Patient Reported Outcome Measures Could Help Transform Healthcare'. *British Medical Journal (Clinical Research Edition)* 346: f167.

Bland, J. Martin and Douglas G. Altman. 1995. 'Multiple Significance Tests: The Bonferroni Method'. *BMJ* 310 (6973): 170.

Bogoch, Earl R., Benjamin G. Escott and Kara Ronald. 2011. 'Hand Appearance as a Patient Motivation for Surgery and a Determinant of Satisfaction with Metacarpophalangeal Joint Arthroplasty for Rheumatoid Arthritis'. *The Journal of Hand Surgery* 36 (6): 1007–14.

Bogoch, Earl R. and Maria G. P. Judd. 2002. 'The Hand: A Second Face?' *The Journal of Rheumatology* 29 (12): 2477–83.

Boone, William J., John R. Staver and Melissa S. Yale. 2014. *Rasch Analysis in the Human Sciences*. Springer.

Bot, Arjan G. J., Jade A. Anderson, Valentin Neuhaus and David Ring. (013. 'Factors Associated with Survey Response in Hand Surgery Research'. *Clinical Orthopaedics and Related Research®* 471 (10): 3237–42.

Bot, Arjan G. J., Stéphanie J. E. Becker, Marianne F. Mol, David Ring and Ana-Maria Vranceanu. 2013. 'Validation of Phone Administration of Short-Form Disability and Psychology Questionnaires'. *The Journal of Hand Surgery* 38 (7): 1383–7.

Braitmayer, Katherin, Caroline Dereskewitz, Cornelia Oberhauser, Klaus-Dieter Rudolf and Michaela Coenen. 2017. 'Examination of the Applicability of the Disabilities of the Arm, Shoulder and Hand (DASH) Questionnaire to Patients with Hand Injuries and Diseases Using Rasch Analysis'. *The Patient: Patient-Centered Outcomes Research* 10 (3): 367–76.

Brazier, J. E., R. Harper, N. Jones, A. O'Cathain, K. Thomas, T. Usherwood and L. Westlake 1992. 'Validating the SF-36 Health Survey Questionnaire: New Outcome Measure for Primary Care'. *BMJ* 305 (6846): 160–4.

Buchbinder, Rachelle, Richard H. Osborne, Peter R. Ebeling, John D. Wark, Peter Mitchell, Chris Wriedt, Stephen Graves, Margaret P. Staples and Bridie Murphy. 2009. 'A Randomized Trial of Vertebroplasty for Painful Osteoporotic Vertebral Fractures'. *The New England Journal of Medicine* 361 (6): 557–68.

BuchbinderRachelle., Richard H. Osborne and David Kallmes. 2010. 'Invited Editorial Presents an Accurate Summary of the Results of Two Randomised Placebo-Controlled Trials of Vertebroplasty'. *The Medical Journal of Australia* 192 (6): 338–41.

Burke, F. D., D. A. McGrouther and P. J. Smith. 1990. *Principles of Hand Surgery*. Edinburgh: Churchill Livingstone.

Calfee, Ryan P. and Amelia A. Adams. 2012. 'Clinical Research and Patient-Rated Outcome Measures in Hand Surgery'. *The Journal of Hand Surgery* 37 (4): 851–5.

Cano, S. J., L. E. Barrett, J. P. Zajicek and J. C. Hobart. 2011a. 'Beyond the Reach of Traditional Analyses: Using Rasch to Evaluate the DASH in People with Multiple Sclerosis'. *Multiple Sclerosis Journal* 17 (2): 214–22.

Cano, S. J., L. E. Barrett, J. P. Zajicek and J. C. Hobart. 2011b. 'Dimensionality is a Relative Concept'. *Multiple Sclerosis Journal* 17 (7): 893–4.

Cano, S. J., J. P. Browne, D. L. Lamping, A. H. N. Roberts, D. A. McGrouther and N. A. Black. 2004. 'The Patient Outcomes of Surgery–Hand/Arm (POS-Hand/Arm): A New Patient-Based Outcome Measure'. *The Journal of Hand Surgery: British & European Volume* 29 (5): 477–85.

Cano, S. J. and J. C. Hobart. 2011. 'The Problem with Health Measurement'. *Patient Prefererence and Adherence* 5: 279–90.

Cano, S. J., A. Klassen and A. L. Pusic (2009). 'The science behind quality-of-life measurement: a primer for plastic surgeons'. *Plastic and Reconstructive Surgery* 123 (3): 98e–106e.

Cella, D., S. Yount, N. Rothrock, R. Gershon, K. Cook, B. Reeve, D. Ader, J. F. Fries, B. Bruce and M. Rose (2007). 'The Patient-Reported Outcomes Measurement Information System (PROMIS): Progress of an NIH Roadmap Cooperative Group During its First Two Years'. *Medical Care* 45 (5 Suppl 1): S3–11.

Chen, Christine C. and Rita K. Bode. 2010. 'Psychometric Validation of the Manual Ability Measure-36 (MAM-36) in Patients with Neurologic and Musculoskeletal Disorders'. *Archives of Physical Medicine and Rehabilitation* 91 (3): 414–20.

Chen, C. C., C. V. Granger, C. A. Peimer, O. J. Moy and S. Wald. 2005. 'Manual Ability Measure (MAM-16): A Preliminary Report on a New Patient-Centred and Task-Oriented Outcome Measure of Hand Function'. *The Journal of Hand Surgery: British & European Volume* 30 (2): 207–16.

Chung, B. T. and S. F. Morris. 2014. 'Reliability and Internal Validity of the Michigan Hand Questionnaire'. *Annals of Plastic Surgery* 73 (4): 385–9.

Chung, B. T. and S. F. Morris. 2015. 'Confirmatory Factor Analysis of the Michigan Hand Questionnaire'. *Annals of Plastic Surgery* 74 (2): 176–81.

Chung, K. C., J. B. Hamill, M. R. Walters and R. A. Hayward. 1999. 'The Michigan Hand Outcomes Questionnaire (MHQ): Assessment of Responsiveness to Clinical Change'. *Annals of Plastic Surgery* 42 (6): 619–22.

Chung, Kevin C., Sandra V. Kotsis, H. Myra Kim, Frank D. Burke and E. F. Shaw Wilgis. 2006. 'Reasons Why Rheumatoid Arthritis Patients Seek Surgical Treatment for Hand Deformities'. *The Journal of Hand Surgery* 31 (2): 289–94.

Chung, Kevin C., Matthew S. Pillsbury, Madonna R. Walters and Rodney A. Hayward. 1998. 'Reliability and Validity Testing of the Michigan Hand Outcomes Questionnaire'. *The Journal of Hand Surgery* 23 (4): 575–87.

Citron, N., C. E. Hulme and N. Wardle. 2007. 'A Self-Administered Questionnaire for Basal Osteoarthritis of the Thumb'. *Journal of Hand Surgery (European Volume)* 32 (5): 524–8.

Cohen, Deborah J. and Benjamin F. Crabtree. 2008. 'Evaluative Criteria for Qualitative Research in Health Care: Controversies and Recommendations'. *The Annals of Family Medicine* 6 (4): 331–9.

Cohen, Jacob. 1988. 'Statistical power analysis for the behavioural sciences'. 2nd Edition, Hillsdale, N.J.:L. Erlbaum Associates, 79-81.

Collins, Debbie. 2003. 'Pretesting Survey Instruments: An Overview of Cognitive Methods'. *Quality of Life Research* 12 (3): 229–38.

Coons, Stephen Joel, Chad J. Gwaltney, Ron D. Hays, J. Jason Lundy, Jeff A. Sloan, Dennis A. Revicki, William R. Lenderking, David Cella and Ethan Basch. 2009. 'Recommendations on Evidence Needed to Support Measurement Equivalence between Electronic and Paper-Based Patient-Reported Outcome (PRO) Measures: ISPOR ePRO Good Research Practices Task Force Report'. *Value in Health* 12 (4): 419–29.

Crawford, John R. and Julie D. Henry. 2003. 'The Depression Anxiety Stress Scales (DASS): Normative Data and Latent Structure in a Large Non-Clinical Sample'. *British Journal of Clinical Psychology* 42 (2): 111–31.

Creswell, John and Vicki Plano Clark. 2011. *Designing and Conducting Mixed Method Research*. 2nd ed. Thousand Oaks, CA.

Cronbach, Lee J. 1951. 'Coefficient Alpha and the Internal Structure of Tests'. *Psychometrika* 16 (3): 297–334.

Dawson, Jill, Helen Doll, Ray Fitzpatrick, Crispin Jenkinson and Andrew J. Carr. 2010. 'The Routine Use of Patient Reported Outcome Measures in Healthcare Settings'. *BMJ* 340: c186.

de Carvalho Leite, Jose C., Christina Jerosch-Herold and Fujian Song. 2006. 'A Systematic Review of the Psychometric Properties of the Boston Carpal Tunnel Questionnaire'. *BMC Musculoskeletal Disorders* **7**: 78.

Dean, Nicola R. and Tamara Crittenden. 2016. 'A Five Year Experience of Measuring Clinical Effectiveness in a Breast Reconstruction Service Using the BREAST-Q Patient Reported Outcomes Measure: A Cohort Study'. *Journal of Plastic, Reconstructive & Aesthetic Surgery* 69 (11): 1469–77.

Detmar, S. B., M. J. Muller, J. H. Schornagel, L. D. Wever and N. K. Aaronson. 2002. 'Health-related quality-of-life assessments and patient-physocian communication: a randomized controlled trial'. *Journal of the American Medical Association* 288 (23): 3027-34.

DeVellis, Robert F. 2006. 'Classical Test Theory'. *Medical Care* 44 (11 Suppl 3): S50–9.

DeVellis, Robert F. 2016. *Scale Development: Theory and Applications*. Sage.

Devlin, Nancy J. and John Appleby. 2010. *Getting the Most out of PROMs: Putting Health Outcomes at the Heart of NHS Decision-Making*. London: King's Fund.

Dias, J. J., and S. Aziz. 2018. 'Fascietcomy for Dupuytren Contracture'. *Hand Clinics* 34 (3): 351-366.

Dias, J. J., B. Bhowal, C. J. Wildin and J. R. Thompson. 2001. 'Assessing the Outcome of Disorders of the Hand: Is the Patient Evaluation Measure Reliable, Valid, Responsive and Without Bias?' *Journal of Bone and Joint Surgery* 83 (2): 235–40.

Döring, Anne-Carolin, Sjoerd P. F. T. Nota, Michiel G. Hageman and David C. Ring. 2014. 'Measurement of Upper Extremity Disability Using the Patient-Reported Outcomes Measurement Information System'. *The Journal of Hand Surgery* 39 (6): 1160–5.

Dowrick, A. S., B. J. Gabbe, O. D. Williamson and P. A. Cameron. 2006. 'Does the Disabilities of the Arm, Shoulder and Hand (DASH) Scoring System Only Measure Disability Due To Injuries to the Upper Limb?' *Journal of Bone and Joint Surgery* 88 (4): 524–7.

Dubert, T. 2014. 'Outcome Measurements in Hand and Upper Limb Surgery'. *Chirurgie de la Main* 33 (4): 235–46.

Dy, Christopher J., Thomas Schmicker, Quynh Tran, Brian Chadwick and Aaron Daluiski. 2012. 'The Use of a Tablet Computer to Complete the DASH Questionnaire'. *The Journal of Hand Surgery* 37 (12): 2589–94.

El-Daly, I., H. Ibraheim, K. Rajakulendran, P. Culpan and P. Bates. 2016. 'Are Patient-Reported Outcome Measures in Orthopaedics Easily Read by Patients?' *Clinical Orthopaedics and Related Research®* 474 (1): 246–55.

Everding, N. G., G. B. Bishop, C. M. Belyea and M. C. Soong. 2015. 'Risk Factors for Complications of Open Trigger Finger Release'. *HAND* 10: 297-300.

Fava, G. A., E. Tomba and N. Sonino. 2012. 'Clinimetrics: The Science of Clinical Measurements'. *International Journal of Clinical Practice* 66 (1): 11–15.

Fayers, P. M., D. J. Hand, K. Bjordal and M. Groenvold. 1997. 'Causal Indicators in Quality of Life Research'. *Quality of Life Research* 6 (5): 393–406.

Fayers, P. M. and D. Machin. 2013. *Quality of Life: The Assessment, Analysis and Interpretation of Patient-Reported Outcomes*. John Wiley & Sons.

FDA (Food and Drug Administration). 2009. 'Guidance for Industry: Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims'. *Federal Register* 74 (235): 65132–3.

Feinstein, A. R. 1987. 'Clinimetric Perspectives'. *Journal of Chronic Diseases* 40 (6): 635–40.

Feinstein, A. R. 1999. 'Multi-Item Instruments vs Virginia Apgar's Principles of Clinimetrics'. *Archives of Internal Medicine* 159 (2): 125–8.

Fisher, Jeffrey D., Marvin Rytting, and Richard Heslin. 1976. 'Hands touching hands: Affective and evaluative effects of an interpersonal touch'. *Sociometry* 1: 416-421

Fitzpatrick, R., C. Jenkinson, A. Klassen and T. Goodacre. 1999. 'Methods of Assessing Health-Related Quality of Life and Outcome for Plastic Surgery'. *British Journal of Plastic Surgery* 52 (4): 251–5.

Forget, Nancy J., Christina Jerosch-Herold, Lee Shepstone and Johanne Higgins. 2014. 'Psychometric evaluation of the Disabilities of the Arm, Shoulder and Hand (DASH) with Dupuytren's contracture: Validity evidence using Rasch modeling'. *BMC Musculoskeletal Disorders* 15: 361.

Forward, D. P., J. S. Sithole and T. R. C. Davis. 2007. 'The Internal Consistency and Validity of the Patient Evaluation Measure for Outcomes Assessment in Distal Radius Fractures'. *Journal of Hand Surgery (European Volume)* 32 (3): 262–7.

Franchignoni, Franco, Andrea Giordano, Francesco Sartorio, Stefano Vercelli, Barbara Pascariello and Giorgio Ferriero. 2010. 'Suggestions for Refinement of the Disabilities of the Arm, Shoulder and Hand Outcome Measure (DASH): A Factor Analysis and Rasch Validation Study'. *Archives of Physical Medicine and Rehabilitation* 91 (9): 1370–7.

Frost M. H., B. B. Reave, A. M. Liepa, J.W. Stauffer, R. D. Hays and the Mayo/FDA Patient-Reported Outcomes Consensus Meeting Group. 2007. 'What is Sufficient Evidence for the Reliability and Validity of Patient-Reported Outcome Measures?'. *Value in Health* 10: S94-S105.

Fujihara, Y., M. J. Shauver, M. E. Lark, L. Zhong and K. C. Chung. 2017. 'The Effect of Workers' Compensation on Outcome Measurement Methods after Upper Extremity Surgery: A Systematic Review and Meta-Analysis'. *Plastic and Reconstructive Surgery* 139 (4): 923–33.

Gabel, C. Philip, Lori A. Michener, Brendan Burkett and Anne Neller. 2006. 'The Upper Limb Functional Index: Development and Determination of Reliability, Validity, and Responsiveness'. *Journal of Hand Therapy* 19 (3): 328–49.

Gabel, Charles Philip, Lori A. Michener, Markus Melloh and Brendan Burkett. 2010. 'Modification of the Upper Limb Functional Index to a Three-Point Response Improves Clinimetric Properties'. *Journal of Hand Therapy* 23 (1): 41–52.

Gibson, Peter D., Garret L. Sobol and Irfan H. Ahmed. 2016. 'Zone II Flexor Tendon Repairs in the United States: Trends in Current Management'. *The Journal of Hand Surgery* 42 (2): e99–e108.

Glickel, Steven Z. 2004. 'A History of Hand Surgery in New York City'. *The Journal of Hand Surgery* 29 (5): 774–84.

Gomes, M., N. Gutacker, C. Bojke and A. Street. 2015. 'Addressing Missing Data in Patient-Reported Outcome Measures (PROMS): Implications for the Use of PROMS for Comparing Provider Performance'. *Health Economics* 25 (5): 515–28.

Gonçalves Bradley, D. C., C. Gibbons, I. Ricci-Cabello, N. J. H. Bobrovitz, E. J. Gibbons, A. Kotzeva, J. Alonso, R. Fitzpatrick, P. Bower and P. J. van der Wees. 2015. 'Routine Provision of Information on Patient-Reported Outcome Measures to Healthcare Providers and Patients in Clinical Practice'. *Cochrane Datebase of Systematic Reviews* 4.

Graham, Brent. 2016. 'Defining and Measuring Patient Satisfaction'. *The Journal of Hand Surgery* 41 (9): 929–31.

Gupta, Amit, Simon P. J. Kay and Luis R. Scheker. 2000. *The Growing Hand: Diagnosis and Management of the Upper Extremity in Children*. Mosby.

Gustafsson, Margareta, Lars-Olof Persson and Anders Amilon. 2000. 'A Qualitative Study of Stress Factors in the Early Stage of Acute Traumatic Hand Injury'. *Journal of Advanced Nursing* 32 (6): 1333–40.

Gustafsson, Margareta, Lars-Olof Persson and Anders Amilon. 2002. 'A Qualitative Study of Coping in the Early Stage of Acute Traumatic Hand Injury'. *Journal of Clinical Nursing* 11 (5): 594–602.

Guyatt, Gordon H., Carol Estwing Ferrans, Michele Y. Halyard, Dennis A. Revicki, Tara L. Symonds, Claudette G. Varricchio, Anna Kotzeva, Jose M. Valderas, Jordi L. Alonso and Clinical Significance Consensus Meeting Group. 2007. 'Exploration of the Value of Health-Related Quality-Of-Life Information from Clinical Research and into Clinical Practice'. *Mayo Clinic Proceedings* 82 (10): 1229–39.

Guyatt, Gordon H., David Osoba, Albert W. Wu, Kathleen W. Wyrwich, Geoffrey R. Norman and Clinical Significance Consensus Meeting Group. 2002. Methods to Explain the Clinical Significance of Health Status Measures. *Mayo Clinic Proceedings* 77 (4): 371–83.

Hackel, Mary E., George A. Wolfe, Sharon M. Bang and Judith S. Canfield. 1992. 'Changes in Hand Function in the Aging Adult as Determined by the Jebsen Test of Hand Function'. *Physical Therapy* 72 (5): 373–7.

Hagquist, Curt, Malin Bruce and J. Petter Gustavsson. 2009. 'Using the Rasch Model in Nursing Research: An Introduction and Illustrative Example'. *International Journal of Nursing Studies* 46 (3): 380–93.

Hahn, Elizabeth A., David Cella, Olivier Chassany, Diane L. Fairclough, Gilbert Y. Wong, Ron D. Hays and Clinical Significance Consensus Meeting Group. 2007. 'Precision of Health-Related Quality-of-Life Data Compared with Other Clinical Measures'. *Mayo Clinic Proceedings* 82 (10): 1244–54.

Hamilton, Clayon B. and Bert M. Chesworth (2013). 'A Rasch-Validated Version of the Upper Extremity Functional Index for Interval-Level Measurement of Upper Extremity Function'. *Physical Therapy* 93 (11): 1507–19.

Harris, Ian, Jonathan Mulford, Michael Solomon, James M. van Gelder and Jane Young. 2005. 'Association Between Compensation Status and Outcome After Surgery: A Meta-Analysis'. *Journal of the American Medical Association* 293 (13): 1644–52.

Harris, Paul A., Robert Taylor, Robert Thielke, Jonathon Payne, Nathaniel Gonzalez and Jose G. Conde. 2009. 'Research Electronic Data Capture (REDCap)—A Metadata-Driven Methodology and Workflow Process for Providing Translational Research Informatics Support'. *Journal of Biomedical Informatics* 42 (2): 377–81.

Harvill, L. M. 1991. 'Standard error of measurement. *Educational Measurement: issues and practice* 10(2): 33-41

Hays, R. D., L. S. Morales and S. P. Reise. 2000. 'Item Response Theory and Health Outcomes Measurement in the 21st Century'. *Medical Care* 38 (9 Suppl): II28.

Hays, Ron D., Karen L. Spritzer, Dagmar Amtmann, Jin-Shei Lai, Esi Morgan DeWitt, Nan Rothrock, Darren A. DeWalt, William T. Riley, James F. Fries and Eswar Krishnan. 2013. 'Upper-Extremity and Mobility Subdomains from the Patient-Reported Outcomes Measurement Information System (PROMIS) Adult Physical Functioning Item Bank'. *Archives of Physical Medicine and Rehabilitation* 94 (11): 2291–6.

Heybeli, N., S. Kutluhan, S. Demirci, M. Kerman and E. F. Mumcu. 2002. 'Assessment of Outcome of Carpal Tunnel Syndrome: A Comparison of Electrophysiological Findings and a Self-Administered Boston Questionnaire'. *The Journal of Hand Surgery: British & European Volume* 27 (3): 259–64.

Hill, R., A. Herstein and K. Walters. 1976. 'Juvenile Rheumatoid Arthritis: Follow-Up into Adulthood—Medical, Sexual and Social Status'. *Canadian Medical Association Journal* 114 (9): 790–6.

Hobart, J. and S. Cano. 2009. *Improving the Evaluation of Therapeutic Interventions in Multiple Sclerosis: The Role of New Psychometric Methods*. Prepress Projects.

Hobart, J. C., S. J. Cano and A. J. Thompson. 2010. 'Effect Sizes Can Be Misleading: Is It Time To Change the Way We Measure Change?' *Journal of Neurology, Neurosurgery & Psychiatry* 81 (9): 1044–8.

Hobby, J. L., C. Watts and D. Elliot. 2005. 'Validity and Responsiveness of the Patient Evaluation Measure as an Outcome Measure for Carpal Tunnel Syndrome'. *The Journal of Hand Surgery: British & European Volume* 30 (4): 350–4.

Hsu, C.-C. and B. A. Sandford. 2007. ‘The Delphi Technique: Making Sense of Consensus’. *Practical Assessment, Research & Evaluation* 12 (10): 1–8.

Hudak, Pamela L., Peter C. Amadio, Claire Bombardier, Dorcas Beaton, Donald Cole,Aileen Davis, Gillian Hawker, et al. 1996. ‘Development of an Upper Extremity Outcome Measure: The DASH (Disabilities of the Arm, Shoulder, and Hand)’. *American Journal of Industrial Medicine* 29 (6): 602–8 **[**Erratum 30 (3): 372].

Huisstede, B. M. A., A. Feleus, S. M. Bierma-Zeinstra, J. A. Verhaar and B. W. Koes. 2009. ‘Is the Disability of Arm, Shoulder, and Hand Questionnaire (DASH) Also Valid and Responsive in Patients with Neck Complaints’. *Spine* 34 (4): E130–8.

Hustedt, Joshua W., Andrew Chung, Daniel D. Bohl, Neil Olmschied and Scott G. Edwards. 2017. ‘Comparison of Postoperative Complications Associated with Anesthetic Choice for Surgery of the Hand’. *The Journal of Hand Surgery* 42 (1): 1–8.e5.

Ipsen, T., L. Lundkvist, T. Barfred and J. Pless. 1990. ‘Principles of Evaluation and Results in Microsurgical Treatment of Major Limb Amputations: A Follow-Up Study of 26 Consecutive Cases 1978–1987’. *Scandinavian Journal of Plastic and Reconstructive Surgery and Hand Surgery* 24 (1): 75–80.

Jabaley, M. E. 1981. ‘Current Concepts of Nerve Repair’. *Clinics in Plastic Surgery* 8 (1): 33–44.

Jebsen, R. H., N. Taylor, R. Trieschmann, M. J. Trotter and L. A. Howard. 1969. ‘An Objective and Standardized Test of Hand Function’. *Archives of Physical Medicine and Rehabilitation* 50 (6): 311.

Johnson, S. P., S. J. Sebastin, S. A. Rehim and K. C. Chung. 2015. ‘The Importance of Hand Appearance as a Patient-Reported Outcome in Hand Surgery’. *Plastic and Reconstructive Surgery–Global Open* 3 (11): e552.

Johnson, Shepard P. and Kevin C. Chung. 2014. ‘Comparative Effectiveness Research in Hand Surgery’. *Hand Clinics* 30 (3): 319–27.

Kallmes, David F., Bryan A. Comstock, Patrick J. Heagerty, Judith A. Turner, David J. Wilson, Terry H. Diamond, Richard Edwards, et al. 2009. ‘A Randomized Trial of Vertebroplasty for Osteoporotic Spinal Fractures’. *The New England Journal of Medicine* 361 (6): 569–79.

Kamal, Robin N. and Hand Surgery Quality Consortium. 2016. ‘Quality and Value in an Evolving Health Care Landscape’. *The Journal of Hand Surgery* 41 (7): 794–9.

Kamburoğlu, H. O., A. E. Aksu, E. Sönmez, E. T. İşçi, T. Şafak and A. Kecik. 2011. 'Which Instrument Should We Use to Assess Hand Function After Digital Replantation?' *Journal of Hand Surgery (European Volume)* 36 (5): 392–5.

Kennedy, C., D. Beaton, S. Solway, S. McConnell and C. Bombardier. 2011. *Disabilities of the Arm, Shoulder and Hand. DASH and QuickDASH Outcome Measure User's Manual Thirds Edition*. Toronto, Ontario: Institute for Work & Health.

Khadka, J., V. K. Gothwal, C. McAlinden, E. L. Lamoureux and K. Pesudovs. 2012. 'The Importance of Rating Scales in Measuring Patient-Reported Outcomes'. *Health and Quality of Life Outcomes* 10 (1): 1.

Khan, Wasim S., Bernice Dillon, Manish Agarwal, Max Fehily and Matthew Ravenscroft 2009. 'The Validity, Reliability, Responsiveness, and Bias of the Manchester-Modified Disability of the Arm, Shoulder, and Hand Score in Hand Injuries'. *Hand* 4 (4): 362–7.

Khan, Wasim S., Rohit Jain, Bernice Dillon, Lawrence Clarke, Max Fehily and Matthew Ravenscroft. 2008. 'The "M2 DASH"—Manchester-Modified Disabilities of Arm Shoulder and Hand Score'. *Hand* 3 (3): 240–4.

Kincaid, J. Peter., Robert P. Fishburne Jr, Richard L. Rogers and Brad S. Chissom. 1975. 'Derivation of New Readability Formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy Enlisted Personnel'. *Institute for Simulation and Training*. 56.

Kirwan, J. and J. Reeback. 1986. 'Stanford Health Assessment Questionnaire Modified to Assess Disability in British Patients with Rheumatoid Arthritis'. *Rheumatology* 25 (2): 206–9.

Klassen, A. F., S. J. Cano, A. Alderman, M. Soldin, A. Thoma, S. Robson, M. Kaur, A. Papas, N. Van Laeken and V. H. Taylor. 2016. 'The BODY-Q: A Patient-Reported Outcome Instrument for Weight Loss and Body Contouring Treatments'. *Plastic and Reconstructive Surgery—Global Open* 4 (4).

Klassen, A. F., S. J. Cano, J. A. Schwitzer, A. M. Scott and A. L. Pusic. 2015. 'FACE-Q Scales for Health-Related Quality of Life, Early Life Impact, Satisfaction with Outcomes, and Decision To Have Treatment: Development and Validation'. *Plastic and Reconstructive Surgery* 135 (2): 375–86.

Klassen, Anne F., Andea L. Pusic, Amie Scott, Jennifer Klok and Stefan J. Cano. 2009. 'Satisfaction and Quality of Life in Women Who Undergo Breast Surgery: A Qualitative Study'. *BMC Women's Health* 9 (1): 1.

Klassen, Anne F., Karen W. Y. Wong Riff, Natasha M. Longmire, Asteria Albert, Gregory C. Allen, Mustafa Asim Aydin, Stephen B. Baker, Stefan J. Cano, A. J. Chan and D. J. Courtemanche et al. 2018. 'Psychometric Findings and Normative Values for the CLEFT-Q Based On 2434 Children and Young Adult Patients with Cleft Lip and/or Palate from 12 Countries'. *Canadian Medical Association Journal* 190 (15): E455–62.

Kus, Sandra, Cornelia Oberhauser and Alarcos Cieza 2012. 'Validation of the Brief International Classification of Functioning, Disability, and Health (ICF) Core Set for Hand Conditions'. *Journal of Hand Therapy* 25 (3): 274–86; quiz 287.

Lehman, Leigh A., Bhagwant S. Sindhu, Orit Shechtman, Sergio Romero and Craig A. Velozo. 2010. 'A Comparison of the Ability of Two Upper Extremity Assessments to Measure Change in Function'. *Journal of Hand Therapy* 23 (1): 31–9; quiz 40.

Lehman, Leigh A., Michelle Woodbury and Craig A. Velozo. 2011. 'Examination of the Factor Structure of the Disabilities of the Arm, Shoulder, and Hand Questionnaire'. *The American Journal of Occupational Therapy* 65 (2): 169–78.

Levine, D. W., B. P. Simmons, M. J. Koris, L. H. Daltroy, G. G. Hohl, A. H. Fossel and J. N. Katz. 1993. 'A Self-Administered Questionnaire for the Assessment of Severity of Symptoms and Functional Status in Carpal Tunnel Syndrome'. *The Journal of Bone & Joint Surgery* 75 (11): 1585–92.

Linacre, J. M. 1994. 'Sample Size and Item Calibration (or Person Measure) Stability'. *Rasch Measurement Transactions* **7**.

Linacre, J. M. 1999. 'Investigating Rating Scale Category Utility'. *Journal of Outcome Measurement* 3: 103–22.

Linacre, J. M. 2002. 'Optimizing Rating Scale Category Effectiveness'. *Journal of Applied Measurement* 3 (1): 85–106.

Linacre, J. M. and B. D. Wright. 1995. 'Assigning Item Weights: Item Replication or Rating Scales?' *Rasch Measurement Transactions* 8 (4): 403.

Lohr, Kathleen N. 2002. 'Assessing Health Status and Quality-of-Life Instruments: Attributes and Review Criteria'. *Quality of Life Research* 11 (3): 193–205.

London, Daniel A., Jeffrey G. Stepan, Martin I. Boyer and Ryan P. Calfee. 2014. 'Performance Characteristics of the Verbal QuickDASH'. *The Journal of Hand Surgery* 39 (1): 100–7.

MacDermid, Joy C. 1996. 'Development of a Scale for Patient Rating of Wrist Pain and Disability'. *Journal of Hand Therapy* 9 (2): 178–83.

MacDermid, Joy C. 2005. ‘Measurement of Health Outcomes Following Tendon and Nerve Repair’. *Journal of Hand Therapy* 18 (2): 297–312.

MacDermid, Joy. 2007. *The Patient-Rated Wrist Evaluation (PRWE) User Manual*. Ontario, Canada: M. U. School of Rehabilitation Science.

MacDermid, Joy C. 2014. ‘Patient-Reported Outcomes: State-of-the-Art Hand Surgery and Future Applications’. *Hand Clinics* 30 (3): 293–304.

MacDermid, Joy C. and Virginia Tottenham. 2004. ‘Responsiveness of the Disability of the Arm, Shoulder, and Hand (DASH) and Patient-Rated Wrist/Hand Evaluation (PRWHE) in Evaluating Change After Hand Therapy’. *Journal of Hand Therapy* 17 (1): 18–23.

Macey, A. C., F. D. Burke, K. Abbott, N. J. Barton, E. Bradbury, A. Bradley, M. J. Bradley, et al. 1995. ‘Outcomes of Hand Surgery’’ *The Journal of Hand Surgery: British & European Volume* 20 (6): 841–55.

Malay, Sunitha and Kevin C. Chung. 2013. ‘How to Use Outcomes Questionnaires: Pearls and Pitfalls’. *Clinics in Plastic Surgery* 40 (2): 261–9.

Marcano Belisario, J. S., J. Jamsek, K. Huckvale, J. O’Donoghue, C. P. Morrison and J. Car 2015. ‘Comparison of Self-Administered Survey Questionnaire Responses Collected Using Mobile Apps Versus Other Methods’. *Cochrane database of systematic reviews*, 7.

Massy-Westropp, Nicoa, Michael Ahern and Jegan Krishnan. 2005. ‘A Visual Analogue Scale for Assessment of the Impact of Rheumatoid Arthritis in the Hand: Validity and Repeatability’. *Journal of Hand Therapy* 18 (1): 30–3.

Mathes, S. J. and V. R. Hentz. 2006. *Plastic Surgery*. Saunders Elsevier.

McCartney, C. J. L., R. Brull, V. W. S. Chan, J. Katz, S. Abbas, B. Graham, H. Nova, R. Rawson, D. J. Anastakis and H. von Schroeder. 2004. ‘Early but No Long-Term Benefit of Regional Compared with General Anesthesia for Ambulatory Hand Surgery’. *Anesthesiology: The Journal of the American Society of Anesthesiologists* 101 (2): 461–7.

McCombe, David B. 2018. ‘Expanding the Possibilities of Hand Surgery Training in Australia’. *Australasian Journal of Plastic Surgery* 1 (2): 16–18.

McGrail, K., S. Bryan and J. Davis. 2010. ‘Let’s All Go to the PROM: The Case for Routine Patient-Reported Outcome Measurement in Canadian Healthcare’. *Healthcare Papers* 11 (4): 8–18; discussion 55–18.

McHorney, C. A., J. E. Ware Jr, J. R. Lu and C. D. Sherbourne. 1994. 'The MOS 36-item Short-Form Health Survey (SF-36): III. Tests of Data Quality, Scaling Assumptions, and Reliability Across Diverse Patient Groups'. *Medical Care* 32 (1): 40–66.

McPhee, Scott D. 1987. 'Functional Hand Evaluations: A Review'. *The American Journal of Occupational Therapy* 41 (3): 158–63.

Meals, Clifton G. and Roy A. Meals. 2007. 'A History of Surgery in the Instrument Tray: Eponymous Tools Used in Hand Surgery'. *The Journal of Hand Surgery* 32 (7): 942–53.

Merle, M. and G. Dautel. 2016. *Emergency Surgery of the Hand*. Elsevier Health Sciences.

Mohan, A., J. Vadher, H. Ismail and D. Warwick. 2014. 'The Southampton Dupuytren's Scoring Scheme'. *Journal of Plastic Surgery and Hand Surgery* 48 (1): 28–33.

Mokkink, Lidwine B., Caroline B. Terwee, Dirk L. Knol, Paul W. Stratford, Jordi Alonso, Donald L. Patrick, Lex M. Bouter and Henrica C. De Vet. 2010. 'The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content'. *BMC Medical Research Methodology* 10: 22.

Mokkink, Lidwine B., Caroline B. Terwee, Donald L. Patrick, Jordi Alonso, Paul W. Stratford, Dirk L. Knol, Lex M. Bouter and Henrica C. W. de Vet. 2010. 'The COSMIN Checklist for Assessing the Methodological Quality of Studies on Measurement Properties of Health Status Measurement Instruments: An International Delphi Study'. *Quality of Life Research* 19 (4): 539–49.

Naidu, S. H., D. Panchik and V. M. Chinchilli. 2009. 'Development and Validation of the Hand Assessment Tool'. *Journal of Hand Therapy* 22 (3): 250–6; quiz 257.

Newcombe, R. G. 1998. 'Two-Sided Confidence Intervals for the Single Proportion: Comparison of Seven Methods'. *Statistics in Medicine* 17 (8): 857–72.

Noback, Peter C., Joseph M. Lombardi, Mani Seetharaman, Donald H. Lee, Robert J. Strauch and Melvin P. Rosenwasser. 2017. 'Development and Validation of a Disease-Specific Questionnaire for Basal Joint Arthritis'. *Journal of Wrist Surgery* 6 (2): 126–33.

Nota, S. P., J. A. Strooker and D. Ring. 2014. 'Differences in Response Rates Between Mail, E-Mail, and Telephone Follow-Up in Hand Surgery Research'. *Hand* 9 (4): 504–10.

Omer Jr, George E. 2000. 'Development of Hand Surgery: Education of Hand Surgeons'. *The Journal of Hand Surgery* 25 (4): 616–28.

Packham, J. C. and M. A. Hall. 2002. 'Long-Term Follow-Up of 246 Adults with Juvenile Idiopathic Arthritis: Social Function, Relationships and Sexual Activity'. *Rheumatology* 41 (12): 1440–3.

Packham, Tara and Joy C. MacDermid. 2013. Measurement Properties of the Patient-Rated Wrist and Hand Evaluation: Rasch Analysis of Responses from a Traumatic Hand Injury Population'. *Journal of Hand Therapy* 26 (3): 216–23; quiz 224.

Pallant, Julie F. 2016. *SPSS Survival Manual.* Allen & Unwin.

Pallant, Julie F. and Alan Tennant. 2007. 'An Introduction to the Rasch Measurement Model: An Example Using the Hospital Anxiety and Depression Scale (HADS)'. *British Journal of Clinical Psychology* 46 (1): 1–18.

Patrick, Donald L., Laurie B. Burke, Chad J. Gwaltney, Nancy Kline Leidy, Mona L. Martin, Elizabeth Molsen and Lena Ring. 2011a. '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 in Health* 14 (8): 967–77.

Patrick, Donald L., Laurie B. Burke, Chad J. Gwaltney, Nancy Kline Leidy, Mona L. Martin, Elizabeth Molsen and Lena Ring. 2011b. '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 in Health* 14 (8): 978–88.

Paulsen, A., S. Overgaard and J. M. Lauritsen. 2012. 'Quality of Data Entry Using Single Entry, Double Entry and Automated Forms Processing—An Example Based on a Study of Patient-Reported Outcomes'. *PloS One* **7** (4): e35087.

Pearl, R. and H. Belcher. 2013. 'Three-Dimensional Assessment of Hand Outcome'. *Annals of the Royal College of Surgeons of England* 95 (6): 421.

Pilkonis, Paul A., Seung W. Choi, Steven P. Reise, Angela M. Stover, William T. Riley, David Cella and PROMIS Cooperative Group. 2011. 'Item Banks for Measuring Emotional Distress from the Patient-Reported Outcomes Measurement Information System (PROMIS®): Depression, Anxiety, and Anger'. *Assessment* 18 (3): 263–83.

Poole, Janet L. 2011. 'Measures of Hand Function: Arthritis Hand Function Test (AHFT), Australian Canadian Osteoarthritis Hand Index (AUSCAN), Cochin Hand Function Scale, Functional Index for Hand Osteoarthritis (FIHOA), Grip Ability Test (GAT), Jebsen Hand Function Test (JHFT), and Michigan Hand Outcomes Questionnaire (MHQ)'. *Arthritis Care & Research* 63 (S11): S189–99.

Pope, C., S. Ziebland and N. Mays. 2000. 'Analysing Qualitative Data'. *BMJ* 320 (7227): 114.

Porter, Michael E. 2010. ‘What is Value in Health Care?’ *The New England Journal of Medicine* 363 (26): 2477–81.

Porter, Michael E. and Elizabeth Olmsted Teisberg. 2006. *Redefining Health Care: Creating Value-Based Competition on Results*. Harvard Business Press.

Pusic, A. L., A. F. Klassen, A. M. Scott, J. A. Klok, P. G. Cordeiro and S. J. Cano. 2009. ‘Development of a New Patient-Reported Outcome Measure for Breast Surgery: The BREAST-Q’. *Plastic and Reconstructive Surgery* 124 (2): 345–53.

Pusic, A. L., V. Lemaine, A. F. Klassen, A. M. Scott and S. J. Cano. 2011. ‘Patient-reported outcome measures in plastic surgery: use and interpretation in evidence-based medicine’. *Plastic and Reconstructive Surgery* 127 (3): 1361–7.

Rank, Benjamin. 1992. ‘Some Early History of Hand Surgery in Australia’. *The Journal of Hand Surgery: British & European Volume* 17 (4): 489–90.

Rasch, G. 1960. *Studies in Mathematical Psychology: I. Probabilistic Models for Some Intelligence and Attainment Tests*. Oxford, England: Nielsen & Lydiche.

Reeve, B. B., R. D. Hays, J. B. Bjorner, K. F. Cook, P. K. Crane, J. A. Teresi, D. Thissen, D. A. Revicki, D. J. Weiss and R. K. Hambleton. 2007. ‘Psychometric Evaluation and Calibration of Health-Related Quality of Life Item Banks: Plans for the Patient-Reported Outcomes Measurement Information System (PROMIS)’. *Medical Care* 45 (5): S22–31.

Reeve, Bryce B., Kathleen W. Wyrwich, Albert W. Wu, Galina Velikova, Caroline B. Terwee, Claire F. Snyder, Carolyn Schwartz, et al. 2013. ‘ISOQOL Recommends Minimum Standards for Patient-Reported Outcome Measures Used in Patient-Centered Outcomes and Comparative Effectiveness Research’. *Quality of Life Research* 22 (8): 1889–1905.

Revicki, Dennis, Ron D. Hays, David Cella and Jeff Sloan 2008. ‘Recommended Methods for Determining Responsiveness and Minimally Important Differences for Patient-Reported Outcomes’. *Journal of Clinical Epidemiology* 61 (2): 102–9.

Reynolds, N. and S. Thirkannad. 2013. ‘The Recall DASH Score—A Novel Research Tool’. *Hand Surgery* 18 (1): 11–14.

Rhee, Peter C., Michelle M. Fischer, Laura S. Rhee, Ha McMillan and Anthony E. Johnson. 2016. ‘Cost Savings and Patient Experiences of a Clinic-Based, Wide-Awake Hand Surgery Program at a Military Medical Center: A Critical Analysis of the First 100 Procedures’. *The Journal of Hand Surgery* 42 (3): e139–e47.

Roche, Nicolas, Helen Reddel, Richard Martin, Guy Brusselle, Alberto Papi, Mike Thomas, Dirjke Postma et al. 2014. ‘Quality standards for real-world research. Focus on

observational database studies of comparative effectiveness.' *Annals of the American Thoracic Society* 11 (2): 99–104.

Rodrigues, J., W. Zhang, B. Scammell, P. Russell, I. Chakrabarti, S. Fullilove, D. Davidson and T. Davis. 2016. 'Validity of the Disabilities of the Arm, Shoulder and Hand Patient-Reported Outcome Measure (DASH) and the Quickdash When Used in Dupuytren's Disease'. *Journal of Hand Surgery (European Volume)* 41 (6): 589–99.

Rolfson, Ola, Roger Salomonsson, Leif E. Dahlberg and Göran Garellick. 2011. 'Internet-Based Follow-Up Questionnaire for Measuring Patient-Reported Outcome After Total Hip Replacement Surgery—Reliability and Response Rate'. *Value in Health* 14 (2): 316–21.

Rose, Matthias and Andrea Bezjak. 2009. 'Logistics of Collecting Patient-Reported Outcomes (PROs) in Clinical Practice: An Overview and Practical Examples'. *Quality of Life Research* 18 (1): 125–36.

Schamber, E. M., S. K. Takemoto, K. E. Chenok and K. J. Bozic. 2013. 'Barriers to Completion of Patient Reported Outcome Measures'. *The Journal of Arthroplasty* 28 (9): 1449–53.

Scholten-Peeters, G. G. M., A. P. Verhagen, G. E. Bekkering, D. A. W. M. van der Windt, L. Barnsley, R. A. B. Oostendorp and E. J. M. Hendriks. 2003. 'Prognostic factors of whiplash-associated disorders: a systematic review of prospective cohort studies'. *Pain* 104 (1): 303–22.

Schoneveld, Karin, Harriet Wittink and Tim Takken. 2009. 'Clinimetric Evaluation of Measurement Tools Used in Hand Therapy to Assess Activity and Participation'. *Journal of Hand Therapy* 22 (3): 221–35; quiz 236.

Scott, A. D., O. Musa, F. Al-Hassani, G. L. Jones, M. B. Hobson and J. G. Miller. 2009. 'Is the Patient Outcomes Of Surgery (POS)-Hand/Arm Questionnaire a Reliable, Valid and Responsive Measurement of Patient-Based Outcomes in Hand and Upper Limb Surgery?' *Journal of Hand Surgery (European Volume)* 34 (4): 530–6.

Scott, Ian A. and Stephen J. Duckett. 2015. 'In Search of Professional Consensus in Defining and Reducing Low-Value Care'. *The Medical Journal of Australia* 203 (4): 179–81.

Shauver, M. J., M. S. Aravind and K. C. Chung. 2011. 'A Qualitative Study of Recovery from Type III-B and III-C Tibial Fractures'. *Annals of Plastic Surgery* 66 (1): 73.

Shauver, Melissa J. and Kevin C. Chung. 2009. 'The Minimal Clinically Important Difference of the Michigan Hand Outcomes Questionnaire'. *The Journal of Hand Surgery* 34 (3): 509–14.

Shauver, Melissa J. and Kevin C. Chung. 2010. ‘A Guide to Qualitative Research in Plastic Surgery’. *Plastic and Reconstructive Surgery* 126 (3): 1089.

Sierakowski, Kyra L., Nicola R. Dean, Andrea L. Pusic, Stefan J. Cano, Philip A. Griffin, Gregory I. Bain, Anne Klassen, Donald Lalonde. 2019. International multiphase mixed methods study protocol to develop a cross-cultural patient-reported outcome and experience measure for hand conditions (HAND-Q)’. *BMJ Open* 9:e025822. Doi:10.1136/bmjopen-2018-025822

Sierakowski, Kyra L., Kathleen A. Evans Sanchez, Rachael A. Damarell, Nicola R. Dean, Philip A. Griffin and Gregory I. Bain. 2018. ‘Measuring Quality of Life and Patient Satisfaction in Hand Conditions: A Systematic Review of Currently Available Patient Reported Outcome Instruments’. *Australasian Journal of Plastic Surgery* 1 (2) 85–99.

Smith, M. V., R. P. Calfee, K. M. Baumgarten, R. H. Brophy and R. W. Wright. 2012. ‘Upper Extremity-Specific Measures of Disability and Outcomes in Orthopaedic Surgery’. *Journal of Bone & Joint Surgery* 94 (3): 277–85.

Snyder, Claire F., Neil K. Aaronson, Ali K. Choucair, Thomas E. Elliott, Joanne Greenhalgh, Michele Y. Halyard, Rachel Hess, Deborah M. Miller, Bryce B. Reeve and Maria Santana. 2012. ‘Implementing Patient-Reported Outcomes Assessment in Clinical Practice: A Review of the Options and Considerations’. *Quality of Life Research* 21 (8): 1305–14.

Snyder, Claire F., Amanda L. Blackford, Aantonio C. Wolff, Michael A. Carducci, Joseph M. Herman and Albert W. Wu. 2013. ‘Feasibility and Value of PatientViewpoint: A Web System for Patient-Reported Outcomes Assessment in Clinical Practice’. *Psycho-Oncology* 22 (4): 895–901.

Snyder, Claire F., Roxanne Jensen, S. Orion Courtin and Albert W. Wu 2009. ‘PatientViewpoint: A Website for Patient-Reported Outcomes Assessment’. *Quality of Life Research* 18 (7): 793–800.

Sollerman, Christer and Arvid Ejeskär. 1995. ‘Sollerman Hand Function Test: A Standardised Method and Its Use in Tetraplegic Patients’. *Scandinavian Journal of Plastic and Reconstructive Surgery and Hand Surgery* 29 (2): 167–76.

Sorensen, Amelia A., Daniel Howard, Weng Hu Tan, Jeffrey Ketchersid and Ryan P. Calfee. 2013. ‘Minimal Clinically Important Differences of 3 Patient-Rated Outcomes Instruments’. *The Journal of Hand Surgery* 38 (4): 641–9.

Spearing, N. M. and L. B. Connelly. 2011. ‘Is Compensation ‘Bad for Health’? A Systematic Meta-Review’. *Injury* 42 (1): 15–24.

Spearman, C. 1904. ‘The Proof and Measurement of Association Between Two Things’. *The American Journal of Psychology* 15 (1): 72–101.

Stamm, T., F. Van der Giesen, C. Thorstensson, E. Steen, F. Birrell, B. Bauernfeind, N. Marshall, B. Prodinger, K. Machold and J. Smolen. 2009. ‘Patient Perspective of Hand Osteoarthritis in Relation to Concepts Covered by Instruments Measuring Functioning: A Qualitative European Multicentre Study’. *Annals of the Rheumatic Diseases* 68 (9): 1453–60.

Stanger, Michael, Christine Morrison, Erdem Yazganoglu and Munjeet Bhalla. 2016. ‘A Study to Assess Patient-Reported Outcome Measures (PROMs) and to Investigate the Practicality of Using PROMs in a Surgical Office’. *British Columbia Medical Journal* 58 (2): 82–9.

Staples J. 2003. ‘Disguise, revelation and copyright: disassembling the South Indian leper’. *Journal of the Royal Anthropological Institute*, 9(2): 295-315

Stratford, P. W. 2001. ‘Development and Initial Validation of the Upper Extremity Functional Index’. *Physiotherapy Canada* 52: 259–67.

Streiner, David., Geoffrey R. Norman, and John Cairney. 2015. *Health Measurement Scales : A Practical Guide to Their Development and Use*. Oxford: Oxford University Press

Szabo, Robert M. 2001. ‘Outcomes Assessment in Hand Surgery: When are they Meaningful?’ *The Journal of Hand Surgery* 26 (6): 993–1002.

Tamai, Susumu. 1982. ‘Twenty Years’ Experience of Limb Replantation—Review of 293 Upper Extremity Replants’. *The Journal of Hand Surgery* 7 (6): 549–56.

Taylor, J. and P. Kersten. 2014. ‘The Patient-Rated Wrist and Hand Evaluation: a systematic review of its validity and reliability’. *New Zealand Journal of Physiotherapy* 42 (3): 117–9.

Tennant, Alan and Philip G. Conaghan. 2007. ‘The Rasch Measurement Model in Rheumatology: What Is It and Why Use It? When Should It Be Applied, and What Should One Look For in a Rasch Paper?’ *Arthritis Care & Research* 57 (8): 1358–62.

Tennant, Alan, Stephen P. McKenna and Peter Hagell. 2004. ‘Application of Rasch Analysis in the Development and Application of Quality of Life Instruments’. *Value in Health* **7**: S22–S26.

Terwee, C. B., C. A. C. Prinsen, A. Chiarotto, M. J. Westerman, D. L. Patrick, J. Alonso, L. M. Bouter, H. C. W. de Vet and L. B. Mokkink. 2018. ‘COSMIN Methodology for Evaluating the Content Validity of Patient-Reported Outcome Measures: A Delphi Study’. *Quality of Life Research* 27 (5): 1159–70.

Thorne, Sally, Sheryl Reimer Kirkham and Janet MacDonald-Emes. 1997. 'Focus on Qualitative Methods. Interpretive Description: A Noncategorical Qualitative Alternative for Developing Nursing Knowledge'. *Research in Nursing & Health* 20 (2): 169–77.

Tong, Allison, Peter Sainsbury and Jonathan Craig. 2007. 'Consolidated Criteria for Reporting Qualitative Research (COREQ): A 32-Item Checklist for Interviews and Focus Groups'. *International Journal for Quality in Health Care* 19 (6): 349–57.

Tsangaris, E., K. W. Y. W. Riff, T. Goodacre, C. R. Forrest, M. Dreise, J. Sykes, T. de Chalain, K. Harman, A. O'Mahony and A. L. Pusic. 2017. 'Establishing Content Validity of the CLEFT-Q: A New Patient-Reported Outcome Instrument for Cleft Lip/Palate'. *Plastic and Reconstructive Surgery—Global Open* 5 (4): e1305.

U.S. Department of Health and Human Services FDA Center for Drug Evaluation and Research; U.S. Department of Health and Human Services FDA Center for Biologics Evaluation and Research; U.S. Department of Health and Human Services FDA Center for Devices and Radiological Health. 2006. 'Guidance for Industry: Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims: Draft Guidance'. *Health and Quality of Life Outcomes* 4: 1–20.

Valderas, Jose M., Jordi Alonso and Gordon H. Guyatt. 2008. 'Measuring Patient-Reported Outcomes: Moving from Clinical Trials into Clinical Practice'. *The Medical Journal of Australia* 189 (2): 93–4.

Valderas, J. M., R. Fitzpatrick and M. Roland. 2012. 'Using Health Status to Measure NHS Performance: Another Step into the Dark for the Health Reform in England'. *BMJ Quality & Safety* 21 (4): 352–3.

Valderas, J. M., A. Kotzeva, M. Espallargues, G. Guyatt, C. E. Ferrans, M. Y. Halyard, D. A. Revicki, T. Symonds, A. Parada and J. Alonso. 2008. 'The Impact of Measuring Patient-Reported Outcomes in Clinical Practice: A Systematic Review of the Literature'. *Quality of Life Research* 17 (2): 179–93.

VanDenKerkhof, E. G., D. H. Goldstein, W. C. Blaine and M. J. Rimmer. 2005. 'A Comparison of Paper with Electronic Patient-Completed Questionnaires in a Preoperative Clinic'. *Anesthesia & Analgesia* 101 (4): 1075–80.

van der Giesen, F. J., R. G. H. H. Nelissen, W. J. van Lankveld, C. Kremers-Selten, A. J. Peeters, E. B. Stern, S. le Cessie and T. P. M. Vliet Vlieland 2010. 'Swan Neck Deformities in Rheumatoid Arthritis: A Qualitative Study on the Patients' Perspectives on Hand Function Problems and Finger Splints'. *Musculoskeletal Care* 8 (4): 179–88.

Van Someren, M., Y. Barnard and J. Sandberg (1994). *'The think aloud method: a practical guide to modelling cognitive processes'*. London: Academic Press.

Van Vliet, M. M., J. A. Maradey, K. A. Homa and C. L. Kerrigan. 2013. 'The Usefulness of Patient-Reported Measures for Clinical Practice'. *Plastic and Reconstructive Surgery* 132 (1): 105–12.

Verdan, C. E. 1960. 'Primary Repair of Flexor Tendons'. *The Journal of Bone & Joint Surgery* 42 (4): 647–57.

Waljee, J. F., H. M. Kim, P. B. Burns and K. C. Chung. 2011. 'Development of a Brief, 12-Item Version of the Michigan Hand Questionnaire'. *Plastic & Reconstructive Surgery* 128 (1): 208–20.

Ware Jr, J. E. and C. D. Sherbourne 1992. 'The MOS 36-Item Short-Form Health Survey (SF-36): I. Conceptual Framework and Item Selection'. *Medical Care* 30 (6): 473–83.

Ware, J. E. 1993. 'Measuring Patients' Views: The Optimum Outcome Measure'. *BMJ* 306 (6890): 1429–30.

Warren A.G. 1972. 'Are deformities stigmatising? A surgeon's approach. *Leprosy review* 43: 74-82

Watchmaker, J. D. and G. P. Watchmaker. 2018. 'Independent Variables Affecting Outcome of Carpal Tunnel Release Surgery'. *HAND* 13 (3): 285-291.

Wehrli, Martina, Stefani Hensler, Stephan Schindele, Daniel B. Herren and Miriam Marks. 2016. 'Measurement Properties of the Brief Michigan Hand Outcomes Questionnaire in Patients with Dupuytren Contracture'. *The Journal of Hand Surgery* 41 (9): 896–902.

Weinberger, Morris, Eugene Z. Oddone, Gregory P. Samsa and Pamela B. Landsman. 1996. 'Are Health-Related Quality-Of-Life Measures Affected by the Mode of Administration?' *Journal of Clinical Epidemiology* 49 (2): 135–40.

Welford, Claire, Kathy Murphy and Dympna Casey (2011). 'Demystifying Nursing Research Terminology. Part 1'. *Nurse Researcher* 18 (4): 38–43.

Wellington, Beverley. 2010. 'Quality of Life Issues for Patients Following Traumatic Brachial Plexus Injury–Part 2 Research Project'. *International Journal of Orthopaedic and Trauma Nursing* 14 (1): 5–11.

WHO (World Health Organization). 1980. *International Classification of Impairments, Disabilities, and Handicaps: A Manual of Classification Relating to the Consequences of Disease—Published in Accordance with Resolution WHA29. 35 of the Twenty-Ninth World Health Assembly, May 1976*. Geneva: WHO.

WHO (World Health Organization). 2001. *The International Classification of Functioning, Disability and Health (ICF)*. Geneva, WHO.

Wild, Diane, Alyson Grove, Mona Martin, Sonya Eremenco, Sandra McElroy, Aneesa Verjee-Lorenz and Pennifer Erikson. 2005. '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 in Health* 8 (2): 94–104.

Williams, Dan 2012. 'The myClinicalOutcomes Website: Providing Real-Time, Patient-Level PROMs Data'. *The Bulletin of the Royal College of Surgeons of England* 94 (1): 20–1.

Wong, Joyce Y. P. 2008. 'Time Off Work in Hand Injury Patients'. *The Journal of Hand Surgery* 33 (5): 718–25.

Wong Riff, Karen W. Y., Elena Tsangaris, Timothy E. Goodacre, Christopher R. Forrest, Jessica Lawson, Andrea L. Pusic and Anne F. Klassen. 2018. 'What Matters to Patients with Cleft Lip and/or Palate: An International Qualitative Study Informing the Development of the CLEFT-Q'. *The Cleft Palate-Craniofacial Journal* 55 (3): 442–50.

Wong Riff, Karen W. Y., Elena Tsangaris, Tim Goodacre, Christopher R. Forrest, Andrea L. Pusic, Stefan J. Cano and Anne F. Klassen. 2017. 'International Multiphase Mixed Methods Study Protocol to Develop a Cross-Cultural Patient-Reported Outcome Instrument for Children and Young Adults with Cleft Lip and/or Palate (CLEFT-Q)'. *BMJ Open* 7 (1): e015467.

Wright, B. D. and J. M. Linacre. 1989. 'Observations Are Always Ordinal; Measurements, However, Must Be Interval'. *Archives of Physical Medicine and Rehabilitation* 70 (12): 857.

Wright, Benjamin D. and Geofferey N. Masters. 1982. *Rating Scale Analysis: Rasch Measurement*. Chicago: MESA.

Wright, James G. and Alvan R. Feinsten. 1992. 'A Comparative Contrast of Clinimetric and Psychometric Methods for Constructing Indexes and Rating Scales'. *Journal of Clinical Epidemiology* 45 (11): 1201–18.

Yafef, Mark, Nitin Goyal, Danial Kokmeyer and Gregory A. Merrel. 2015. 'The Use of an iPad To Collect Patient-Reported Functional Outcome Measures in Hand Surgery'. *Hand* 10 (3): 522–8.

Appendices

Appendix A.1: *Australasian Journal of Plastic Surgery* Paper

Removed due to copyright restriction.

Appendix A.2: Search Strategy

The table shows the detailed search strategy used for MEDLINE. The search was run on 19 June 2017. OVID MEDLINE(R) includes Epub Ahead of Print, In-Process and Other Non-Indexed Citations, Ovid MEDLINE(R) Daily, Ovid MEDLINE and Versions(R).

#	Searches	Results
1	hand/ or fingers/ or thumb/ or metacarpus/ or wrist/ or carpal bones/ or carpometacarpal joints/ or metacarpophalangeal joint/ or wrist joint/	91334
2	hand injuries/ or finger injuries/ or wrist injuries/	22396
3	exp Hand Deformities/	6816
4	(finger? or thumb? or wrist* or metacarp* or carpus* or carpal or carpometacarp* or metacarpophalangeal or phalangeal).tw,kw.	130599
5	hand.ti,kw,jw. or hand.ab. /freq=2	104428
6	or/1-5	249362
7	orthopedics/ or surgery, plastic/	44481
8	surgical procedures, operative/ or orthopedic procedures/ or reconstructive surgical procedures/	116070
9	Surgery, Plastic/	25292
10	Postoperative Complications/ or Postoperative Care/ or Postoperative Period/	410172
11	(surg* or postsurg* or operat* or postoperati* or orthopedic* or orthopaedic*).tw,kw.	2428554
12	or/7-11	2598679
13	“surveys and questionnaires”/ or health care surveys/ or health surveys/ or health status indicators/ or “Severity of Illness Index”/ or sickness impact profile/ or self report/	670834
14	Patient Outcome Assessment/	3254
15	Psychometrics/	65555
16	(questionnaire* or survey* or instrument? or tool? or score? or scoring or scale? or subscale? or index or indices or psychometric* or domain* or multidomain* or dimension* or multidimension* or sphere* or criterion).tw,kw.	3816409
17	Patient reported outcome measures/	451
18	(patient reported outcome* or PRO or PROs or PROM or PROMs).tw,kw.	170013
19	((patient or self) adj (rated or report*)).tw,kw.	152408
20	(standard* adj3 (question* or response*)).tw,kw.	21505
21	(item adj (bank or response or generation or reduction)).tw,kw.	3770
22	((outcome* or evaluation*) adj measure*).tw,kw.	197947
23	(patient evaluation measure* or PEM).tw,kw.	2599

24	(michigan hand questionnaire or MHQ or briefMHQ).tw,kw.	250
25	(DASH or qDASH or QuickDASH or "Disab* of the arm shoulder hand").tw,kw.	5227
26	(patient rated wrist evaluation or patient rated hand wrist evaluation or PRWE or PRHWE).tw,kw.	253
27	(Patient-Reported Outcomes Measurement Information System or PROMIS).tw,kw.	888
28	POS hand-arm.tw,kw.	2
29	“patient outcomes of surgery hand*”.tw,kw.	1
30	Duruoz hand index.tw,kw.	31
31	(“Modern Activity Subjective Survey” or MASS07).tw,kw.	3
32	or/13-31	4332843
33	Validation Studies/ or “Reproducibility of Results”/	399215
34	(Valid* or reliability or responsiveness or calibrat* or reproducib* or pretest* or pre-test* or test-retest or internal consistency or rasch).tw,kw.	941735
35	or/33-34	1145556
36	health status/ or "quality of life"/	212837
37	(“quality of life” or QoL or QL or HRQoL or HRQL or health status or quality adjusted life year* or functional status or wellbeing or well-being).tw,kw.	333029
38	“Activities of Daily Living”/ or “Recovery of Function”/	98676
39	(“activities of daily living” or ADL or AODL or “daily living” or mobilit* or dexter* or (activ* adj1 (limit* or ceas* or reduc* or stop*)) or disabilit* or impair* or function*).tw,kw.	3677974
40	Self Care/	30142
41	(burden or self efficac* or self care or self manag* or autonom* or mastery or independence).tw,kw.	340068
42	Return to Work/	1362
43	("return to work" or work capacity).tw,kw.	11934
44	disability evaluation/ or work capacity evaluation/	47297
45	Pain/ or Pain Measurement/ or Pain Perception/	179314
46	(Side effect* or symptom* or pain* or dysfunction* or sensation*).tw,kw.	1910709
47	Patient Satisfaction/	72230
48	Patient participation/	22280
49	(patient* adj (satisf* or participat*)).tw,kw.	42826
50	(unmet need* or care need* or healthcare need* or patient need*).tw,kw.	25722
51	Esthetics/ or Body image/ or Self concept/	73773
52	(aesthetic* or esthetic* or body image or self image or self concept or self esteem or appearance).tw,kw.	269852
53	or/36-52	5889195
54	6 and 12 and 32 and 35 and 53	810
55	limit 54 to English language	753

Appendix B.1: Ethics Approval 404.16

Content removed for privacy reasons.

Appendix B.2: Disabilities of the Arm, Shoulder and Hand

Removed due to copyright restriction.

Appendix B.3: Patient-Rated Wrist/Hand Evaluation

Removed due to copyright restriction.

Appendix B.4: Michigan Hand Outcomes Questionnaire

Removed due to copyright restriction.

Appendix C.1: *BMJ Open* HAND-Q Protocol

Removed due to copyright restriction.

Appendix C.2: Australia Ethics Phase I

Content removed for privacy reasons.

Appendix C.3: Canada Ethics Phase I

Content removed for privacy reasons.

Appendix C.4: Individual Participant Demographics for Qualitative Study

Participant reference	Country	Gender	Age	Primary condition
1	Australia	Female	38	Other—distal radioulnar joint instability
2	Australia	Male	62	Dupuytren's contracture
3	Australia	Female	65	Osteoarthritis
4	Australia	Female	78	Osteoarthritis
5	Australia	Male	46	Dupuytren's contracture
6	Australia	Female	62	Trauma—fracture proximal phalanx
7	Australia	Male	71	Dupuytren's contracture
8	Australia	Male	47	Post-trauma—volar plate repair
9	Australia	Male	62	Dupuytren's contracture
10	Australia	Male	78	Dupuytren's contracture
11	Australia	Female	64	Osteoarthritis
12	Australia	Female	57	Dupuytren's contracture
13	Australia	Female	73	Dupuytren's contracture
14	Australia	Male	72	Trigger finger
15	Australia	Male	62	Dupuytren's contracture
16	Australia	Female	60	CTS
17	Australia	Male	64	CTS
18	Australia	Female	54	Trigger finger
19	Australia	Female	57	CTS
20	Australia	Female	59	Osteoarthritis
21	Australia	Male	71	Dupuytren's contracture
22	Australia	Female	75	Osteoarthritis
23	Australia	Female	58	CTS
24	Australia	Female	64	Osteoarthritis
25	Australia	Female	70	Osteoarthritis
26	Australia	Female	57	Osteoarthritis
27	Australia	Female	54	Trigger finger
28	Australia	Male	66	Dupuytren's contracture
29	Australia	Male	54	CTS
30	Australia	Male	76	Trigger finger
31	Australia	Male	59	Dupuytren's contracture
32	Australia	Female	67	Other—mucous cyst
33	Australia	Male	66	Trauma—tendon reconstruction
34	Australia	Female	68	CTS
35	Australia	Male	64	Post-trauma—tenolysis

36	Australia	Male	71	Post-trauma—tendon transfer
37	Australia	Female	63	CTS
38	Australia	Female	54	CTS
39	Australia	Female	73	Trauma—extensor tendons and nerve injury
40	Australia	Male	76	Rheumatoid arthritis
41	Canada	Female	54	CTS
42	Canada	Female	75	CTS
43	Canada	Male	85	Dupuytren’s contracture
44	Canada	Male	83	CTS
45	Canada	Male	70	Trigger finger
46	Canada	Male	70	CTS
47	Canada	Female	86	CTS
48	Canada	Female	58	CTS
49	Canada	Male	61	CTS
50	Canada	Female	68	CTS
51	Canada	Male	77	Dupuytren’s contracture
52	Canada	Female	83	Trigger finger
53	Canada	Male	28	Post-trauma—tendon transfer
54	Canada	Male	52	Dupuytren’s contracture
55	Canada	Male	59	Trigger finger
56	Canada	Male	53	CTS
57	Canada	Female	86	Trigger finger
58	Canada	Female	66	CTS
59	Canada	Female	72	Other—tendon transfer
60	Canada	Female	77	Other—DeQuervain’s tenosynovitis
61	Canada	Female	55	CTS
62	Canada	Female	62	CTS

Appendix D.1: United States Ethics Cognitive Interviews

Content removed for privacy reasons.

Appendix D.2: Individual Participant Demographics for Cognitive Interviews

Participant reference	Qualitative study	Country	Gender	Age	Primary condition
2	YES	Australia	Male	62	Dupuytren's contracture
6	YES	Australia	Female	62	Trauma—fracture proximal phalanx
8	YES	Australia	Male	47	Post-trauma—volar plate repair
11	YES	Australia	Female	64	Osteoarthritis
24	YES	Australia	Female	64	Osteoarthritis
26	YES	Australia	Female	57	Osteoarthritis
27	YES	Australia	Female	54	Trigger finger
39	YES	Australia	Female	73	Trauma—extensor tendons and nerve injury
40	YES	Australia	Male	76	Rheumatoid arthritis
48	YES	Canada	Female	58	CTS
49	YES	Canada	Male	61	CTS
50	YES	Canada	Female	68	CTS
51	YES	Canada	Male	77	Dupuytren's contracture
58	YES	Canada	Female	66	CTS
59	YES	Canada	Female	72	Other—tendon transfer
61	YES	Canada	Female	55	CTS
63	NO	US	Female	61	Trigger finger
64	NO	US	Male	70	CTS
65	NO	US	Male	32	Trauma
66	NO	US	Female	35	CTS

Appendix D.3: HAND-Q Field-Test Scales

HAND-Q©

Field-Test Version

Post-Surgery Module

Assigned Study ID: _____

Hand Treatment Centre/Hospital: _____

Today's date: __ __ __ __ __ __

day month year

THANK YOU FOR PARTICIPATING IN THE DEVELOPMENT OF THE HAND-Q: A QUESTIONNAIRE TO MEASURE OUTCOMES FROM THE PATIENT'S VIEW.

Please complete the following questions as honestly as possible. If you feel that some of the questions make you uncomfortable, please leave them blank.

1. **Where do you LIVE?**
   ☐ Australia ☐ Canada ☐ USA ☐ UK
2. **What is your AGE?** _____ years
3. **What is your GENDER?**
   ☐ Male ☐ Female ☐ Transgender ☐ Other ☐ Prefer not to disclose
4. **What is your OCCUPATION?**
   ☐ Office work (e.g., administrative, secretarial) ☐ Student
   ☐ Professional (e.g., teacher, lawyer, nurse) ☐ Home duties
   ☐ Manual labour (e.g., cleaner, gardener) ☐ Retired
   ☐ Trades (e.g., plumber, electrician, mechanic) ☐ Unemployed
   ☐ Unable to work ☐ Volunteer
   ☐ Other, please specify: _____ ☐ Carer
5. **Is your hand problem covered by work-related INSURANCE (Work Cover/Workers Compensation)?**
   ☐ Yes ☐ No ☐ Unsure
6. **What is the highest level of EDUCATION you have completed?**
   ☐ Primary School ☐ High School
   ☐ Further education (college, university or similar)
   ☐ Other, please specify: _____
7. **What <u>side</u> is your DOMINANT hand (i.e., are you left handed or right handed)?**
   ☐ Left ☐ Right ☐ Both
8. **What <u>side</u> is your HAND PROBLEM?**
   ☐ Left ☐ Right ☐ Both
9. **Is your hand currently in a plaster or splint/brace?**
   ☐ Yes ☐ No
10. **(a) What is your HAND PROBLEM? Please circle all that apply by rating how severe the problem is.**

My hand problem is	Mild	Moderate	Severe
A. Carpal tunnel	**1**	**2**	**3**
B. Dupuytren's contracture	**1**	**2**	**3**
C. Trigger finger	**1**	**2**	**3**
D. Osteoarthritis	**1**	**2**	**3**
E. Rheumatoid arthritis	**1**	**2**	**3**
F. Soft tissue injury (please describe) _____	**1**	**2**	**3**
G. Fracture (please describe) _____	**1**	**2**	**3**
H. Other hand problem (please describe) _____	**1**	**2**	**3**
I. Not sure	**1**	**2**	**3**

(b) Which hand problem(s) are you here about today?_____

(c) Approximately how long have you had your current hand problem(s)?_____ weeks OR _____ months OR _____ years

11. How MANY times have you come to see the surgeon and/or hand therapist in the past 3 months?

☐ This is my first visit ☐ 1 to 4 visits ☐ 5 to 8 visits ☐ More than 8 visits

12. (a) Have you had SURGERY for your current hand problem(s)?

☐ Yes ☐ No (Please continue on the next page)

(b) Approximately HOW LONG ago was your most recent hand surgery?_____ weeks OR **_____** months OR **_____** years

(c) Do you need MORE SURGERY for your hand problem?

☐ Yes ☐ No ☐ Unsure

HOW DO YOUR HANDS LOOK? Please answer thinking of how your hands look NOW.

NOTE: If your hands look different from each other, answer each question thinking about the hand you are least satisfied with.

How dissatisfied or satisfied are you with:

	Very dissatisfied	Somewhat dissatisfied	Somewhat satisfied	Very satisfied
1. How your hands look from far away?	1	2	3	4
2. How the palms of your hands look?	1	2	3	4
3. How straight your fingers and thumbs look?	1	2	3	4
4. The size of your fingers and thumbs?	1	2	3	4
5. The shape of your fingers and thumbs?	1	2	3	4
6. How your fingers and thumbs line up with each other?	1	2	3	4
7. How well your fingers match each other?	1	2	3	4
8. How your fingernails look?	1	2	3	4
9. How your knuckles look?	1	2	3	4
10. The size of your knuckles?	1	2	3	4
11. The shape of your knuckles?	1	2	3	4
12. How your hands look when you rest your palms on a table?	1	2	3	4
13. How your hands look when you wave at someone?	1	2	3	4
14. How your hands look when you hold a glass?	1	2	3	4
15. How your hands look compared with other people's hands?	1	2	3	4
16. How normal your hands look?	1	2	3	4

	Very dissatisfied	Somewhat dissatisfied	Somewhat satisfied	Very satisfied
17. How masculine or feminine your hands look?	**1**	**2**	**3**	**4**
18. How well proportioned your hands look (i.e., all parts look the right size and shape)?	**1**	**2**	**3**	**4**
19. How the veins on the back of your hands look?	**1**	**2**	**3**	**4**
20. How noticeable the veins on the back of your hands are?	**1**	**2**	**3**	**4**
21. How the tendons on the back of your hands look? Note: Tendons are the cords on the back of your hands that straighten out your fingers.	**1**	**2**	**3**	**4**
22. How visible the tendons on the back of your hands are?	**1**	**2**	**3**	**4**
23. How the skin on your hands looks?	**1**	**2**	**3**	**4**
24. How taut (i.e., firm) the skin on the back of your hands looks?	**1**	**2**	**3**	**4**
25. How smooth the skin on the back of your hands looks?	**1**	**2**	**3**	**4**
26. How blemish-free the skin on the back of your hands looks?	**1**	**2**	**3**	**4**
27. How youthful your hands look?	**1**	**2**	**3**	**4**
28. The age your hands look?	**1**	**2**	**3**	**4**
29. How your hands look from close up?	**1**	**2**	**3**	**4**
30. How your hands look overall?	**1**	**2**	**3**	**4**

HOW DIFFICULT IS IT TO USE YOUR HANDS? Please answer thinking of the PAST WEEK.

NOTE: If your hand problem affects one hand more than the other, answer each question for the hand that is worse. If you use gadgets or devices to help perform tasks, please answer according to how your hand(s) work without using them.

With your hand problem in mind, how difficult would these tasks have been:

	Not at all difficult	A little difficult	Moderately difficult	Extremely difficult
1. Placing your palms flat on a table?	1	2	3	4
2. Making a fist with your hand(s)?	1	2	3	4
3. Shaking someone's hand?	1	2	3	4
4. Clapping your hands?	1	2	3	4
5. Holding a phone to your ear?	1	2	3	4
6. Holding a book to read?	1	2	3	4
7. Holding a bag of groceries?	1	2	3	4
8. Plugging a cord into a socket?	1	2	3	4
9. Using a TV remote control?	1	2	3	4
10. Gripping handles (e.g., tennis racket, golf club, broom)?	1	2	3	4
11. Picking up a coin?	1	2	3	4
12. Taking things out of a pocket?	1	2	3	4
13. Turning a door knob?	1	2	3	4
14. Turning a key in a lock?	1	2	3	4
15. Turning a tap (i.e., faucet)?	1	2	3	4
16. Writing with a pen or pencil?	1	2	3	4
17. Typing?	1	2	3	4
18. Opening a jar?	1	2	3	4

	Not at all difficult	A little difficult	Moderately difficult	Extremely difficult
19. Opening a small lid (e.g., water or other beverage bottle)?	**1**	**2**	**3**	**4**
20. Washing the dishes?	**1**	**2**	**3**	**4**
21. Preparing food (e.g., peeling, cutting)?	**1**	**2**	**3**	**4**
22. Eating with cutlery (e.g., fork, spoon, knife)?	**1**	**2**	**3**	**4**
23. Eating with your hand(s)?	**1**	**2**	**3**	**4**
24. Holding a glass?	**1**	**2**	**3**	**4**
25. Scratching an itch?	**1**	**2**	**3**	**4**
26. Washing your hands?	**1**	**2**	**3**	**4**
27. Brushing your teeth?	**1**	**2**	**3**	**4**
28. Clipping your fingernails?	**1**	**2**	**3**	**4**
29. Buttoning a shirt or coat?	**1**	**2**	**3**	**4**
30. Doing up a zipper?	**1**	**2**	**3**	**4**
31. Tying shoelaces?	**1**	**2**	**3**	**4**
32. Cleaning (e.g., wiping) yourself after a bowel movement?	**1**	**2**	**3**	**4**
33. Putting on or taking off clothes?	**1**	**2**	**3**	**4**
34. Showering?	**1**	**2**	**3**	**4**
35. Personal grooming (e.g., shaving, putting on make-up)?	**1**	**2**	**3**	**4**

HOW DO YOUR HANDS FEEL (SYMPTOMS)? Please answer thinking of the PAST WEEK.

NOTE: If the symptoms you feel in each hand differ, answer each question thinking about the hand that bothers you the most.

Please rate the severity of each of these symptoms:

	None	Mild	Moderate	Severe
1. Hands feeling itchy?	1	2	3	4
2. Hands feeling numb (i.e., less feeling)?	1	2	3	4
3. Tingling in your hands (i.e., pins and needles feeling)?	1	2	3	4
4. Hands feeling sensitive (i.e., too much feeling)?	1	2	3	4
5. Hands feeling stiff?	1	2	3	4
6. Swelling or puffiness?	1	2	3	4
7. Cramping in your hands?	1	2	3	4
8. Hands feeling hotter or colder than normal?	1	2	3	4
9. Hands feeling weak (i.e., lack of strength)?	1	2	3	4
10. Hands feeling achy?	1	2	3	4
11. Throbbing pain in your hands?	1	2	3	4
12. Stinging or burning pain in your hands?	1	2	3	4
13. Pain when you use your hands?	1	2	3	4
14. Pain when your hands are at rest?	1	2	3	4
15. Pain when your hands are touched?	1	2	3	4
16. Pain when the weather changes?	1	2	3	4
17. Hands feeling dry?	1	2	3	4
18. Hands feeling moist?	1	2	3	4

	None	Mild	Moderate	Severe
19. Clumsiness (e.g., dropping or spilling things)?	**1**	**2**	**3**	**4**
20. Hand tremors (i.e., shaking)?	**1**	**2**	**3**	**4**
21. Hand symptoms (e.g., pain, numbness) disturbing your sleep?	**1**	**2**	**3**	**4**
22. Hands that are worse in cold weather?	**1**	**2**	**3**	**4**

DOES YOUR HAND PROBLEM AFFECT HOW YOU FEEL? Please answer thinking of the PAST WEEK.

With your <u>hand problem</u> in mind, how often have you felt:

	Never	Sometimes	Often	Always
1. Frustrated?	**1**	**2**	**3**	**4**
2. Upset?	**1**	**2**	**3**	**4**
3. Worried?	**1**	**2**	**3**	**4**
4. Concerned?	**1**	**2**	**3**	**4**
5. Sorry for yourself?	**1**	**2**	**3**	**4**
6. Depressed?	**1**	**2**	**3**	**4**
7. Irritated?	**1**	**2**	**3**	**4**
8. Angry?	**1**	**2**	**3**	**4**
9. Embarrassed?	**1**	**2**	**3**	**4**
10. Self-conscious?	**1**	**2**	**3**	**4**
11. Anxious?	**1**	**2**	**3**	**4**
12. Fed-up?	**1**	**2**	**3**	**4**
13. Overwhelmed?	**1**	**2**	**3**	**4**
14. Annoyed?	**1**	**2**	**3**	**4**
15. Stressed?	**1**	**2**	**3**	**4**
16. Unattractive?	**1**	**2**	**3**	**4**
17. Useless?	**1**	**2**	**3**	**4**
18. Hopeless?	**1**	**2**	**3**	**4**
19. Desperate?	**1**	**2**	**3**	**4**

DOES YOUR HAND PROBLEM AFFECT YOUR QUALITY OF LIFE? Please answer thinking of the PAST WEEK.

How much has your <u>hand problem interfered</u> with the following:

	Not at all	A little bit	Quite a bit	Very much
1. Being physically active?	**1**	**2**	**3**	**4**
2. Taking a bath or shower?	**1**	**2**	**3**	**4**
3. Being able to relax?	**1**	**2**	**3**	**4**
4. Sleeping at night?	**1**	**2**	**3**	**4**
5. Doing activities you enjoy?	**1**	**2**	**3**	**4**
6. Your emotional wellbeing?	**1**	**2**	**3**	**4**
7. Your mood?	**1**	**2**	**3**	**4**
8. Your ability to enjoy life?	**1**	**2**	**3**	**4**
9. Your social life?	**1**	**2**	**3**	**4**
10. Your close relationships?	**1**	**2**	**3**	**4**
11. Your ability to be independent?	**1**	**2**	**3**	**4**

DOES YOUR HAND PROBLEM AFFECT YOUR SOCIAL LIFE? Please answer thinking of the PAST WEEK.

With your <u>hand problem</u> in mind, how much do you disagree or agree with each statement:

	Definitely disagree	Somewhat disagree	Somewhat agree	Definitely agree
1. I stayed at home more than I would have liked.	1	2	3	4
2. I found it hard to get out and meet people.	1	2	3	4
3. I felt embarrassed about my hands.	1	2	3	4
4. I cut down on social activities I enjoy.	1	2	3	4
5. I saw friends less than I would have liked.	1	2	3	4
6. I missed out on social events.	1	2	3	4
7. I felt like I was a burden to family or friends.	1	2	3	4
8. I felt isolated from family or friends.	1	2	3	4
9. I felt that people did not understand what I go through with my hand problem.	1	2	3	4
10. I covered up or hid my hand(s).	1	2	3	4
11. My hand problem interfered with my ability to enjoy life.	1	2	3	4
12. I felt self-conscious about my hands around other people.	1	2	3	4
13. I avoided greetings (e.g., waving or shaking hands).	1	2	3	4

How many nights did your hand(s) interfere with your SLEEP in the past week?

☐ **Never** ☐ **Sometimes (1–2 times a week)**

☐ **Often (3–4 times a week)** ☐ **Very Often (5–7 times a week)**

If you answered "Never" then please continue on the next page.

DOES YOUR HAND PROBLEM AFFECT YOUR SLEEP? Please answer thinking of the PAST WEEK.

With your hand problem in mind, how often have you:

	Never	Sometimes	Often	Always
1. Had trouble falling asleep?	**1**	**2**	**3**	**4**
2. Had trouble staying asleep?	**1**	**2**	**3**	**4**
3. Had trouble finding a comfortable position to sleep in?	**1**	**2**	**3**	**4**
4. Woken up at night?	**1**	**2**	**3**	**4**
5. Not had enough sleep?	**1**	**2**	**3**	**4**
6. Taken medication to help you sleep?	**1**	**2**	**3**	**4**
7. Had symptoms (e.g., pain, numbness) from your hands disturb your sleep?	**1**	**2**	**3**	**4**
8. Felt tired during the day?	**1**	**2**	**3**	**4**

Do you have a long-term hand problem?

☐ **Yes** ☐ **No**

By long-term, we mean a hand problem that has lasted at least 6 months and is still a problem for you. If yes, please complete the questions below. If no, please continue on the next page.

WHAT IS IT LIKE TO LIVE WITH YOUR HAND PROBLEM?

How much do you disagree or agree with each statement:

	Definitely disagree	Somewhat disagree	Somewhat agree	Definitely agree
1. I have learned to live with my hand problem.	1	2	3	4
2. My hand problem has become part of my life.	1	2	3	4
3. I have accepted my hand problem.	1	2	3	4
4. I get on with my life as best I can.	1	2	3	4
5. If my hand problem does not improve, I will be okay.	1	2	3	4
6. I have a positive attitude towards my hand problem.	1	2	3	4
7. I am fine with my hand problem.	1	2	3	4

Does your hand problem affect your sex life?

☐ **Yes** ☐ **No** ☐ **I prefer not to answer**

If yes, please complete the questions below.

If no, or you prefer not to answer, please continue on the next page.

DOES YOUR HAND PROBLEM AFFECT YOUR SEX LIFE?

With your <u>hand problem</u> in mind, how much are you <u>bothered</u> by the following:

	Not at all bothered	A little bothered	Moderately bothered	Extremely bothered
1. How your hands look?	**1**	**2**	**3**	**4**
2. Being able to use your hands in tender ways (e.g., touch, hold)?	**1**	**2**	**3**	**4**
3. Limitations in hand function that can interfere with sexual activity (e.g., grip, strength)?	**1**	**2**	**3**	**4**
4. Symptoms you feel in your hands that can interfere with sexual activity (e.g., pain, numbness, tingling)?	**1**	**2**	**3**	**4**
5. Being aware of your hands during sexual activity?	**1**	**2**	**3**	**4**
6. Your hand problem affecting how much you enjoy sexual activity?	**1**	**2**	**3**	**4**
7. Your hand problem being a distraction during sexual activity?	**1**	**2**	**3**	**4**
8. Your hand problem interfering with your ability to give pleasure?	**1**	**2**	**3**	**4**
9. Your partner seeing your hands during sexual activity?	**1**	**2**	**3**	**4**

Did you work at a job in the past 3 months?

☐ **Yes** ☐ **No**

If yes, please complete the questions below. If no, please continue on the next page.

How much did you work? ☐ **Full-time** ☐ **Part-time** ☐ **Other**

Are you self-employed? ☐ **Yes** ☐ **No**

How important are your hands to doing your job?

☐ **Not important** ☐ **A little important**

☐ **Moderately important** ☐ **Very important**

DOES YOUR HAND PROBLEM AFFECT YOUR WORK LIFE? Please answer thinking of when you were last working.

With your <u>hand problem</u> in mind, how much do you disagree or agree with each statement:

	Definitely disagree	Somewhat disagree	Somewhat agree	Definitely agree
1. I was conscious of my hand(s) at work.	1	2	3	4
2. I worried about missing work.	1	2	3	4
3. I had to reduce the amount of work I do in a day.	1	2	3	4
4. It was hard for me to keep up with my work.	1	2	3	4
5. I had trouble performing my job.	1	2	3	4
6. I had to change how I do my job.	1	2	3	4
7. My work made my hand(s) worse.	1	2	3	4
8. I worried about losing my job.	1	2	3	4
9. The quality of my work has gone down.	1	2	3	4
10. I thought about quitting work.	1	2	3	4
11. I was not able to do my job.	1	2	3	4

Did you have hand surgery within the last 3 months? If yes, please complete the questions below. If no, please continue onto page 20.

Which type of anaesthesia did you have for your most recent hand surgery?

☐ **General anaesthesia where you are completely asleep**

☐ **Sedation where you are partly asleep**

☐ **Local anaesthesia where you are wide awake but the area has been numbed**

☐ **I don't know**

HOW WAS THE ANAESTHESIA YOU HAD FOR YOUR HAND SURGERY?

With the anaesthesia you had in mind, how bothered were you by the following:

	Not at all bothered	A little bothered	Moderately bothered	Extremely bothered
1. Time spent to prepare for the anaesthesia (e.g., tests or appointments, forms, travel)?	1	2	3	4
2. Any preoperative anxiety about having an anaesthetic?	1	2	3	4
3. Thoughts of embarrassing yourself during anaesthesia (e.g., saying something inappropriate)?	1	2	3	4
4. The amount of anaesthetic you might be given (e.g., if the operation took longer than normal)?	1	2	3	4
5. The chance that something could go wrong during the anaesthesia?	1	2	3	4
6. That you might feel pain during surgery (i.e., if the anaesthetic is not effective)?	1	2	3	4
7. The affect the anaesthesia might have on your health?	1	2	3	4
8. The number of needles you had in total (i.e., for blood tests and anaesthetic needles you felt during surgery)?	1	2	3	4
9. Any pain caused by the needle(s) used to give you the anaesthetic?	1	2	3	4
10. Any discomfort caused by the tight armband used during surgery (i.e., tourniquet)?	1	2	3	4
11. How long it took to recover from the anaesthetic?	1	2	3	4
12. How long you had to wait in total at the hospital or clinic on the day of your surgery?	1	2	3	4

	Not at all bothered	A little bothered	Moderately bothered	Extremely bothered
13. The impact of the anaesthesia on your productivity that day?	**1**	**2**	**3**	**4**
14. The impact of the anaesthesia on your ability to do your usual activities that day?	**1**	**2**	**3**	**4**

HOW DID THE ANAESTHESIA FOR YOUR HAND SURGERY MAKE YOU FEEL?

These questions ask about your recovery from the anaesthesia.

With your anaesthesia mind, how severe was each post-surgery symptom for you:

	None	Mild	Moderate	Extreme
1. Nausea?	1	2	3	4
2. Vomiting?	1	2	3	4
3. Difficulty passing urine?	1	2	3	4
4. Constipation or diarrhoea?	1	2	3	4
5. Feeling sleepy?	1	2	3	4
6. Feeling tired or exhausted?	1	2	3	4
7. Feeling down or depressed?	1	2	3	4
8. Feeling irritable?	1	2	3	4
9. Feeling unwell?	1	2	3	4
10. Problems thinking clearly?	1	2	3	4
11. Trouble remembering?	1	2	3	4
12. Pain caused by the anaesthesia (e.g., use of needles, breathing tube, arm or leg band)?	1	2	3	4
13. Numbness of the arm?	1	2	3	4

Did you have hand surgery within the last 3 months using a local anaesthesia to numb your hand so you could be wide awake during surgery? If yes, please complete the questions below. If no, please continue on the next page.

HOW SATISFIED ARE YOU WITH YOUR HAND SURGERY?

These questions ask about your experience of hand surgery using a local anaesthetic to numb your hand so you could be wide awake during surgery.

How dissatisfied or satisfied were you with the following:

	Very dissatisfied	Somewhat dissatisfied	Somewhat satisfied	Very satisfied
1. Information you were given about how your surgery would be done?	1	2	3	4
2. Being awake during your surgery?	1	2	3	4
3. How the local anaesthetic injection(s) felt?	1	2	3	4
4. How good the local aneasthetic was at preventing pain?	1	2	3	4
5. How your surgery felt while it was taking place?	1	2	3	4
6. Being able to ask questions during your surgery?	1	2	3	4
7. Being able to take part in conversation during your surgery?	1	2	3	4
8. Noises from your surgery (e.g., cutting into the hand)?	1	2	3	4
9. The amount of blood you saw?	1	2	3	4
10. How comfortable the surgical team made you feel?	1	2	3	4
11. The confidence you felt in the surgical team?	1	2	3	4
12. The room where you had your surgery (e.g., sterile, comfortable)?	1	2	3	4
13. How long your surgery took?	1	2	3	4

	Very dissatisfied	Somewhat dissatisfied	Somewhat satisfied	Very satisfied
14. How long you had to wait after your surgery before you could leave the hospital or clinic?	**1**	**2**	**3**	**4**
15. The total amount of time you spent at the clinic or hospital on the day of your surgery?	**1**	**2**	**3**	**4**
16. How long it took for the local anaesthetic to wear off?	**1**	**2**	**3**	**4**
17. Information you were given about how to care for your hand at home?	**1**	**2**	**3**	**4**

HOW SATISFIED ARE YOU WITH THE INFORMATION YOU RECEIVED AND HOW IT WAS GIVEN?

These questions ask about information you received from the hand care team (e.g., doctor, nurse, hand therapist) about your hand surgery and recovery.

How dissatisfied or satisfied were you with the information and how it was given:

	Very dissatisfied	Somewhat dissatisfied	Somewhat satisfied	Very satisfied
1. Options for how the surgery could be done?	1	2	3	4
2. Who would be involved in your care (e.g., doctor, nurse, hand therapist)?	1	2	3	4
3. How much pain you might feel during your recovery?	1	2	3	4
4. What to do if you have a complication (e.g., infection, bleeding)?	1	2	3	4
5. How to care for your hand(s) when you bathe or shower?	1	2	3	4
6. How your surgery would be done?	1	2	3	4
7. The amount of time it would take to heal and recover?	1	2	3	4
8. How much you could use your hands during your recovery?	1	2	3	4
9. Knowing what activities you should avoid (e.g., vigorous activity)?	1	2	3	4
10. How much your hands would change with surgery?	1	2	3	4
11. How to change behaviours that affect hand healing (e.g., smoking, diet)?	1	2	3	4
12. How well your questions were answered?	1	2	3	4
13. The written information you were given?	1	2	3	4
14. How easy it was for you to ask questions?	1	2	3	4

	Very dissatisfied	Somewhat dissatisfied	Somewhat satisfied	Very satisfied
15. How easy it was to understand the information you were given?	**1**	**2**	**3**	**4**
16. The timing of when you were given information (i.e., told you what you needed to know at the right time)?	**1**	**2**	**3**	**4**
17. How likely the surgery would help you to achieve the goals you have for your hands?	**1**	**2**	**3**	**4**
18. The amount of time you had to discuss the information you were given?	**1**	**2**	**3**	**4**
19. That the information you were given by different members of the healthcare team was the same (i.e., did not contradict each other)?	**1**	**2**	**3**	**4**
20. That the information given to you helped you to have realistic expectations about how your hands would change after surgery?	**1**	**2**	**3**	**4**

HOW SATISFIED ARE YOU WITH YOUR HAND SURGEON?

These questions ask about the hand surgeon. Please answer thinking of your recent appointments. Did you feel that your hand surgeon:

	Definitely disagree	Somewhat disagree	Somewhat agree	Definitely agree
1. Made you feel comfortable?	1	2	3	4
2. Acted in a professional manner?	1	2	3	4
3. Was friendly and kind?	1	2	3	4
4. Was easy to talk to?	1	2	3	4
5. Talked to you in a way that was easy to understand?	1	2	3	4
6. Answered all your questions?	1	2	3	4
7. Treated you with respect?	1	2	3	4
8. Listened to you and understood your concerns?	1	2	3	4
9. Involved you in the decisions about your treatment?	1	2	3	4
10. Was attentive to your needs?	1	2	3	4
11. Tailored treatment to address your concerns?	1	2	3	4
12. Helped you figure out what was best for you?	1	2	3	4
13. Was available when you had concerns?	1	2	3	4
14. Spent enough time with you?	1	2	3	4
15. Made sure to protect your privacy?	1	2	3	4
16. Really cared about you?	1	2	3	4
17. Looked after your hand(s) carefully?	1	2	3	4
18. Knew your medical history?	1	2	3	4

	Definitely disagree	Somewhat disagree	Somewhat agree	Definitely agree
19. Knew the history of your hand problem?	**1**	**2**	**3**	**4**
20. Was knowledgeable about hand problems?	**1**	**2**	**3**	**4**
21. Had the right amount of experience?	**1**	**2**	**3**	**4**
22. Knew what they were doing?	**1**	**2**	**3**	**4**
23. Inspired hope that your hand problem would improve with treatment?	**1**	**2**	**3**	**4**
24. Shared your information with other members of the healthcare team who needed it (e.g., hand therapists, nurses)?	**1**	**2**	**3**	**4**
25. Consistently provided a high level of care?	**1**	**2**	**3**	**4**

Did you see a surgeon at their office? If yes, please complete the questions below. If no, please continue on the next page.

HOW SATISFIED ARE YOU WITH THE OFFICE STAFF?

These questions ask about members of the office staff (e.g., secretaries, receptionists) at the location where you saw a hand surgeon. Please answer thinking of your recent appointments. Did you feel that the office staff:

	Definitely disagree	Somewhat disagree	Somewhat agree	Definitely agree
1. Treated you with respect?	1	2	3	4
2. Made you feel comfortable?	1	2	3	4
3. Were knowledgeable?	1	2	3	4
4. Were attentive to your needs?	1	2	3	4
5. Were thorough?	1	2	3	4
6. Worked together as a team?	1	2	3	4
7. Welcomed you at the front desk?	1	2	3	4
8. Answered all your questions?	1	2	3	4
9. Were available when you had concerns?	1	2	3	4
10. Were friendly and kind?	1	2	3	4
11. Acted in a professional manner?	1	2	3	4
12. Treated you with respect over the phone?	1	2	3	4
13. Made sure to protect your privacy?	1	2	3	4
14. Were caring?	1	2	3	4

Did you see a surgeon at a hospital hand clinic? If yes, please complete the questions below. If no, please continue on the next page.

HOW SATISFIED ARE YOU WITH THE HAND CLINIC?

These questions ask about the hand clinic. Please answer thinking of your <u>recent appointments</u>. Did you feel that the hand clinic:

	Definitely disagree	Somewhat disagree	Somewhat agree	Definitely agree
1. Had a nice atmosphere (e.g., welcoming, calm)?	1	2	3	4
2. Welcomed you at the front desk?	1	2	3	4
3. Was clean and sterile?	1	2	3	4
4. Was well organised?	1	2	3	4
5. Made it easy to book an appointment?	1	2	3	4
6. Kept your appointment as scheduled (i.e., did not cancel or change)?	1	2	3	4
7. Was on time (i.e., did not make you wait)?	1	2	3	4
8. Had enough healthcare staff?	1	2	3	4
9. Had consistent healthcare staff (i.e., not constantly changing)?	1	2	3	4
10. Had healthcare staff that were knowledgeable about hand problems?	1	2	3	4
11. Was a place you would recommend to other people with hand problems?	1	2	3	4
12. Protected your healthcare information?	1	2	3	4
13. Provided a phone number you could use outside of office hours?	1	2	3	4

Did you see a <u>hand therapist</u> for your hand problem? If yes, please complete the questions below. If no, please continue on the next page.

HOW SATISFIED ARE YOU WITH YOUR HAND THERAPIST(S)?

These questions ask about the <u>hand therapist(s)</u>. Please answer thinking of your most <u>recent appointments</u>. Did you feel that your hand therapist(s):

	Definitely disagree	Somewhat disagree	Somewhat agree	Definitely agree
1. Acted in a professional manner?	1	2	3	4
2. Were friendly and kind?	1	2	3	4
3. Were easy to talk to?	1	2	3	4
4. Talked to you in a way that was easy to understand?	1	2	3	4
5. Listened to you and understood your concerns?	1	2	3	4
6. Answered all your questions?	1	2	3	4
7. Treated you with respect?	1	2	3	4
8. Spent enough time with you?	1	2	3	4
9. Involved you in the decisions about your treatment?	1	2	3	4
10. Were attentive to your needs?	1	2	3	4
11. Were caring?	1	2	3	4
12. Were knowledgeable about hand problems?	1	2	3	4
13. Had the right amount of experience?	1	2	3	4
14. Knew what they were doing?	1	2	3	4
15. Saw you at the scheduled time?	1	2	3	4
16. Looked after your hand(s) carefully?	1	2	3	4
17. Inspired hope that your hand problem would improve with treatment?	1	2	3	4

	Definitely disagree	Somewhat disagree	Somewhat agree	Definitely agree
18. Knew the history of your hand problem?	**1**	**2**	**3**	**4**
19. Consistently provided a high level of care?	**1**	**2**	**3**	**4**

HOW SATISFIED ARE YOU WITH THE OUTCOME OF YOUR HAND TREATMENT?

We would like to know how you feel about your most recent hand treatment. Please indicate how much you disagree or agree with each statement:

	Definitely disagree	Somewhat disagree	Somewhat agree	Definitely agree
1. I am glad that I had the hand treatment.	1	2	3	4
2. I am satisfied with the results.	1	2	3	4
3. Having the hand treatment changed my life for the better.	1	2	3	4
4. The outcome of my hand treatment met my expectations.	1	2	3	4
5. I would recommend the hand treatment I had to others.	1	2	3	4
6. The hand treatment was worth the time and effort it took.	1	2	3	4
7. The results of my hand treatment turned out great.	1	2	3	4
8. If necessary I would have this hand treatment again without any hesitation.	1	2	3	4
9. I am pleased with the outcome of my hand treatment.	1	2	3	4

Did you use a hand splint or brace in the last 3 months? If yes, please complete the questions below. If no, please continue on the next page.

HOW SATISFIED ARE YOU WITH YOUR HAND SPLINT OR BRACE?

These questions ask about your satisfaction with the hand splint or brace you most recently used.

How dissatisfied or satisfied were you with:

	Very dissatisfied	Somewhat dissatisfied	Somewhat satisfied	Very satisfied
1. How comfortable the splint was to wear?	1	2	3	4
2. How easy the splint was to put on?	1	2	3	4
3. How easy the splint was to remove?	1	2	3	4
4. How often you needed to replace the splint?	1	2	3	4
5. How the splint looked?	1	2	3	4
6. How much the splint cost?	1	2	3	4
7. Your ability to be physically active with the splint on?	1	2	3	4
8. Your ability to sleep with the splint on?	1	2	3	4
9. Your ability to socialise with the splint on?	1	2	3	4
10. Your ability to enjoy life with the splint on?	1	2	3	4
11. Your ability to dress yourself with the splint on?	1	2	3	4
12. Your ability to care for your hand with the splint on?	1	2	3	4

Appendix E.1: Australia Ethics Phase II

Office for Research

Flinders Medical Centre
Ward 6C, Room 6A219
Flinders Drive, Bedford Park SA 5042
Tel: (08) 8204 6453
E: Health.SALHNOfficeforResearch@sa.gov.au

Government of South Australia
SA Health
Southern Adelaide Local Health Network

Final Approval for Ethics Application

15 March 2018

Dr Nicola Dean
Plastic & Reconstructive Surgery
Flinders Medical Centre
BEDFORD PARK SA 5042

Dear Dr Dean

OFR Number: 23.18
HREC reference number: HREC/13/SAC/223
Study title: Developing the HAND-Q: Phase 2 Field Test
Project title: **Chief Investigator**: Dr Nicola Dean

Ethics Approval Period: 14 February 2018 – 14 February 2019

The Southern Adelaide Clinical Human Research Ethics Committee (SAC HREC EC00188) have reviewed and provided approval for this application which meets the requirements of the *National Statement on Ethical Conduct in Human Research (2007).*

You are reminded that this letter constitutes **Ethics** approval only. **Ethics approval is one aspect of the research governance process**.

You must not commence this research project at any SA Health sites listed in the application until a Site Specific Assessment (SSA), or Access Request for data or tissue form, has been approved by the Chief Executive or delegate of each site.

Public health sites approved under this application: Flinders Medical Centre

The below documents have been reviewed and approved:

- HREA form v 1.3 dated 19 January 2018
- Project description v2 dated 31 January 2018
- Invitation letter v1 dated 31 January 2018
- Participant information sheet and consent form – main v4 dated 14 February 2018
- Participant information sheet and consent form – Group B v4 dated 14 February 2018
- Hand-Q post-surgery module v1 dated 22 January 2018
- Flyer wording v1 dated 31 January 2018

Terms and Conditions Of Ethics Approval:

It is essential that researchers adhere to the conditions below and with the *National Statement chapter 5.5.*

Final ethics approval is granted subject to the researcher agreeing to meet the following terms and conditions:

1. The approval only covers the science and ethics component of the application. A SSA will need to be submitted and authorised before this research project can commence at any of the approved sites identified in the application.
2. If University personnel are involved in this project, the Principal Investigator should notify the University before commencing their research to ensure compliance with University requirements including any insurance and indemnification requirements.
3. Compliance with the *National Statement on Ethical Conduct in Human Research* (2007) & the *Australian Code for the Responsible Conduct of Research* (2007).
4. To immediately report to SAC HREC anything that may change the ethics or scientific integrity of the project.
5. Report Significant Adverse events (SAE's) as per SAE requirements available at our website.
6. Submit an annual report on each anniversary of the date of final approval and in the correct template from the SAC HREC website.
7. Confidentiality of research participants MUST be maintained at all times.
8. A copy of the signed consent form must be given to the participant unless the project is an audit.
9. Any reports or publications derived from the research should be submitted to the Committee at the completion of the project.
10. All requests for access to medical records at any SALHN site must be accompanied by this approval email.
11. To regularly review the SAC HREC website and comply with all submission requirements, as they change from time to time.
12. Once your research project has concluded, any new product/procedure/intervention cannot be conducted in the SALHN as standard practice without the approval of the SALHN New Medical Products and Standardisation Committee or the SALHN New Health Technology and Clinical Practice Innovation Committee (as applicable). Please refer to the relevant committee link on the SALHN intranet for further information.

For any queries about this matter, please contact The Office for Research on (08) 8204 6453 or via email to Health.SALHNOfficeforResearch@sa.gov.au

Yours sincerely

A/Professor Bernadette Richards
Chair, SAC HREC

Appendix E.2: Canada Ethics Phase II

RÉSEAU DE SANTÉ Horizon HEALTH NETWORK

Research Ethics Board
Horizon Health Network, 2CN SJRH,
400 University Avenue, Saint John, New Brunswick, Canada E2L 4L2
Research Office - 506-648-6094 Tel 506-648-7734 Fax

March 1, 2018

Dr. Donald Lalonde
600 Main Street Suite C204
Saint John NB E2K 1J5

Dear Dr. Lalonde

Re: Developing the HAND-Q: Phase 2 Field Test
Principal Investigator: Dr. Donald Lalonde
Protocol Number: N/A
Our File #: 2017-2574

This will acknowledge receipt of the information requested by the REB in the February 23, 2018 Conditional Approval letter.

This information has been reviewed and **Final Approval** is now granted effective March 1, 2018. Please note that if your study is ongoing annual re-approval should be sought prior to March 31, 2019.

REVIEWED:

- **Research Study Application:** signed and dated February 14, 2018
- **Protocol:** version undated
- **Horizon General Information Informed Consent Form:** version undated
- **HAND-Q Pre-Surgery Module:** version undated
- **HAND-Q Post-Surgery Module:** version undated

Also received and on file:

- Letter of Support: signed and dated February 13, 2018
- Letter of Support: signed and dated February 15, 2018

The Research Ethics Board for Horizon Health Network is organized and operates according to the principles of the ICH Harmonized Tripartite Guidelines: Good Clinical Practice, Tri-Council Policy Statement and Division

Research Ethics Board (REB)
(Address correspondence c/o Ethics Services)

Membership

Dr. Marc Smith, PhD
Ethics Expert
Chair
Dr. Glendon Sullivan
Medical
Dr. Kenneth Obenson
Medical
Dr. William Cook
Medical
Ameena Meerasa
Content Expert
Dr. Tsetan Dolkar
Content Expert
Gisia Pisegna
Pharmacist/Content Expert
Dr. Timothy Christie, MHSc, PhD
Ethics Expert
Esther Akinkugbe
Law
Kevin Toner
Law
Dr. Yu Chen
Methodology Expert
Claire Wright
Methodology Expert
Corinne Shea
Non-Scientific Member
David Ross
Community Member

2017-2574
Dr. Donald Lalonde
February 23, 2018

5 of the Food and Drug Regulations. The Horizon Health Network REB carries out its functions in a manner consistent with Good Clinical Practices; and has reviewed and approved the clinical trial protocol and informed consent form for the trial which is to be conducted by the qualified investigator named above at the specified clinical trial site. This approval and the views of this Research Ethics Board have been documented in writing.

With kind regards,

Dr. Marc Smith
Chairman – Research Ethics Board
Horizon Health Network

MS/jr

Appendix E.3: Preliminary HAND-Q Scales

HAND-Q©

Preliminary Scales

1. **Appearance**
2. **Function**
3. **Symptoms**
4. **Psychological Impact**
5. **Quality of Life**
6. **Sleep**
7. **Sexual Life**

HOW DO YOUR HANDS LOOK? Please answer thinking of how your hands look NOW.

NOTE: If your hands look different from each other, answer each question thinking about the hand you are least satisfied with.

How dissatisfied or satisfied are you with:

	Very dissatisfied	Somewhat dissatisfied	Somewhat satisfied	Very satisfied
1. How your hands look when you wave at someone?	1	2	3	4
2. How well proportioned your hands look (i.e., all parts look the right size and shape)?	1	2	3	4
3. The shape of your fingers and thumbs?	1	2	3	4
4. How your fingers and thumbs line up with each other?	1	2	3	4
5. How your knuckles look?	1	2	3	4
6. How your hands look when you rest your palms on a table?	1	2	3	4
7. How your hands look when you hold a glass?	1	2	3	4
8. How normal your hands look?	1	2	3	4
9. How straight your fingers and thumbs look?	1	2	3	4
10. How your hands look compared with other people's hands?	1	2	3	4

HOW DIFFICULT IS IT TO USE YOUR HANDS? Please answer thinking of the PAST WEEK.

NOTE: If your hand problem affects one hand more than the other, answer each question for the hand that is worse. If you use gadgets or devices to help perform tasks, please answer based on how your hand(s) work without using them.

With your hand problem in mind, how difficult would these tasks have been:

	Not at all difficult	A little difficult	Moderately difficult	Extremely difficult
1. Brushing your teeth?	1	2	3	4
2. Eating with yours hand(s)?	1	2	3	4
3. Using a TV remote control?	1	2	3	4
4. Personal grooming (e.g., shaving, putting on make-up)?	1	2	3	4
5. Holding a phone to your ear?	1	2	3	4
6. Holding a glass?	1	2	3	4
7. Typing?	1	2	3	4
8. Clipping your fingernails?	1	2	3	4
9. Plugging a cord into a socket?	1	2	3	4
10. Writing with a pen or pencil?	1	2	3	4
11. Picking up a coin?	1	2	3	4
12. Buttoning a shirt or coat?	1	2	3	4
13. Turning a tap (i.e., faucet)?	1	2	3	4
14. Taking things out of a pocket?	1	2	3	4
15. Holding a bag of groceries?	1	2	3	4
16. Gripping handles (e.g., tennis racket, golf club, broom)?	1	2	3	4
17. Preparing food (e.g., peeling, cutting)?	1	2	3	4
18. Opening a jar?	1	2	3	4

HOW DO YOUR HANDS FEEL (SYMPTOMS)? Please answer thinking of the PAST WEEK.

NOTE: If the symptoms you feel in each hand differ, answer each question thinking about the hand that bothers you the most.

Please rate the severity of each of these symptoms:

	None	Mild	Moderate	Severe
1. Hands feeling weak (i.e., lack of strength)?	1	2	3	4
2. Pain when you use your hands?	1	2	3	4
3. Hands feeling stiff?	1	2	3	4
4. Swelling or puffiness?	1	2	3	4
5. Clumsiness (e.g., dropping or spilling things)?	1	2	3	4
6. Hands feeling numb (i.e., less feeling)?	1	2	3	4
7. Hand symptoms (e.g., pain, numbness) disturbing your sleep?	1	2	3	4
8. Throbbing pain in your hands?	1	2	3	4
9. Tingling in your hands (i.e., pins and needles feeling)?	1	2	3	4
10. Pain when your hands are at rest?	1	2	3	4

DOES YOUR HAND PROBLEM AFFECT HOW YOU FEEL? Please answer thinking of the PAST WEEK.

With your <u>hand problem</u> in mind, how often have you felt:

	Never	Sometimes	Often	Always
1. Frustrated?	**1**	**2**	**3**	**4**
2. Annoyed?	**1**	**2**	**3**	**4**
3. Fed-up?	**1**	**2**	**3**	**4**
4. Irritated?	**1**	**2**	**3**	**4**
5. Upset?	**1**	**2**	**3**	**4**
6. Stressed?	**1**	**2**	**3**	**4**
7. Anxious?	**1**	**2**	**3**	**4**
8. Depressed?	**1**	**2**	**3**	**4**
9. Overwhelmed?	**1**	**2**	**3**	**4**
10. Sorry for yourself?	**1**	**2**	**3**	**4**

DOES YOUR HAND PROBLEM AFFECT YOUR QUALITY OF LIFE? Please answer thinking of the PAST WEEK.

How much has your <u>hand problem interfered</u> with the following:

	Not at all	A little bit	Quite a bit	Very much
1. Doing activities you enjoy?	**1**	**2**	**3**	**4**
2. Being physically active?	**1**	**2**	**3**	**4**
3. Your ability to be independent?	**1**	**2**	**3**	**4**
4. Taking a bath or shower?	**1**	**2**	**3**	**4**
5. Your mood?	**1**	**2**	**3**	**4**
6. Being able to relax?	**1**	**2**	**3**	**4**
7. Your social life?	**1**	**2**	**3**	**4**
8. Your close relationships?	**1**	**2**	**3**	**4**

DOES YOUR HAND PROBLEM AFFECT YOUR SLEEP? Please answer thinking of the PAST WEEK.

With your <u>hand problem</u> in mind, how often have you:

	Never	Sometimes	Often	Always
1. Had trouble finding a comfortable position to sleep in?	**1**	**2**	**3**	**4**
2. Woken up at night?	**1**	**2**	**3**	**4**
3. Had symptoms (e.g., pain, numbness) from your hands disturb your sleep?	**1**	**2**	**3**	**4**
4. Not had enough sleep?	**1**	**2**	**3**	**4**
5. Had trouble falling asleep?	**1**	**2**	**3**	**4**

DOES YOUR HAND PROBLEM AFFECT YOUR SEX LIFE?

With your <u>hand problem</u> in mind, how much are you <u>bothered</u> by the following:

	Not at all bothered	A little bothered	Moderately bothered	Extremely bothered
1. Limitations in hand function that can interfere with sexual activity (e.g., grip, strength)?	**1**	**2**	**3**	**4**
2. Being able to use your hands in tender ways (e.g., touch, hold)?	**1**	**2**	**3**	**4**
3. Being aware of your hands during sexual activity?	**1**	**2**	**3**	**4**
4. Your hand problem interfering with your ability to give pleasure?	**1**	**2**	**3**	**4**
5. Symptoms you feel in your hands that can interfere with sexual activity (e.g., pain, numbness, tingling)?	**1**	**2**	**3**	**4**
6. Your hand problem being a distraction during sexual activity?	**1**	**2**	**3**	**4**
7. Your hand problem affecting how much you enjoy sexual activity?	**1**	**2**	**3**	**4**