ASSESSING THE OUTCOME OF SURGERY FOR DUPUYTREN'S DISEASE OF THE HAND

JEREMY NEIL RODRIGUES

BSc (Honours); MB, ChB (Honours); MRCS; MSc (Distinction)

Thesis submitted to The University of Nottingham for the degree of Doctor of Philosophy

JULY 2016

Abstract

Dupuytren's disease (DD) is a fibroproliferative disorder causing deformity of fingers and disability. Different treatments exist ranging from dividing cords (needle aponeurotomy) through disease excision (fasciectomy) to disease excision with resurfacing with a skin graft (dermofasciectomy). A range of outcome measurements has been used in DD, including angular measurements of finger joints and patient-reported outcome measures (PROMs).

This thesis hypothesised that the leading candidate outcome measures are inadequate (subject to bias, invalid, and/or uninterpretable) and that currently there is insufficient evidence to inform patient-centred treatment choice in DD.

To investigate this, existing evidence was appraised and studies of validity and interpretability of outcomes were conducted. This comprised a systematic review and meta-analysis of surgical trials, a systematic review of interpretability of outcome measures, cross sectional studies of the validity of leading candidate outcome measures (joint angles, the Disabilities of the Arm, Shoulder and Hand tool (DASH), and the Unité Rhumatologique des Affections de la Main scale (URAM)), and a prospective cohort study of outcome interpretability and variables associated with functional outcome.

Key findings:

- There were too few trials comparing treatments to inform practice in DD, and methodological quality was generally poor.

- There were limited interpretability data to guide the design of future studies.
- Dynamism was present in 89% of digits, with mean MCPJ dynamism of 6^{o} and PIPJ dynamism of 14^{o}; 11% of digits exhibited over 30 degrees of dynamism.
- Patients had virtually unique goals for surgery, with 26% captured by the URAM
- The DASH and the URAM were not structurally valid in factor analysis
- The DASH was uninterpretable; the URAM's minimal important change was 10.5
- The factors associated with poor functional outcome differ from those associated with recurrence.

Future work should examine validity for other outcome measures; qualitative investigation of patients' experiences; and patient-centred high quality randomised controlled trials.

Publications and Presentations

The following publications have arisen from the work in this thesis:

Rodrigues JN, Becker GW, Ball C, Zhang W, Giele H, Hobby J, Pratt AL, Davis T. Surgery for Dupuytren's contracture of the fingers.

Cochrane Database of Systematic Reviews 2015, Issue 11.

Art. No.: CD010143. DOI: 10.1002/14651858.CD010143.pub2

Rodrigues J, Zhang W, Scammell B, Russell P, Chakrabarti I, Fullilove S, Davidson D, Davis TRC. 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) Epub ahead of print 2015.

PMID: 26307142

Warwick D, Worsley P, Graham D, **Rodrigues J**, Zhang W, Scammell B, Davis T, Akhavani M, Muir L. Re: Akhavani MA, McMurtrie A, Webb M, Muir L. A review of the classification of Dupuytren's Disease. J Hand Surg Eur. 2015, 40: 155–65 and Rodrigues JN, Zhang W, Scammell BE, Davis TRC. What patients want from the treatment of Dupuytren's Disease – is the Unité Rhumatologique des Affections de la Main (URAM) scale relevant? [CORRESPONDENCE]

Journal of Hand Surgery (European Volume) (2015) 40(5): 544-5. **PMID: 26002706**

Rodrigues J, Mabvuure N, Nikkhah D, Shariff Z, Davis TRC. Minimal clinically important changes and differences in elective hand surgery.
Journal of Hand Surgery (European Volume) Epub ahead of print. **PMID: 25320122**

Rodrigues J, Zhang W, Scammell BE, Davis TRC. Re: What patients want from the treatment of Dupuytren's disease – is the Unité Rhumatologique des Affections de la Main (URAM) scale relevant? [CORRESPONDENCE]
The Journal of Hand Surgery (European Volume) (2014) 39(6): 673-675. **PMID: 24939553**

Rodrigues J, Zhang W, Scammell BE, Davis TRC. The dynamism of Dupuytren's contractures
The Journal of Hand Surgery (European Volume) Epub ahead of print. **PMID: 24664160**

Rodrigues J, Zhang W, Scammell BE, Davis TRC. What patients want from the treatment of Dupuytren's disease – is the Unité Rhumatologique des Affections de la Main (URAM) scale relevant?

***The Journal of Hand Surgery (European Volume)*.** Epub ahead of print. **PMID: 24563499**

Davis T, Becker G, **Rodrigues J**, Ball C, Giele H, Hobby J, Zhang W. C245-P Surgery for Dupuytren's contractures of the fingers (Protocol).

Cochrane Database of Systematic Reviews (2012) Issue 10.

Art. No.: CD010143. DOI: 10.1002/14651858.CD010143.

[PROTOCOL]

The following oral presentations have arisen from the work in this thesis:

Rodrigues J, Becker G, Ball C, Zhang W, Giele H, Hobby J, Davis TRC. The effect of postoperative splinting in Dupuytren's disease: a meta-analysis of randomised controlled trials. ***Joint BSSH-IHSS Autumn Meeting***, London, October 2014.

Rodrigues J, Zhang W, Scammell B, Davis TRC. Recovery, responsiveness and minimal important differences after fasciectomy and dermofasciectomy. ***Joint BSSH-IHSS Autumn Meeting***, London, October 2014.

Rodrigues J, Mabvuure N, Nikkhah D, Shariff Z, Davis TRC. Minimal important differences in elective hand surgery: a systematic review. ***Joint BSSH-IHSS Autumn Meeting***, London, October 2014

Rodrigues J, Zhang W, Scammell BE, Chakrabarti I, Davidson D, Russell P, Fullilove S, Davis TRC. The correlation between the DASH and the QuickDASH in Dupuytren's disease.

BSSH Spring Meeting, Newcastle, May 2014.

Rodrigues J, Zhang W, Scammell BE, Davidson D, Fullilove S, Russell P, Chakrabarti I, Davis TRC. Complication rates following different Dupuytren's disease treatments, and associated factors. A multi-centre case-control study

BSSH Spring Meeting, Newcastle, May 2014.

Rodrigues J, Zhang W, Scammell BE, Russell P, Davidson D, Chakrabarti I, Fullilove S, Davis TRC. Validating patient-reported outcome measures (PROMs) in Dupuytren's disease

Joint BAPRAS-IAPS Winter Scientific Meeting, Dublin, Ireland, November 2013.

Rodrigues J, Zhang W, Scammell BE, Fullilove S, Russell P, Chakrabarti I, Davidson D, Davis TRC. Which is the most valid patient-reported outcome measure for studying Dupuytren's disease?

BSSH Autumn Meeting, London, October 2013.

Rodrigues J, Zhang W, Scammell BE, Davis TRC. How dynamic are Dupuytren's contractures?

BSSH Autumn Meeting, London, October 2013.

Rodrigues J. Linking NICE and hand surgery: underused opportunities from NICE's Fellows and Scholars Programme.

BSSH Spring Meeting, Harrogate, April 2013.

Rodrigues J, Scammell B, Davis T. The true indications for Dupuytren's disease surgery: a study of patient determined goals.

BSSH Autumn Meeting, York, October 2012.

The following poster presentations have arisen from the work in this thesis:

Rodrigues J, Lawton G. The uptake of NICE's Fellows and Scholars Programme by plastic surgeons.

Joint BAPRAS-IAPS Winter Scientific Meeting, Dublin, Ireland, November 2013.

Rodrigues J, Zhang W, Scammell BE, Fullilove S, Chakrabarti I, Russell P, Davidson D, Davis TRC. Five-year outcomes and associated factors following Dupuytren's disease surgery: A multi-centre retrospective cohort study in the UK.

BSSH Autumn Meeting, London, October 2013.

Declaration & Acknowledgements

This project was designed and conducted during my appointment as the British Society for Surgery of the Hand (BSSH) Research Fellow at The University of Nottingham.

Background work for the project conducted and submitted for the degree of MSc in Surgical Sciences at The University of Edinburgh has been included for discussion only, and has been acknowledged as such. All other work has been performed specifically and only for this thesis.

Some of this work had to be performed in collaboration with others. However, for all such work, the candidate was the primary author and primary contributor. In addition to the candidate's formal supervisors, the following people collaborated with work in this project:

Chapter	Collaborators
Systematic Review of Trials of Surgery	Mr Giles Becker (2nd person to review abstracts) Ms Cathy Ball (2nd person to extract data) Mr Jonathan Hobby (contributed to search strategy)

	Mr Henk Giele (contributed to search strategy) Ms Anna Pratt (contributed to meta-analysis of trials of postoperative plinting)
Systematic review of Interpretability	Dr Nigel Mabvuure (2nd person to review abstracts and extract data) Mr Dariush Nikkhah (resolved disparities between candidate and 2nd reviewer) Mr Zakir Shariff (contributed to analysis of results)
Studies involving cross sectional data (validity of PROMs and late outcome studies)	Mr Peter Russell (Principal investigator at Royal Derby Hospital) Mr Indranil Chakrabarti (Principal investigator at Rotherham General Hospital) Mrs Dominique Davidson (Principal investigator at St John's Hospital at Howden) Ms Sue Fullilove (Principal investigator at Derriford Hospital)

I would like to acknowledge the tremendous support and encouragement that I have received from my supervisors Professor Tim RC Davis, Professor Brigitte Scammell and Dr Weiya Zhang. The word ‘supervision’ grossly misrepresents the significance of contribution they have made to this work, and to my development as an academic surgeon.

Numerous people have supported this work across the United Kingdom, including collaborators from the multi centre study. These include Mrs Audrey Parks, Mrs Melanie Arundell, Ms Sandra Owdziej and Mr Carlos Heras-Palou at the Pulvertaft Hand Centre, Mr David Myers, Dr Sally Atkinson and Mr Indranil Chakrabarti at Rotherham General Hospital, Ms Karole Champion and Mrs Dominique Davidson at St John’s Hospital at Howden and Miss Sue Fullilove at Derriford Hospital in Plymouth.

I would also like to acknowledge the advice received from Mr Nick Downing, Mr John Oni, Professor Alan Montgomery, Mr Barrie White and Professor Eamonn Ferguson.

My family, my parents in particular, have also supported me and encouraged me. They have often made sacrifices so that I could pursue my ambitions and so that I would have the best opportunities. For this, I will always be grateful.

Dr Anshu Sengupta is a close friend, who has taken considerable time out of his own doctoral studies to proof read and provide invaluable feedback.

Acknowledgement of Sources of Support

A BSSH Research Fellowship, a Nottingham Hospitals Charity grant, and a Nottingham Orthopaedic Walk grant supported this work financially. Academic and indirect financial support was also provided by a Scholarship from the National Institute for Health and Care Excellence (NICE).

Abbreviations

Abbreviation	Full term
95% CI	95% Confidence Intervals
95%CIs	95% Confidence Intervals
AGREE II	Appraisal of Guidelines, Research and Evaluation Instrument II
BEST	BSSH Evidence for Surgical Treatment Guideline
BNI	British Nursing Index
BSSH	British Society for Surgery of The Hand
CCG	Clinical Commissioning Group
CCGOIS	Clinical Commissioning Group Outcomes Indicator Set
CENTRAL	The Cochrane Central Register of Controlled Trials
CINAHL	Cumulative Index of Nursing and Allied Health Literature
COSMIN	Consensus based standards for the selection of health status measurement instruments
CRPS	Complex Regional Pain Syndrome
DASH	Disability of the Arm, Shoulder and Hand patient-reported outcome measure
DH	Derriford Hospital, Plymouth
DIPJ	Distal interphalangeal joint
ECM	Extracellular matrix

EFA	Exploratory factor analysis
EQ5D™	EuroQol 5D PROM
GP	General Practitioner
GRADE	Grading of Recommendations Assessment, Development and Evaluation
GRC	Global Rating of Change PROM
GRC	Global Rating of Change PROM
HES	Hospital Episode Statistics
IASP	International Association for the Study of Pain
ICER	Incremental cost effectiveness ratio
IPAC	Interventional Procedures Advisory Committee (of NICE)
LILACS	Latin American and Caribbean Health Sciences
MCPJ	Metacarpophalangeal joint
MHQ	Michigan Hand Questionnaire PROM
MMP	Matrix metalloproteinase
MYMOP	Measure Yourself Medical Outcome Profile PROM
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
OPCS-4	Office of Population Censuses and Surveys Classification of Interventions and Procedures Version 4
OR	Odds Ratio

PEM	Patient Evaluation Measure PROM
PHC	Pulvertaft Hand Centre within Royal Derby Hospital, Derby
PI	Principal Investigator
PIPJ	Proximal interphalangeal joint
PROM	Patient-reported outcome measure
QALY	Quality adjusted life year
QMC	Queen's Medical Centre, Nottingham
RCT	Randomised controlled trial
RGH	Rotherham General Hospital, Rotherham
SF-36	Short Form 36 PROM
SJH	St John's Hospital at Howden, Livingston
TGFβ1	Transforming growth factor beta 1
TIMP	Tissue inhibitor of matrix metalloproteinase
TPED	Total Passive Extension Deficit
UK	United Kingdom
URAM	Unité Rhumatologique des Affections de la Main PROM
VA	Veterans' Affairs
VAS	Visual analogue scale
α-SMA	Alpha smooth muscle actin

Table of Contents

List of Tables

List of Figures

List of Appendices

1 Introduction

This chapter will describe the key features of Dupuytren's disease, and explain why this project was required.

1.1 What Dupuytren's disease is and why it matters

Dupuytren's disease is a condition in which fibrous changes occur in association with the palmar fascia, a subcutaneous layer in the hand. This forms palpable nodules, which are believed to progress into cords and eventually flexion contractures of the fingers (Luck, 1959), preventing full extension at their metacarpophalangeal joints (MCPJs) and the proximal interphalangeal joints (PIPJs), and more rarely at their distal interphalangeal joints (DIPJs). This results in impairment of function. The reported prevalence of the condition varies widely. Given that the prevalence of Dupuytren's disease in the United Kingdom (UK) may be as high as 18% for men and 9% for women (Early, 1962), it may account for considerable morbidity.

Furthermore, Hospital Episode Statistics (HES) data reveal that 12 191 palmar or digital fasciectomy surgeries were performed on National Health Service (NHS) inpatients in England alone in 2009-2010 (Hospital Episode Statistics, 2011), thus management of Dupuytren's disease constitutes a significant workload to healthcare services in the UK. Data based on NHS tariffs estimate the cost of this treatment at £41 576 141 in 2010-2011 in England

alone (Gerber et al., 2011). The actual number of procedures carried out for Dupuytren's disease is probably much higher, as the HES data figure only captures inpatient operations from England, and not minor outpatient procedures, or data from other UK countries. Thus the burden of the condition for the NHS across the UK is probably much greater. The true socioeconomic 'cost' of Dupuytren's disease is not described by tariff alone, and the cost of ramifications of the condition and its treatment are almost certainly much greater than the figure described by Gerber and colleagues. Functional impairment in the hand may affect employability of working age patients, and may increase dependence and the need for social care services in the elderly. These effects are difficult to capture accurately, and are not assessed in current UK national cost effectiveness analyses (NICE, 2008).

As Dupuytren's disease is a progressive condition, recurrence or progression are commonly encountered, even after successful treatment. Both may result in a return to disability and/or need for further treatment for some patients. Additionally, different treatments exist, and vary in terms of recovery time, complication rate, early and late outcome, as well as "procedure" cost. High quality comparative data are urgently required to inform treatment choice, and to establish the cost effectiveness of treating the condition at all.

Published data do exist describing recurrence after individual treatment types. These typically are descriptions of return of angular deformity in non-comparative studies (Armstrong et al., 2000, van Rijssen and Werker, 2006, Zachariae, 1969). In comparative studies, either a similar definition was used

(van Rijssen et al., 2012), or the definition was not provided (Ullah et al., 2009). However, work that the candidate previously conducted for his MSc degree dissertation, and the findings of two different systematic reviews, demonstrated that there is no consistent definition of recurrence, and no consensus on when it should be assessed, between different studies (Ball et al., 2013, Becker and Davis, 2010, Rodrigues, 2010). The two most likely candidates for primary outcome measure in future trials of Dupuytren's disease treatment are: (a) an assessment of finger joint angles, as have been used in previous randomised controlled trials (Hurst et al., 2009, van Rijssen et al., 2006, van Rijssen et al., 2012); or, (b) a patient-reported outcome measure (PROM) assessing impairment of hand function, of which several have been previously used (Ball et al., 2013). Angular deformity does not correlate well with the most commonly used of these measures of functional impairment (Engstrand et al., 2009, Jerosch-Herold et al., 2011, Zyluk and Jagielski, 2007), and whether either provides meaningful quantification of treatment efficacy is unclear. This limits the applicability of study outcomes, prevents comparison of findings between studies and procedures, and compounds the paucity of high quality comparative trials in this condition.

With new techniques such as collagenase injection being developed and launched to treat this common condition (FDA, 2010), there is an urgent need for robust and meaningful comparative data.

In this project, the existing data comparing different surgical interventions for Dupuytren's disease will be appraised, along with the tools used to assess

outcomes studied. Outcome measure selection, early functional recovery, treatment failure rate, late outcome and complication rate will be considered for the most commonly employed surgical treatments of Dupuytren's disease.

1.2 Anatomy and Patho-anatomy

There are important aspects of the anatomy of the hand and other bodily sites affected by Dupuytren's disease and its associated conditions that must be appreciated. The complexity of the anatomy of the hand itself is also of relevance to treatment and expected outcome of treatment at different sites within the hand and even within a single digit.

1.2.1 Palmar fascia anatomy

Dupuytren's disease is associated with the palmar fascia. This consists of a continuum that spreads out distally from the tendon of palmaris longus at the wrist, forming a specific three-dimensional arrangement. The components of these normal arrangements are referred to as fascia, ligaments or bands (Hurst, 2010). In contrast, diseased tissue is referred to as nodules or cords (Hurst, 2010). Understanding the normal anatomical arrangement of structures aids the surgeon in identifying and treating the disease.

In the previous (fifth) edition of Green's Operative Hand Surgery, McGrouther classifies these arrangements based on fibre orientation. There are transverse fibres of Skoog crossing the distal palm, which are the deepest of all the fascial structures. The neurovascular bundles are deep to them, and thus sheltered by them in the palm.

Distal to Skoog's fibres are longitudinal fibres, classified into three layers by McGrouther (McGrouther, 2005), referring to his own review of anatomical

descriptions of the palm, and McFarlane's review of the digit, both in the same book (McFarlane, 1990, McGrouther, 1990). These fan out distally when viewed in a sagittal plane through the hand. The deepest ones pass lateral to both the flexor tendon and the bone around the metacarpophalangeal joint (MCPJ) level.

The most superficial layer forms weak insertions into the palmar skin at proximal skin creases of the fingers, with some fibres continuing distally into the finger as the pretendinous band.

The second layer also passes deep, around the side of the flexor tendon and deep to the neurovascular bundle. These form the spiral bands, passing around the neurovascular bundle, and further distally to this, continue as retrovascular bands deep to the neurovascular bundle in the finger.

The third layer passes deepest close to the MCPJ to insert into the extensor apparatus.

The structural pattern summarised by McGrouther has implications for the surgeon. For example, as the longitudinal fibres rest on the superficial surface of Skoog's transverse fibres, with the neurovascular bundles deep to all of them, a longitudinal cord can be peeled off the transverse fibres without risking damage to the bundles. Furthermore, as one moves into the finger, the wraparound relationship of the spiral band with the neurovascular bundle

means that as the contracture develops, the bundle will be displaced towards the midline and will become more superficial (McFarlane, 1974).

In contrast to McGrouther, Hurst's chapter in the most recent edition of Green's Operative Hand Surgery divides the fascia into three regions based on surface landmarks. Passing proximally to distally, these are the palm, the palmar-digital junction, and the digit (Hurst, 2010). By classifying the disease into discrete surface anatomy regions, Hurst's approach focuses the surgeon's mind on the specific pathological structures likely to be relevant when managing disease at a particular site clinically. Within the palm, central cords are encountered and extend distally through the other regions, with natatory cords at the palmar-digital junction, and retrovascular cords and lateral digital cords in the digit. Spiral cords extend through the regions in a similar fashion to central cords (Hurst, 2010).

It is apparent from both of these descriptions that Dupuytren's disease is not a single anatomical entity. At any site passing from the palm to the digit, one or more different cords may be encountered, in different planes, with different consequences in terms of contracture. Furthermore, treating only some of the Dupuytren's tissue present in a ray may or may not be adequate to release the contracture, depending on the pattern of disease present. As a result, it is conceivable that the 'same' operation performed by different surgeons may have very different outcomes. Furthermore, the exact pattern of disease

present in a ray may not necessarily be describable by surface clinical examination alone.

1.2.2 Finger joint anatomy

The joints involved in contractures must also be considered. The MCPJs of the digits are synovial joints between the metacarpal head and the proximal phalanx. The metacarpal head is not perfectly round, instead it is narrower dorsally, with a volar prominence (Tubiana et al., 1996). Consequently, the radial and ulnar collateral ligaments that augment the lateral aspects of the joint capsule are relatively lax when the joint is in extension, allowing medial and lateral deviation (clinically this is abduction and adduction of the digits) and some degree of rotation. They tighten when the joint flexes. Thus, the joint only becomes stable when flexed. An increasing contracture of the MCPJ (as occurs in Dupuytren's disease) tightens the collateral ligaments and maintains their length. As a result, when treatment releases the Dupuytren's cord, the joint range of motion of the joint itself has been maintained by the increased tension on the ligaments, and the contracted joint can be straightened (Figure 1.1).

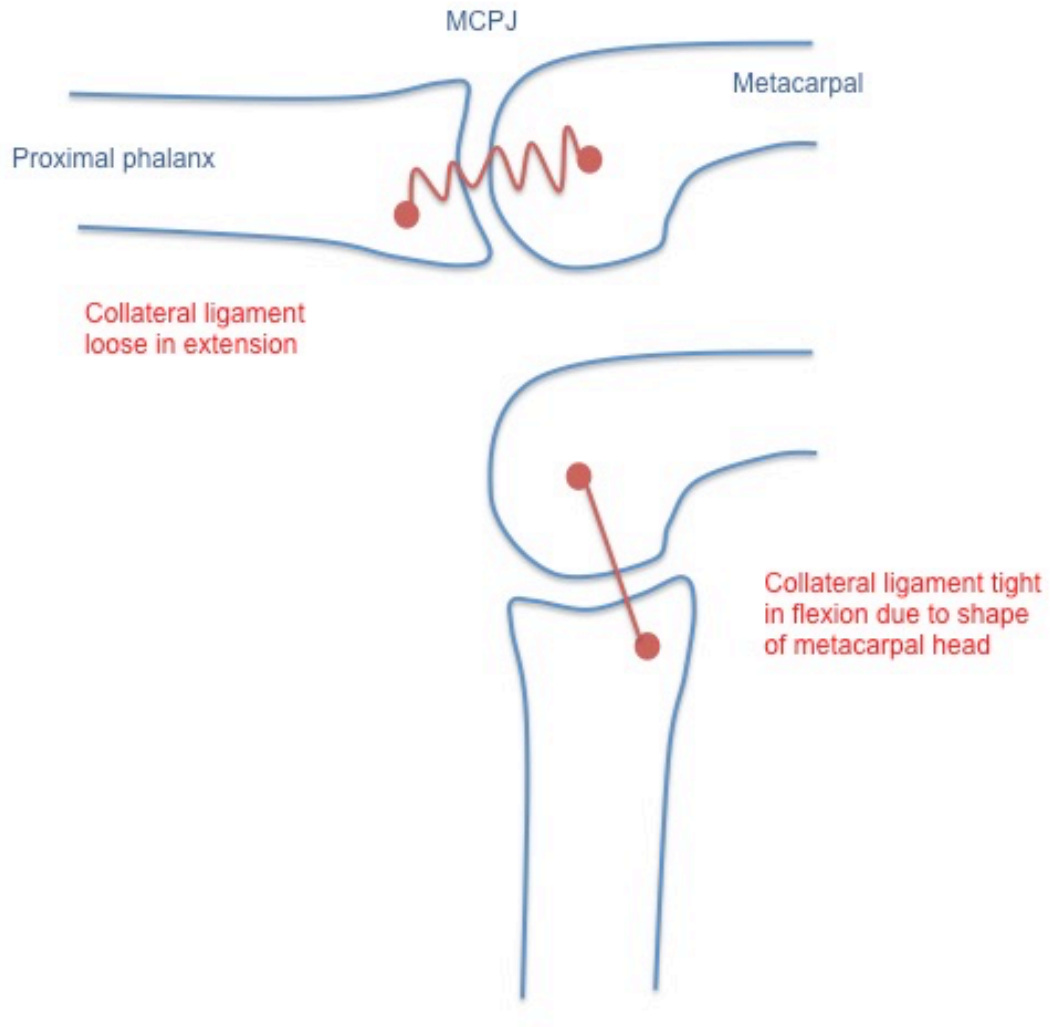

Figure 1.1: Metacarpophalangeal joint throughout its arc of motion

In contrast, the proximal interphalangeal joint (PIPJ) is stable throughout its range of motion; hence it does not allow abduction or adduction. This is achieved by maintaining even tension in the collateral ligaments throughout flexion and extension (Tubiana et al., 1996). Furthermore, the volar plate (the palmar augmentation of the joint capsule) structure differs at the PIPJ from the MCPJ. At the PIPJ, it inserts proximally into the flexor sheath via two strong 'check rein ligaments', one on each side (Tubiana et al., 1996). When the

joint is held flexed for a prolonged period (as in Dupuytren's disease across the joint), the ligaments remain as tight as in any other position, but the volar plate contracts through its check rein ligaments. Consequently, even if the Dupuytren's cord is adequately treated at the PIPJ, the joint itself may not straighten until a 'check rein ligament release' has been performed. As the proximal extent of the flexor sheath at the A1 pulley is level with the MCPJ, the MCPJ's volar plate does not have proximal check rein ligaments. As a result, MCPJs the mechanism of stiffness in flexion that occurs at the PIPJ does not occur at the MCPJ. Such stiffness at the PIPJ is often incorrectly considered a collateral ligament effect, when it is instead a volar plate effect (Figure 1.2). This explains why Dupuytren's contractures of the PIPJ are harder to successfully correct than contractures of the MCPJ (Ritchie et al., 2004, Smith and Breed, 1994).

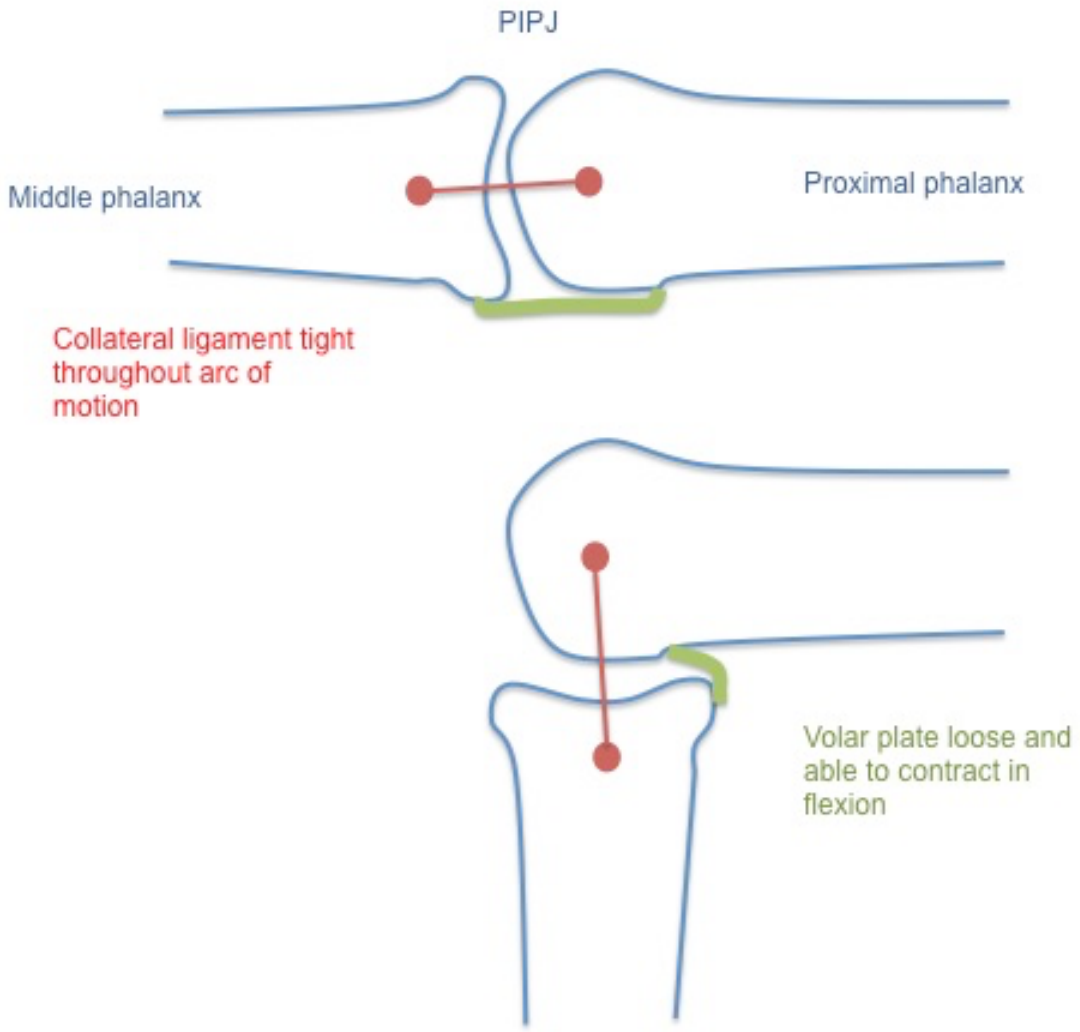

Figure 1.2: Proximal interphalangeal joint throughout its arc of motion

SUMMARY

Dupuytren's disease is associated with the palmar fascia of the hand. As there is a complex three-dimensional pattern to this structure, Dupuytren's disease may present as a range of different cords in a digit, and more than one cord may be present in a digit, with implications for both clinical assessment and treatment. The MCPJs and PIPJs respond differently to prolonged contracture and this may lead to differences in the expected treatment outcome of MCPJ and PIPJ disease.

1.3 Pathophysiology, aetiology and epidemiology in relation to outcome

Dupuytren's disease is characterised by progressive fibroproliferative changes associated with the palmar fascia, or aponeurosis, of the hand (Hurst, 2010). McGrouther points out that this entity is neither truly palmar, nor is it strictly fascia (McGrouther, 2005). Instead, it comprises the subcutaneous fibrofatty layer of the hand, deep to the glabrous skin and superficial to deeper structures such as the flexor tendons. The three-dimensional structure of the palmar fascia that influences the pattern of Dupuytren's disease has been discussed in the anatomy section. There are aspects of the pathology, aetiology and epidemiology that may influence the risk of developing Dupuytren's disease, the rate of progression and severity of any contractures that do occur. Numerous studies have investigated the influence of variables on prevalence of Dupuytren's disease as summarised by Hart and Hooper (Hart and Hooper, 2005), usually through cross-sectional and case-control study designs. However, only a few studies have investigated factors influencing a high risk of recurrence after treatment (Abe et al., 2004, Hindocha et al., 2006, van Rijssen et al., 2012). Identification of relevant factors that affect clinical outcome following treatment are important in trial design, as steps may then taken to match intervention and control cohorts, or to control for them. A selection of particular factors putatively influencing treatment outcome will be discussed here to identify which will be studied in this project. The likelihood that they influence functional outcome is

summarised (Table 1.1: Summary of likelihood that factors influence functional outcome).

Throughout any discussion of such scientific work, it must be appreciated that this is a very common, but late onset, condition. This poses a particular challenge for researchers conducting case-control studies. Obtaining true 'controls' from within a population itself at significant risk of developing the condition later is problematic – how does one know that a member of the 'control' group is not actually a pre-clinical 'case'? The answer may lie in obtaining large enough sample sizes such that these latent 'cases' within the 'control' group become a small minority. This may be possible for large, multi-centre studies investigating human genetics (Dolmans et al., 2011). However, much laboratory work is labour-intensive, and only capable of producing small sample sizes. This issue must be borne in mind when analysing studies involving such case-control comparison with a small sample size.

1.3.1 Studies of Dupuytren's diathesis

The logical starting point for identifying factors influencing outcome is to review studies of Dupuytren's diathesis. The term diathesis was first coined by Hueston (Hueston, 1963). It describes a group of factors that are associated with a higher likelihood of recurrence or extension of Dupuytren's disease after treatment. Recurrence has been defined as the reappearance of disease within the treated field, and extension as the appearance of disease in a neighbouring previously normal area (Leclercq, 2000). The

original factors proposed by Hueston included early onset of disease, ethnicity, family history of the condition, bilateral hand disease, and 'ectopic' lesions outwith the glabrous palmar skin. Different groups have conducted further analysis of diathesis factors. Abe and colleagues studied the factors associated with progression to revision surgery (Abe et al., 2004). In their study, early onset of disease, bilateral hand disease, ectopic lesions, radial sided disease and little finger surgery were associated with increased risks of recurrence and extension requiring revision surgery in their Japanese population. However, family history was not associated with either. The authors went on to generate a formula for calculating a diathesis score to risk-stratify preoperative patients. There are some criticisms of this paper. Firstly, it involved relatively small sample sizes (18 recurrences and 47 with no recurrence), and was based on retrospective case note review. Therefore, data completion rate and accuracy may have been compromised. The use of linear discriminant analysis is dependent on normal distributed data, in contrast to binary logistic regression (Spicer, 2004). Abe and colleagues do not define their data distribution sufficiently for us to be sure of the legitimacy of this approach, an omission that limits the reader's ability to support their conclusions. Importantly, the use of revision surgery to dichotomise recurrence/extension may not be suitable. Not all patients may choose to undergo revision surgery, even if their condition recurs, and some may not be fit for further surgery for unrelated reasons. Furthermore, the generalizability of a small Japanese study to Northern European population may not be

appropriate. Indeed, there is growing evidence of an underlying genetic predisposition to Dupuytren's disease amongst Northern Europeans (Dolmans et al., 2012). In keeping with this, Hindocha and colleagues' UK-based study identified family history as a factor associated with recurrence (defined as reappearance of palpable disease) (Hindocha et al., 2006). This larger study of 322 patients also identified male gender and age of onset below 50 years as factors associated with recurrence. Whilst this study may be more generalisable to the rest of UK NHS, there are still some limitations to the findings. The endpoint of reappearance of palpable disease does not describe a clinically meaningful endpoint, as will be discussed below. No patient would undergo further surgery simply for the reappearance of a nodule, in the absence of evidence that this reliably predicts future functional impairment or need for further treatment. Furthermore, it is not applicable to treatment modalities involving division of disease tissue (such as needle aponeurotomy) rather than excision of disease tissue. The particular surgical procedures that the patients had undergone are not discussed in the paper. If any had undergone needle or blade aponeurotomy, then this may compromise the validity of their findings. It is likely that the majority underwent fasciectomy, and while this is an excisional procedure, it is unclear whether recurrent disease activity can be reliably distinguished from postoperative scar tissue. Indeed, hand surgeons' diagnoses of Dupuytren's disease outstrip those by other doctors (Noble et al., 1984). Hindocha and colleagues' work thus may be of limited value when informing clinical practice.

A recent randomised controlled trial comparing fasciectomy to needle aponeurotomy from the Netherlands also included an assessment of factors influencing outcome (van Rijssen et al., 2012). In this paper, the definition of recurrence involved recurrent angular deformity, in contrast to reoperation as used by Abe and colleagues, or palpable disease reappearance as used by Hindocha and colleagues. The only variable associated with greater recurrence was young age (<50) at time of treatment. All other diathesis factors were not significantly associated with recurrence. These data are perhaps the most useful to date. However, the relatively small sample size of the study (93 patients) may have rendered their logistic regression underpowered to identify relevant factors associated with recurrence. Van Rijssen and colleagues' data still do not describe the factors associated with poor functional outcome after treatment, which might be expected to be most relevant to patients, treatment providers and treatment commissioners.

1.3.2 Disease stage

The formation of additional tissue in cases of Dupuytren's disease, rather than simply contraction of existing tissue, was first identified in the 19th century by Goyrand (Goyrand, 1833). When clinically detectable upon inspection and/or palpation, this tissue is described as nodules or cords. Inconsistently, at a microscopic level, the term 'nodule' is also used to describe a whorl of cells in diseased tissue (McGrouther, 2005). In the past half-century, significant resource and energy have been focussed on better understanding the biology

of this novel Dupuytren's disease tissue. Early in this line of work, Luck reviewed histology of tissue from over 200 affected hands in his widely-cited 1959 paper (Luck, 1959). His work remains the cornerstone of textbook discussions (McGrouther, 2005). He proposed a sequence of progression from a nodule to a cord, and he classified the disease into three stages: proliferative, involutional and residual. In the proliferative phase, nodules are composed primarily of disorganised fibroblasts, with little collagen. During the involutional phase, the fibroblasts align in the direction of the predominant tension. The cells mature, collagen production increases and contraction occurs, resulting in the classical joint contractures. At the same time, the nodule becomes less well defined, eventually completing its involution to leave a fibrous and relatively acellular cord in the residual phase. This sequence implies that cords may be relatively inert compared to the developing nodule.

The purported culprit cell, Gabbiani's 'modified fibroblast' (Gabbiani et al., 1971), is today referred to as a myofibroblast, although its absolute distinction from smooth muscle cells still remains poorly defined. Numerous markers have been suggested to separate the pair. However, as Hinz summarises, all have failed to stand up to rigorous examination (Hinz, 2007). Amongst others, the capacity to produce laminin was considered to be a smooth muscle cell capability, but the myofibroblast has since been shown capable of this *in vivo* too (Berndt et al., 1994).

The influence of disease stage on clinical outcome may be complex. Whilst logic suggests that those with more advanced contractures may have more severe disease, contracture severity did not correlate with genetic burden in a large study (Dolmans et al., 2012). Instead, preoperative contracture severity may be influenced by a variety of factors, including patient behaviour in terms of delayed presentation to health services. Furthermore, angular deformity does not correlate well with functional impairment (Engstrand et al., 2009, Jerosch-Herold et al., 2011, Zyluk and Jagielski, 2007). However, there are anatomical reasons why severe contractures may achieve poorer clinical outcomes in terms of contracture correction. This will be discussed in the anatomy section.

1.3.3 Age

Early age of onset of Dupuytren's disease has been proposed as a factor contributing to Dupuytren's diathesis in terms of recurrent surgery and recurrent palpable disease (Abe et al., 2004, Hindocha et al., 2006). Additionally, this is associated with carriage of high genetic risk alleles (Dolmans et al., 2012). However, it is unclear how reliable the data are. When a diabetologist assessed the prevalence of Dupuytren's disease amongst a British diabetes clinic population, it was reported to be half of the prevalence obtained when assessments were made by a hand surgeon (Noble et al., 1984). It might be expected that prior to diagnosis, patients' own identification of early disease (and thus their estimation of their age at onset) will be at least as inaccurate as that of a non-specialist doctor, and probably

much worse. As they may be much older by the time of questioning, they may also have issues with recall of their age at onset. As a result, this may be an unreliable factor to attempt to study. In contrast, age at surgery can be studied very accurately, and older age at surgery was associated with lower recurrence of angular deformity in a recent trial (van Rijssen et al., 2012). Patient expectation and level of activity is likely to vary with age, and so younger patients may be expected to have greater functional outcome demands. Understanding the age at treatment might also be of more use in estimating outcome when a patient attends for surgery. As a result, age at treatment will be studied here.

1.3.4 Previous Surgery

Attending for repeated surgery might influence outcome in several ways. Patients who have undergone surgery for Dupuytren's disease to a site may be expected to have more severe disease, which has progressed more rapidly requiring re-intervention. Moreover, further surgery may compound risks of complications and potentially reduce hand function. It is possible that requiring revision surgery at the same centre may also indicate that the first operation performed was inadequate, and therefore revision surgery by the same surgeon may pose a greater risk of inadequacy than average. However, a different surgeon at the same centre may have performed the first operation. Alternatively, the patient may have changed to a different service provider, either due to previous dissatisfaction, or simply due to changes in

referral pathways and waiting times. As such, it is likely to be a complex variable to interpret, but may be important to consider. There are also aspects of basic science that may be relevant to the influence of previous surgery on outcome from current surgery.

The sources of the transforming growth factor beta-1 (TGFβ1) that may drive myofibroblast differentiation in Dupuytren's disease are likely to include a variety of immune system cells, as reviewed by Al-Qattan (Al-Qattan, 2006). Langerhans cells, the skin's antigen-presenting cell, may coordinate this. Langerhans cells may be stimulated by interleukin-1 to migrate into Dupuytren's tissue from the epidermis and orchestrate this effect (Qureshi et al., 2001).

The role of TGFβ1 itself in the development of myofibroblasts has also been investigated. Whilst one study demonstrated elevated levels of TGFβ1 in Dupuytren's tissue compared to control tissues (Baird et al., 1993), examination of common polymorphisms in the TGFβ1 gene did not demonstrate significant differences between Dupuytren's disease patients and controls (Bayat et al., 2002), nor did examination of a novel polymorphism by the same group (Bayat et al., 2002). However, other *in vitro* work has suggested that Dupuytren's tissue fibroblasts may be more sensitive to the effects of TGFβ1 (Alioto et al., 1994). Furthermore, it has been proposed that recurrence of the condition after surgery may result from stimulation of quiescent fibroblasts by TGFβ1 (Bisson et al., 2003). Thus, it may be that the

fibroblasts and myofibroblasts of Dupuytren's patients that differ in their response to TGFβ1 compared to controls, rather than a difference in quantity or quality of the TGFβ1 itself in Dupuytren's patients. Furthermore, the potential mechanism by which surgery may actually stimulate recurrence may be of relevance to rapid recurrence observed in some cases.

1.3.5 Family history and surrogate markers of genetic risk

A genetic component to the aetiology of Dupuytren's disease is increasingly supported. A large multi-centre, international, genome-wide association study has implicated nine different genome loci as being associated with Dupuytren's disease (Dolmans et al., 2011). Six of these nine are involved in the Wnt signalling pathway (Dolmans et al., 2011). Wnt proteins are thought to contribute to processes as varied as embryogenesis and carcinogenesis, as reviewed by Logan and Nusse (Logan and Nusse, 2004). It is striking that all such situations involve cell proliferation and its regulation, so it is feasible that Wnt dysfunction might contribute to Dupuytren's disease development. It has been proposed that in Dupuytren's disease, Wnt's action on frizzled receptor proteins drives beta-catenin production, which may stimulate cell proliferation (Dolmans et al., 2011). Thus a genetic-based mechanism for driving fibroblast proliferation may be present. Further work by the same group has suggested that patients with an early age of onset (<50), knuckle pads, and a family history of Dupuytren's disease are more likely to carry these risk alleles (Dolmans et al., 2012). This suggests that such clinical features may be

surrogate markers of high genetic risk of developing Dupuytren's disease. As has been discussed, such features have been shown to be associated with recurrent surgery (Abe et al., 2004), and recurrence of palpable disease (Hindocha et al., 2006), but only age at treatment was associated with recurrence of angular deformity in a recent randomised controlled trial (van Rijssen et al., 2012). It is not clear whether genetic factors would influence functional outcome.

1.3.6 Ethnicity

The fact that the prevalence of Dupuytren's disease is reported to be highest in men of Northern European origin is widely quoted. Identification of this is attributed to Early's 1962 paper, where he attributes its geographical distribution to the historical migration of people of Norse origin (Early, 1962). Other factors such as climate, or behaviour might be similar between such places and account for any higher prevalence. Conversely, if it were the Norse genetics that were causative, then the term 'Northern European' may not be suitable, as it would include other genetic lineages beyond simply Viking genetics. Furthermore, the condition is not purely one seen in those of Norse origin. Dupuytren's disease has been reported in other geographical areas, such as Bosnia, (Srivastava et al., 1989, Zerajic and Finsen, 2004). Such data demonstrate its occurrence in a population with different genetic heritages, as well as different climates and behavioural norms. Some variation between ethnic groups within a region may be present. For example,

Zerajic's group investigated the incidence of Dupuytren's disease in a sample of 1207 members of the population in Bosnia (Zerajic and Finsen, 2004). There was a significantly lower prevalence of Dupuytren's disease in Muslim men compared to Serbs or Croats. Whether this truly represents a genetic difference is not clear. The same difference was not seen in the women evaluated. This may represent a genetic variation in susceptibility to the influence of androgens in driving Dupuytren's. Certainly, a role for androgens at a molecular level has been investigated (Pagnotta et al., 2002, Pagnotta et al., 2003). However, there may be other explanations. There may be socioeconomic variation between men of different groups, for example in terms of rates of heavy manual occupation. Alternatively, there may be cultural differences. For example, alcohol consumption may be higher in Serb and Croat men compared to Muslims and women. Such confounding variables limit the reliability of the findings of epidemiological work in general.

Srivastava and colleagues described Dupuytren's disease in a case series of ten people of Indian origin (Srivastava et al., 1989). At the time, this was the first report of the condition on Indian Asians. However, the cohort was all treated in the West Midlands in the United Kingdom. Whilst this might demonstrate the presence of a genetic cause of Dupuytren's disease in Indians, which has only been recognised in a free at the point of use healthcare service in the United Kingdom, alternatively it may be simply under-reported in India. Another explanation is that it may support a strong

influence of climate, or cultural or social change, in terms of manual work or alcohol consumption.

Saboeiro's work standardised variables to allow inter-racial comparison. The group compared the rate of surgery for Dupuytren's disease between ethnicities in Veterans Affairs (VA) hospitals in the United States of America (Saboeiro et al., 2000). This demonstrated that surgery rates were highest in Caucasians, followed by Hispanics with lowest rates in African Americans. Whilst this appears to be a controlled comparison on the basis of ethnicity, there are numerous confounding factors. Whilst the VA services are free at the point of use, there may have been cultural and socioeconomic differences between racial groups affecting their uptake of services by veterans with Dupuytren's disease. Finally, there is no fixed indication for surgery for Dupuytren's disease. This racial variation could be accounted for by institutional bias in treatment on the basis of race. Consequently, whilst this study design provides data on a large cohort, and eliminates some sources of bias, it is affected by several other potential biases. Thus, the epidemiology of Dupuytren's disease is not as clear as may be thought.

1.3.7 Gender

It is accepted that Dupuytren's disease is more common in men, and that its prevalence increases with age. Early's paper suggests that the prevalence in British people over the age of 75 may be as high as 18.1% in men, and 9% in women (Early, 1962). Certainly, a difference in prevalence between sexes

would fit with Pagnotta and colleagues' work on the potential role of androgens in driving Dupuytren's disease at a cellular level (Pagnotta et al., 2002, Pagnotta et al., 2003). This difference between sexes balanced out in old age in one widely-cited study (Ross, 1999). In contrast, whilst the difference between sexes was also seen in Zerajic's study from Bosnia, the difference was seen in every decade, and was still marked in those over eighty (Zerajic and Finsen, 2004). This may be a racial difference or there may be other contributory factors. Studying Dupuytren's disease in very elderly patients is not straightforward; Gudmundsson's group demonstrated increased mortality in Dupuytren's disease patients, even once several confounders (such as smoking) were accounted for (Gudmundsson et al., 2002). Their study of Iceland's death registry suggests that there may be an increased cancer incidence in those with the condition, which might explain such a finding. The mechanistic explanation of this is not known. Nevertheless, if increased mortality were proportional to Dupuytren's disease severity, or androgen-dependence, then the difference in Dupuytren's prevalence between men and women might appear to decrease with age, due to relatively higher survival of women with the condition, whilst men with condition are more likely to have died. Whilst gender differences in incidence exist and can potentially be explained, it is unclear whether gender would influence functional outcome specifically.

1.3.8 Diabetes, alcohol and smoking

Other factors may also contribute to pathophysiology. The presence of free radicals may stimulate fibroblast proliferation, which have been demonstrated to increase fibroblast density in Dupuytren's tissue cultured *in vitro* (Murrell et al., 1987). Free radicals could arise from the impact of exogenous factors, such as alcohol, trauma and diabetes mellitus, as well as from ageing. In his review article covering pathogenic factors, Al-Qattan suggests that such agents may give rise to free radicals by stimulating the conversion of xanthine dehydrogenase to xanthine oxidase (Al-Qattan, 2006). No evidence is cited to support this suggestion. If such an effect does exist, it might provide a mechanism for the potential (and controversial) aetiological effect of such factors. There are numerous potential associations with Dupuytren's disease that may relate to free radical stress: diabetes mellitus, tobacco and alcohol usage, and even epilepsy. Not only might this free radical pathway affect disease progression, but also it may influence other aspects of functional outcome and recovery.

Geoghegan and colleagues have shown that diabetes mellitus is associated with Dupuytren's contracture, even following logistic regression to adjust for the influence of consulting behaviour (Geoghegan et al., 2004). The association between Dupuytren's disease and diabetes mellitus has also been suggested in numerous earlier studies, such as that by Chammas and colleagues (Chammas et al., 1995).

In Geoghegan's study, the strongest association was between insulin therapy and Dupuytren's contracture, compared to other diabetes treatments. By demonstrating that medically controlled diabetics are more likely to have Dupuytren's disease than diet-controlled diabetics, Geoghegan suggests that diabetes severity may contribute to the likelihood of developing Dupuytren's disease (Geoghegan et al., 2004).

If studying the relationship between Dupuytren's disease and well-defined comorbidities such as diabetes mellitus is challenging, then studying some of the other purported risk factors for Dupuytren's may prove even more difficult.

Tobacco and alcohol use correlate with each other, as reviewed by Bien (Bien and Burge), thus population studies that attempt to examine either individually must account for this. Alternatively, both might be examined together, given that they are thought to contribute to Dupuytren's pathogenesis through the same mechanism of free radical generation (Murrell et al., 1987).

Godtfredsen and colleagues recruited patients from the Copenhagen Heart study to investigate the relationship between alcohol and tobacco intake and Dupuytren's disease. They found a dose-dependent association for both variables (Godtfredsen et al., 2004). However, their data were collected by nurses and medical students, and so may be subject to the same poor sensitivity as that described by Saboeiro and colleagues (Saboeiro et al., 2000). Furthermore, the non-expert assessors were trained specifically to identify contractures affecting the ring and little fingers. This strategy might

overlook patients whose only present disease was radial, as is widely known to occur in diabetics. Consequently, there is a risk that the accuracy of the data obtained was doubly inaccurate. Also, the measure of alcohol and tobacco intake was based on self-reports by patients. There is thus no guarantee that this variable has been accurately determined. This raises the issue of reliable assessment of variables with potential stigma attached to them, such as alcohol and tobacco consumption, where a social desirability response may occur.

1.3.9 Epilepsy

Epilepsy has been suggested to be an association since the 1940s (Lund, 1941). This has remained controversial since. For example, one large study of Dupuytren's patient from Germany has suggested that Dupuytren's patients with epilepsy are more likely to develop worse contractures (Loos et al., 2007). In contrast, no increased odds ratio of having Dupuytren's disease was seen in epileptics in Geoghegan and colleagues' study of the United Kingdom General Practice Research database (Geoghegan et al., 2004). Furthermore, epilepsy itself may not cause Dupuytren's disease, but may be a confounding factor; certain antiepileptic medications are purported to contribute to is development based on different studies (Al-Qattan, 2006, Lund, 1941), in particular phenobarbitone (Critchley et al., 1976).

Geoghegan and colleagues' work looked at epilepsy-associated factors. They demonstrated no association of specific antiepileptic medications being

associated with Dupuytren's disease. Lund had identified phenobarbitone, a barbiturate, as being particularly associated with Dupuytren's (Lund, 1941). Geoghegan did not demonstrate statistical significance of such an effect, and could not find any mechanistic explanation as to why it might contribute to Dupuytren's pathogenesis, and so disregards it (Geoghegan et al., 2004).

However, Al-Qattan's later review of pathogenesis does provide a potential mechanism. It suggests that it may be due to increased lysophosphatidic acid levels from altered cholesterol metabolism, although without adequate justification of this suggestion (Al-Qattan, 2006). An *in vitro* study by Rayan and colleagues provides the explanatory link missing from Al-Qattan's algorithm (Rayan et al., 1996). They studied Dupuytren's myofibroblast contraction using a collagen matrix – a similar technique to that used to study contraction by Brown and colleagues (Brown et al., 1998). They demonstrated a dose-dependent myofibroblast contraction in response to lysophosphatidic acid (Rayan et al., 1996). Additionally, a study by Tripoli and colleagues considered cases of contractures in patients treated with phenobarbitone (Tripoli et al., 2011). Three cases are considered that illustrate a potential dose-dependent effect of phenobarbitone on recurrence following surgery for Dupuytren's contracture. Regression of early recurrence was observed in two patients who had a dose reduction of the drug and changed to carbamazepine. The third patient, who continued phenobarbitone, continued to develop recurrence (Tripoli et al., 2011). Unfortunately, as a very small and retrospective observational study, little value can be placed on

these findings alone. Phenobarbitone represented the highest odds ratio for association with Dupuytren's disease (2.67 with 95% confidence intervals of 0.93 and 7.69). It is highly likely that this study was underpowered given the small sample size, as evidenced by the broad confidence intervals, but it may merit further attention.

Furthermore, many of the more specific problems with the larger studies that have been discussed all arise due to frequent basic flaws. One is that the datasets being used to study Dupuytren's disease were actually gathered for another purpose. This can be seen with Saboeiro's study investigating race (Saboeiro et al., 2000), in which data were gathered from assessments by different clinicians in Veterans' Affairs hospitals. It is also the case with Geoghegan and colleagues' work (Geoghegan et al., 2004), which used GP records. Godtfredsen and colleagues' work used assessment of nurses and medical students trained to identify disease in the fourth and fifth rays (Godtfredsen et al., 2004). In these three studies, diagnostic inaccuracy becomes problematic, as has been demonstrated by Noble and colleagues (Noble et al., 1984). Also, Gotfredsen's paper used patients from the Copenhagen City Heart Study (Godtfredsen et al., 2004). This gives rise to issues with selection bias, as well as inaccurate diagnosis of Dupuytren's disease (with its resulting impact on false negatives within the control cohort). One clear way to avoid such issues would be to carry out a well-designed, multi-centre population study specifically intended to study Dupuytren's disease.

1.3.10 Occupation

Manual labour has also been controversially associated with the development of Dupuytren's disease. The first suggestions of this came from the historical works of Cooper and Dupuytren himself, both of whom considered occupational factors as contributory to the pathogenesis (Cooper, 1822, Dupuytren, 1834). Examination of this potential association extends beyond medical interest alone; a causative association between specific occupations and Dupuytren's disease would be of political and economic interest. Indeed, a United Kingdom government report examined the potential link a century ago, and found no evidence of an association (Collis and Eatock, 1912). As a result, Dupuytren's disease was not entered onto the list of occupational diseases for which compensation could be claimed. Proving or disproving a genuine causative association with manual work may be challenging: the type and duration of 'trauma' experienced by the glabrous skin of the hand is likely to vary between specific occupations, and between workers within a particular group. Additionally, pastimes and changes in occupation may act as confounding factors. The growing evidence of genetic associations with the disease provides an alternative, albeit a non-mutually exclusive, aetiology. However, the association with other fibroproliferative conditions such as Ledderhose's disease in the soles of the feet and Peyronie's disease in the tunica albuginea of the corpora cavernosa in the penis (Leclercq, 2000), suggests that a systemically-acting causative factor is at least partially responsible, rather than simply manual labour alone.

1.3.11 Summary of candidate factors

There is not currently evidence to support the roles of all of the factors that comprise the Dupuytren's diathesis. Furthermore, some of the factors that are associated with disease development, progression or even recurrence, may not contribute to poor outcome when defined in terms of function. The factors reviewed are summarised in Table 1.1.

Candidate factor	Likelihood of influencing functional outcome	Comment
Disease stage	+/-	Poor correlation between deformity and existing function PROMs. Advanced disease may be less functionally limiting.
Age	+	Younger patients expected to have higher functional demands
Previous surgery	+	Expected to adversely affect function
Family history	-	May influence incidence, but unlikely to influence functional outcome. Surrogate markers such as knuckle pads relatively straightforward to assess.
Ethnicity	-	Difficult to define in a cosmopolitan society, and to distinguish from environmental and social confounders. Both this and family history may be phenotypes of underlying genetic tendencies,

		which are becoming more apparent.
Gender	-	Gives rise to differences in incidence, but not for difference in functional outcome
Diabetes mellitus	+	Likely to affect disease progression, complications and thus function.
Alcohol intake	+	Likely to affect disease progression, complications and thus function. May be subject to social desirability response posing a challenge to accurate recording
Smoking status	+	Likely to affect disease progression, complications and thus function. May be subject to social desirability response posing a challenge to accurate recording.
Epilepsy	-	Unclear whether this influences disease incidence, and probably does not influence functional outcome
Occupation	-	Unclear whether this does influence disease progression. It is likely to be confounded by differences

	between occupations and by other activities (e.g. hobbies). Unlikely to affect functional outcome in a predictable manner

Table 1.1: Summary of likelihood that factors influence functional outcome

Key:

+ Probably influences functional outcome in reliable manner

- Probably does not influence functional outcome in reliable manner

+/- May influence functional outcome, but not in a reliable manner

SUMMARY

Whilst some studies have investigated factors that might influence Dupuytren's diathesis, there are no data describing which factors influence functional outcome following surgery. Of the candidate factors discussed here, some are more likely to affect function in a predictable manner, and are amenable to being assessed. Some factors that are traditionally associated with disease development or progression, such as ethnicity, may not affect functional outcome in those with Dupuytren's disease.

Others, such as diabetes, may affect disease progression in terms of postoperative recurrence, but also affect risk of complications and so are likely to affect functional outcome.

1.4 Clinical Assessment & Outcome Measures

Dupuytren's disease is diagnosed and assessed through history and clinical examination. There are no further investigative tests specific to making this diagnosis, or to assessing severity, in either clinical practice or research at present. However, there are different approaches to the clinical assessment of Dupuytren's disease. As a result, there are several potential measures that could constitute the optimal technique for assessing outcome following treatment.

1.4.1 Context of assessment

Making the diagnosis involves history and examination, to identify the development and presence of palpable nodules and cords, with or without resulting joint extension deficits (Hurst, 2010). Rarely, imaging techniques such as x-ray or magnetic resonance imaging may be used when associated conditions, such as arthritis, are suspected (Hurst, 2010), but no specific diagnostic tests exist. Once the diagnosis is made, an assessment of the 'severity' of the condition could be made using a number of strategies, ranging from patient-reported history of limitation, to objective measures of angular deformity. The measures used will be required to assess early outcome following treatment, and also long-term outcome, particularly as recurrence may occur following even 'successful' treatment. Both are of importance for research purposes as well as clinical assessment of follow up cases. As a condition that is not terminal, other commonly used biomedical research

endpoints, such as mortality, are not relevant to Dupuytren's disease. Instead, assessments of disease state, morbidity and cost effectiveness are required. However, Dupuytren's disease is a slowly progressive condition, and so identifying a late outcome such as recurrence itself may require an impractically long follow-up for a study. Thus, an outcome measure that predicts later deterioration might be advantageous, but this would need to be appropriately validated.

The situation is made more complicated by several variables. Patients' opinion and experience may influence behaviour. As a non-terminal condition, some patients may elect to decline treatment for primary disease. Alternatively, patients who experience recurrence or complications may refuse further intervention.

Furthermore, surgeons' behaviour may influence treatment. Treatment might be employed to prevent loss of function, rather than to treat it. For example, it is considered that outcome from treating PIPJ contractures is often worse than MCPJ contractures, possibly due to the anatomical differences in ligament arrangements of these joints, as previously discussed. There is a lower recommended threshold for treating PIPJ contractures, with surgery being considered for even mild contracture (BSSH, 2008). Thus the surgeon might offer treatment at an early stage for a PIPJ contracture, to prevent subsequent worsening functional impairment that cannot be corrected. As a result, if a patient subsequently develops a contracture, it is not classified as a

'recurrence', as he or she is experiencing a first episode of function loss. However, it is still not clear whether a significant proportion of patients undergoing NHS treatment at present are being treated 'prophylactically' based on medical advice rather than symptoms. This will be investigated as part of this thesis. A further question is whether such prophylactic treatment is cost effective, and therefore whether it should be funded.

1.4.2 Primary outcome measure for future research

In a randomised controlled trial of Dupuytren's disease surgery, a variety of assessments would be included as secondary outcome measures. However, there is a need to select an appropriate primary outcome measure. This will be the measure for which the study would be powered to identify a clinically significant difference. This thesis will investigate some of the relevant parameters of validity and reliability to assist the selection of a future primary outcome measure.

The main long-term outcome described in Dupuytren's disease is 'recurrence' of the condition. However, the definition of this is not standardised and varies widely (Ball et al., 2013, Becker and Davis, 2010, Rodrigues, 2010). Recurrence rate is reported in terms of recurrence per joint treated, per ray operated upon, per hand, and in others per patient.

Furthermore, the length of follow up employed in studies is not standardised and has been shown to vary greatly (Rodrigues, 2010). As Dupuytren's

disease is a progressive condition, a gradual increase in 'recurrence rate' is to be expected with a longer follow up. At the same time, lengthy follow up studies will delay translation into clinical practice, and may affect factors such as attrition from participant dropout.

'Recurrence' in Dupuytren's disease typically refers to the proportion of patients successfully treated at the early assessment, that has gone on to experience a deterioration. However, it should be borne in mind that a proportion of patients experience treatment failure at the early assessment, and that these patients are not accounted for in recurrence rates. This proportion can be large – in one recent randomised controlled trial of collagenase therapy for Dupuytren's disease, 36% of those treated failed to reach the primary endpoint (Hurst et al., 2009). Any subsequent description of this patient cohort should account for the fact that over a third of those treated had a suboptimal initial result, and will thus continue to have a poor outcome at the late follow up, regardless of what the recurrence rate is for those successfully treated.

Thus, there is currently no accepted primary outcome measure for future research, and no consensus for when it should be assessed. Furthermore, in future, it may be more meaningful to consider the rate of poor outcome following intervention rather than recurrence in those successfully treated.

The choice of primary outcome measure is inextricably linked to the main question that a study would set out to answer. Previous studies have

measures such as initial correction of angular deformity as the primary early outcome measure, with recurrent deformity as the primary late outcome measure (Rodrigues, 2010). Such measures can answer questions about whether treatment options straighten a digit effectively, and how durable this straightening is. However, this may not correlate with improvement in either hand function or quality of life. Such data are required to for calculating quality adjusted life years (QALYs), which are central to the calculation of cost effectiveness endpoints such as the incremental cost effectiveness ratio (ICER) of introducing a novel treatment such as collagenase. As a result, these metrics are of limited use in the kind of cost effectiveness analysis currently being employed in the UK by NICE (NICE, 2008). For such processes, measures that assess health-related quality of life and hand function would be more useful, and answer questions investigating whether a treatment option has improved function or quality of life. As discussed above, it is more informative to describe such changes across an entire study population, rather than just those for whom a good initial outcome was obtained. Asking this genre of question, and employing the appropriate outcome measures needed to do so, might not only facilitate successful approvals for cost-effective Dupuytren's disease interventions, but would also provide clinically relevant data that could inform patient choice.

1.4.3 Assessment strategies

A range of different options could be chosen as the primary outcome measure for future research. As has been discussed, a range of strategies have been

employed previously (Becker and Davis, 2010). During work that the candidate conducted for an MSc previously, these were analysed for 33 studies describing outcome from surgery for Dupuytren's disease (Rodrigues, 2010). This dissertation was submitted in 2010, with the search performed that year. The primary outcome measures in these studies are shown in Figure 1.3. All studies considered 'recurrence', but in over a third, no definition of recurrence was given. The most popular definition was reappearance of palpable disease in the operated field, which was employed by just under a third of studies. This definition is cited from two different sources, most frequently Leclercq's chapter in Tubiana's book from 2000 (Leclercq, 2000), and also Hueston's 1984 article (Hueston, 1984). As many citations refer to Leclercq's definition, it shall be referred to as this here. Besides these options, all other studies described some form of loss of extension or range of motion, be it self-report by the patient using diagrams of increasing deformity (Dias and Braybrooke, 2006), or angular measurements using a goniometer. The relative merits and disadvantages of each of these, as well as other options, must be appreciated.

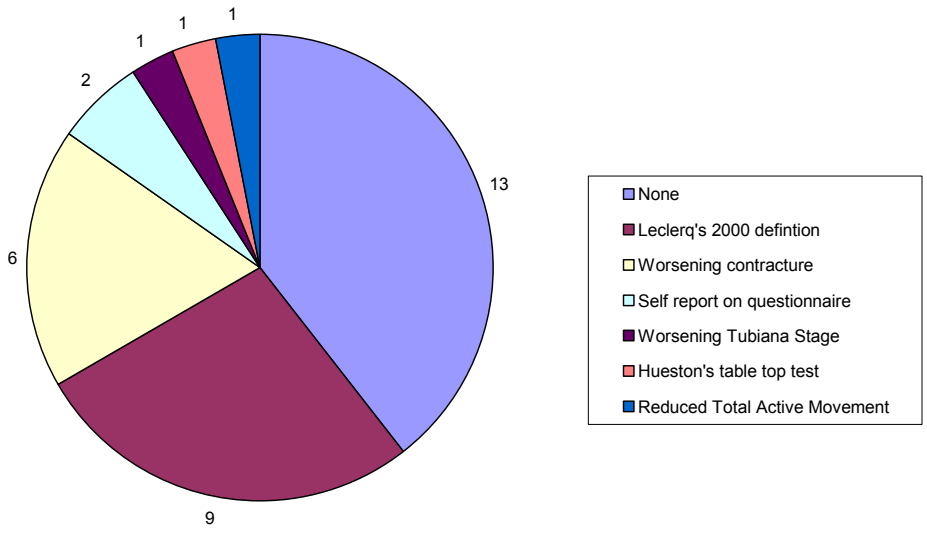

Figure 1.3: Definitions of recurrence used as primary outcome measure in studies of surgery for Dupuytren's disease.

Reproduced from MSc dissertation (Rodrigues, 2010)

1.4.4 Assessing palpable disease

Hueston and Leclercq define recurrence as the reappearance of Dupuytren's disease in a zone previously operated on (Hueston, 1984, Leclercq, 2000). In contrast, Leclercq defines extension as the appearance of Dupuytren's disease in a zone previously unaffected (Leclercq, 2000). This accounted for the primary outcome measure in 27% of studies included in the literature review discussed already (Rodrigues, 2010). However, in even more studies, no definition of recurrence was provided at all. A major limitation of Leclercq's

definition is that it is not applicable to techniques that involve division of Dupuytren's tissue (such as needle aponeurotomy or collagenase) rather than its excision, in that the diseased tissue is not removed, and so remains palpable throughout, even in 'successfully' treated cases. Thus, this definition lacks rigour for future comparative studies involving these techniques. Furthermore, where does the 'operated field' end? This might be apparent for a dermofasciectomy where a scar can be seen at the junction of the graft and the glabrous skin, but is not so after fasciectomy, where the skin has been elevated and undermined and closed back in place. How far does the treated subcutaneous field actually extend? Another issue with Leclercq's definition is whether it is a relevant outcome at all. No patient should be re-operated on simply for the reappearance of nodules in the operated field alone. Reappearance of nodules might predict subsequent functional deterioration, thus Leclercq's definition might provide a 'marker' of impending poor outcome. However, this assumption is not proven. Finally, Leclercq's definition is not robust enough for high quality research. It is possible that it was never meant to be; the definition is taken from the prose of a textbook, from the very opening paragraph on recurrence, rather than from the validation studies that other assessment tools have been subjected to. Identifying recurrent nodular change of disease in a scarred post-operative field is not straightforward, and distinguishing this from acceptable post-operative scarring is likely to be biased and subjective. As different procedures result in distinct patterns of scarring, it is very difficult to blind an assessor to the intervention that the

patient has undergone without covering the scarring on the palm and digit. This in turn may influence how readily recurrence is palpable. This may contribute to the wide variation in reported recurrence rates seen in different studies. Other issues may contribute to this apparent difference, as have been discussed already.

1.4.5 Assessing loss of extension

In clinical practice, loss of extension is typically assessed as angular deformity using goniometry, or Hueston's tabletop test (see page 56). Surgery tends to be performed to correct this loss of extension (BSSH, 2008). An assessment of the redevelopment of extension loss is a common endpoint to use for research purposes (Rodrigues, 2010). There are different ways to quantify this: angular deformity can be measured at joints, or linear loss of extension of the entire digit can be assessed.

1.4.5.1 Assessing angular deformity of joints

Common practice includes using goniometry to quantify the degree of joint contracture, which is expressed as the angle in degrees away from full extension at the joint concerned (Tubiana, 2000). This can be performed in many different ways. Indeed, a review of outcome measures in Dupuytren's disease surgery has shown that 17 categories of angle-based measure have been reported in 91 studies in the literature (Ball et al., 2013). Some of these measures account for dynamism, which may be observed if contractures span

more than one joint of a digit (Hurst, 2010), but others do not. Dynamism may be an important influence to

consider when interpreting angle-based data. If a patient has a contracture of both the MCPJ and PIPJ of the same finger, and the PIPJ angle is being measured, the position of the MCPJ will affect the angle of the PIPJ. If the MCPJ is held in maximal extension, then the volar soft tissues (including the Dupuytren's cord) will be relatively tightened, flexing the PIPJ. However, if the MCPJ is held flexed, the volar soft tissues are relatively lax, allowing more correction of the PIPJ contracture. This principle is illustrated in Figure 1.4.

Amongst the range of objective measures reported, total passive extension deficit, often abbreviated to TPED, can be obtained from the sum of the passive extension deficit of the joints of the digit. Dupuytren's disease typically affects the MCPJ and PIPJ, but may rarely affect the distal interphalangeal joint (DIPJ) as well. Whether the MCP and PIP joints have been assessed, or whether all three finger joints are measured, is not always specified. Indeed, one of the most recent randomised controlled trials of treatment discusses total passive extension deficit of each joint individually (van Rijssen et al., 2012). Furthermore, some goniometers may not be able to span the distal interphalangeal joint due to the relatively short length of the distal phalanx.

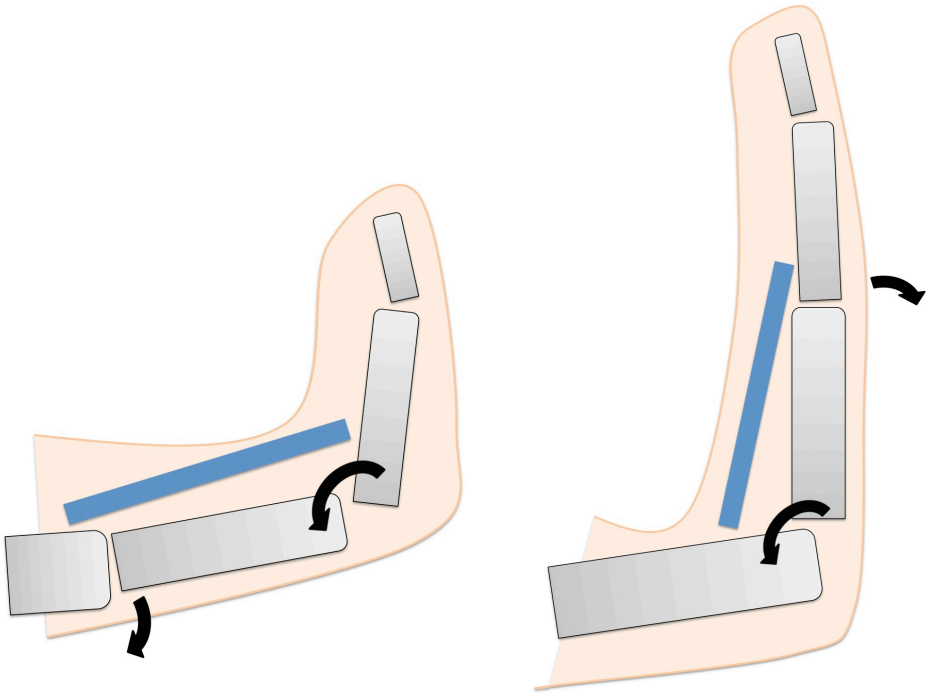

Figure 1.4: Schematic diagram of contracture dynamism

The Dupuytren's cord is illustrated as the blue strip, and is inelastic. On the left, the MCPJ is extended, resulting in increased flexion at the PIPJ. On the right, the MCPJ is flexed, allowing correction of the PIPJ contracture

Tubiana's classification uses TPED to identify four stages of increasing contracture. However, as TPED is achieved by adding separate, optimised passive MCPJ and PIPJ contracture angles, the TPED itself is minimised, as the 'best possible' passive angles are obtained for each joint. The patient cannot actually straighten the digit as much as this value suggests. It is thus not a 'real-world' measurement, and not necessarily clinically relevant. Furthermore, TPED only accounts for extension loss at the MCPJ and PIPJ. If

the patient develops a stiff DIPJ, which might affect grip, this is not incorporated in the measurement.

Angular deformity can also be assessed without goniometry, by using pictures of contracted digits. Patients then select the image they believe corresponded to their preoperative state and the image that corresponds to their current state. This approach was adopted for a postal evaluation of surgery outcome for Dupuytren's disease published by Dias and Braybrooke (Dias and Braybrooke, 2006). This avoids some of the disadvantages of measuring angles with goniometers or other devices. It avoids the inconvenience and cost of a clinic appointment for assessment. It also allows retrospective data collection, with the patient self-reporting their preoperative state. It also attempts to assess a more clinically relevant endpoint – the active extension that the patient can achieve (rather than the passive extension obtained by an assessor). However, the reliability of the data obtained with this strategy is unclear. How accurately and objectively patients assess their current state is not clear, and recall bias may be an issue when attempting to measure preoperative state. Given the lack of validation data describing this technique, it has not been studied here.

Besides TPED, the range of motion can be calculated as the difference between the joint angle in maximal flexion and this extension angle. This also provides an assessment of flexion. The importance of this should not be underestimated. Whilst treating Dupuytren's disease aims to restore loss of

extension, if this has the side-effect of significantly affecting flexion function, then the patient may experience a deterioration in hand function overall.

An assessment of angular loss of extension is often the primary endpoint of research. This is demonstrated by considering two of the most recent high profile randomised controlled trials in Dupuytren's disease treatment. In van Rijssen and colleagues' trial of limited fasciectomy versus needle aponeurotomy, TPED was used (van Rijssen et al., 2012). Hurst and colleagues used reduction in individual joint contracture to less than five degrees (Hurst et al., 2009). The accuracy of these measurements needs to be assured. Goniometry has been demonstrated as having a resolution of 5 degrees (Boone et al., 1978). A similar level of accuracy was found in a study of reproducibility of goniometry performed by surgeons, conducted for the candidate's MSc thesis (Rodrigues, 2010). There are disadvantages to the use of goniometry. Firstly, whilst the measurements can be documented, they cannot be reassessed for validation purposes. The assessment is a one-off, and the contracture is likely to change with time, and so the exact circumstances in which the recording is made will have changed. Other factors may also impair goniometry results, such as infection, misshapen hands or post-operative swelling (Hamilton and Lachenbruch, 1969). One solution that allows reassessment is to obtain digital photographs of the hand from a lateral aspect and then perform angular measurements on the photograph. This has been described by Georgeu and colleagues (Georgeu et al., 2002). However, this particular study used a large jig to fix both the

hand being measured, and also the camera itself. As such, the equipment does not appear to be easily portable from the photograph shown in the paper. The measurements themselves required specific computer software, adding to costs. It is also unclear how measurements of joint angles can be made on the middle and ring fingers, when these may be obstructed from view by the index and little digits. A further issue applies to all such angular measurements; whilst a change in contracture of greater than 5 degrees might be detectable by a goniometer, it may not represent recurrence rather than extension. This may not matter. If hand function deteriorates following treatment, it is unlikely that the semantics of this distinction will concern the patient's appraisal of their condition.

It is not clear whether angular deformity is truly clinically relevant; function may not necessarily be impaired by the development of a contracture of an arbitrarily defined severity, such as 20 degrees. Draviaraj and Chakrabarti reported that improvement in contracture correlated with improvement in hand function as assessed using the Sollerman score, and the correlation between functional improvement and PIPJ angular improvement in particular was most pronounced (Draviaraj and Chakrabarti, 2004). However, a systematic review of outcome assessment in Dupuytren's disease demonstrated that this particular outcome tool has only been used in Dupuytren's disease by this team, and in only one other Dupuytren's study from the same institution (Ball et al., 2013, Sinha et al., 2002). Its validation is not as well established as that of other measures (Ball et al., 2013). In other studies, angular deformity only

correlated weakly with the most commonly employed patient-reported outcome measure used in Dupuytren's disease research: the Disabilities of the Arm, Shoulder and Hand (DASH) tool (see page 59) (Ball et al., 2013, Jerosch-Herold et al., 2011). A common interpretation of this was to suggest that the DASH is not a valid assessment of functional limitation in Dupuytren's disease (Packham, 2011). However, the relationship between angular deformity of the digits and functional limitation may not be linear. It is possible that maximal limitation is the result of having approximately 90 degrees of extension deficit, and that having the fingertips closer to the palm may be associated with less functional impairment, as illustrated in Figure 1.5.

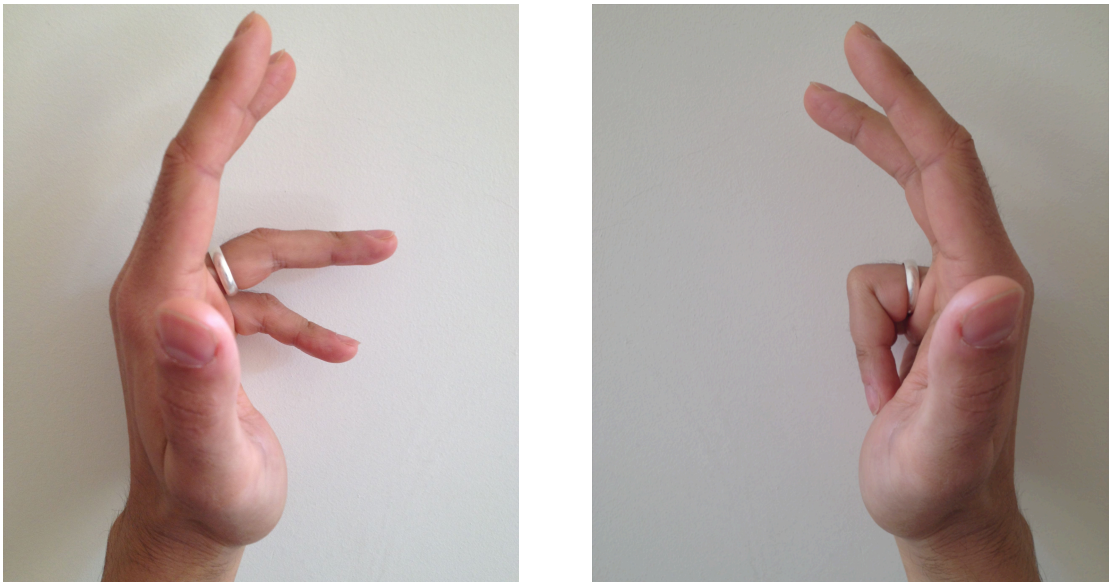

Figure 1.5: Photographs illustrating 'moderate' and 'severe' contractures

The contracture on the left is 'less severe' than that on the right, as defined by extension deficit. However, hand function may be better with the posture on the right, as the ring and little fingers can be bypassed to allow an effective tripod grip using the thumb, index and middle fingers.

Furthermore, attempting to fully correct angular deformity may cause complications, such as a reduction in active flexion, and so might paradoxically impair hand function. As a result, the conclusion that angular deformity is the gold standard for approximating functional impairment is not currently supported by evidence. Comparing a range of measures, rather than just two, is required to arrive at such a conclusion. When the two measures differ, as was the case is the studies discussed, it is not clear which is 'right' and which is 'wrong'.

In addition, functional impairment may arise from Dupuytren's disease that does not cause angular deformity of MCPJs or PIPJs. For example, a thumb-index web cord may limit thumb abduction and thus grasp. Alternatively, natatory cords preventing finger abduction may limit the span of the palm. This may be of consequence to those who use alphabetical or musical keyboards. Neither of these problems is necessarily represented by angular deformity.

1.4.5.2 Assessing linear loss of extension

Mäkelä and colleagues described an approach measuring loss of extension as a linear measurement perpendicular to the metacarpal, though no detailed method for performing the measurement was described (Makela et al., 1991). A portable jig was designed to do this as part of previous work by the candidate, and is shown in Figure 1.6 (Rodrigues, 2010). Mäkelä and

colleagues also quantified flexion function by measuring active flexion as the distance that the patient can flex to, short of the pulp of the fingertip reaching the distal palmar crease. This provides a mirror 'loss of flexion' measurement as well as the loss of extension measurement. As discussed with angular range of motion, flexion function is also likely to be important in hand function. One advantage of Mäkelä and colleagues' strategy is that it may minimise the impact of dynamism of a contracture. If a contracture crosses both the MCPJ and the PIPJ, then the position of one joint will affect the passive extension of the other (Hurst, 2010). For this reason, goniometry should always be performed with the other joint held in maximum passive flexion, as discussed. This in turn gives the least contracted value for the joint being measured. However, whether all assessors reliably perform this is not clear. Additionally, the passive angle obtained is not what the patient can achieve actively. By measuring the entire digit's active loss of extension as a linear measure, the influence of contracture dynamism is standardised. Indeed, as the digits are extended actively, it standardises dynamism at a clinically relevant point, unlike measures such as TPED. Thus, a more clinically realistic measure is obtained by using loss of active extension rather than passive angular deformity, as assessed by goniometry.

There are still some issues with using linear loss of extension as an outcome. As with angular deformity, some patterns of Dupuytren's disease, such as thumb-index web contracture, or limited palmar span are not quantifiable. Furthermore, some patients with a PIPJ contracture develop compensatory

hyperextension of the MCPJ and distal interphalangeal joint, possibly reducing the functional impact of the contracture. Mäkelä and colleagues do not discuss whether this is accounted for in their measurements. With the jig in Figure 1.6, it is not.

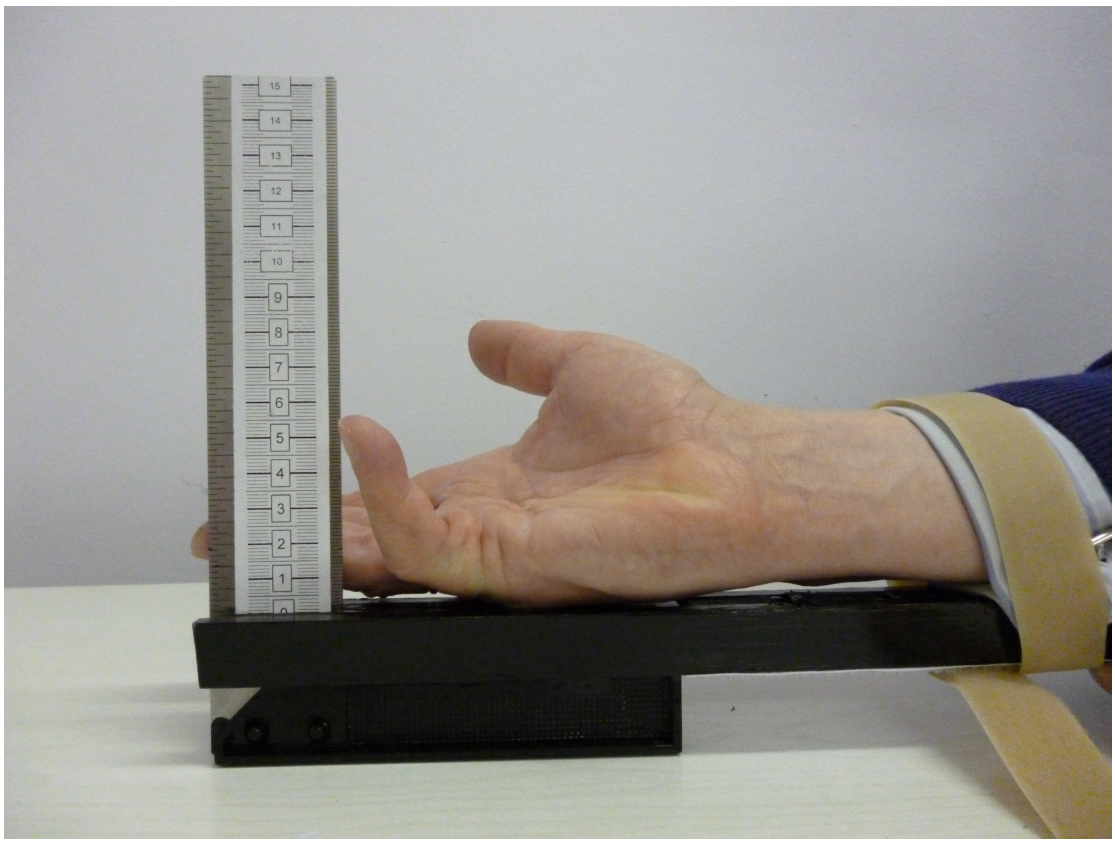

Figure 1.6: Typical study photograph from MSc project

From the candidate's MSc dissertation (Rodrigues, 2010).

1.4.5.3 Hueston's tabletop test

The basis of Hueston's tabletop test is loss of extension in general. However, it is neither a true measure of angular deformity, nor is it a perpendicular loss of extension. This test was proposed as a simple method for selecting patients with a severe enough contracture to merit surgery, by assessing whether they can place the palm of the hand flat on a table or not due to loss of extension (Hueston, 1974). Whilst it is dichotomous rather than quantitative, it has been used in one study to assess recurrence (Rodrigues, 2010). Although this only provides binary data regarding recurrence, it may provides a clinically relevant outcome measure, as it is also used as the indication for surgery in many instances. Thus, a positive test at follow up represents recurrence, extension or treatment failure severe enough to merit reoperation. However, whether the indication for surgery should be an objective test of this nature that does not assess impact on function or quality of life in any way is contentious.

1.4.6 Patient Reported Outcome Measures

Another approach to quantifying Dupuytren's disease in research is to use patient reported outcome measure (PROM) tools, of which there are many. One database lists 580 different tools, excluding foreign language translations of tools (MAPI, 2013), and careful choice of appropriate and validated PROMs is important.

Such scales aim to assess patient impact of the condition or its treatment on function, activity or health-related quality of life. In Dupuytren's disease, they may be able to summarise the impact of disease at multiple sites, as they are not limited by pattern of disease. Thus, they may provide a clinically relevant assessment. For example, if a patient experiences functional impairment due to a loss of palmar span, then this might be detected, whereas it will be overlooked when looking at joint angular deformity. Also, assessing the impact of treatment at different sites is complex. For example, if a patient with combined MCPJ and PIPJ contracture has simply the former component corrected, the residual PIPJ contracture may be functionally compensated for by hyperextension at the MCPJ. This again may not be reflected in objective assessments, but may be of importance to the patient.

Furthermore, these tools can often be administered as a questionnaire, without requiring specialist assessment. As a result, they may be performed with less cost than measuring palpable disease or loss of extension: the questionnaire may be posted to patients to return, without incurring the cost of and inconvenience of additional outpatient clinic attendances for assessment. Additionally, in the current economic climate, it is likely that evidence of direct patient benefit will be required to justify expenditure in health systems for treatments. The use of patient-reported outcome has grown in recent years and is promoted directly in the UK by the Department of Health (Department of Health, 2010).

PROMs may be broadly separated into those that assess health-related quality of life non-specifically, those that are domain-specific, e.g. assess the upper limb as a whole, and those that are disease-specific (Szabo, 2001). A variety have been used in studies of Dupuytren's disease, including domain-specific PROMs and a disease-specific one (Ball et al., 2013). Since then, a further disease-specific PROM has been published (Mohan et al., 2014). However, no report of the use of generic PROMs was found (Ball et al., 2013).

One common problem for the development of all such tools is defining health-related quality of life, or hand function, objectively. What constitutes acceptable hand function to a nursing home resident may be considerably different to what is acceptable to a concert pianist. Quality of life for patients with Dupuytren's disease has been explored recently (Wilburn et al., 2013). However, neither concept can be measured directly.

1.4.6.1 Generic PROMs

Non-specific questionnaires assessing general quality of life allow comparison between very different conditions. An example is the EuroQol 5D™ (EQ5D), which is the preferred measure for assessing health related quality of life by NICE as part of its cost effectiveness analyses (NICE, 2008). In theory these allow comparison to any other condition or treatment, at any anatomical site. However, they may not be sensitive enough to detect subtle differences in effect between different interventions for Dupuytren's disease itself. Instead they may experience ceiling or floor effects, whereby the pre-treatment and

post-treatment scores are both either very high, with little difference between them, or both are low due to the impact of comorbidities. Furthermore, their lack of specificity may mean that they are influenced by other comorbidities. The latter may be of particular relevance given that Dupuytren's disease becomes more prevalent with increasing age.

1.4.6.2 Domain-specific PROMs

Domain-specific tools are more specific. They include the DASH (Hudak et al., 1996). This 30-question tool is considered one of the most widely validated upper limb scales, with numerous studies assessing it in different conditions and languages, as summarised in a review of outcome measures by Dorwick and colleagues (Dowrick et al., 2005). Furthermore, it has been the most popular PROM used in Dupuytren's disease research to date (Ball et al., 2013), and has been previously used in studies of different aspects of Dupuytren's disease, including treatment outcome and also for clinicopathological correlation (Degreef et al., 2009, Dias and Braybrooke, 2006). Whilst the DASH is supposed to be specific to upper limb function, there is evidence that it is responsive to disease at sites outwith the upper limb (Dowrick et al., 2006). Its validity for comparing different interventions for Dupuytren's disease has been questioned in one study of patient quality of life (Wilburn et al., 2013), where a review was cited to suggest that the DASH's validation studies were flawed, and that its dimensionality was unproven. However, whilst this statement does appear in the review article's abstract, the

main text (which considers PROMs for shoulder conditions only) actually concludes that the DASH is the best existing PROM (Bot et al., 2004). Several of the questions in the DASH assess pain symptoms in the limb. As a typically painless condition, it is feasible that pain scores in Dupuytren's disease may actually be worse soon after surgery. It might be expected that this be the case for more invasive and extensive surgery, which might have lower long-term recurrence than less invasive treatments. This may give an inappropriately negative evaluation of a treatment that gives patients long-term recurrence-free function. Although this might suggest that scores like DASH are not suitable for assessing a condition like Dupuytren's disease, it should still be borne in mind that if a patient were to develop chronic pain following treatment, then this might reduce function. Thus, appropriate timing of assessment using tools like DASH is important.

As with quality of life scores, comorbidities may influence the score obtained for domain-specific tools, for example due to coexisting shoulder or elbow disease. The impact of comorbidities can be neutralised by considering the change in the score after treatment compared to that before the treatment – the difference could be assumed to be the result of the treatment itself, as the comorbidity will remain constant. Whilst this is valid for early outcome, it is not true for late outcomes, as the comorbidities cannot be assumed to have remained constant: degenerative conditions such as arthritis may worsen during the intervening period, or indeed might be treated themselves.

1.4.6.3 Disease-specific PROMs

A disease-specific scoring system for Dupuytren's disease has been recently developed and validated by the Unité Rhumatologique des Affections de la Main (URAM) (Beaudreuil et al., 2011). It is perhaps of note that Pfizer, who recently launched collagenase for the treatment of Dupuytren's disease, funded the study and employed two of the named authors (Beaudreuil et al., 2011). The theoretical advantage of such a scale is that it has been designed to sensitively assess the pertinent problems of this condition, which might be expected to differ from other upper limb conditions, such as arthritis. Furthermore, the impact of comorbidities on the score should be minimised, and the effect of less relevant upper limb symptoms, such as pain or paraesthesia avoided. Indeed, the URAM authors highlight the apparent weakness of using scores that comprise an assessment of pain, such as DASH for measuring Dupuytren's disease outcome (Beaudreuil et al., 2011). However, as has been discussed already, it seems inappropriate to ignore the fact that the treatment itself may result in chronic postoperative pain that in turn reduces post-operative function. If Dupuytren's disease is painless both before and after surgery, then 'pain questions' should not unduly influence the score achieved, so why not ask about pain? A major disadvantage of disease-specific scoring systems is their specificity itself. In an era in which health economics is increasingly important, it may be necessary to demonstrate how effective or cost-effective treating Dupuytren's disease is compared to treating other conditions, such as carpal tunnel syndrome, for

example. Whilst scores like URAM might allow a detailed comparison between different Dupuytren's treatments, demonstrating that treating Dupuytren's disease improves a disease-specific score is not of use when comparing to other conditions. This would be possible when using domain-specific and quality of life scores. The design of the URAM involved item generation by French doctors and patients. Once a list of tasks made troublesome by Dupuytren's disease was created, this was narrowed by eliminating tasks that either correlated very closely to other tasks, or were 'never' performed by over 5% of the patients involved. The latter may affect the generalisability of the tool. Its cultural sensitivity for use in populations other than the French group may need investigation, as the tasks included may be less relevant to other populations, and some of those excluded may be major issues elsewhere.

1.4.6.4 Patient rated change PROMs

A variation of PROM not discussed in the review by Szabo (Szabo, 2001) is patient rated change. Several such tools exist, including the Measure Yourself Medical Outcome Profile (MYMOP) (Paterson, 1996) and Global Rating of Change scales (GRC), such as that used Jaeschke and colleagues (Fitzpatrick et al., 1998, Jaeschke et al., 1989). Such patient rated change measures may be of use in analyses of interpretability of more specific quantitative tools, as well as being options for stand-alone measures.

1.4.6.5 Patient-specific PROMs

All of the above PROMs involve fixed scale, whereby all patients complete the same items. An alternative strategy is to use patient-specific PROMs. Rather than having rigid scales, these typically involve the patient specifying the items based on their own symptoms and goals, and then scoring these before and after treatment. Examples do exist for use in niche areas of hand surgery (Law et al., 1990), but have not been used in Dupuytren's disease (Ball et al., 2013).

1.4.6.6 Validity and Reliability of PROMs

There are issues with the development and use of function and quality of life tools. One is demonstrating validity. An international standard has been set for the study of patient-reported outcome, via a Delphi consensus. The resulting Consensus-based Standards for the Selection of Health Status Measurement Instruments (COSMIN) inform the design of studies of validity, responsiveness and, to a lesser extent, interpretability (Mokkink et al., 2010).

These define different aspects of the validity of PROMs:

- Content validity assesses whether the items that comprise a PROM are an adequate reflection of what is trying to be measured. It involves assessing the relevance and comprehensiveness of the items in a PROM.
- Construct validity examines hypotheses about the PROM. Such hypotheses may relate to its structural validity (internal relationships

between items), hypothesis testing (assessing its relationship with other PROMs) and differences between patient groups (cross-cultural validity).

- Internal consistency assesses how related the items in a PROM are to each other. This assumes that all of the items that contribute to a summary score actually assess the same underlying entity, or factor (e.g. impairment of structures in the hand versus restriction of activity involving the shoulder), i.e. they are 'unidimensional'.
- Criterion validity tests a PROM against a 'gold standard'. The only accepted methodology for this is the comparison of a shortened PROM against the long version (e.g. the QuickDASH against the DASH).

Responsiveness is the ability to detect change over time, as opposed to the single time point score assessed in 'validity'.

Interpretability is concerned with being able to interpret and understand the relevance of a change score.

Much validation of older PROMs falls short of the standards set by COSMIN. For example, the Short Form 36's (SF-36) validation in the UK involved comparison to the Nottingham Questionnaire (Brazier et al., 1992). This strategy is limited, as the 'gold standard' that has been used as the reference is the existing flawed measure. Indeed, this is explicitly condemned by COSMIN.

There is limited validation of the DASH in Dupuytren's disease specifically (Kennedy et al., 2011),

Validation of the URAM scale involved construct validity assessment against existing functional measures such as DASH, and also to Tubiana stage (Beaudreuil et al., 2011). The URAM authors point out that their scale correlates better with the Tubiana stage than the DASH does. This is cited as an advantage. Furthermore, incorporating patient opinion as an outcome measure is currently considered important. However, in this situation correlation with just one standard seems inadequate for an ideal test in the absence of a gold standard. Indeed, if an outcome test only correlates well with one standard, such as Tubiana stage improvement, then there may be little incremental gain from this 'functional assessment', over and above the baseline clinical assessment except perhaps a façade of assessing patient opinion.

The authors of the paper describing URAM do, however, make the good point that it can be conducted without a specialist present, as it involves a simple questionnaire. This may allow simplification of assessing outcome in research studies of Dupuytren's treatment.

Even for an individual, there will be variability in functional compromise depending on the dominance and the affected hand. It might seem that this can be tackled by subgroup analysis of dominant hand disease versus non-dominant hand disease. However, patients cannot be categorised so easily.

If the dominant hand is affected, this is likely to have more impact than the non-dominant hand being affected. However, the ability of the patient to compensate for reduced function of their dominant hand will be affected by other variables. For example, a right-handed patient with right-handed severe Dupuytren's disease, but an unaffected left hand, might cope better than a patient with bilateral severe disease or a comorbid condition compromising the left hand. The prevalence of comorbidities in the typical older Dupuytren's patient may also affect this. For example, if a walking aid is used, and one or both hands have reduced function, then independence might be reduced more than expected.

Function and quality of life tools may effectively standardise this variation by quantifying the improvement or deterioration relative to the patient's own baseline function, in the same way that this might seem to be a valid way of standardising for the influence of comorbidities on domain-specific or quality of life scores. Assessing 'change in hand function score' would provide an elegant solution to standardising PROM changes from the treatment and might be easily applicable to other hand conditions, such as arthritis. However, in addition to the problems arising from changes in comorbidities and their treatment that have already been discussed, there are also issues pertaining to Dupuytren's treatment itself that make this more complicated. Dupuytren's disease is often treated in a prophylactic manner (unlike arthritis, for example). Thus surgery may not actually change function relative to the patient's preoperative state. Function may even worsen with time, but this

might still represent a success – treatment may have prevented a more rapid future functional decline.

'Change in hand function score' might still be a valid outcome measure, even if hand function tends to decline after treatment for the reasons discussed above, so long as it were interpreted appropriately. If the rate of decline in hand function after a particular trial treatment were slower than after standard treatment, then this would demonstrate the superiority of the trial treatment over standard treatment (despite both cohorts experiencing a decline in hand function over time). Indeed standardised change in hand function (relative to the patient's own baseline function) might allow the generation of Kaplan-Meier survival curves of hand function following intervention. Changes in comorbidities (either deterioration or treatment) would still affect these, but the overall pattern may still prove useful.

If this continuum of decline in hand function were to be used, several issues would need to be resolved, such as defining what constitutes a clinically important drop in function.

1.4.7 Assessing cost effectiveness

Another strategy in assessing the impact of Dupuytren's disease would be to assess cost effectiveness of treatment. This is likely to be increasingly important when attempting to justify future resource allocation to a particular

treatment or condition, particularly when the British National Health Service is undergoing cost-saving changes.

Common cost effectiveness measures are available. In particular, NICE employs the incremental cost effectiveness ratio (ICER). The ICER amounts to the cost per quality adjusted life years (QALYs) gained. Although Dupuytren's disease is not a terminal condition, QALYs can still be calculated. Alternatively, change in hand function might be used as the denominator, but this is also more complicated.

Other approaches have been used. Some data describing the cost of treatment in England has been published (Gerber et al., 2011). Like the URAM validation work, Pfizer also funded this study, and three of the four authors declared conflicting interests with Pfizer. Data describing the number of treatments performed were multiplied by the cost described in The National Schedule of Tariff. Whilst this crudely estimates the cost of treatment, it cannot comment on cost effectiveness. Furthermore, the true cost/benefit of treatment would need to incorporate far wider reaching data, such as return to work or change in occupation following treatment, as these factors also affect cost to society.

Given the advantages and considerable disadvantages of all methods of assessing Dupuytren's disease, it is likely that no single method will suffice. Considering the particular unsuitability of the previously most popular technique (Leclercq's definition of palpable recurrence) for use in fasciotomy,

a new standard is needed. The ideal would be consensus on what techniques should be used, and timing of use following the procedure.

SUMMARY

Research on Dupuytren's disease treatments requires an appropriate choice of outcome and a means to measure this. There are many different ways to approach this. The most common strategy has been to assess recurrence and/or extension. At present there is no consensus on the definition of recurrence, or when it should be assessed. The most commonly used strategy of palpating the reappearance of nodules is no longer relevant. Other approaches, including assessing angular deformity, are not patient-specific. Measuring hand function or quality of life is probably more meaningful, and lends itself to cost effectiveness analysis. The most popular PROM used is the DASH, though the URAM has been recently developed as a Dupuytren's disease specific PROM that could be used. Further work is required to identify the most appropriate outcome measure to do this, in terms of validity, responsiveness and interpretability.

1.5 Treatment

A range of treatment options is currently available for Dupuytren's disease. These include strategies that divide cords of disease, with the bevel of a hypodermic needle in an aponeurotomy, with a scalpel blade, or with enzymatic degradation of collagen via an injection of collagenase. Alternatively, disease may be removed surgically with skin retained in a fasciectomy, or skin replaced with skin graft in a dermofasciectomy. Comparison of these using the outcome measures discussed is necessary to establish the appropriate role of each.

1.5.1 Observation

As a condition that is not life threatening, observation of disease is an option. This is particularly for early or mild disease that is not limiting the patient functionally, and in the presence of significant comorbidities that would increase the risk of treatment.

There is relatively little literature describing the progression of disease in patients who are observed rather than treated. Progression from nodules to cords occurs in the majority of patients (Luck, 1959), but does not necessarily reach a severity for which surgery is indicated (Reilly et al., 2005). A small minority of patients in the latter study experienced spontaneous resolution of a nodule.

1.5.2 Selection for Surgery

The aim of treatment is to manage Dupuytren's disease, as curative treatment does not currently exist. Additionally, the optimal timing for treatment is yet to be established, but Hueston's table top test has been used as a simple means of determining whether a contracture is severe enough to merit intervention (Hueston, 1974). This may be an oversimplification as some patients may experience functional limitation prior to this point, whilst others with more severe contractures may not wish to undergo treatment. More recent guidance from the British Society for Surgery of the Hand suggests that the minimum criteria for which intervention should be considered are functional problems, MCPJ contractures over 30 degrees, any PIPJ contracture or a first web contracture (BSSH, 2008). Whilst these criteria seem reasonable, there is no evidence cited for the algorithm. Such guidance is provided primarily to assist clinical practice, but indications for surgery need to be standardised for prospective trials. Valid clinical indications for surgery such as 'functional problems' in the absence of a moderate or severe contracture may not be appropriate for inclusion in a trial, particularly if the primary study endpoint is return of an angular contracture. Cases with a milder contracture initially may achieve a better postoperative correction, whereas those with a severe contracture that is only partially correctable may be at a disadvantage. As with the variation seen in outcome measures and endpoints, this aspect of research might benefit from a consensus agreement on an international standard for Dupuytren's disease trials. However, any such consensus would

need to be informed by further research to establish the relationship between structural impairment, such as loss of extension, hand function limitation, and activity and participation restriction with its impact on health-related quality of life.

Once active management has been deemed necessary, the specific treatment modality can be considered. Often textbooks classify options as 'operative' and 'non-operative' strategies (McGrouther, 2005). However, this may be misleading. Some 'non-operative' techniques, such as needle aponeurotomy or collagenase therapy, require aseptic technique and are often performed in an operating theatre. They also carry risks such as nerve or tendon injury, which are similar to 'operative' techniques. Here, strategies for treating contractures will be classified into those aiming to remove disease, those aiming to divide disease and those aiming to prevent the formation or progression of disease.

1.5.3 Removal of disease

Surgery to excise disease may involve excision of nodules and cords with preservation of the overlying skin (fasciectomy), or excision of the Dupuytren's tissue with the overlying skin *en masse* (dermofasciectomy). Collagenase enzyme therapy lyses collagen, and thus may remove a segment of disease, rather than simply divide it. However, its effect on the myofibroblasts within the cord is unclear, whereas other forms of disease removal, such as fasciectomy, will remove collagen and myofibroblasts en masse. In the

absence of these concepts being formally investigated and documented, collagenase will be discussed as a treatment that divides disease.

Fasciectomy is the most common procedure performed for the management of Dupuytren's disease, accounting for 93% of day case procedures and 91.6% of inpatient procedures for the condition in England in 2007-2008 (including revision surgery) (Gerber et al., 2011). However, the term describes a continuum of excisional procedures that occupy a spectrum between an open fasciotomy at the more conservative end and dermofasciectomy at the more extensive end. Indeed, common terminology used to sub-classify 'fasciectomy' includes segmental, limited, regional, radical and total fasciectomies, in order of increasing extent of tissue excision. Alternatively, fasciectomy may be classified on the basis of the skin incision, and skin closure, used to approach the disease.

Skin incisions have evolved from those used for very rapid access prior to the availability of anaesthesia, through a myriad of different historical options, to several commonly used today (Tubiana, 2000). There is little evidence base for many historical incisions, and as they are not in common usage today, they will not be discussed. One common technique is to perform a longitudinal incision along the axis of the ray, extending from the palm into the digit. Closing such a longitudinal incision might result in scar contracture, again limiting extension of the digit, so several incisions are made from the side of it, which can then be closed as a series of z-plasties, breaking up the straight

line scar and recruiting additional skin laxity into the straightened finger (Tubiana, 2000). Alternatively, a zig-zag approach can be used, as suggested by Bruner (Bruner, 1951). By making small darts from the apices of the triangular flaps, and then allowing these 'Y' shapes to open into 'V' shapes, additional skin length can also be recruited into the wound closure at the expense of skin width (Tubiana, 2000). A prospective randomised controlled trial comparing these two approaches has been published (Citron and Nunez, 2005). This showed no significant difference in recurrence rate as assessed by a surgeon who had not performed the initial operation, using Leclercq's definition (Leclercq, 2000), by 2 years. However, no power calculation was provided. The intention to treat analysis revealed that several patients were excluded from the Bruner cohort after randomisation due to administrative issues, such as having previously been enrolled for contralateral surgery. However, if intention to treat analysis is performed using Fisher's Exact test and including those patients initially excluded, the difference in recurrence remains statistically insignificant ($p=0.44$, versus initial result of $p=0.20$).

Another skin management strategy is the open palm technique, described by McCash in a 43-case series (McCash, 1964). This is used for palmar disease, often involving two or more finger rays. The palmar Dupuytren's cords are excised through a transverse incision in the distal palmar crease. This is left open to heal by secondary intention when there is inadequate skin laxity to close it, or when closing other incisions in the digits requires all available skin laxity. Indeed, similarities between this and Dupuytren's original description of

operative technique are highlighted by McCash (McCash, 1964). A prospective non-randomised study investigated wound healing and early complications, comparing the open palm technique to z-plasty and zig-zag closures (Gelberman et al., 1982). The study concluded that the open palm technique was the 'most dependable' of the three, though it took the longest to heal, and no statistical analysis was performed. Along with the lack of randomisation, this reduces the value of these data. A higher incidence of some transient early complications, such as delayed wound healing, might be tolerated, if the long term outcome was improved by adopting a particular operative strategy. However, no high quality comparative data of recurrence rates are available. A retrospective study by Foucher and colleagues showed a 34% rate of recurrence at an average of 5.6 years follow up for open palm surgery, which they deemed comparable to other published series of limited fasciectomy (Foucher et al., 1992). However, recurrence was not defined in the study.

Once the skin incision has been made, the next area of variation in fasciectomy practice is the extent of excision. More conservative strategies include segmental fasciectomy. This was popularised by Moermans, who prospectively studied his first 213 consecutive procedures (Moermans, 1991). The operative strategy involved short curvilinear incisions through which short sections of cord were excised. In so doing, gaps in the cord were formed, creating a theoretical advantage over aponeurotomy. The recurrence rate found by Moermans and colleagues was 35.7% at a mean of 2.6 years of

follow up. Limited fasciectomy involves greater excision than segmental fasciectomy. Typically this is the excision of all involved fascia (McGrouther, 2005). It has been popularised by Hueston following his publication of the technique and a series of 96 procedures with early and late (albeit a mean of less than 2 years) outcomes (Hueston, 1961). A more aggressive approach is the radical fasciectomy. This involves excision of uninvolved fascia from the palm in addition to diseased tissue, but still only diseased tissue from the digits (McGrouther, 2005). McGrouther suggests it is more popular in central Europe (McGrouther, 2005). Exactly where limited fasciectomy ends has been discussed (Zachariae, 1969). A study comparing extensive and limited fasciectomy found similar long-term outcomes, though extensive surgery was associated with longer surgery and postoperative recovery (Zachariae, 1967). Thus, more extensive surgery may not necessarily be better.

Dermofasciectomy may also be considered as a heterogeneous group of procedures. Hueston proposed the concept of deliberate excision of the overlying skin, when he reported a series of 38 patients for whom he had performed this [sic] (Heuston, 1969). All his skin grafts were full thickness skin grafts. This concept built upon his previous observation that recurrence did not seem to occur under skin grafts in a series of 8 patients followed up for two years who had previously undergone multiple revision fasciectomies (Hueston, 1962). Skin grafts may be used to resurface large areas of the digit, often running from the metacarpophalangeal skin crease to the distal interphalangeal skin crease and from the radial midlateral line to ulnar

midlateral line, thus replacing the entire glabrous skin of one or more digits (Seah et al., 2012). Alternatively, smaller grafts may be placed strategically in an attempt to interrupt the development of a recurrent cord. The latter are referred to as 'firebreak' grafts, a term coined by Hueston (Hueston, 1984). A prospective randomised controlled trial conducted by Ullah and colleagues compared the three-year outcome between patients who had z-plasty skin closure at PIPJ level with those who received firebreak skin grafts (Ullah et al., 2009). They found a low recurrence rate of 12.5% - 15% per hand by 3 years, with recurrence defined as progressive angular deformity. There was no significant difference between the cohorts in an appropriately powered study. Despite this being a relatively rare example of high quality evidence in Dupuytren's disease, there are still some issues. Precisely what constituted 'progressive angular deformity' was not specified in their paper, nor was there any attempt to blind the assessor to the outcome or surgical procedure that had been performed. The authors acknowledged this limitation. Although they stated that the follow-up period used was adequate to detect recurrence, this may not have been enough to allow for the deterioration in a contracture that their vague definition of recurrence alluded to. This might account for their reported low rate of recurrence. Ullah and colleagues also pointed out that their findings were not applicable to more extensive skin grafting procedures. An observational study of outcome following more extensive dermofasciectomy has been published by Armstrong and colleagues (Armstrong et al., 2000). This reviewed 143 dermofasciectomies in 103

patients, out of a total of 135 patients who had undergone this procedure over a nine-year period. Although not explicitly stated, the description of the study design suggests that it was retrospective. The dermofasciectomy procedure was well described and was extensive, stretching from the radial to the ulnar midlateral line, even when the cord being excised was more localised. A recurrence rate of 8.4% per ray, and 11.6% per patient was reported after a mean follow up of 5.2 years. Given the longer follow up than in Ullah and colleagues' study, this suggests a lower recurrence for more extensive dermofasciectomy. However, there are caveats in this study. Armstrong and colleagues did not explicitly specify the definition of recurrence that they used. Their discussion of identifying nodules and cords implies the use of Leclercq's definition of 'the return of palpable disease within the operated field' as their primary endpoint. However, they also noted one case that progressed to re-contracture in the results, but the severity of this was not discussed. How this was determined as re-contracture, rather than incomplete initial correction is not clear. It might be expected that palpable disease would appear prior to a progressive deterioration in angular deformity, as was assessed by Ullah and colleagues. Thus, the true difference in recurrence rate between the two studies may be even more dramatic. Including a fasciectomy cohort in Armstrong's work would have been most appropriate, and would have avoided the need to extrapolate and compare between studies. There are other limitations to Armstrong and colleagues' study. The length of follow up was not standardised. Whilst the mean follow up of 5.2 years seems of reasonable

length, it conceals a range of follow up of 2.1 to 11.5 years. This breadth of follow up renders the recurrence rate from this study much less useful, particularly as the distribution of follow up lengths is not shown. There are no well designed prospective studies comparing dermofasciectomy to other forms of treatment, though at present dermofasciectomy is still widely considered to have a lower recurrence rate than other forms of treatment.

Excisional strategies may need to be combined with secondary procedures, such as check rein ligament release, to tackle underlying joint contractures. This piece will focus on the main treatment modalities for tackling the Dupuytren's tissue itself, but will also discuss these other procedures, as they have implications for the design of studies in Dupuytren's disease.

1.5.4 Division of disease

This may involve dividing cords with a blade, as in the original historic descriptions by Astley-Cooper, Goyrand and Dupuytren. More modern approaches tend to use the bevel of a hypodermic needle as a 'blade' to mechanically divide cords, or employ collagenase enzyme therapy for chemical cord division. Other suggested approaches to disrupting cords include using shockwave therapy, though at present this remains a hypothesis under investigation (Knobloch et al., 2011).

Open fasciotomy is the current version of the very original Dupuytren's disease surgery, which has been historically termed the 'Astley Cooper

procedure' (Leclercq, 2000). Small skin incisions allow the cord to be divided under direct vision without any excision. It therefore might be considered between needle aponeurotomy and segmental fasciectomy on the spectrum of invasiveness. As with fasciectomy, numerous different skin incisions have been adopted to approach the cord, some with theoretical advantages, but often without evidence for their use. A prospective non-randomised study has compared two techniques used for fasciotomy of cords causing MCPJ contractures: direct closure of a transverse incision versus z-plasty closure of a longitudinal incision (Citron and Hearnden, 2003). This small study was abandoned early as it showed that z-plasty closure was associated with markedly less recurrence than fasciotomy alone at 2 year follow up, but this did not reach statistical significance (p is quoted simply as <0.1). With the absence of a power calculation for the study it is unclear whether this lack of significance reflects a tendency towards significance in an underpowered study, or a true lack of difference between the cohorts. In spite of this lack of conventional statistical significance, a conclusion is made that the data support the theory that the development of Dupuytren's disease may represent a response to skin tension. However, it is notable that Leclercq's definition of recurrence as palpable disease within the operated field was used as the primary endpoint, despite this being a trial of fasciotomy. Thus it might be expected that only a relatively small gap in palpable disease would have been achieved by the procedure. As the fasciotomy cohort will have a transverse scar overlying this gap in the cord, distinguishing recurrence from

skin scar may not be easy. In contrast, the z-plasty cohort received a longitudinal incision, and the use of a z-plasty would rotate the scar out of the direction of the cord, but without rotating it through 90 degrees, i.e. the z-plasty cohort would not be expected to have a scar running transversely across the gap in the cord, nor would they be expected to have one running parallel to the cord either. This in turn might make scar-related lumpiness less likely to be palpated after a z-plasty, without actually affecting true recurrence. Such an effect may have been compounded by changes in assessors – five unblinded assessors were used over the course of the study. Whilst blinding may be difficult in this context given the visible distinct appearance of the scars from the two closures, it would not be impossible as the patient's hand could be covered with a latex glove. Laying aside these concerns, if there is a true difference between closure of a transverse incision and z-plasty, it is not clear whether this truly represents the effect of reducing tension. The approach to the fasciotomy differed between the cohorts; the z-plasty requires a longitudinal approach, as the geometry of the procedure then allows skin laxity to be recruited in this axis. However, the longitudinal approach may have allowed for better visualisation. This may have facilitated a more effective fasciotomy in this cohort, regardless of skin tension in closure.

Fasciotomy may be performed using other techniques as well. The concept of a closed fasciotomy is not new. Indeed, needle aponeurotomy has recently increased in popularity, and has been studied in large volume (Pess et al., 2012). It involves a closed fasciotomy, in which the cord is divided by

repeatedly passing a hypodermic needle through it or across it. The bevel of the needle divides fibres, and eventually breaks the cord. This may then be combined with a steroid injection into the operated field, in an attempt to soften the residually diseased tissue (Badois et al., 1993). Reported recurrence rates may be as high as 65% at 32 months (van Rijssen and Werker, 2006). In this study, recurrence was defined as a reduction in TPED of 30 degrees or more. As has been discussed, it is a more appropriate endpoint when investigating fasciotomy than Leclercq's definition. Furthermore, it represents an indication for re-operation (BSSH, 2008), and thus is a clinically relevant endpoint. The same team have conducted a prospective randomised controlled trial comparing needle aponeurotomy to limited fasciectomy with a 5 year follow up. The early outcomes of this trial demonstrate better outcome in terms of TPED improvement for limited fasciectomy compared to needle aponeurotomy in severe contractures, but similar outcomes for mild contractures (van Rijssen et al., 2006). Early patient satisfaction was better in the aponeurotomy cohort, and their DASH scores were significantly lower throughout the first six weeks of follow up. However, recurrence rate was higher after aponeurotomy by 5 years, by which time patient satisfaction had reversed; it was higher after fasciectomy (van Rijssen et al., 2012). Needle aponeurotomy has been the subject of an Interventional Procedure Overview by NICE, published in 2003 (NICE, 2003). These overviews, by NICE's Interventional Procedures Advisory Committee (IPAC) provide a rapid assessment of literature describing a new procedure, so that a

statement can be made about the procedure's safety and efficacy only. They do not aim to review the literature systematically, or to provide any guidance describing the role of that procedure in the management of a particular condition. Seven case series were included in the review, and the specialist advisors noted variation in outcome measures. However, it was acknowledged that it is an established practice. It was considered that there was adequate safety and efficacy evidence to support the use of needle aponeurotomy in NHS practice.

A more recent development has been the use of collagenase enzyme therapy, marketed as Xiapex™ in Europe. Attempting to divide cords by enzymatic dissolution is not new (Hueston, 1971); however, it has recently undergone considerable development by Badalamente and Hurst, who have brought collagenase into clinical practice. Their development process started with *in vitro* experimentation (Starkweather et al., 1996), and progressed through phase two trials to the Collagenase Option for the Reduction of Dupuytren's (CORD) 1 study. This was a phase three, double blind, placebo-controlled randomised controlled trial (Hurst et al., 2009), which demonstrated a highly significant difference between collagenase-treated contracted joints and placebo treated joints, in terms of the primary endpoint (correction to within 5 degrees of full extension). There were two flexor tendon ruptures and one case of chronic regional pain syndrome in the 203 joints treated with collagenase.

Data describing recurrence after collagenase treatment do exist. Recurrence (defined as deterioration in contracture per joint of 20 degrees of more) in joints 'successfully treated' (corrected to 0-5 degrees following treatment) was 35% at 3-year follow up (Peimer et al., 2013). However, deterioration of contracture by 20 degrees in other joints that showed only partial initial response to treatment (contracture improved by 20 degrees or more, but not to the 0-5 degrees cut-off defined as 'success') was 50%. This demonstrates the difficulty of extrapolating comparisons between studies, as there have often been variations in the definitions of early successful treatment, late recurrence or poor outcome, and the length of follow up has varied (Becker and Davis, 2010). Indeed, the need for standardisation of definitions has been highlighted (Werker et al., 2012). Other follow up studies have also demonstrated encouraging results (Warwick et al., 2014). However, no trials comparing collagenase to needle aponeurotomy or open surgery have yet been performed. Until such data become available, and the cost effectiveness of the agent compared to other treatment options is determined, the role of the former in managing Dupuytren's disease cannot be established.

It is also notable that collagenase treatment has a specific treatment protocol that must be adhered to. The drug is injected directly into the Dupuytren's cord using a hypodermic needle. The clinician administering the injection must have received specific training in the procedure. The patient then returns for a manipulation of the digit (in the initial studies this was 24 hours, though longer intervals are being studied), at which point an attempt to break

the cord is made in order to correct the contracture (Hurst et al., 2009). Therefore, although the injection itself may be performed on an outpatient basis, the patient will need to return to the treating centre the next day for manipulation. This may have implications when treating patients who live a long distance away from the treating centre. The cost impact of this may only be partially captured by existing cost effectiveness analysis methodologies, which tend to assess direct healthcare costs only (NICE, 2008).

1.5.5 Prevention of disease formation or progression

1.5.5.1 Radiotherapy

Numerous strategies have been considered to try to prevent Dupuytren's disease from developing, progressing or recurring. The most widely discussed is radiotherapy. As with needle aponeurotomy, this has been the subject of an Interventional Procedure Overview by NICE's IPAC (NICE, 2011). The studies considered here were all published in English, though it was noted that other literature exists, particularly in German. The publications included two describing one randomised controlled trial and two case series. The five specialist advisors who contributed to the overview were divided in opinions: two considered radiotherapy to be novel with unproven safety and efficacy, and two considered it to be established. The fifth opinion is not described. The NICE guidance provided is based on the same publications and acknowledges the lack of evidence available describing efficacy, mechanism of action and safety (NICE, 2011). Consequently, it is limited to

concluding that radiotherapy 'should only be used with special arrangements for clinical governance, consent and audit or research'.

The randomised controlled trial cited in both the overview and guidance involved randomising patients with "early" Dupuytren's disease to receive one of two protocols of radiotherapy (totalling 30 Grays in one group and 21 Grays in the other), without any control or placebo group (Seegenschmiedt et al., 2001). No description is provided for the method of randomisation, and it is not stated whether any blinding was used. The adverse events are described as total numbers observed in the whole trial, rather than by treatment group. The primary endpoint used was 'subjective improvement in symptoms', which is vague and does not fit with commonly used outcome measures for Dupuytren's disease. Follow up was for a minimum of one year, which is likely to be inadequate to assess long-term efficacy on disease progression and safety. Radiotherapy poses additional long-term risks to the patient, including dry palmar skin, but also secondary malignancy. If a potentially harmful therapeutic modality is to be used to try to prevent a non-terminal condition like this, then obtaining more high quality evidence to support both its efficacy and safety profile are important.

1.5.5.2 Steroid injection

Attempts to modify the disease with steroid injections into cords have been reported by Ketchum and Donohue, with the aim of softening cords in a manner akin to that seen when keloid scars are therapeutically injected with

corticosteroid (Ketchum and Donahue, 2000). This cohort study reported a high level of subjectively assessed 'regression' of nodules following serial triamcinolone injection, though no controls were available and no attempt was made to assess improvement in angular contracture or hand function. These data in isolation are not adequate to determine the effectiveness of this strategy.

1.5.5.3 Splinting and Traction

Extension splinting is often used following surgery as an adjunct. Different regimens may be employed, including night only splinting, or day and night splinting. The period of splintage may also vary. Splints may need modification of revision during the rehabilitation period to maintain their effect (Jerosch-Herold et al., 2008). However, relatively high quality studies have failed to identify clear evidence of benefit from this (Collis et al., 2013, Jerosch-Herold et al., 2011, Kemler et al., 2012). Splinting as primary therapy has been studied, and may provide some benefit in terms of delaying deterioration when used in conjunction with soft tissue mobilisation (Larocerie-Salgado and Davidson, 2012).

Other strategies may be adopted to correct contractures without direct effect on the cord. Bone traction using Kirschner wire-based devices has been described (Craft et al., 2011, Messina and Messina, 1993), by which gradual correction is achieved with skin elongation of 2 millimetres per day (Messina and Messina, 1993). This may be used as a definitive strategy in isolation, or prior to other treatment techniques. The developers claimed that the gradual

elongation of the contracture achieved with this technique might reduce the strain placed on neurovascular bundles and joints that might occur with sudden operative release. Furthermore, the developers claim that this strategy may flatten nodules and induce regression of cords, although the method used to assess this was not described in detail and appears to be subjective (Messina and Messina, 1993). This approach does result in a protracted course of treatment in comparison to other treatment modalities. However, it is possible that this time commitment would be offset if rehabilitation is accelerated and avoids more invasive surgery. Furthermore, there may be complications from this strategy. In one study, active flexion of the finger was completely lost at the time of removal of the device, and returned following a course of hand therapy (Messina and Messina, 1993). Although function was restored by therapy, the participants are likely to have experienced a further period of considerably limited hand function even after the distraction procedure has been completed.

1.5.5.4 Other modalities

A review article raises the possibility of using interferon gamma to modulate the disease process as a scientific hypothesis, but does not provide any evidence of experimentation to support this (Tomasek et al., 1999). Similarly, there are also other potential treatments currently lacking evidence for their efficacy. These include ultrasound and massage (McGrouther, 2005).

1.5.6 Correction of associated pathology

Appropriately treating the Dupuytren's disease cord or nodule may not necessarily achieve correction of the finger deformity, due to changes in other soft tissue structures. As discussed already, the PIPJ may develop a joint contracture due to its arrangement of volar plate check reins. This may limit correction of joint angles when treating disease at the PIPJ (Bryan and Ghorbal, 1988). Release of structures besides the cord of Dupuytren's disease may be performed. However, this requires an open surgical approach, and so cannot be performed with minimally invasive techniques. Previous studies but PIPJ release may involve release of a range of the structures illustrated in Figure 1.7. A typical strategy comprises identification and protection of the proper collateral ligament, then sequential division of the flexor sheath distal to the A2 pulley, the check rein ligaments proximal to the volar plate, the accessory collateral ligament and the remainder of the proximal part of the volar plate, until contracture release is achieved (Beyermann et al., 2004).

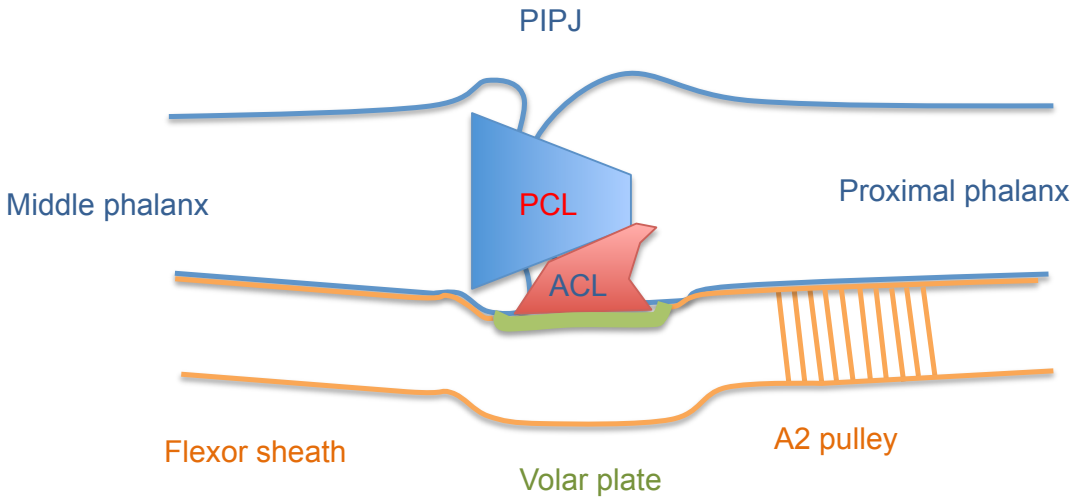

Figure 1.7: Schematic representation of structures at proximal interphalangeal joint

During proximal interphalangeal joint (PIPJ) release, the proper collateral ligament (PCL) is preserved, and sequential release of other structures is performed, including the flexor sheath at the distal end of the A2 pulley, the accessory collateral ligament (ACL) and the volar plate.

Several studies have investigated the value of PIPJ release in the treatment of Dupuytren's disease, with some identifying improved outcome following joint release (Ritchie et al., 2004), whereas others did not demonstrate improved angular deformity outcome (Beyermann et al., 2004, Misra et al., 2007). Furthermore, joint release may pose further risk of complications such as injury to the digital nerves, joint instability, or further scarring and stiffness. The functional outcome of joint release may reflect a balance of the benefit of additional correction versus increased risk of complications. Joint release has

been compared to bone traction using the Digit Widget device (Craft et al., 2011).

The different behaviour of disease affecting the PIPJ compared to cords affecting the MCPJ influences current recommendations regarding clinical practice, with the indications for surgery typically involving a 30 degree contracture at the MCPJ, or any contracture at the PIPJ (BSSH, 2008), as PIPJ contractures are harder to correct. As alluded to earlier in the discussion of Selection for Surgery, the evidence supporting these particular cut-offs as the indications for treating Dupuytren's disease is not robust.

The role of joint release in treating PIPJ contractures requires further investigation. However, it is likely that contractures of the PIPJ pose a greater challenge for surgical correction than do MCPJ contractures. As a result, comparative clinical studies may need to match treatment arms in terms of proportions of participants with PIPJ contractures, the severity of such contractures, and the length of time that the digit has been deformed. This may be particularly important when comparing different treatment modalities; needle aponeurotomy is often performed on MCPJ contractures, with more invasive fasciectomy and dermofasciectomy performed in cases with PIPJ contractures.

1.5.7 Complications

A range of complications can occur following treatment for Dupuytren's disease (Hurst, 2010). These can be categorised as early or late onset complications. Early complications include infection, haematoma, wound

healing problems, tendon injury or nerve injury. Some early complications persist for a prolonged period or are permanent, such as cold intolerance and altered sensation. Symptoms following nerve injury may improve if the injury is identified and repaired. However, even if a technically acceptable nerve repair is performed, then symptoms may persist indefinitely. Furthermore, this requires an open surgical approach, and so may require a secondary operation if a percutaneous treatment such as aponeurotomy or collagenase. This will result in further morbidity at the time of treatment of the complication.

Late complications arise at an interval after treatment and include recurrence, which is the focus of much of the literature describing outcome in Dupuytren's disease (Crean et al., 2011). However, there are other common complications. The median reported rate of neurapraxia was 22% after excisional surgery, and 3% after aponeurotomy (Crean et al., 2011).

A recent review identified that many complications are more common following more invasive treatment (Crean et al., 2011), which may influence functional outcome independently from recurrence. Indeed, the reported rate of nerve injury after dermofasciectomy ranged from 41% to 51%, compared to 0.8% to 5.3% after aponeurotomy (Crean et al., 2011). As a result, the benefit of treatments such as dermofasciectomy in terms of low recurrence rate may be offset by increased risk of complications when hand function is considered.

1.5.8 Relative risks and benefits of different treatments

The theoretical pros and cons of different treatments suggest that different treatment modalities may have specific indications. However, limited

comparative data exists to confirm the roles of particular treatments. The current theoretical benefits and limitations of the most common treatments discussed are summarised in Table 1.2.

Treatment type	Treatment Sub groups	Benefits	Limitations	Possible role
Dermofasciectomy	Firebreak skin graft 'Conventional' dermofasciectomy	Low recurrence rate (11.6%/patient at 5.2 years for conventional dermofasciectomy (Armstrong et al., 2000))	Higher rates of complications (e.g. 46% nerve injury vs 8.6% for fasciectomy (Crean et al., 2011). Replaces glabrous skin with less resilient non-glabrous skin	Used in severe or recurrent disease, or if strong risk of diathesis and therefore likely to recur (Armstrong et al., 2000)
Fasciectomy	Segmental fasciectomy	Most commonly used procedure so most	Longer recovery period than aponeurotomy (van	Currently most commonly used procedure in the UK

	Limited fasciectomy Radical fasciectomy	surgeons are experienced (Hospital Episode Statistics, 2011). Higher late patient satisfaction than aponeurotomy (van Rijssen et al., 2012). Lower recurrence rate than aponeurotomy (van Rijssen et al., 2012).	Rijssen et al., 2006). Lower early patient satisfaction than aponeurotomy (van Rijssen et al., 2006)	(Hospital Episode Statistics, 2011)
Aponeurotomy	Needle aponeurotomy Blade	Shorter recovery period than fasciectomy (van Rijssen et al., 2006).	Higher recurrence rate than fasciectomy. Cannot perform	Suitable for MCPJ disease with a well-defined cord (van Rijssen et al., 2006).

	aponeurotomy	Lower complication rate than fasciectomy or dermofasciectomy (Crean et al., 2011).	secondary joint release as minimally invasive. Unable to treat painful nodules without contracture as disease not excised.	
Collagenase		Benefits unclear as no comparative data available	Second patient contact required for manipulation. No comparative data describing long term outcome compared to other treatments.	Role not defined, as no comparative data.

Table 1.2: Theoretical benefits and limitations of different common treatment options

SUMMARY

Fasciectomy remains the most popular form of treatment for Dupuytren's disease. However, this term covers a range of different surgical approaches in terms of skin incision and management, and extent of excision of disease. Dermofasciectomy, with skin grafting is considered to have the lowest recurrence rate, though is more invasive, and probably poses a greater risk of complications. Other strategies currently gaining popularity include needle aponeurotomy and collagenase therapy. Whilst radiotherapy has been considered to attempt to slow disease progression or recurrence, its efficacy and safety are not universally accepted. More high quality evidence is needed to determine the appropriate roles of each of these therapies in the management of Dupuytren's disease. In particular, the impact of other complications in addition to recurrence on function needs to be considered.

1.6 Adoption of high quality evidence

Once high quality evidence that reveals best practice has been generated, then it should be adopted. Conventionally this may occur through presentation of research findings at scientific meetings, journal publications and through incorporation into clinical guidelines that are then implemented. However, adoption of best practice may not always occur. Furthermore, guidelines may vary according to the quality of the process used in their development, and implementation of robust guidelines may be inconsistent. In England and Wales, the National Institute for Health and Clinical Excellence (NICE) was established with one of its objectives being to reduce variation in quality of clinical practice. With The Health and Social Care Act 2012 coming into force (Health and Social Care Act c. 7, 2012), NICE's remit expanded to include social care, and so its title changed to the National Institute for Health and Care Excellence. This Act of Parliament also initiated changes to the architecture of the National Health Service (NHS) in England and Wales, some of which have the potential to facilitate adoption of best practice and to standardise the availability of evidence-based treatments.

1.6.1 Changes to the National Health Service Architecture

The organisation of secondary care, such as hand surgery, has changed since the introduction of the changes in the Health and Social Care Act. Consortia based in primary care, entitled Clinical Commissioning Groups (CCGs), now commission services. To incentivise appropriate commissioning, NICE has been charged with establishing a product that will assist commissioning, and

act as benchmark. This has been entitled the Clinical Commissioning Group Outcome Indicator Set (CCGOIS). The CCGOIS must be developed by assessments of practice that are derived from high quality evidence. NICE is also developing a series of Quality Standards to serve this purpose. Quality Standards are based upon NICE's own clinical guidance, but also on clinical guidance that meets current international standards (NICE, 2013).

1.6.2 Guideline development

NICE guidelines cover a broad range of areas of clinical practice. However, it is acknowledged that other organisations may also produce robust guidance. If NICE were to also produce guidance for the same situation, then duplication would arise. Furthermore, it would be unachievable for NICE to attempt to produce clinical guidance for every candidate topic across all aspects of clinical practice.

As a result, NICE's roles include appraisal of the guideline development processes of other organisations, in addition to developing NICE guidance. The standard for the development of clinical guidelines used to appraise processes is adapted from the Appraisal of Guidelines for Research and Evaluation (AGREE) II instrument (Brouwers M et al., 2010). This comprises 23 items spread across six domains that aim to comprehensively cover all areas of good practice in guideline development. Processes submitted for appraisal by NICE that meet its standards receive NICE Accreditation, and can be used to inform Quality Standards (NICE, 2011). Thus, a pathway is currently evolving by which an organisation, such as the British Society for

Surgery of the Hand, could systematically review high quality evidence and develop appropriate clinical guidance. This guidance could then be implemented nationally through the process described that links Quality Standards to the CCGOIS. This mechanism may become an important audience for high quality clinical outcome research. Alongside the project, the candidate developed a process for the development of clinical guidelines that meet the standards set by NICE and specified in AGREE II.

SUMMARY

A process is emerging by which evidence from high quality studies of clinical practice, such as the treatment of Dupuytren's disease could be adopted and implemented nationally. The input to this process is through NICE Accredited guidance, and ensuring that clinical outcome evidence can be incorporated into such guidance is likely to become increasingly important in the United Kingdom.

2 Aim and Objectives

2.1 Aims

This thesis hypothesises that the leading candidate outcome measures are inadequate for future use in Dupuytren's disease research, and that currently there is insufficient evidence to inform patient-centred treatment choice in DD.

In particular, existing candidate outcome measures are hypothesised as being subject to bias, invalid, and/or uninterpretable to different extents, such that one or more may not be suitable for future use.

The aim of this study was to investigate these hypotheses as part of the preparation to design future high quality clinical studies in Dupuytren's disease. Furthermore, this needs to be achieved in the context of the changes to NHS architecture brought about by The Health and Social Care Act 2012 (Health and Social Care Act c. 7, 2012) (Figure 2.1).

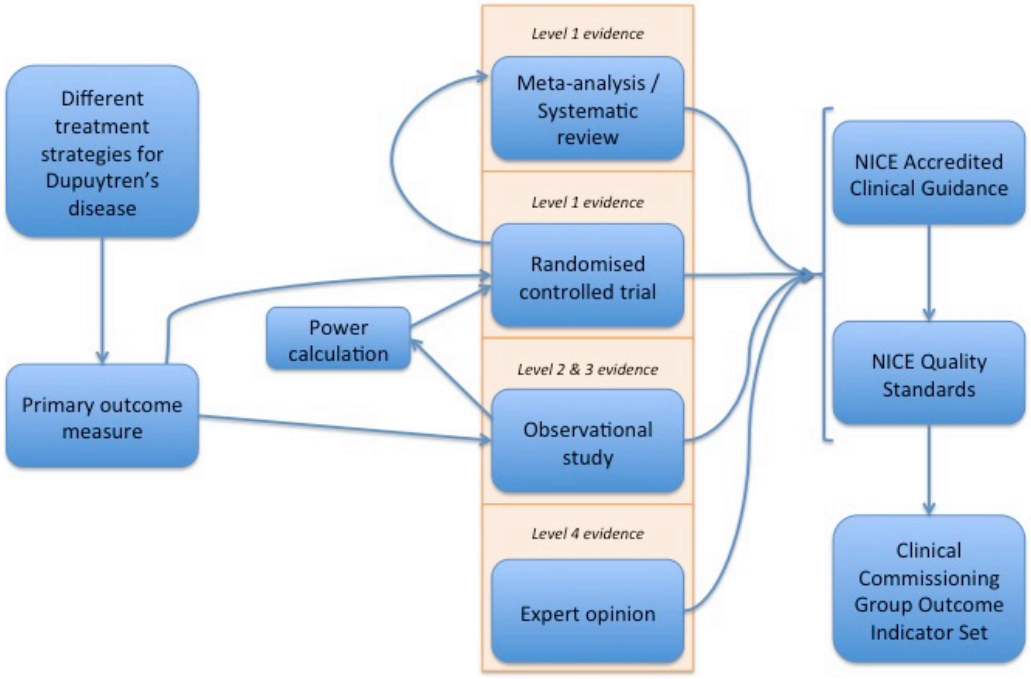

Figure 2.1: Simplified research pathway

2.2 Objectives

To achieve the aims, several objectives have to be achieved.

1. Establish which surgical interventional trials have been conducted in Dupuytren's disease and what outcomes were reported in them, by conducting a systematic review

Firstly, existing evidence needs to be appraised so that the state of research comparing treatment options in Dupuytren's disease is fully understood, and the need for further work established prior to continuing with other elements of the project. By doing so, the outcome measures used to date in randomised controlled trials can be identified and compared to those used across all clinical studies of Dupuytren's disease, as recently reviewed elsewhere (Ball et al., 2013).

Therefore, this first review will identify high quality studies comparing treatments for Dupuytren's disease to date, and identify which outcome measures have been reported in these trials.

2. Investigate the validity of leading candidate outcome measures, by performing a cross sectional study

Once the need for further clinical research in Dupuytren's disease is confirmed by achieving Objective 1, then the behaviour of candidate outcome measures needs to be investigated to inform the choice of the primary outcome measure for future research. The outcome measures studied will be based on the

findings of Objective 1, and a recent systematic review performed by others (Ball et al., 2013)

Specifically, the following uncertainties need to resolved regarding existing outcome measures:

a. Is dynamism in Dupuytren's disease is significant enough to constitute a source of bias for passive angular measurements?
b. What are the goals of British patients for the treatment of Dupuytren's disease, and is the URAM scale cross culturally sensitive against British patients' goals?
c. Do the DASH and QuickDASH tools demonstrate acceptable structural validity when used in Dupuytren's disease specifically?

3. Study the interpretability of leading outcome measures in Dupuytren's disease, by conducting a systematic review and cohort study.

As interpretability is a different domain of the behaviour of outcome measures, it will be studied separately here. Firstly, the existing data describing the interpretability of different outcome measures in Dupuytren's disease will be appraised by conducting a systematic review. Alongside the systematic review, a cohort study will provide original data, to resolve the following uncertainties:

a. Is 6 weeks after surgery too early to measure outcome from surgery for Dupuytren's disease?

b. What is the responsiveness and interpretability of the DASH and URAM when used in Dupuytren's disease surgery?

4. Estimate the late functional outcome after surgery for Dupuytren's disease and the factors associated with poor functional outcome, by performing a cross sectional study

Having studied the validity and interpretability of potential outcome measures to be used for future research, estimating functional outcome is also necessary as part of the preparation for future research. As well as estimating the outcome itself, this will also allow investigation of the factors associated with poor outcome. Such factors may need to be recorded in future studies to ensure cohorts are comparable. This objective will clarify the following uncertainties:

a. What is the late functional outcome of surgery for Dupuytren's disease from a cross sectional study?
b. What are the complication rates following the treatment of Dupuytren's disease by needle aponeurotomy, fasciectomy and dermofasciectomy?
c. Which factors discussed in the introduction are associated with poor functional outcome and complications, rather than with the tendency to recur, or 'diathesis'?

3 General Methods

The methods for general data collection are described here. Further methods specific to individual chapters are described in the relevant chapters.

3.1 Studies in the project

Two groups of patients were recruited. Early functional recovery was studied in a group of patients recruited in Nottingham prior to surgery, and then followed for 1 year after the procedure, as a cohort study. A second group of patients was recruited from five centres including Nottingham, who had undergone treatment 1 year or 5 years previously. The outcome of their treatment was evaluated as a cross sectional study.

The candidate personally conducted all assessments of patients for both studies.

3.2 Centre Enrolment for cross sectional study

Centres were identified and approached on the basis of perception of the estimated number of Dupuytren's cases treated per year, willingness to support the project, and their treatment of choice for Dupuytren's disease (with the aim of recruiting centres with a range of preferred treatments). At each centre, a local consultant surgeon was appointed to act as principal investigator (PI) for that site by local consensus.

In addition to the Queen's Medical Centre (QMC) (Nottingham University Hospitals NHS Trust), where the candidate was employed, other centres approached that met the above criteria were:

- Rotherham General Hospital (RGH)
  (Rotherham NHS Foundation Trust)
- The Pulvertaft Hand Centre within The Royal Derby Hospital (PHC)
  (Derby Hospitals NHS Foundation Trust)
- Derriford Hospital (DH)
  (Plymouth Hospitals NHS Trust)
- St John's Hospital at Howden (SJH)
  (NHS Lothian)

3.3 Generic Approvals & Nottingham Approval

The candidate registered with the Information Commissioner's Office (www.ico.gov.uk) as a Registered Data Controller (Z2200077), with approval for the specific purposes of the project (approval included: holding identifiable details including patient names and addresses, demographics, hand function, and digital images; transporting them between centres; storing paper copies of data capture in a secure environment in QMC; and storing electronic copies on a fully encrypted device). However, the processes used for the project were designed to considerably reduce the risk of data protection being breached.

The original project processes were reviewed and approved as service evaluation by Dr Brian Thomson (Director of Research) and by Charlotte Davies (Operations Manager), of the Nottingham University Hospitals Research & Innovation team (see Appendix 1).

As an employee of Nottingham University Hospitals NHS Trust, the candidate completed all required pre-employment checks mandated for work in the NHS. Employment as a Clinical Fellow in Trauma & Orthopaedics constituted membership of the 'usual clinical care team' for hand surgery patients. Photographic identification in the form of Nottingham University Hospitals NHS photographic ID was worn at all times.

3.4 Cross sectional study centre approval

With assistance from the PI at each centre, the project received approval from Research & Development departments (R&D) and Information Governance departments (IG).

At one centre (SJH), IG referred the project for further approval by The Caldicott Guardian. Data collection and handling processes were revised to meet The Caldicott Guardian requirements at this centre.

One centre (PHC) was able to provide administrative support from a dedicated Research & Postgraduate Secretary, and the processes used at this centre were modified to account for this.

Patient recruitment commenced at a centre once local approval was obtained.

3.5 Cross sectional study patient identification & contact

The PI at each centre formally approached his or her clinical coding department. Patients who had undergone treatment for primary Dupuytren's disease either 1 year earlier or 5 years earlier were identified using the Office of Population Consensuses and Surveys Classification of Interventions and Procedures version 4 (OPCS-4) codes (Table 3.1).

CODE	DESCRIPTION
T521	Palmar fasciectomy
T522	Revision of palmar fasciectomy
T525	Digital fasciectomy
T526	Revision of digital fasciectomy
T541	Division of palmar fascia
T548	Other specified
T561	Dermofasciectomy
T562	Revision of dermofasciectomy

Table 3.1: OPCS-4 codes used to identify prospective study participants

Lists were screened either by clinical coding staff, or by other workers at the local centre, to identify and exclude those who met the following criteria:

- Deceased
- Address over 20 miles from local centre by road
- Nursing home residents

3.6 Cross sectional study recruitment at Nottingham

Patients who were not excluded from Nottingham were contacted by telephone to invite them to participate in the service evaluation. An appointment was scheduled with those who agreed to participate in principle, at a mutually agreeable time and location. At the start of the appointment, patients were provided with written information about the project (Appendix 2) and given the opportunity to withdraw their consent for participation.

3.7 Cross sectional study recruitment at other centres

The local centre PI wrote to all patients who were not excluded, using local centre NHS Trust headed paper to invite them to participate in the project. R&D at Rotherham General Hospital approved the content of the invitation letter (Appendix 3). Thereafter, each centre's consultant made only minor stylistic changes to the content of the letter. A consent form (adapted from that used by R&D at RGH, and approved at each centre (Appendix 4) was enclosed with the letter of invitation. A stamped addressed envelope for the return of the consent form was also enclosed.

The return address for the consent form was determined by local IG approval and the presence of administrative support, as listed in Table 3.2.

Centre	Return Address	Rationale
PHC	PHC	Administrative support available
RGH	QMC	Approved by local centre
DH	QMC	Approved by local centre
SJH	SJH	Stipulated by Caldicott Guardian

Table 3.2: Return postal arrangements for consent forms for different local centres

Patients who consented to participate were contacted by telephone by the author to arrange an appointment at the local centre, under the nominal supervision of the local centre consultant hand surgeon. If patients could not attend the local centre and yet wished to participate, an appointment at another location of their choice was arranged.

All assessments were performed within a two-month window of the anniversary of their surgery (i.e. 10-14 months or 58-62 months post-surgery). If an appointment could not be made within the time frame, the patient was excluded from the study.

3.8 Cross sectional study data capture

During the single appointment, data were captured by targeted history and examination using a proforma approved by all centres (Appendix 5) and completion of PROMs. Data captured included patient demographics,

purported risk factors for the development of Dupuytren's disease, measures of outcome, and complications.

Captured items included:

- procedure type
- hands and digits treated
- age at surgery
- postoperative splinting
- gender
- hand dominance
- previous Dupuytren's disease surgery
- present occupation
- main previous occupation and occupational impact of Dupuytren's disease.

Procedures were divided into aponeurotomy (with needle or blade), fasciectomy and dermofasciectomy. Collagenase was not studied as part of this project, due its United States Food and Drug Administration (FDA) approval only being granted 1 year prior to the project commencing (FDA, 2010). Procedure type was determined by description of the procedure by the patient supported by clinical examination of scarring on the hand by the author, followed by reference to the OPCS-4 code for the procedure and the case notes if required.

Information was recorded on procedures to all hands and digits treated on the date concerned. If both hands were treated with the same procedure (e.g.

needle aponeurotomy to right and left ring fingers in one operation), then the one treated digit on each hand was assessed, as a separate entity.

If more than one digit on a hand was treated with the same procedure on the date concerned, then only one digit was assessed. The digit selected in such cases was the one that, at follow up, had the worst total active flexion deficit as assessed by eye. For example, for a single procedure comprising fasciectomies to right ring and little fingers, after which the ring finger had remained straight, but little finger had contracted towards palm, then only little finger would be assessed.

If more than one procedure was performed on a hand then both procedures were recorded as separate entities, for example a fasciectomy to the left ring finger and revision dermofasciectomy to the left little finger in one operation would be recorded independently. Whether both events were used, or the patient was excluded from analysis, depended on the particular analysis concerned. Such information is provided in the methods section of each chapter.

Age at surgery was determined from the patient's date of birth at the operation date provided by clinical coding.

Whether the patient had undergone Dupuytren's disease surgery previously was determined by a combination of patient recall, examination of the hands for additional scarring, and reference to case notes if uncertainty remained.

Patients provided details of their current and significant previous occupational histories. They were asked whether their occupation changed once

Dupuytren's disease had become apparent, and whether it had changed following the surgical procedure under assessment or following any previous surgery.

Risk factors included:

- known family history
- current smoking status
- current alcohol consumption
- epilepsy, and medications if currently treated
- diabetes mellitus, and current treatment
- knuckle pads
- Ledderhose's disease

A positive family history was recorded only if the patient was confident that a biological relative had had Dupuytren's disease, regardless of whether the relative was of a more senior or more junior generation, and irrespective of whether it was medically diagnosed or treated.

Positive smoking status was only recorded if the patient described himself or herself as a current active smoker. Previous smoking history was not recorded.

Alcohol consumption was the patient's self-report of the UK units consumed per week on average. If the patient could not specify the number of units, then the product and volume were recorded, and number of units calculated using the tool available from the NHS Choices website (NHS, 2013).

Epilepsy and diabetes mellitus were recorded if the patient reported a diagnosis of either condition made by a doctor. Diabetes mellitus was then divided into type 1 or type 2 diabetes. Current treatments were recorded.

Specific clinical examination was carried out to identify knuckle pads. Knuckle pads were considered to be present if there was at least one subcutaneous firm nodule separate from the skin overlying the PIPJ. Cutaneous thickening of the skin alone was not counted as a knuckle pad.

Ledderhose's disease was described to patients as involving similar nodules affecting the soles of the feet. If patients had clearly noted such lumps, this was recorded. If they were unsure, the feet were examined to confirm the presence of Ledderhose's disease.

Outcomes included:

- passive extension deficit at MCPJ and PIPJ by goniometry
- one or more PROMs
- further surgery for Dupuytren's disease, either to the same digit or to the same hand

Passive extension was measured at each joint individually, with the other joint of the ray held in passive flexion.

All patients completed the DASH questionnaire. The URAM questionnaire was published during the course of the project (Beaudreuil et al., 2011), and was added to the data capture once it had been reviewed. After completing all PROMs, patients were asked to specify which PROM they considered to have been the most appropriate assessment of their hand function.

DASHs with more than 3 missing answers were excluded from analysis (in keeping with the developer's instructions (Kennedy et al., 2011)). URAMs with any missing answers were excluded from analysis.

Complications included:

- loss of active flexion at MCPJ and PIPJ
- loss of active flexion recorded as fingertip-distal palmar crease distance in the same digit
- altered two point discrimination sensation in the distribution of the radial digital nerve and the ulnar digital nerve
- trauma-induced cold associated symptoms (cold intolerance)
- complex regional pain syndrome (CRPS)

CRPS was defined using the clinical criteria devised for use in orthopaedic surgery (Atkins, 2003), that are modified from those established by the International Association for the Study of Pain (IASP) (Stanton-Hicks et al., 1995). These comprise the presence of the following symptoms, none of which arise in uncomplicated Dupuytren's disease (Atkins, 2003):

- neuropathic pain
- vasomotor instability
- reduced joint mobility
- soft tissue contractures

Trauma-induced cold associated symptoms (cold intolerance) were identified using the scale described by Campbell and Kay (Campbell and Kay, 1998). This scale recognises five types of cold associated symptom: pain, numbness,

tingling, stiffness, and colour change. Symptoms severity is scored as summarised in Table 3.3.

GRADE	DESCRIPTION
1	Severe symptoms
2	Troublesome symptoms
3	Minor symptoms
4	Symptoms present, but cause no trouble
5	No symptoms

Table 3.3: Campbell's scale of cold-associated symptom severity

3.9 Cohort study recruitment

Patients booked for fasciectomy or dermofasciectomy at QMC were invited to participate in the cohort study when attending the preadmission clinic. Written information about the project was provided (Appendix 2) and patients were allowed ample time to read it. If patients wished to participate, verbal consent was taken in keeping with local Nottingham University Hospitals NHS Trust policy for service evaluation.

3.10 Cohort study data capture

Patients who consented to participate were reviewed during their preadmission clinic visit. A standardised proforma was used to capture details of patient demographics, planned procedure and risk factors for poor early outcome (Appendix 6).

A single observer measured finger joint angles. When each MCPJ and PIPJ was measured, the other joint in the ray was held in maximum passive flexion. Patients completed PROMs to describe their preoperative state.

Once each patient's planned operation date had passed, confirmation that the operation had proceeded as planned was obtained through the Trust's electronic appointment system. If surgery was cancelled, this was noted and the patient excluded from further contact. If surgery was postponed, confirmation that surgery had taken place on the revised date was ensured in the same way. The specific procedure performed during the surgery was also identified from this system, in case a planned fasciectomy had been converted to a dermofasciectomy intraoperatively, and *vice versa*.

DASH PROMs were posted to patients to be completed three weeks and six weeks after surgery and returned by post using a stamped addressed envelope enclosed. The PROMs were posted to patients within the week preceding the three or six week time point to ensure receipt prior to the date concerned. Patients were requested to complete the PROM on the appropriate date and to return to QMC using the stamped addressed envelope.

Those who returned both three and six week PROMs, and those who completed the six week PROM but not the three week PROM, were sent a further DASH at one year following surgery. Enclosed with the DASH was a PROM that assesses change in state: a Global Rating of Change (GROC) PROM. Again, a stamped addressed envelope was enclosed to return

completed PROMs. Clinical contact at Nottingham University Hospitals NHS Trust with patients who had returned one year PROMs was reviewed using the electronic appointments system, to identify those who had undergone further hand surgery since the fasciectomy or dermofasciectomy a year earlier.

3.11 Data handling and analysis

All patients from both the cross sectional study and the cohort study were allocated a study reference number during initial contact for the cross sectional study, and during initial data capture for the cohort study. This reference number, pseudo-anonymised the patients. Names and hospital numbers were stored with the study code at the local centre where the patient was recruited.

Data were entered into password-protected Microsoft Excel® spread sheets (Microsoft for Mac 2011 version, Microsoft®, Redmond, USA, 2010) with each patient identifiable by their study code. The spreadsheets were stored on a laptop computer with an encrypted hard drive. Paper copies of data collection sheets were stored securely at local centres. Study code allocations for patients from each centre were also stored at that centre.

Data analysis was performed using SPSS® Statistics version 21 (IBM® Software, Armonk, USA, 2012), and Prism 6.0 for Mac OS X (GraphPad® Software, La Jolla, USA, 2012).

As data analyses were specific to the individual components of the project, they are discussed in the methods section of each chapter individually.

3.12 Accuracy of data entry

Prior to data analysis, a blinded independent observer assessed accuracy of data entry into Microsoft Excel® spreadsheets. The independent observer was an orthopaedic surgery registrar from the hand surgery firm, and was therefore part of the usual clinical care team for hand surgery patients at Nottingham University Hospitals NHS Trust at the time concerned. A target of less than 2% inaccurate data entry was set prior to the audit being carried out. If more than 2% of data entries were inaccurate, then all data entry from all centres would be audited. If less than 2% of data entries were inaccurate, then this would be tolerated.

An electronic random number generator provided five numbers. The patient data corresponding to those row numbers in the master spreadsheet of patients recruited in Nottingham were extracted. These five patient datasets comprised a total of 720 data items. The independent observer reviewed the spreadsheet data against the original paper copies of data capture sheets on 17th April 2013. The observer identified one error of data entry, giving a data entry error rate of 0.14%. Therefore, data entry was deemed accurate, and analysis proceeded as planned.

4 Systematic Review and Meta-analysis of Surgical Trials in Dupuytren's Disease

4.1 Preface

The previous chapter described a new guideline development process to be used by the United Kingdom's national hand surgery society, the British Society for Surgery of the Hand (BSSH). It is likely that there is insufficient high quality comparative evidence to be able to produce a clinically useful guideline for the treatment of Dupuytren's disease at present. Furthermore, it is necessary to appraise the existing high quality studies prior to designing further research. In addition to avoiding unintentional repetition of research, it will also identify aspects of trial design in Dupuytren's disease that contribute to achieving high study quality, as well as aspects of design that have downgraded the quality of the study, based on the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach (Atkins et al., 2004). This may be particularly important, as minimising risks of bias such as performance bias, may be difficult to achieve in surgical trials where blinding surgeons is difficult. This chapter describes a systematic review of trials of the surgical treatment of Dupuytren's disease, including a meta-analysis of trials investigating the role of postoperative splinting.

4.2 Background

Comparative studies of the different treatment options for Dupuytren's disease are needed to determine the role of different procedures. In particular, whether the benefits of more invasive procedures, such as dermofasciectomy, that are believed to have lower "recurrent contracture" rates outweigh potentially higher rates of adverse events (complications) or a less acceptable rehabilitation period. Whilst comparison of different operative techniques is important, it must be recognised that surgery is only part of a complex intervention for the treatment of Dupuytren's contracture. The outcome of this may not be exclusively determined by what surgery is performed, but also by the post-operative rehabilitation regimen (splinting and hand therapy) and other treatment factors such as patient selection and site of contracture (metacarpophalangeal joint alone, proximal interphalangeal joint alone, or both joints together). Also the outcome of Dupuytren's surgery is usually defined by the "recurrent contracture rate", in contrast to "disease extension" to other digits, the rate of which is unaffected by surgery. Only a few studies have assessed outcome in a patient-centred manner, or investigated the severity and length of the post-operative recovery from the surgery.

4.3 Objectives

The aim of this review was to assess the effects (benefits and harms) of different surgical procedures for the treatment of Dupuytren's contracture of the index, middle, ring and little fingers (the thumb will be excluded, where

disease is rarer, and where cords form on the radial aspect of the thumb, and thus are not readily accessible in terms of angular deformity.).

4.4 Methods

4.4.1 Criteria for considering studies for this review

Randomised controlled trials (RCTs) and controlled clinical trials (CCTs) were included, irrespective of sample size or language of publication.

Participants could be adult men and women of any ethnicity, with or without risk factors for Dupuytren's disease, who had undergone surgery for Dupuytren's contracture of one or more of the index, middle, ring and little fingers.

Interventions studied were any surgical intervention, including:

- Percutaneous Needle Fasciotomy (Aponeurotomy)
- Very Limited Fasciectomy
- Limited Fasciectomy
- Dermofasciectomy

Comparators that would be included could vary from:

- Different Surgical Procedures
- Placebo/Sham Surgery
- Other Active Non-Surgical Treatments (Collagenase Injection, Physiotherapy, Radiotherapy, Hand Therapy)

It was not anticipated that there would be studies comparing treatment to observation alone. If such studies were identified, they were discussed.

The validity and reliability of any outcome measure commonly used in Dupuytren's disease has not been well studied. Anticipated outcomes have been listed below. Hand function was selected as the key primary outcome, as this represents an important patient-centred measure. In contrast, angular measurements are surgeon-centred measurements.

4.4.2 Major outcomes

Change in hand function

This could be determined by any appropriate assessment, such as the Disabilities of the Arm, Shoulder and Hand (DASH) patient-reported outcome measure (PROM) (Hudak et al., 1996), PEM (Macey et al., 1995), grip strength, Jebsen-Taylor Hand Function Test (Jebsen et al., 1969). It was uncertain which standardised outcome instruments would be encountered but all were reported.

Patient satisfaction and other patient rated outcomes

This could include pain or health-related quality of life (HRQoL).

Early angles outcomes and other objective outcomes

Possible outcomes in this group include 1) change in contracture after surgery - the difference between the finger angle measurement immediately after surgery and the preoperative finger angle measurements, or 2) residual contracture after surgery - as assessed by angle measurement (goniometry), or 3) early result (as above) at another relevant time point, such as time of discharge from care. Active or passive angles may be reported. Angles may

be presented per joint, or per ray (e.g. total extension deficit across the MCPJ and the PIPJ, or across the MCPJ, the PIPJ and the DIPJ.

Recurrence of Dupuytren's disease/contracture where previously treated

Given that recurrence is time-dependent, length of follow-up has not been standardised and there is no consensus definition of recurrence, recurrence rates and length of follow-up were described in narrative form. Where appropriate data existed, time-to-event analyses would be performed. However, it was not expected that suitable data would be available. Meta-analyses would only have been performed for studies with similar definitions of recurrence and recurrence data at similar follow-up times after surgery. The minimum length of follow-up for eligibility in analyses of recurrence was 18 months. This was chosen for two reasons: firstly, shorter follow-up gives insufficient time for recurrence, and secondly, there was no current consensus to define minimum length of follow-up, and this varies widely, from 3 weeks to 13 years (Becker and Davis, 2010).

Adverse effects

Anticipated adverse effects included loss of finger flexion, loss of finger sensation due to digital nerve injury, vascular compromise, delayed healing and infection. As the completeness of reported adverse events was unknown, all adverse effects data were collected and reviewed. This review focussed on key adverse events should there prove to be extensive reporting of numerous adverse effects.

4.4.3 Minor outcomes

Cost effectiveness

Where provided this would be assessed using the total cost of the procedure and rehabilitation. Where time-to-recurrence was documented, cost per year of recurrence-free survival would be calculated. However, it was not anticipated that these data would be available.

4.4.4 Search methods for identification of studies

All searches were performed on 17 September 2012, and re-run on 10th March 2014 to update the results.

The following databases were searched to identify reports of relevant RCTs and CCTs:

- The Cochrane Central Register of Controlled Trials (CENTRAL), *The Cochrane Library*, Issue 8 2012
- BNI (British Nursing Index and Archive) - 1985 to September 2012
- CINAHL - 1981 to September 2012
- EMBASE - 1980 to September 2012
- LILACS (Latin American and Caribbean Health Sciences) - 1982 to September 2012
- Ovid MEDLINE - 1948 to September 2012
- Ovid MEDLINE - In-Process and other Non-Indexed Citations - 1948 to September 2012
- Proquest (ABI/INFORM Global and Dissertations & Theses)- all entries to September 2012

- ISI Web of Science
- clinicaltrials.gov

The full search strategy for CENTRAL was:

#1 MeSH descriptor Dupuytren explode all trees

#2 MeSH descriptor Fibromatosis explode all trees

#3 Dupuytren*:ti, ab, kw

#4 (#1 OR #2 OR #3)

#5 MeSH descriptor surgery explode all trees

#6 MeSH descriptor fascia explode all trees

#7 (#5 OR #6)

#8 (#4 OR #7)

#9 MeSH descriptor Dupuytren surgery explode all trees

#10 (dupuytren NEAR/3 surgery*): ti, ab, kw

#11 (contracture NEAR/3 surgery*): ti, ab, kw

#12 (#8 OR #9 OR #10 OR #11)

The Ovid MEDLINE search was combined with the Cochrane Highly Sensitive Search Strategy for identifying randomised trials in MEDLINE: sensitivity- and precision-maximizing version (2008 revision) (Lefebvre et al., 2011):
Database: Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations and Ovid MEDLINE(R) <1948 to Present>Search Strategy:

---

1 randomised controlled trial.pt. (319496)

2 controlled clinical trial.pt. (83719)

3 randomised.ab. (224758)

4 placebo.ab. (129278)

5 clinical trials as topic.sh. (158838)

6 randomly.ab. (161802)

7 trial.ti. (96632)

8 1 or 2 or 3 or 4 or 5 or 6 or 7 (742481)

9 exp animals/ not humans.sh. (3701210)

10 8 not 9 (685232)

11 exp Dupuytren Contracture/ (2035)

12 exp Fibroma/ (10932)

13 Fibromatosis.tw. (2305)

14 exp Fascia/ (8039)

15 Fibroblasts/ (90308)

16 (palmar adj3 fascia).tw. (186)

17 Dupuytren*.tw. (1902)

18 (palmar adj3 fibromatosis).tw. (63)

19 (viking adj3 disease).tw. (1)

20 or/11-19 (111278)

21 10 and 20 (689)

The EMBASE and CINAHL searches were combined with the trial filters developed by the Scottish Intercollegiate Guidelines Network (SIGN). There were no language or date of publication restrictions.

Variations of the Ovid MEDLINE search strategy were used to search the other databases.

The reference lists of shortlisted articles were screened to identify further suitable studies and Web of Science was searched to identify studies citing the items in the shortlist. No language restrictions were applied and potentially eligible foreign language studies were obtained and translated using electronic web-based translation.

4.4.5 Selection of studies

From the title, abstract or descriptor, two reviewers (the candidate and a second author) independently screened all abstracts to identify potential studies for review using a checklist of the criteria for inclusion:

Q1. Does the paper report the outcome of a clinical study? (i.e. not a review article or just a paper describing an operative technique description)?

Q2. Have participants had a surgical intervention for Dupuytren's contracture of a finger?

Q3. Did the study report either the short term or long term outcomes (recurrence) of the surgery?

Q4. Did participants receive an intervention compared to a control group or

were at least two interventions compared?

Q5. Was the study randomised or quasi-randomised?

The two authors compared their lists of potential studies and produced a consensus shortlist. Copies of full papers on the shortlist were obtained.

Two review authors (the candidate and a second author) independently reviewed the full text of the abstracts of the "agreed shortlist" papers and identified those suitable for inclusion using the checklist above. Disagreements were resolved by referral to a third author. No masking of titles of journals, names of authors or supporting institutions was performed.

4.4.6 Data extraction and management

Data describing source, study design, intervention, population and outcomes were extracted using a piloted form by two authors (the candidate and a second author) independently. Disagreements were resolved by consensus after an additional review by a third author.

4.4.7 Assessment of risk of bias in included studies

Two authors (the candidate and a second author) independently used The Cochrane Collaboration's tool for assessing risk of bias (Higgins JPT and Green S, 2011). All seven domains (sequence generation, allocation concealment, blinding of participants and personnel, blinding of outcome assessment, incomplete outcome data, selective outcome reporting and other

issues) were assessed and classified as 'high risk of bias', 'low risk of bias' or 'unclear risk of bias'.

Each outcome was judged upon the majority of the seven domains fitting either high or low risk. Disagreements were resolved by referral to a third author (the supervisor). As it was anticipated that few studies would employ blinding, the use of blocked randomisation in unblinded studies was assessed as a source of "other" bias risk.

4.4.8 Measures of treatment effect

If appropriate, standardised mean differences (SMD) were used to combine different outcome measures from different trials (Hedges, 1982).

If studies reported dichotomous data, risk ratios with 95% confidence intervals would be calculated. For rare events (< 10%), Peto odds ratios with 95% confidence intervals were calculated.

4.4.9 Unit of analysis issues

It was expected that the hand would be used as the unit of randomisation in most studies. Assessing outcomes such as hand function would not be possible if individual digits on the same hand were used as the unit of randomisation. The unit of randomisation (patient, hand, finger or unclear) was recorded for each included study.

It was not expected that there would be any crossover studies, given that the interventions here are definitive treatments. Cluster randomised studies were not expected to be encountered.

4.4.10 Dealing with missing data

Two types of missing data were anticipated: unreported and withdrawn. If there were unreported data in included studies, the authors were contacted for assistance. No imputation was attempted.

4.4.11 Assessment of heterogeneity

If appropriate, statistical heterogeneity was tested using visual inspection of graphs, Chi square and I square statistic tests. A Chi square test result was considered significant when $p < 0.10$. An I square test result greater than 50% was considered to demonstrate substantial heterogeneity.

4.4.12 Assessment of reporting biases

To reduce the risk of reporting bias, multiple sources were searched, including Proquest (ABI/INFORM Global and Dissertations & Theses), to identify all unpublished results where possible.

Funnel plots were drawn to assess the risk of publication bias.

ISI Web of Science was searched to identify relevant results that had not been published. If such work were identified, the authors were contacted and asked to provide a copy of the data.

4.4.13 Data synthesis

Extracted data from included studies were compared using the Cochrane Collaboration's statistical software, Review Manager. If studies were sufficiently similar, meta-analysis was undertaken.

When the same outcome measures were assessed with different scales, standardised mean differences (SMDs) were used.

However, it was expected that data from different studies would be difficult to compare, such that meta-analysis might not be possible. This would particularly be the case for "recurrence", because of variation in follow-up length (the recurrence rate increases longer follow-up) and differences in the definition of "recurrence" between studies.

A random-effects model was used to combine data if outcomes were homogeneous. If the results were heterogeneous, reasons for heterogeneity were identified through subgroup analysis. If significant heterogeneity was found, the data was not pooled and a narrative (qualitative) summary of methodological quality and results was presented

4.4.14 Subgroup analysis and investigation of heterogeneity

Reasons for heterogeneity were considered using subgroup analysis. Subgroup analysis was considered with regard to:

- Length of time to follow-up
- PIPJ and MCPJ outcomes separately

  As it is well recognised that contractures of the latter correct better than those of the former

- Severity of disease prior to operation

  Where provided, it was expected that this would be in the form of total passive extension deficit at the MCPJ and PIPJ

- Number of joints involved

- Postoperative treatment offered

4.4.15 Sensitivity analysis

Outcome measures, such as the definition of recurrence, have been defined differently (Becker and Davis, 2010). If required, sensitivity analyses were be performed to examine whether the result varied based on definitions.

We also undertook sensitivity analysis for missing data, for example intention to treat versus per protocol analysis, to examine variations between different analytic approaches.

4.5 Results

4.5.1 Description of studies

4.5.1.1 Search Results

The search strategy retrieved 2382 references. One hundred and three duplicates were removed prior to abstract screening.

For three studies, inadequate detail was available from the abstract and reference to make a decision, and the original paper could not be obtained (Gazdzik and Wasilewski, 1997, Slullitel, 1987, Yoshida et al., 1998). They were excluded on this basis. Two could not be classified as the full text article did not adequately describe the study design, such that it was not clear if they were randomised or pseudorandomised (Hazarika et al., 1979, Ward, 1976). Both of these articles are over 35 years old, and clarification could not be obtained from the authors.

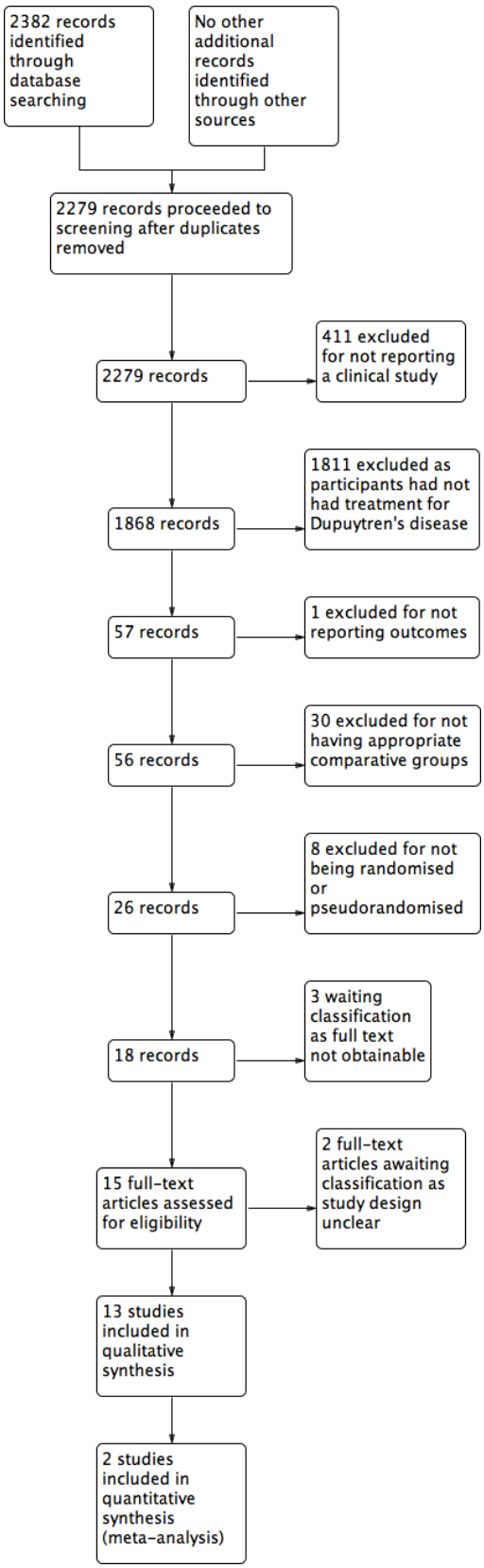

Figure 4.1: Flow chart of article selection

Of the remaining 2279, 13 were included in the review (Bhatia et al., 2002, Bulstrode et al., 2004, Chignon-Sicard et al., 2012, Citron and Hearnden, 2003, Citron and Nunez, 2005, Collis et al., 2013, Howard et al., 2009, Jerosch-Herold et al., 2011, Kemler et al., 2012, McMillan and Binhammer, 2012, Ullah et al., 2009, van Rijssen et al., 2006, van Rijssen et al., 2012), and 2264 were excluded (see Figure 4.1).

4.5.1.2 Included studies

Ten studies were single centre studies, of which seven were based in the United Kingdom, one each in Canada, France and New Zealand and two were reports of one trial based in the Netherlands. There were two multi-centre studies (Jerosch-Herold et al., 2011, Kemler et al., 2012). All of the five centres in the former were from the UK. Both centres in the latter were in the Netherlands. All studies were published in English. One study had an associated publication describing the trial protocol (Jerosch-Herold et al., 2008).

The 13 studies included 910 participants, of which the 93 reported in a 5-year outcome paper were participants in the same trial also described in the early outcome paper (van Rijssen et al., 2006, van Rijssen et al., 2012). Thus 817 individual patients were recruited across all studies.

Three studies described the outcomes of two trials that compared different surgical procedures (Ullah et al., 2009, van Rijssen et al., 2006, van Rijssen et

al., 2012). One trial compared using "firebreak" full thickness skin grafts (a form of dermofasciectomy) to z-plasty closure of a limited fasciectomy. The authors refer to the original description of 'firebreak' grafts (Hueston, 1984). Here 'firebreak' grafts are described as small grafts strategically placed at flexion creases. In contrast, traditional dermofasciectomy may involve resurfacing much larger areas of palmar skin (Seah et al., 2012). This was achieved by conducting a limited fasciectomy, and then excising palmar skin to accommodate the skin graft in those randomised to this cohort. The other two papers reported early and late outcomes respectively for a trial comparing needle fasciotomy to limited fasciectomy.

Four of the other ten studies compared surgical incision and wound management options (Bhatia et al., 2002, Citron and Hearnden, 2003, Citron and Nunez, 2005, Howard et al., 2009). These included studies that compared staple closure against suture closure, and absorbable versus no-absorbable suture closures respectively, both in limited fasciectomy. The other two studied types of incision for limited fasciectomy.

Two publications studied adjunctive treatments to surgery: one investigated bathing the operation site in 5-fluorouracil prior to closure, versus saline (Bulstrode et al., 2004), and one compared the use of steroid injections in conjunction with needle fasciotomy versus no adjunctive treatment (McMillan and Binhammer, 2012).

The four others studied non-invasive adjuncts to surgery, including the use of postoperative splints versus no splint, the use of a fibrin-and platelet-rich fibrin

plug as a primary dressing to open palm surgery versus conventional low adherence dressing, and comparing postoperative intermittent pneumatic compression to standard elevation of the limb (Chignon-Sicard et al., 2012, Collis et al., 2013, Jerosch-Herold et al., 2011, Kemler et al., 2012).

These different interventions can be used to classify studies into:

- Those studying different treatment options;
- Those refining a treatment option (e.g. limited fasciectomy incisions, closure types, invasive adjuncts or equipment usage);
- Those refining rehabilitation.

4.5.1.3 Inclusion and exclusion criteria

Criteria were not always specified. Two articles did not describe either inclusion or exclusion criteria (Bhatia et al., 2002, Howard et al., 2009). Of the other eleven, four specified age-related cut offs for recruitment, of age under 70 years (Bulstrode et al., 2004), and over 18 years (Chignon-Sicard et al., 2012, Jerosch-Herold et al., 2011, Kemler et al., 2012). One study did not describe the ratio of participant genders (Howard et al., 2009). None of the other studies explicitly excluded potential participants based on gender, though one study comprised only male subjects (Bulstrode et al., 2004).

Patients undergoing revision surgery were excluded in three studies (Citron and Hearnden, 2003, Citron and Nunez, 2005, McMillan and Binhammer, 2012). Of these, one also excluded patients who had previously undergone other types of hand surgery (McMillan and Binhammer, 2012).

Site of disease within the hand was specified in three studies. One study only included patients with palmar disease affecting the MCPJ (Citron and Hearnden, 2003). In contrast, another only included those with at least 30 degrees of contracture at the PIPJ (Ullah et al., 2009). One splinting study excluded thumb and first web space treatments (Jerosch-Herold et al., 2011). Another recruited participants with Dupuytren's disease in one ray only (Citron and Nunez, 2005).

Some studies used exclusion criteria relating to comorbidities that might influence outcome, such as bleeding tendencies (Citron and Nunez, 2005, Ullah et al., 2009, van Rijssen et al., 2006, van Rijssen et al., 2012). Diabetes mellitus was an exclusion criterion (Chignon-Sicard et al., 2012, McMillan and Binhammer, 2012).

Some specific criteria were related to study design. In one study, patients had to receive treatment for two rays in the one procedure, as one was randomised to receive 5-fluorouracil, and the other to receive the control treatment of normal saline (Bulstrode et al., 2004). In the Dutch papers describing fasciectomy versus aponeurotomy, participants had to have well-defined cords of disease (van Rijssen et al., 2006, van Rijssen et al., 2012). This is a requirement for suitability for needle fasciotomy.

4.5.1.4 Unit of analysis

The predicted unit of analysis was that studies would be randomised by 'hand'. This could lead to the same patient being enrolled twice, for surgery to each hand on separate occasions in some studies (van Rijssen et al., 2006,

van Rijssen et al., 2012). In contrast, one patient with bilateral disease was entered only once in another trial (McMillan and Binhammer, 2012). Similarly, six patients presented for randomisation twice in another study (Citron and Nunez, 2005), for the treatment of bilateral disease. They were only enrolled once in the trial. In these studies, the unit of randomisation was by 'patient'. Other studies have specific individualised methodologies, such as using an internal control, with one digit on a hand randomised to treatment and another to control (Bulstrode et al., 2004).

Reporting recurrence varied, including presenting the number of hands in the number of patients who had developed over 20 degrees of deformity in one joint (van Rijssen et al., 2012). Other studies presented recurrence per patient (Citron and Hearnden, 2003, Citron and Nunez, 2005). Only one study investigated recurrence (Ullah et al., 2009), in which recurrence was described as the percentage of fingers that showed recurrence, rather than the proportion of hands or patients.

4.5.1.5 Outcome measures

The outcomes measured varied between studies. Specific outcomes were used for particular studies.

Length of follow-up varied between papers. Those investigating rehabilitation and early recovery varied from two weeks follow up (Bhatia et al., 2002, Howard et al., 2009), to six weeks follow-up (van Rijssen et al., 2006). Late outcome papers varied in length of follow-up from two years (Citron and

Hearnden, 2003, Citron and Nunez, 2005) to five years (van Rijssen et al., 2012).

Hand function

Several trials presented patient-reported outcomes. These included previously published PROMs such as the DASH (Collis et al., 2013, Jerosch-Herold et al., 2011, van Rijssen et al., 2006, van Rijssen et al., 2012) or the PEM (Ullah et al., 2009). The design of studies with two digits on the same hand randomised to different groups (Bulstrode et al., 2004) would have prevented meaningful interpretation of hand patient-reported outcomes such as hand function.

Patient satisfaction and other PROMs

Patient satisfaction was reported in studies comparing procedure types (van Rijssen et al., 2006, van Rijssen et al., 2012). However, whilst statistical significance was presented, full data were not. Furthermore, the development, validity and reliability of the tools used to assess satisfaction were not described or referenced. Satisfaction was also assessed in other studies (Collis et al., 2013, Jerosch-Herold et al., 2011), as was patient-perceived change (Kemler et al., 2012). Some studies included self-reported pain assessed using a visual analogue scale (VAS) (Bhatia et al., 2002, Howard et al., 2009, Kemler et al., 2012). Patient-assessed wound appearance was also reported in one study (Bhatia et al., 2002), though the development, validity or reliability of the tool used was not described or referenced.

Early angles and other objective outcomes

Angular deformity was presented in different ways. Some presented active finger angles (Bulstrode et al., 2004, Collis et al., 2013, Jerosch-Herold et al., 2011, Kemler et al., 2012, McMillan and Binhammer, 2012), whilst others presented passive angles (van Rijssen et al., 2006, van Rijssen et al., 2012). In other studies, it was not clear whether the angles presented were active or passive (Citron and Hearnden, 2003, Citron and Nunez, 2005, Ullah et al., 2009). The presentation of angular measurements varied between the early and late outcomes of the same clinical trial (van Rijssen et al., 2006, van Rijssen et al., 2012). Other objective outcomes measured included timings, such as time taken to perform key tasks involved in surgery or postoperative care (Bhatia et al., 2002, Howard et al., 2009, Ullah et al., 2009), time to healing (Bulstrode et al., 2004, Chignon-Sicard et al., 2012), and an analysis of grip strength (Collis et al., 2013, Ullah et al., 2009).

Recurrence

The studies that reported comparisons of operative technique considered recurrence in late outcome papers (Citron and Hearnden, 2003, Citron and Nunez, 2005, Ullah et al., 2009, van Rijssen et al., 2012), and extension deficit at early outcome points (Ullah et al., 2009, van Rijssen et al., 2006). The definition of recurrence varied from reappearance of palpable disease in the operated field (Citron and Hearnden, 2003, Citron and Nunez, 2005) to recurrent angular deformity (Ullah et al., 2009, van Rijssen et al., 2012).

Within the studies comparing different procedures, recurrence was defined as an increase in a joint angle, MCPJ or PIPJ, of 20 degrees or more (van Rijssen et al., 2012), or was not explicitly defined (Ullah et al., 2009). However, the latter paper does discuss "progressive recurrence of contracture of the PIP joint", suggesting that angular deformity was employed for this, rather than the reappearance of palpable disease.

Adverse effects

Adverse effect reporting varied from not studying complications in a study of rehabilitation adjuncts (Jerosch-Herold et al., 2011), through reporting "no intraoperative complications" (Bulstrode et al., 2004), or "no complications" (Citron and Hearnden, 2003), to describing and attempting to quantify specific complications (Chignon-Sicard et al., 2012). There was no standardisation of the choice of adverse effects studied, even between similar papers.

Cost effectiveness

No included studies presented formal cost effectiveness analysis, though it was noted that several articles did assess cost effectiveness. However, none were randomised or pseudo-randomised studies, and so were excluded. Three studies presented analyses of time taken to perform key tasks involved in surgery or postoperative care (Bhatia et al., 2002, Howard et al., 2009, Ullah et al., 2009), which would be expected to have cost effectiveness implications.

SUMMARY

The primary objective of this review was to study trials comparing different treatment options. This group comprises only three papers describing two trials, and these two trials compared different interventions using different inclusion and exclusion criteria. One compared small 'firebreak' full thickness skin grafting to z-plasty closure of limited fasciectomy for contracture involving the PIPJ (Ullah et al., 2009). The other two described a trial comparing needle fasciotomy to limited fasciectomy, with inclusion criteria including contractures that may not necessarily affect the PIPJ (van Rijssen et al., 2006, van Rijssen et al., 2012).

Amongst trials refining intraoperative techniques, all compared different interventions. Amongst the trials refining rehabilitation adjuncts, two investigated the same intervention using comparable measures and timepoints, allowing meta-analysis (Collis et al., 2013, Jerosch-Herold et al., 2011).

The outcome measures studied varied and included a range of PROMs. Of these, the DASH was one such PROM. It was also identified as the most commonly reported PROM in Dupuytren's disease research (Ball et al., 2013). This supports further research into the DASH's behaviour in Dupuytren's disease.

4.6 Risk of bias in included studies

Risks of bias for each study are summarised in Figure 4.2.

4.6.1 Allocation (selection bias)

Two trials did not explain the randomisation process used, and one used alternation. The remaining nine provided information on appropriate randomisation processes, albeit with poor descriptions of allocation concealment in all but three studies that adequately described secure processes.

Allocation concealment processes were robust in three studies only (Chignon-Sicard et al., 2012, Citron and Nunez, 2005, Jerosch-Herold et al., 2011). The former two described sealed, sequentially numbered opaque envelopes, whereas the latter used telephone randomisation from another site. Three other studies also used numbered sealed envelopes, but did not describe whether these were opaque or not (Ullah et al., 2009, van Rijssen et al., 2006, van Rijssen et al., 2012). Similar inadequate detail of concealment was provided in other studies.

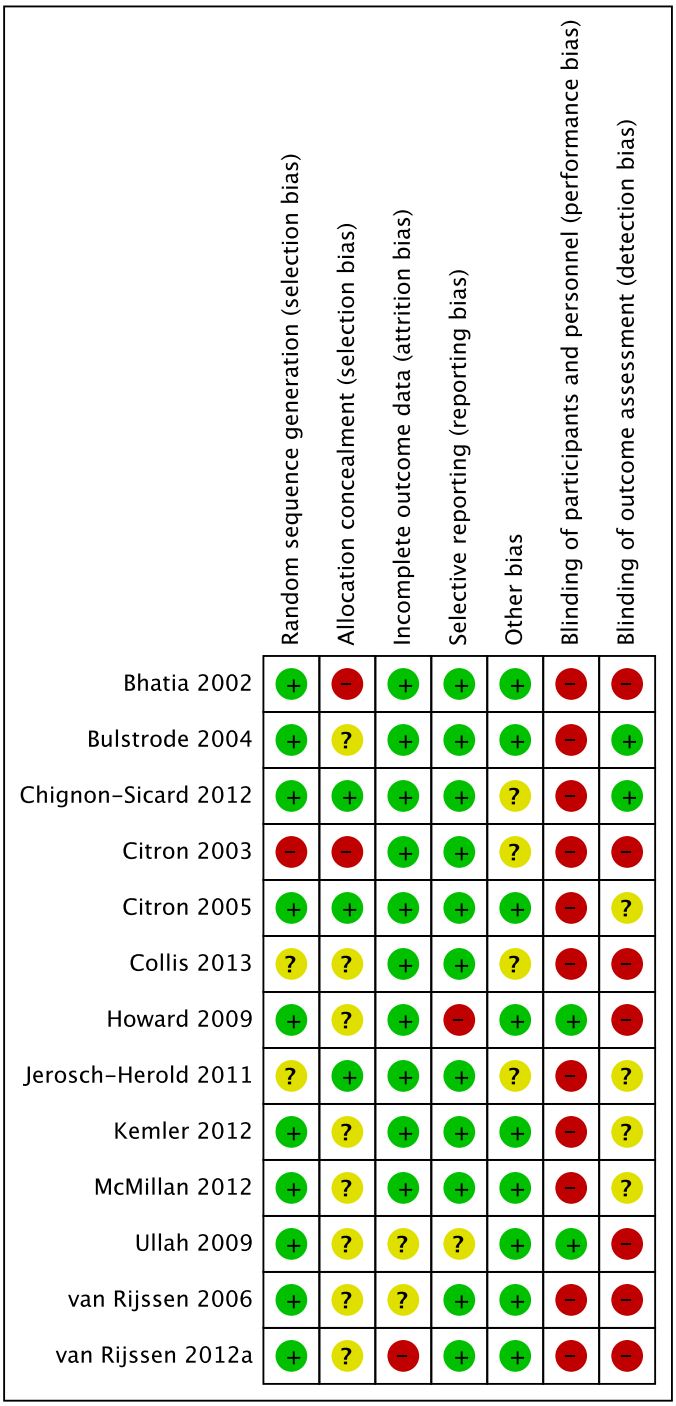

Figure 4.2: Risks of bias in included studies

+ indicates low risk of bias; - indicates high risk of bias; ? indicates inadequate detail to make judgement, or inconclusive

4.6.2 Blinding (performance bias and detection bias)

As the treatment involved is a surgical procedure, it is acknowledged that many trials are likely to be at high risk of performance bias, as the surgical team performing the procedure cannot always be blinded. However, trials of wound closure and adjuncts could defer randomisation until after the corrective element of the procedure has been completed. This was only done in one study (Ullah et al., 2009). Several other studies (Bhatia et al., 2002, Bulstrode et al., 2004, Chignon-Sicard et al., 2012, Howard et al., 2009, McMillan and Binhammer, 2012) could have deferred randomisation in this way to reduce the impact of performance bias on other parts of the procedure, but did not.

Few studies explicitly described blinding of assessment. In one double blinding of the patient in addition to the assessor was employed (Bulstrode et al., 2004). It is acknowledged that such blinding may be difficult to achieve when comparing procedures that leave distinctive and very different scar patterns on the hand (such as needle fasciotomy and fasciectomy). One also described blinding of outcome assessment (Chignon-Sicard et al., 2012).

4.6.3 Incomplete outcome data (attrition bias)

Attrition was not formally described in several studies. The study with the longest follow-up period (van Rijssen et al., 2012), had significantly different levels of attrition between the groups, which could have been influenced by the outcome of the treatment. 'Unclear' studies did not explicitly describe levels of attrition.

4.6.4 Selective reporting (reporting bias)

In one study (Howard et al., 2009), outliers were excluded despite formally testing the normality of the distribution of the data and electing to use non-parametric statistics. It is possible that the primary conclusion of the study could become invalid with these outliers included in the analysis. No protocol for their exclusion was described. One further study (Ullah et al., 2009) listed the highest number of secondary outcomes, but did not describe them all in detail, with some represented only graphically.

4.6.5 Other potential sources of bias

As many studies were expected to be unblinded, the risk from blocked randomisation in such studies was considered.

SUMMARY

Minimising risk of bias in trials of surgery of this nature can be difficult to achieve. However, there are aspects of the included studies here that demonstrate that most common sources of bias could be controlled in a future trial of Dupuytren's disease treatment.

4.7 Effects of interventions

4.7.1 Comparing procedure types

4.7.1.1 Comparing needle fasciotomy with fasciectomy

One trial compared these procedures, with early and late outcomes reported separately (van Rijssen et al., 2006, van Rijssen et al., 2012). The early outcome article reported 125 hands in 121 patients. The late outcome article had 93 participants from the original cohort. This comparison involved low quality evidence due to study design limitations and imprecision.

Hand function

There was low quality evidence that hand function, as determined by the DASH, may be significantly lower after needle fasciotomy than fasciectomy at all time points up to 5 weeks following surgery (van Rijssen et al., 2006). However, only 97/121 (80%) patients completed the PROM tool adequately to allow analysis. There was no evidence describing later functional outcome (van Rijssen et al., 2012).

Patient satisfaction and other patient rated outcomes

The comparison of satisfaction between fasciotomy and fasciectomy was only described in terms of p values, without effect sizes. There was low quality evidence that patient satisfaction may be significantly higher for the needle fasciotomy group compared to the fasciectomy group at 6 weeks ($p = 0.003$)

(van Rijssen et al., 2006). Satisfaction reversed by five years; it was significantly higher for fasciectomy ($p < 0.001$) (van Rijssen et al., 2012). There was low quality evidence that, overall, satisfaction was lower in patients with recurrence ($p < 0.001$) (van Rijssen et al., 2012). However, the tools used may not have been robustly developed or validated, as has been discussed already, and not all data was presented.

Early angles and other objective outcomes

There was low quality evidence that correction in total passive extension deficit was not different between procedures for milder contractures (Tubiana stage I and II, which equates to total passive extension deficit across all joints of less than 90 degrees) by six weeks, but limited fasciectomy achieved significantly better correction for more severe contractures (Tubiana III and IV, i.e. over 90 degrees of total passive extension deficit).

Recurrence

There was low quality evidence that recurrence may be significantly higher five years after needle fasciotomy (84.9% of hands after fasciotomy versus 20.9% of hands after fasciectomy, $p < 0.001$) (van Rijssen et al., 2012).

Adverse effects

There was low quality evidence that complication rates were similar between procedures in terms of infection, haematoma, wound slough, skin fissure, sympathetic dystrophy, altered sensation, digital nerve injury, tendon injury, revision surgery. The incidence of paraesthesia was significantly higher after limited fasciectomy than after fasciotomy.

Cost effectiveness

There was no evidence identified comparing cost effectiveness of needle fasciotomy and fasciectomy.

SUMMARY

There was evidence that needle fasciotomy delivered better satisfaction and function at early outcome than fasciectomy, though poor rates of completion of the PROM were an issue. Fasciectomy was more effective at correcting severe disease. Recurrence was higher at five years after needle fasciotomy, though functional outcome had not been described. The cost effectiveness of performing multiple needle fasciotomies over a period compared to a single fasciectomy had not been studied. At present, there is insufficient evidence in key areas to determine which is a superior treatment overall. These on going uncertainties justify further research into the treatment of Dupuytren's disease.

4.7.1.2 Comparing dermofasciectomy with fasciectomy

One study compared firebreak skin grafting to direct closure of fasciectomy (Ullah et al., 2009), which provided low quality evidence. There were 79 participants. Much of the data were only presented graphically.

Hand function

Hand function, as determined by the PEM, was not different between fasciectomies and firebreak dermofasciectomies at 36 months. Earlier time points were presented graphically only.

Patient satisfaction and other patient rated outcomes

There was no evidence identified comparing these outcomes for fasciectomies and firebreak dermofasciectomies.

Early angles and other objective outcomes

Grip strength, angular deformity and motion at the PIPJ all correlated at 36 months. No differences between groups were seen throughout the study, though the data supporting this were presented graphically only.

Recurrence

There was no difference between firebreak dermofasciectomies and fasciectomies in terms of recurrence, defined as progressive contracture, and time to recurrence.

Adverse effects

There was no difference between the procedures in terms of antibiotic requirement, skin necrosis, wound dehiscence, radial hypoaesthesia or reflex

sympathetic dystrophy. There was a significantly greater incidence of ulnar hypoaesthesia after firebreak dermofasciectomy.

Cost effectiveness

There was no formal cost effectiveness analysis evidence. However, firebreak dermofasciectomy took significantly longer to perform than a fasciectomy involving a z-plasty closure (79 versus 66 minutes, p = 0.01).

SUMMARY

There was high quality evidence that firebreak dermofasciectomy and fasciectomy with z-plasty closure performed similarly. Given that firebreak dermofasciectomy took longer to perform, there was no evidence to support its routine use. However, this conclusion cannot be extended to other approaches to dermofasciectomy involving larger skin grafts. No evidence was identified to compare other approaches to dermofasciectomy with fasciectomy.

4.7.2 Technical refinements

4.7.2.1 Type of incision

Incisions were studied in two articles with 30 and 100 participants respectively. The first compared z-plasty closure to direct closure of a transverse incision for fasciectomy (Citron and Hearnden, 2003). The second compared a zig-zag (Bruner's) incision with direct closure to a longitudinal incision with z-plasty closure for fasciectomy (Citron and Nunez, 2005). They provided low quality evidence due to study design limitations and imprecision.

Hand function

There was no evidence identified to describe this.

Patient satisfaction and other patient rated outcomes

There was no evidence identified to describe this.

Early angles and other objective outcomes

There was no difference between a zig-zag incision and a z-plasty closure in terms of deformity or extension.

Recurrence

There was evidence that z-plasty closure of a palmar fasciectomy had significantly less recurrence (reappearance of palpable disease) than direct closure of a transverse incision ($p < 0.01$ when trial recruitment was stopped at the interim analysis point (Citron and Hearnden, 2003)). There was no difference in recurrence defined this way between a zig-zag incision and a z-plasty closure.

Adverse effects

In one study, no complications were encountered (Citron and Hearnden, 2003). In the other, total complications, rates of algodystrophy and incidence of digital nerve injury were no different between a zig-zag incision and a z-plasty, and digital nerve injury not different between a zig-zag incision and z-plasty (Citron and Nunez, 2005).

Cost effectiveness

There was no evidence identified to describe this.

SUMMARY

There was evidence supporting z-plasty closure over direct closure of transverse incisions for MCPJ cords. There was no difference between a zig-zag incision and a z-plasty closure for fasciectomy. However, the evidence was of poor quality.

4.7.2.2 Wound closure

Two studies investigated wound closure (Bhatia et al., 2002, Howard et al., 2009). One compared staple closure to non-absorbable suture closure in 31 participants (Bhatia et al., 2002). The other compared absorbable suture closure with non-absorbable suture closure for fasciectomy in 62 participants (Howard et al., 2009). Both provided level low quality evidence due to study design limitations and imprecision.

Hand function

There was no evidence identified to describe this.

Patient satisfaction and other patient rated outcomes

There was no difference in patient reported wound appearance at two weeks between those who received staples and those who received non-absorbable sutures (Bhatia et al., 2002). Staples were more painful to remove than non-absorbable sutures ($p = 0.008$) (Bhatia et al., 2002). There was no difference in pain VAS at the first postoperative visit between absorbable and non-absorbable suture groups (Howard et al., 2009).

Early angles and other objective outcomes

Staple closure was quicker to perform than non-absorbable suture closure ($p < 0.001$) (Bhatia et al., 2002). Absorbable suture closure incurred less clinic time for management, once outliers were excluded, than non-absorbable suture closure ($p = 0.003$) (Howard et al., 2009). However, excluding outliers from a non-parametric analysis may not have been appropriate, and may invalidate this finding.

Recurrence

There was no evidence identified to describe this.

Adverse effects

No formal analysis of differences in complications between groups was performed in either study. However, complication rates were low for both groups in both studies.

Cost effectiveness

There was no formal cost effectiveness analysis evidence. The analysis of timings presented may not be robust, as explained above.

SUMMARY

Staple closure may be quicker to perform than suture closure, and may achieve a comparable early wound appearance. However, staples may be more painful to remove.

Absorbable sutures achieved comparable early outcomes to non-absorbable sutures, and were reported as requiring less clinic time for postoperative management. However, the evidence was incomplete and of poor quality evidence. In particular the main conclusion may not be valid, as the statistical analysis performed may have been inappropriate.

4.7.2.3 Intraoperative adjuncts

Two studies investigated intraoperative adjuncts to surgery. Each considered a different intervention. One investigated bathing a fasciectomy wound in 5-fluorouracil compared to control, with 15 participants and provided low quality evidence (Bulstrode et al., 2004). The second compared a postoperative series of steroid injections as an adjunct to needle fasciotomy versus no injections in 47 participants and provided low quality evidence due to study design limitations and imprecision (McMillan and Binhammer, 2012).

Hand function

There was no evidence identified to describe this.

Patient satisfaction and other patient rated outcomes

There was no evidence identified to describe this.

Early angles and other objective outcomes

There was no difference between 5-fluorouracil treatment and control treatment in terms of healing time, total active motion and loss of extension (Bulstrode et al., 2004).

A series of steroid injections resulted in significantly greater percentage improvement in total active extension deficit at all time points (McMillan and Binhammer, 2012).

Recurrence

There was no difference between 5-fluorouracil treatment and control treatment in terms of loss of extension or total active motion at 18 months (Bulstrode et al., 2004).

There was a significantly greater percentage improvement in total active extension deficit at six months (65% correction for steroid group versus 41% for control group, p = 0.04), and in MCPJs and PIPJs when considered separately at six months (McMillan and Binhammer, 2012).

Adverse effects

Both studies reported no complications (Bulstrode et al., 2004, McMillan and Binhammer, 2012).

Cost effectiveness

There was no evidence identified to describe this.

SUMMARY

There was no evidence of benefit or harm from the addition of a 5-fluorouracil bath at the completion of a fasciectomy, though function and long-term outcome (beyond 18 months) have not been studied. There was evidence that the addition of a series of steroid injections following needle fasciotomy may achieve and maintain better correction of contractures than needle fasciotomy alone, though no long term data were identified, and the quality of evidence available was poor.

4.7.3 Rehabilitation adjuncts

The aim of this review had been to study operative techniques primarily. However, as rehabilitation adjuncts are components of hand therapy in

Dupuytren's disease, these trials met the inclusion criteria specified in the protocol, and so were included.

Four studies were identified that investigated adjuncts that might aid rehabilitation. Two studies compared three months of static postoperative splinting to no postoperative splinting (Collis et al., 2013, Jerosch-Herold et al., 2011). One other study also investigated postoperative splinting, providing low quality evidence, though the intervention protocol differed from the former two studies: the arm underwent day and night splinting for a month then night splinting for a further two months (Kemler et al., 2012). One hundred and fifty four participants were reported. The final study investigated the application of fibrin- and platelet-rich fibrin plug to open palmar wounds after fasciectomy, to identify whether this improved healing (Chignon-Sicard et al., 2012). These studies all provided low quality evidence due to study design limitations and imprecision.

Not all relevant data in one study of splinting were available from the published article (Collis et al., 2013). The authors were contacted, and provided the necessary data for per protocol analyses.

Hand function

Hand function, as assessed with the DASH PROM, was not affected by postoperative splinting at three months, six months and twelve months (Jerosch-Herold et al., 2011). No effect was also found in the other comparable splinting study (Collis et al., 2013), where individual time points were analysed up to three months postoperatively. In the latter, time points

were also combined based on there being no difference in a mixed effect model.

Meta-analysis of these two studies demonstrated no significant heterogeneity at baseline between studies (see Figure 4.3). There was also no difference in function between splint and no splint groups at three month follow up, in an intention to treat analysis (Figure 4.4). In keeping with all intention to treat analyses, those presented in both papers had some limitations. It had been considered ethically inappropriate to withhold a splint from patients who develop early deterioration in contracture. Therefore, patients in the 'no splint' group' who experienced early recontracture were then given a splint. Furthermore, some patients in the splint group were not compliant with splinting (defined as self report of <50% compliance). No difference between 'splint' and 'no splint' groups was also seen when the data were analysed per protocol (Figure 4.5).

Patient satisfaction and other patient rated outcomes

Patient satisfaction was not different in those receiving postoperative splinting compared to patients not receiving splinting, as assessed using an 11-point verbal rating scale (Jerosch-Herold et al., 2011). However, the validity and reliability of this scale is not described or cited. Patient perceived change was also not significantly different between groups in one of the other splinting studies (Kemler et al., 2012).

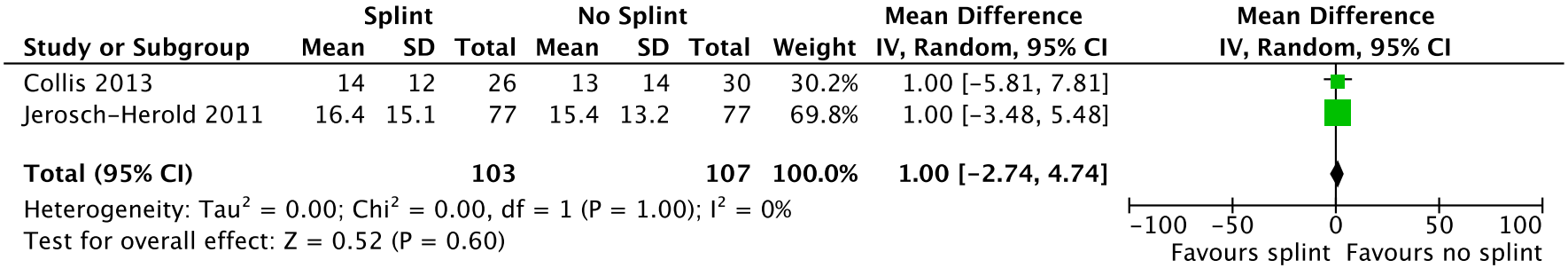

Figure 4.3: Forest plot of baseline (preoperative) DASH summary score (scale: 0-100, higher score indicates worse function)

SD – standard deviation, IV – inverse variance statistical method, 95% CI – 95% confidence intervals, random – random effects model

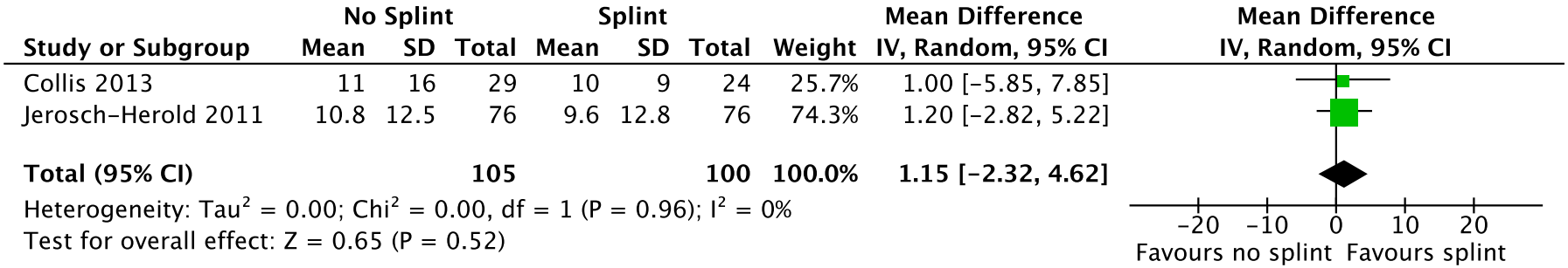

Figure 4.4: Forest plot of 3-month follow up DASH assessments of postoperative night splinting (Intention to treat) (scale: 0-100, higher score indicates worse function)

SD – standard deviation, IV – inverse variance statistical method, 95% CI – 95% confidence intervals, random – random effects model

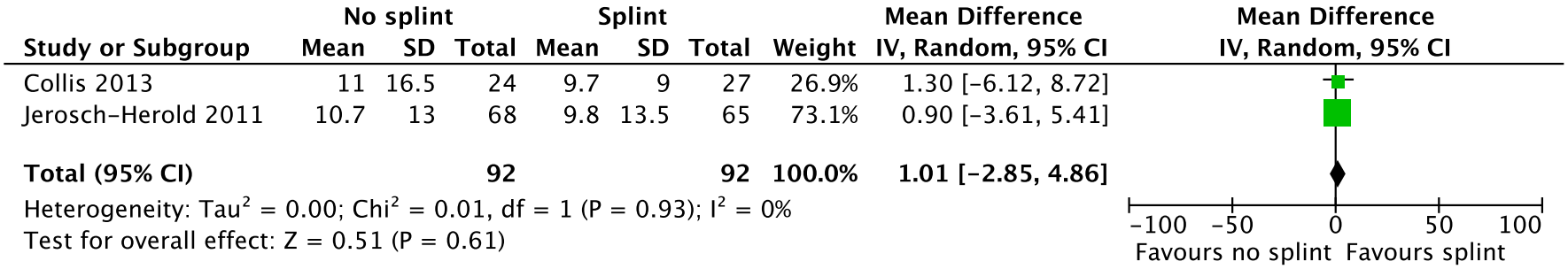

Figure 4.5: Forest plot 3-month follow up DASH assessments following postoperative night-splinting (Per protocol) (scale: 0-100, higher score indicates worse function)

SD – standard deviation, IV – inverse variance statistical method, 95% CI – 95% confidence intervals, random – random effects model

Early angles and other objective outcomes

As with hand function, there was no significant heterogeneity between studies at baseline, in terms of total active extension across the MCPJ, PIPJ and DIPJ (see Figure 4.6), and also in terms of total active flexion across all three joints in the ray (see Figure 4.7). Total active flexion and total active extension were not different in those who received postoperative splinting, at three, six or twelve months (Jerosch-Herold et al., 2011). No differences in total active extension or flexion were found in the second paper either (Collis et al., 2013). When the 3-month follow up results from both studies were meta-analysed, there was no difference in total active extension between splint and no splint groups (Figure 4.8). However, there was a significant difference in total active flexion, with splint group participants achieving 8.42 degrees less total active flexion than no splint group participants (Figure 4.9). As discussed with hand function, the intention to treat analyses were potentially affected by some of the 'no splint' group were given a splint if they experienced early re-contracture, and some of the 'splint' group being non compliant. When meta-analyses were performed based on per protocol data, there was still no difference between groups in terms of total active extension (Figure 4.10), but the significant difference in total active flexion seen in the intention to treat analysis was more pronounced (Figure 4.11).

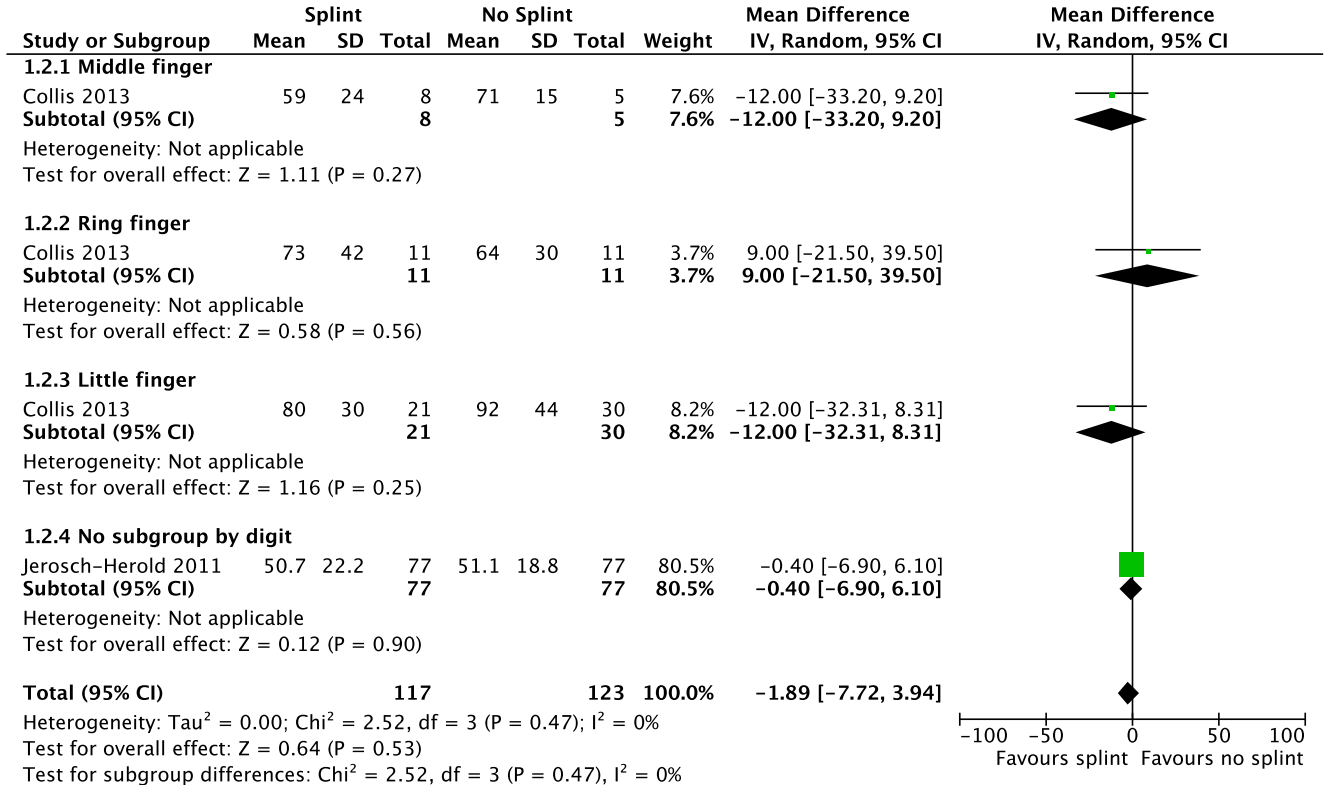

Figure 4.6: Forest plot of baseline (preoperative) total active extension in degrees (MCPJ+PIPJ+DIPJ) (higher value indicates more impairment)

SD – standard deviation, IV – inverse variance statistical method, 95% CI – 95% confidence intervals, random – random effects model

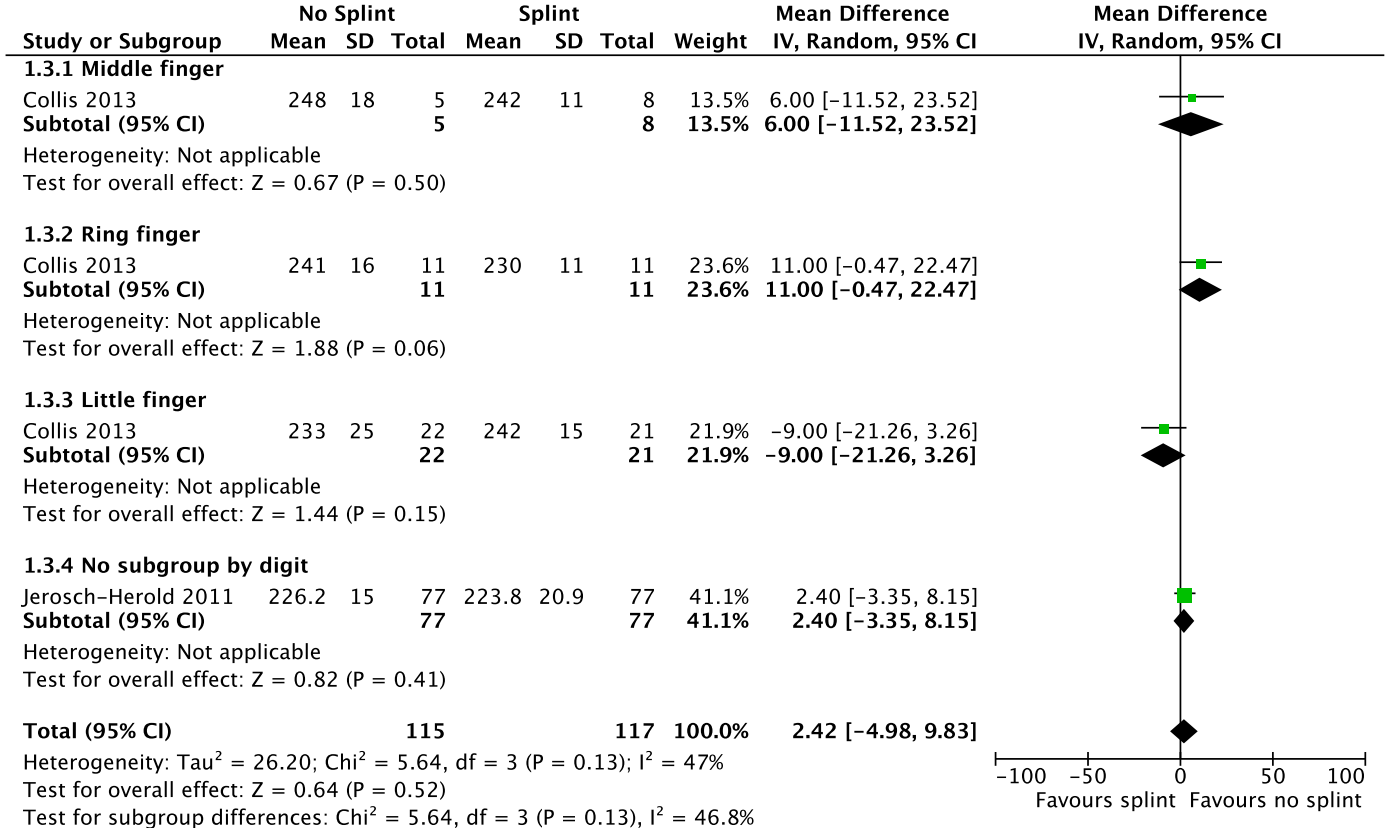

Figure 4.7: Forest plot of baseline (preoperative) total active flexion in degrees (MCPJ+PIPJ+DIPJ) (higher value indicates less impairment)

SD – standard deviation, IV – inverse variance statistical method, 95% CI – 95% confidence intervals, random – random effects model

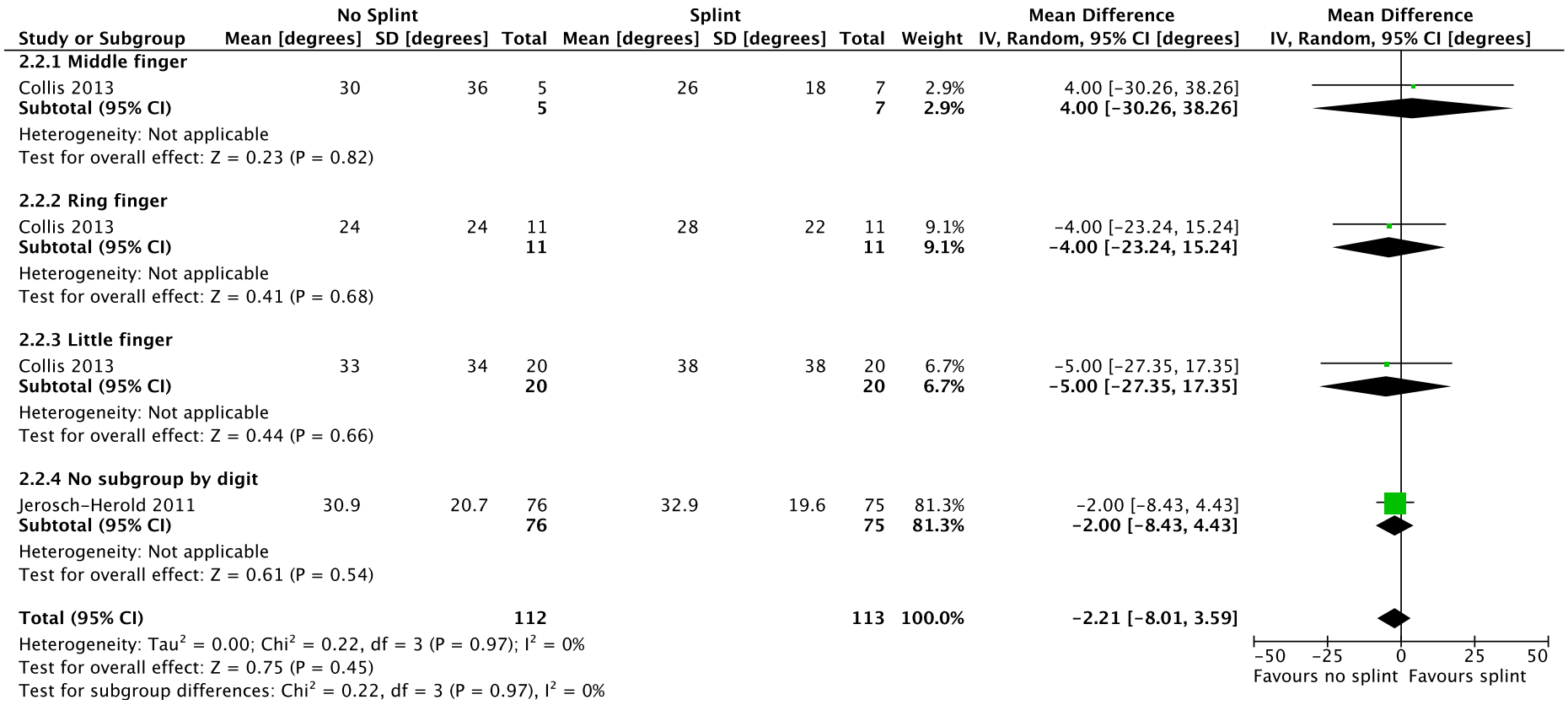

Figure 4.8: Forest plot of 3-month follow up total active extension in degrees (MCPJ+PIPJ+DIPJ) following postoperative night splinting (Intention to treat) (higher value indicates more impairment)

SD – standard deviation, IV – inverse variance statistical method, 95% CI – 95% confidence intervals, random – random effects model

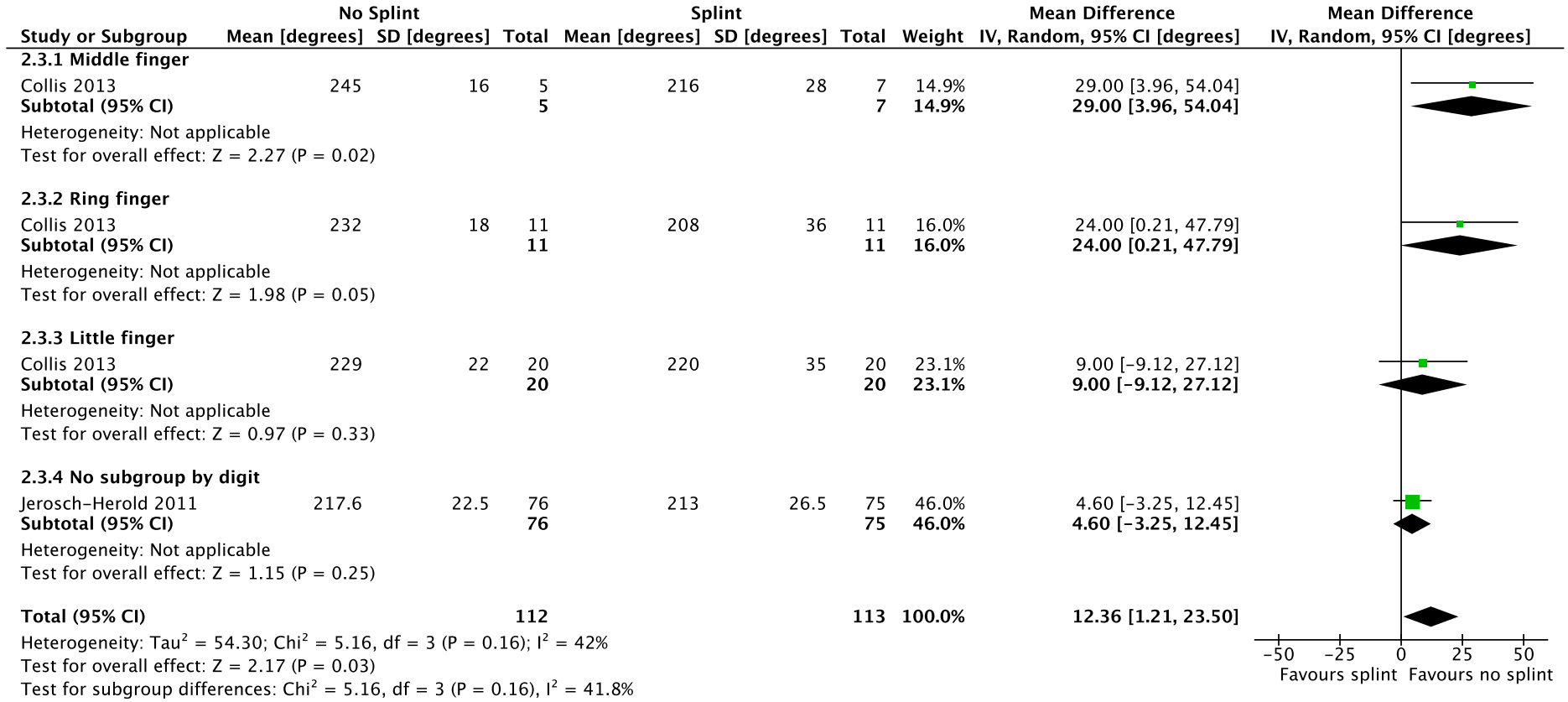

Figure 4.9: Forest plot of 3-month follow up total active flexion in degrees (MCPJ+PIPJ+DIPJ) following postoperative night splinting (Intention to treat) (higher value indicates less impairment)

SD – standard deviation, IV – inverse variance statistical method, 95% CI – 95% confidence intervals, random – random effects model

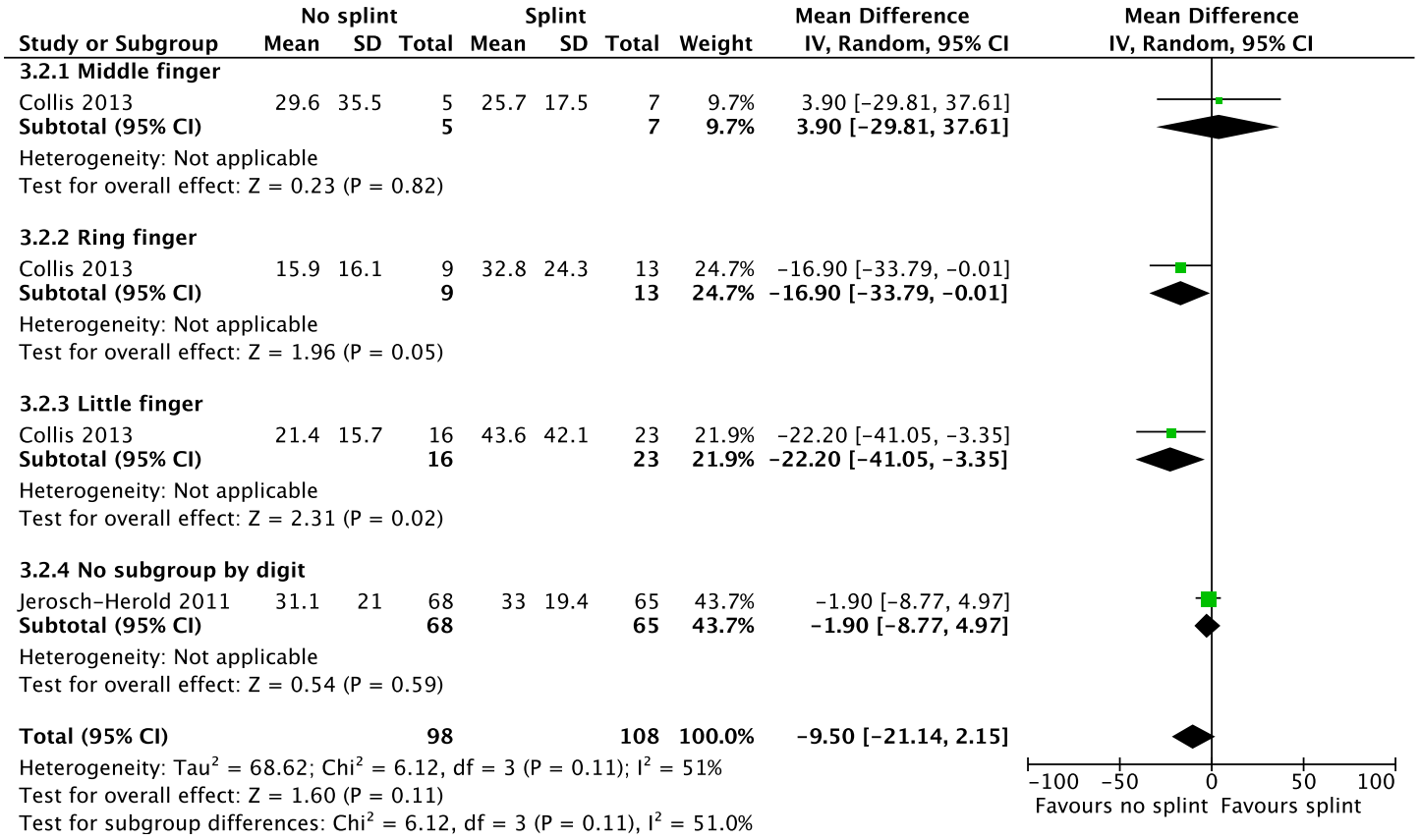

Figure 4.10: Forest plot of 3-month follow up total active extension in degrees (MCPJ+PIPJ+DIPJ) following postoperative night splinting (Per protocol) (higher value indicates more impairment)

SD – standard deviation, IV – inverse variance statistical method, 95% CI – 95% confidence intervals, random – random effects model

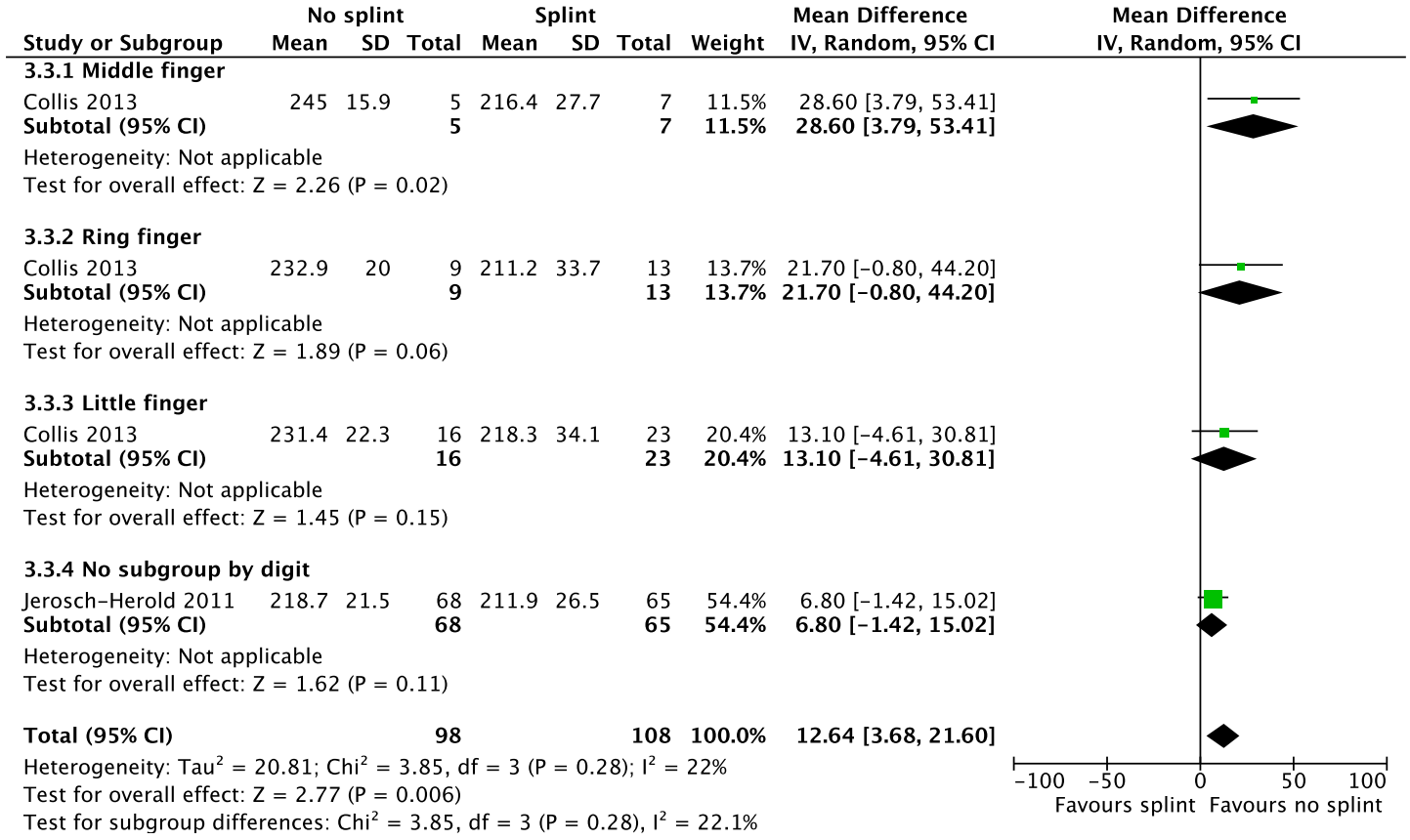

Figure 4.11: Forest plot of 3-month follow up total active flexion in degrees (MCPJ+PIPJ+DIPJ) following postoperative night splinting (Per protocol) (higher value indicates less impairment)

SD – standard deviation, IV – inverse variance statistical method, 95% CI – 95% confidence intervals, random – random effects model

There was a statistically significantly shorter healing delay from the application of fibrin- and platelet-rich fibrin plug to the open palmar wound compared to control group (median 24 days versus median 29 days), but no difference in secondary endpoints of pain, bleeding or wound exudate (Chignon-Sicard et al., 2012).

Recurrence

This was not assessed (Chignon-Sicard et al., 2012, Collis et al., 2013, Jerosch-Herold et al., 2011, Kemler et al., 2012).

Adverse effects

Bleeding and exudate were studied in one paper, but were not significantly different between intervention and control groups (Chignon-Sicard et al., 2012). Other adverse events, including wound and chest infections were rare, but were discussed.

Cost effectiveness

There was no formal cost effectiveness analysis evidence in any article.

SUMMARY

There is no evidence that postoperative splinting improves rehabilitation after fasciectomy or dermofasciectomy, but the meta-analysis conducted here has provided the first evidence that three months of night splinting reduces total active flexion at three months.

Intermittent pneumatic compression may reduce postoperative oedema, though the quality of evidence to support this was poor, and the risks and benefits of this adjunct have not been studied.

4.8 Discussion

4.8.1 Present study findings

4.8.1.1 Study questions

This systematic review has demonstrated that there are very few high quality studies of Dupuytren's disease surgery. Amongst the trials included, fewer still compared different procedures, with others studying refinements of practice. Despite many current treatment options being available for years or decades, the paucity of studies suggests that research in this field lacks direction.

4.8.1.2 Outcomes measured

Variation in primary outcome measures assessed was found between studies. Whilst some of this might be expected, for example studying time taken for staple removal, rather than recurrence as a long-term outcome, the extent of variation within groups of studies limits the utility of the data presented, and their interpretation. Such variation is seen across lower quality studies as well. Providing no definition of recurrence is commonplace (Becker and Davis, 2010), and limits interpretation of data.

Defining recurrence as the reappearance of palpable disease, as used in two papers here (Citron and Hearnden, 2003, Citron and Nunez, 2005), is acknowledged as generating qualitative data (Werker et al., 2012). When detection bias is a risk, the combination of an unblinded assessor defining outcome in a binary but subjective manner might be expected to be unsound.

This was the case in one study here (Citron and Hearnden, 2003). This outcome may not be applicable to fasciotomy, as palpable disease is not cleared from the treated field.

Even within studies that use angular deformity to define outcome and recurrence, wide variation is seen in what exactly is measured, and how it is described and presented (Ball et al., 2013). Angles may be presented as the passive angle obtained by an assessor, or active extension, achieved unsupported by the patient. Such angles may again differ, and influence study outcome.

Whilst meta-analysis could not be performed to compare operation types due to the paucity of comparable trials, it is probable that even if more data were to be generated in future, without standardisation of follow-up length and outcome measures used, meta-analysis would still not be possible. As Dupuytren's disease is a slowly progressive condition (Reilly et al., 2005), the incidence of recurrence is likely to rise with longer periods of follow-up. Furthermore, the natural history of the condition suggests that recurrence defined as reappearance of palpable disease is likely to be encountered earlier than a deteriorating angular deformity (Luck, 1959). The combination of these two inconsistencies contributes to the wide variation in reported recurrence rates in the existing literature, of 0% to 71% (Becker and Davis, 2010). Others have also highlighted the need for clarity and consistency (Werker et al., 2012). This review reiterates this, but also calls for more detailed study of the validity and reliability of outcome measures, to ensure

that the most appropriate outcomes are assessed, and at consistent and meaningful time points. Furthermore, recurrence of palpable disease or angular deformity may not be truly relevant endpoints. Instead, assessment of hand function as a patient-reported and patient-centred measure may be more appropriate. Patient-reported outcomes were used as a secondary endpoint in some of the studies included in this review.

There was marked variation in what was reported by study authors treated as secondary outcomes. In part this related to the study question, though outcomes handled as 'secondary' measures in studies were classified as appropriate primary outcome measures in this review, for example, patient reported hand function. However, trials of procedure type and technical refinement still varied between recording numerous secondary measures (Ullah et al., 2009), and recording virtually none other than complications (McMillan and Binhammer, 2012). The value of capturing all of these outcomes is not clear. One secondary outcome measure that is of importance is a measure of health-related quality of life. Systems for analysing cost effectiveness are informed by data describing this, captured using PROMs such as the EuroQol-5D (NICE, 2008). Determining functional outcome describes a patient-centred outcome, and may also be of pragmatic interest to commissioners of healthcare. The use of patient-reported data has been promoted nationally in the UK (Darzi, 2008, Department of Health, 2010). The variation in PROMs used in Dupuytren's disease in general (Ball et al., 2013), and in the studies included in this review, demonstrates that there is no

consensus on which is most appropriate for use in future research. Three recent reviews have called for further study of outcome measures (Ball et al., 2013, Becker and Davis, 2010, Werker et al., 2012). To date, only a few studies included patient-reported hand function. The data captured by the Patient Evaluation Measure in the study comparing firebreak dermofasciectomy to fasciectomy was not described in detail (Ullah et al., 2009). The DASH has been the most popular measure across all studies of Dupuytren's disease (Ball et al., 2013), and was used in studies included in this review (van Rijssen et al., 2012, van Rijssen and Werker, 2006). The DASH data presented in the early outcome study of this trial did not support the same conclusions as measuring angles; needle fasciotomy fared better throughout early rehabilitation in terms of DASH scores, despite fasciectomy arguably providing better correction of angular deformity in general. Thus, the conclusions drawn from this paper are likely to vary considerably depending on which outcome is considered to be of primary importance. Corresponding late outcome function data were not presented (van Rijssen et al., 2012). Given the value of health-related quality of life data in accepted cost effectiveness analysis methodology (NICE, 2008), future pragmatic trials might consider the use of PROMs as the primary outcome measures, with joint angles demoted in importance. Furthermore, such a change might support the design of pragmatic studies. A variety of secondary outcomes were reported in the studies included. However, separating the primary outcome from complications may limit the clinical applicability of research

findings. If an intervention achieves low rates of recurrence, but does so with significant risks of complications such chronic regional pain syndrome, cold intolerance, and loss of grip strength or flexion, then it may still fail to achieve meaningful clinical improvement for patients, and cost effectiveness for commissioning bodies. Patient satisfaction was considered in the reports of early and late outcome of the trial comparing fasciectomy and needle fasciotomy (van Rijssen et al., 2012, van Rijssen and Werker, 2006), though the validity and reliability of these assessments was not clear. As already discussed for angular measurements, further work to establish the validity and reliability of candidate patient-reported outcome measures is urgently required.

4.8.1.3 Comparing procedure types

The only trial comparing aponeurotomy with fasciectomy demonstrated that needle fasciotomy may achieve comparable angular correction to limited fasciectomy for milder Tubiana I and II contractures, but inferior correction for Tubiana III and IV contractures (van Rijssen and Werker, 2006). However, it caused less functional impairment in the early postoperative phase (up to five weeks), earlier recovery, and higher early patient satisfaction. By five years it had significantly higher recurrence and lower satisfaction than fasciectomy. However, attrition bias may have affected late outcomes, with late functional outcome not recorded. Patient satisfaction is an important outcome to measure, but perhaps ought to be measured in addition to, rather than instead of, valid measures of hand function, as the former might be influenced by

factors besides the functional efficacy of the treatment. It is also not clear whether it is reasonable to expect needle fasciotomy, a demonstrably less invasive procedure that can be repeated for recurrent disease (van Rijssen and Werker, 2012), to achieve a comparably durable effect as the more invasive fasciectomy. A more pragmatic study might be to consider early and late functional outcome in groups randomised to receive one fasciectomy or multiple needle fasciotomies over a period of years, with cost effectiveness then calculated based on functional outcome and treatment pathway expense. The comparison of fasciectomy with z-plasty closure to firebreak skin grafting found no differences between groups, other than prolonged operation time for skin grafting (Ullah et al., 2009). This suggests that 'firebreak' skin grafting may not prevent recurrence compared to fasciectomy. However, dermofasciectomy comprises a spectrum, with small skin grafts used as 'firebreaks' at one end (Ullah et al., 2009), and much more extensive skin grafts at the other end (Seah et al., 2012). Thus, further comparison between limited fasciectomy and dermofasciectomy is still required.

4.8.1.4 Trials investigating postoperative splinting

This was the only area in which meta-analysis was possible. It is noted that the earlier publication of a trial protocol facilitated standardisation (Jerosch-Herold et al., 2008). Indeed, this was the only published trial protocol identified. Advanced publication of trial protocols is encouraged as this may facilitate future standardisation of outcome assessment.

The functional outcome studied here was the absolute DASH score at three months (rather than change in DASH from preoperative to postoperative state), as preoperative DASH scores were not different between splint and no splint groups in either of the included studies (Collis et al., 2013, Jerosch-Herold et al., 2011). Furthermore, this final result represents the participants' functional performance at that time. All three of the individual trial results showed no beneficial or adverse effect from postoperative splinting, but meta-analysis demonstrated a statistically significant loss of flexion at three months from splinting. Whether the magnitude of the difference seen is of clinical significance, or whether it persists later in the rehabilitation period is unclear. However, given the potential resource utilisation in producing and maintaining splints, their on going routine use is not supported.

As discussed in the results, some of the 'no splint' group patients who developed early recontracture were issued with a splint. The effect of this depends on the outcome considered. If splinting improves loss of extension, then t might improve the average total active extension in the 'no splint' group. However, if splinting impairs total active flexion, then this outcome may be adversely affected in the 'no splint' group as a result. Due to the potential for the above antagonistic effects, it is difficult to predict the effect of the protocol on functional outcome.

4.8.1.5 Trials investigating other questions

The primary objective of this review was to identify trials comparing types of procedure. However, our methodology has identified other trials within

Dupuytren's disease surgery, which have been grouped into those investigating technical refinements of procedures, and those investigating rehabilitation adjuncts. Although these studies might be considered tangential to the central aim of this review, appraising these studies is important. It has ensured that this review has been comprehensive, and aspects of study methodology and reporting have contributed to conclusions that can be made in this review. In particular, analysing these studies informs the implications for future research in this field. For example, only one study included adequately described a randomisation process using envelopes that had adequate allocation concealment (Citron and Nunez, 2005). As with studies comparing types of procedure, the lack of comparable studies limited the performance of meta-analysis.

4.8.1.6 Quality of evidence

The overall quality of methodology varied between studies. More modern studies were generally at less risk of bias. The current assessment of performance bias might be controversial. To minimise the potential for the surgeon to influence the quality of the procedure, blinding of the surgeon during the procedure was included as a desirable feature. Achieving this blinding may be extremely challenging or even impossible in certain studies. However, clear efforts were made in some studies to standardise the surgical procedure as far as possible, with randomisation performed intraoperatively rather than preoperatively (Ullah et al., 2009), unlike other studies investigating an intervention of relevance to the closing stages of the

procedure. By doing so, the excision of disease that might be considered to be the "correction" part of the surgery was not subject to lack of blinding, and only the wound closure was unblinded. Taking such steps where possible may limit the effect of performance bias.

In addition to risks of bias that related to study design limitations, the quality of evidence was further downgraded due to imprecision, with most comparisons based on one or two studies, with small sample sizes and wide confidence intervals.

4.8.2 Limitations

There were limited deviations from the protocol to the final review. These differences were minor and did not influence the outcome of the review. Whilst explicit steps were taken to review conference proceeding abstracts, it is possible that unpublished data were missed. However, given the paucity of trial data identified across all sources, it is unlikely that there is a significant volume of relevant data that has not been published.

4.9 Conclusions of Systematic Review of Trials

There is a marked paucity of randomised controlled trials in Dupuytren's disease surgery. The quality of design and reporting of trials in Dupuytren's disease surgery remains generally poor, with only some robust examples of good practice. This is the case for comparing different treatment procedures, of which several are in current use, including needle fasciotomy, fasciectomy, dermofasciectomy and collagenase therapy. The effectiveness and role of each of these treatments is currently based on poor quality evidence.

The meta-analysis performed here questions the routine use of splinting following surgery for Dupuytren's disease, and this warrants further research. It is possible that splinting may impair outcome, though this is not certain from the data presented here. Furthermore, given the unclear role of splinting in early recontracture, splinting should still be considered on an individual patient basis until further evidence is available.

Future trials should ensure that risks of bias are minimised. As acknowledged, performance bias may prove difficult to minimise in some studies. Certain components of the studies included here set precedents for processes by which risk of bias in random sequence generation, allocation concealment and outcome detection can be minimised. Future studies should endeavour to employ such robust processes, and to report them clearly.

However, prior to embarking on the trials needed in this area, it is clear that further study of outcome measures to establish their validity and reliability for use in Dupuytren's disease. Once this has been done, consensus and

consistency of outcome choice and time point of assessment is needed, to ensure standardisation with other studies.

SUMMARY

This chapter described a systematic review of trials of the treatment of Dupuytren's disease, including a meta-analysis of the effect of postoperative splinting as an adjunct to surgery.

There have been relatively few randomised controlled trials to describe the roles for different surgical procedures in the treatment of Dupuytren's disease.

Most have been at risk of bias that could have been further minimised.

The primary outcomes studied have often involved surgeon-centred objective measures, such as angular loss of extension, rather than patient-centre function.

None of the studies included in this review included cost effectiveness analyses.

The effect of postoperative splinting was meta-analysed here, and these analyses confirmed the findings of individual studies, in that it does not improve outcome, but this analysis demonstrates that impairs active flexion.

This chapter demonstrates that further studies are urgently needed in Dupuytren's disease. Currently, there is insufficient high quality comparative data to inform the development of a meaningful clinical guideline in Dupuytren's disease, or to best utilise the changes in NHS architecture that might promote standardisation of practice.

This chapter also highlighted the variation in outcome measures used to study Dupuytren's disease. To be able to design the required comparative study or studies, and to be able to interpret existing published data, better understanding of the most appropriate PROMs is required. Within the PROMs previously used, the DASH was used in one trial comparing operative techniques. This is the most common PROM in Dupuytren's disease research (Ball et al., 2013). Further study of the DASH is required.

The following chapters will investigate aspects of validity, reliability and responsiveness of the DASH and other outcome measures that may be used as the primary outcome in future studies.

5 The Dynamism of Dupuytren's Contractures

5.1 Preface

The previous chapter reviewed existing "interpretability" data for relevant clinical outcome measures used in Dupuytren's disease. Such data facilitate the interpretation of outcome measurements, by establishing what constitutes a clinically important change or difference. The review demonstrated a paucity of such studies, which has implications for future study design, data interpretation and guideline development. Passive extension deficit is an example of a measure that has been used as the primary outcome in randomised controlled trials of treatment of Dupuytren's disease (van Rijssen et al., 2006, van Rijssen et al., 2012), but for which no interpretability data could be identified. There may be limitations to the stability of this as an outcome measure, which have been previously overlooked. Dynamism of contractures is discussed in the introduction, whereby for cases of Dupuytren's disease in which a cord crosses more than one joint, the passive extension deficit at one joint is influenced by the position of other joints in the ray. This is because the cord of Dupuytren's disease is inelastic and so may limit the simultaneous extension of both joints (Hurst, 2010). This could constitute a source of bias for passive finger joint angles in trials, and has not been studied.

If dynamism is frequently encountered and is of significant magnitude, this would support the adoption of a different primary outcome measure in future research, and caution the interpretation of previous studies reporting such angles. This chapter examines the prevalence and severity of dynamism in preoperative Dupuytren's contractures to determine whether this might constitute a source of bias that has been overlooked.

5.2 Introduction

Systematic reviews of Dupuytren's disease have demonstrated no consistent definition of recurrence and that length of follow up varies between studies (Becker and Davis, 2010, Werker et al., 2012). Common practice in studies of surgery for Dupuytren's disease is to report 'recurrence rates' after treatment, but often without providing a definition of recurrence (Becker and Davis, 2010).

When recurrence is defined, a commonly employed definition is the recurrence of angular deformity. Seventeen variations, such as an angular deformity of 30°, are reported (Ball et al., 2013). Improvement in angular deformity was the primary endpoint of the two most recent and high profile randomised controlled trials (RCTs) of Dupuytren's disease treatment (Hurst et al., 2009, van Rijssen et al., 2006). Return of angular deformity was used to assess recurrence as the primary outcome in the 5-year follow-up of the latter trial (van Rijssen et al., 2012).

If a Dupuytren's contracture spans both the MCP and PIP joints, the severity of the passive flexion deformity of one of these joints may be influenced by the position of the other when the measurement is taken, through what is referred to as dynamism (Hurst, 2010). Thus, if the metacarpophalangeal joint (MCPJ) is held in maximal passive flexion, then the flexion deformity at the proximal interphalangeal joint (PIPJ) may be partially or completely relieved. Conversely, if the MCPJ is held in maximal passive extension, then the PIPJ flexion deformity may be exaggerated (Figure 5.1). To minimise the effect of

such dynamism, observers may hold the other joints of the finger in maximum passive flexion when measuring a flexion deformity of a joint, and thus report a minimised passive extension deficit, rather than the active deficit encountered by the patient, which might be larger due to dynamism.

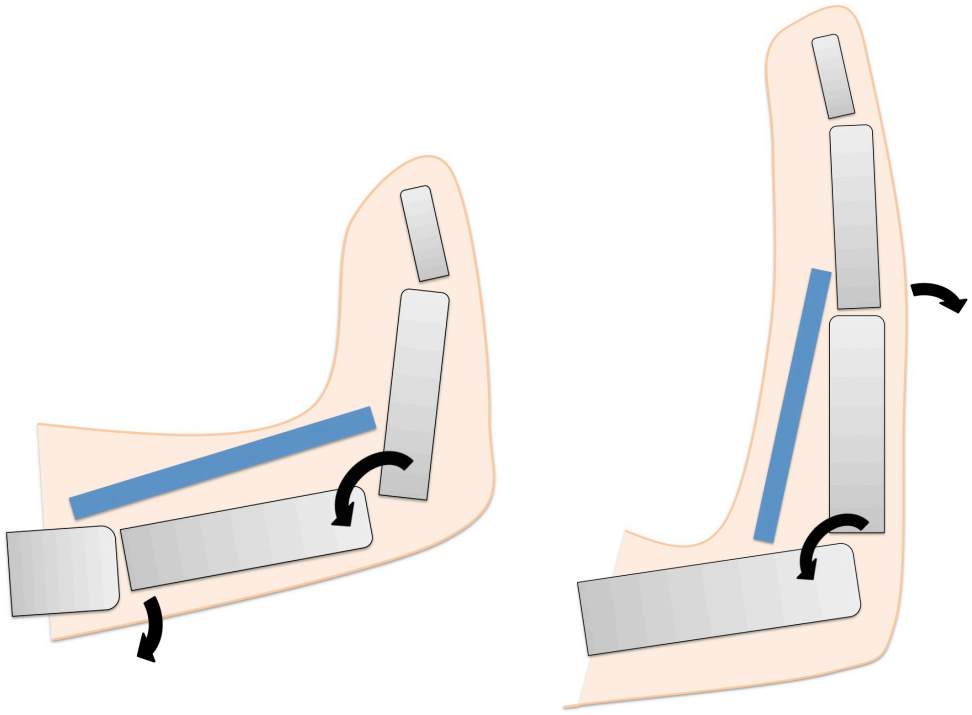

Figure 5.1: Schematic representation of dynamism

The Dupuytren's cord (shown in blue) is not elastic and has fixed attachments to the palmar aponeurosis in the palm and flexor sheath in the middle pulp space. On the left, extension of the MCPJ tightens the cord at the PIPJ, pulling it into flexion, and generating the "worst" PIPJ angle. Alternatively, passive flexion of the PIPJ relaxes the cord and allows increased extension at the MCPJ, giving the "best" MCPJ angle.

On the right, flexion of the MCPJ relaxes the same cord at the PIPJ and allows increased extension, generating the "best" PIPJ angle. Alternatively, passive extension of the PIPJ tightens the cord, and pulls the MCPJ into flexion, giving the "worst" MCPJ angle.

Dynamism may also reduce the correlation between passive angles measured by an observer and the patient's experience of the contracture, as the patient cannot achieve the milder contracture obtained by measuring passive extension angles at the MCPJ and PIPJ separately.

We are not aware of any reports on the potential impact of dynamism on angular joint measurements. Goniometry in general has good inter- and intra-observer reproducibility when assessed in controlled situations. When performed by hand therapists to measure joint angles in normal digits held by thermoplastic splints, inter-observer reproducibility was found to be 7°-9°, and intra-observer reproducibility was 4°-5° degrees (Ellis and Bruton, 2002). However, if common and significant, dynamism may introduce an important source of bias and error, which could influence reported recurrence rates.

This study aimed to establish the extent of dynamism on Dupuytren's contractures affecting PIPJs awaiting surgery (with or without MCPJ involvement).

5.3 Methods

These data are derived from the baseline data collection from a study of early functional recovery following Dupuytren's disease surgery. They comprise a minor element of the study. The local research and development department approved the overall project as service evaluation. In keeping with National Research Ethics Service guidance, research ethics committee approval for this project component was not required (NRES, 2012).

Data collection took place between February 2012 and May 2013. The inclusion criteria were patients meeting of all of the following:

- Awaiting fasciectomy or dermofasciectomy at a single United Kingdom hand surgery service
- Primary or recurrent Dupuytren's disease of one or more fingers
- Extension deficit affecting the PIPJ of one of the digits when actively extending this digit
- Able to attend a suitable pre-operative assessment clinic appointment scheduled for the candidate to perform the measurements

Exclusion criteria were:

- An isolated MCPJ contracture
- Cognitive impairment preventing informed consent
- Refusal of invitation to participate

Contractures that appeared to affect the PIPJ only were included, as in these cases there may have been MCPJ dynamism within the range of hyperextension. However, as PIPJs do not typically hyperextend passively, MCPJ only contractures were excluded.

Eligible patients who attended a routine orthopaedic preadmission clinic were issued with a letter explaining the project and inviting them to participate on a voluntary basis. The candidate assessed all patients who consented to participate. Demographic details, including previous surgery to the finger involved, were recorded. Goniometry was performed to measure the extension deficits of the MCPJ and PIPJ of all digits that were to undergo surgery. This was initially performed with the other joint of the digit held in maximum passive flexion, thus minimising the extension deficit measured and giving the 'best' measure. It was then repeated with the other joint of the digit in maximum passive extension (thus maximising the extension deficit measured, giving the 'worst' measure). When assessing an MCPJ in which hyperextension was present, this was recorded as 0°, in keeping with the methodology used in previous studies of goniometry in Dupuytren's disease (Jerosch-Herold et al., 2011).

Analysis was performed using Prism 6.0 for Mac OS X (GraphPad® Software, La Jolla, USA, 2012).

5.4 Results

Eighty-four patients were considered for inclusion in this study but 14 were excluded as they had MCPJ only contractures. A further six declined to participate and one was excluded as she had Alzheimer's disease that prevented informed consent. This left 85 digits on 72 hands from the 70 patients, all of whom consented to participate.

Of the 70 patients, 55 were men (79%) and the mean age was 67 years. The digits studied were predominantly little fingers, followed by ring, middle and only three index fingers (Table 5.1).

DIGIT	SAMPLE SIZE (n=85)	DIFFERENCE OF 'WORST'-'BEST' ANGLES AT PIPJ Mean (95%CIs)
Little	55 (65%)	11^o (9^o to 14^o)
Ring	19 (22%)	20^o (12^o to 29^o)
Middle	8 (9%)	16^o (5^o to 27^o)
Index	3 (4%)	64^o (30^o to 99^o)

Table 5.1: Differences between 'worst' and 'best' angles at PIPJ by digit

Thirty-five of 85 digits studied (41%) had PIPJ only contractures. In these, the 'best' and 'worst' MCPJ angles obtained were both within hyperextension. By

finger, the proportions of PIPJ only contractures were significantly different (p=0.0225, Chi square test):

- Index: 0/3 (0%)
- Middle: 1/8 (13%)
- Ring: 5/19 (26%)
- Little: 29/55 (53%)

Seventy-six digits showed dynamism at either the MCPJ or PIPJ (89%). Seventy-four showed dynamism at the PIPJ (87%), but only 35 (41%) MCPJs exhibited dynamism. Other MCPJs may have had dynamism within the range of hyperextension, which was not captured here.

In nine digits there was more than 30° of dynamism at one joint (11%). Overall, the mean 'best' MCPJ angle was 19° and the mean 'worst' MCPJ angle was 25°. The mean range of dynamism for MCPJs was thus 6°, which represented a highly significantly difference between best and worst angles ($p<0.0001$, paired t test). The mean dynamism seen increased to a mean of 9° (95%CIs: 7-12) when those digits in which both 'worst' and 'best' MCPJ angles were 0° were excluded (as dynamism within the range of hyperextension might have been present in these, but was not assessed in this study). The mean 'best' PIPJ angle was 49° and the mean 'worst' PIPJ angle was 63°. Thus the mean dynamism for PIPJs was 14° (95%CI: 11° - 17°) which also represented a highly significant difference between best and worst angles ($p<0.0001$, paired t test). The results are shown in Table 5.2.

JOINT	BEST ANGLE Mean (95%CI)	WORST ANGLE Mean (95%CI)	DYNAMISM Mean (95%CI)
MCPJ	19^o (14^o-25^o)	25^o (19^o-31^o)	5^o (4^o-7^o)
PIPJ	49^o (44^o-54^o)	63^o (58^o-68^o)	14^o (11^o-17^o)

Table 5.2: Best and worst angles (95% confidence intervals are in brackets)

Twenty-one fingers had undergone previous Dupuytren's surgery (25%). Fingers that had already had surgery for Dupuytren's disease had comparable ranges of dynamism to those that had not undergone previous surgery. For the MCPJ, the mean range of dynamism was 4.9^o (95%CIs: 3.8^o – 5.9^o) for digits that had previously undergone surgery and 5.5^o (95%CIs: 1.8^o – 9.2^o) for those that had not had previous surgery. For the PIPJ the mean range of dynamism was 12.6^o (95%CIs: 9.7^o – 15.5^o) for fingers that had previously undergone surgery and 14.6^o (95%CIs: 11.2^o – 18.0^o) for those that had not had previous surgery.

Results from different digits are presented in Table 5.3. The difference between 'worst' and 'best' MCPJ angles did not differ significantly between digits ($p=0.137$, one-way ANOVA). However, the extent of PIPJ dynamism varied by digit ($p<0.0001$, one-way ANOVA). When Tukey's multiple comparisons test was applied, highly significant differences ($p<0.0001$) were seen between the index finger and all other fingers. The only other significant

difference was between the ring and little fingers. There was a difference in the proportion of PIPJ-only contractures between the ring and little fingers (ring: 5/19 (26%) versus little: 29/55 (53%), (p=0.0464, Chi square test), which may have contributed to this. However, there were very few index fingers in the study, and so this finding merits further investigation.

	Middle	**Ring**	**Little**
Index	48^o (70^o to 26^o) p<0.0001	44^o (64^o to 23^o) p<0.0001	53^o (72^o to 33^o) p<0.0001
Middle		4^o (-9^o to 18^o) NS	5^o (17^o to -8^o) NS
Ring			8^o (18^o to 0^o) p=0.042

Table 5.3: Differences in dynamism at the PIPJ between different digits [mean (95%CIs), significance levels from Tukey's multiple comparisons test]

5.5 Discussion

5.5.1 Present study findings

These data demonstrate that dynamism is common. Frequently it can be of a large enough magnitude to potentially reclassify a contracture from "sufficiently large contracture (i.e. >30 degrees) to warrant surgical treatment" to "insufficient contracture (i.e. < 30 degrees) to warrant surgery" or from "recurred" to "not recurred" based on commonly used cut offs (i.e. 30 degrees) in Dupuytren's disease research. Therefore, it may constitute a source of inadvertent measurement error or systematic bias in clinical trials of Dupuytren's disease management.

Differences in technique for measuring angular deformity may result in some measurements being taken with passive flexion of the adjacent joint, such that dynamism may reduce the measured angular deformity of a joint. In contrast, if the joint angle were measured with passive extension of the adjacent joint, then this would increase the joint angle measurement worse. Furthermore, if observers were not blinded and were able to recognise the treatment that the patient has received, conscious or subconscious bias may influence the angular measurement. Different treatment modalities for Dupuytren's disease, such as needle aponeurotomy, fasciectomy and dermofasciectomy leave very visibly distinct scar patterns on the hand, which can be recognised by observers, even if they had been blinded to the treatment group allocation. As goniometry may be considered a learned skill, acceptable reliability is likely to only be achieved if experienced observers perform assessments. Such

observers are likely to be aware of the different treatment options and their scar patterns. Dynamism was frequently capable of producing a 5°-10° change in a PIPJ measurement, and thus could move a finger from a “no recurrent contracture” to a “recurrent contracture” status, or vice-versa, if a definition of recurrence such as “20° or more increase in angular deformity” is used.

In this study, index fingers showed the most dynamism. However, there were only three index fingers in the series, and so this finding is probably not robust, and should be interpreted with caution.

5.5.2 Limitations

One limitation of the present study is that only preoperative and established recurrent contractures were measured. In contrast, the results of trials typically also involve assessing early postoperative angular deformities to determine residual contractures after treatment. However, 21 of the 85 fingers studied had undergone previous Dupuytren’s disease surgery, and these exhibited similar ranges of dynamism to the others. Thus, we believe our findings can be reasonably extrapolated to postoperative situation.

The results of the present study represent the maximum range of dynamism, given that the ‘worst’ angles were obtained by passively manipulating the digit to exaggerate a contracture at a particular joint. This strategy was intentional, as the aim was to establish the maximum amount of dynamism achievable. However, it is unlikely that the maximum effect of dynamism identified here would be encountered in a trial, unless bias was very marked, or some

observers had been trained to measure angular deformities in a manner which, through dynamism, either maximised or minimised the angular deformity. Nonetheless, this work demonstrates just how large the impact dynamism might have on individual joint measurements. Even if only half the range of dynamism demonstrated in the present study actually occurred in routine clinical assessments of joint angles, altering nearly 90% of PIPJ measurements in a trial by a mean of 7° might create important shifts in the balances of success/failure or recurrence/no recurrence.

The sub group analysis by digit involved small numbers of radial digits (index and middle fingers). Consequently, this analysis is underpowered to make reliable comparisons between digits, as indicated by the broad 95% confidence intervals in Table 5.3. However, the sub group data have been presented for completeness and to direct potential further study. Disease affecting the radial side of the hand, including the index finger is relatively uncommon, and occurs more frequently in diabetics. It is possible that there are differences between contractures affecting the radial digits compared to the ulnar digits.

5.5.3 Relationship to existing literature

Accurate intra- and inter-observer reproducibility of goniometry measurements taken by therapists have been previously reported (Ellis and Bruton, 2002). However, the model used in that study did not reproduce the clinical situation as measurements were taken from the fingers of a normal hand that was held in a fixed flexed posture by a thermoplastic splint. Thus there was no

possibility of measurement errors or differences due to dynamism or other patient and contracture factors, and the estimates of reproducibility are probably over-optimistic.

In recent research in Dupuytren's disease, unblinded surgeons performed the goniometric measurements (van Rijssen et al., 2006, van Rijssen et al., 2012). Such studies defined treatment success as correction of angular deformity at single joint level to less than 5°, with recurrence defined as a subsequent increase in angular deformity of 20° or more (Hurst et al., 2009, van Rijssen et al., 2012). The 14° mean range of dynamism at the PIPJ, or even just 6° at the MCPJ demonstrated in the present study, could have markedly altered the proportions of patients classified as "successes" or "failures", or "recurrences" or "no recurrences".

Reporting active (rather than passive) extension deficit might minimise the impact of dynamism on clinical outcome assessment. Alternatively, using appropriate patient-reported outcome measures may provide a more meaningful description of hand function and the benefit of surgery, and align trial design with the growing use and study of patient-reported outcome measure (PROM) data across in the United Kingdom National Health Service and other countries (Darzi, 2008, Mokkink et al., 2010). The correlation between angular deformity and patient-reported hand function in Dupuytren's disease is typically poor. It has been studied for the most popular PROM in Duputren's disease (Ball et al., 2013), the Disabilities of the Arm, Shoulder and Hand (DASH), and no meaningful correlation was identified (Engstrand et

al., 2009, Jerosch-Herold et al., 2011, Zyluk and Jagielski, 2007). This may be interpreted as evidence that the DASH is not valid (Packham, 2011). However, such a conclusion is not supported by a lack of correlation between the two. As discussed by the Consensus-based Standards for the Selection of Health Status Measurement Instruments (COSMIN) collaboration, assessing a PROM against a 'gold standard' of hand function is not robust, as no such gold standard exists (Mokkink et al., 2010). To conclude that the DASH is not valid based on poor correlation with angular deformity would only be appropriate if that the latter were the gold standard. Instead, the performance of PROMs such as the DASH, or the recently developed Unité Rhumatologique des Affections de la Main (URAM) Dupuytren's disease-specific PROM (Beaudreuil et al., 2011), needs to investigated using accepted standards for validity (Mokkink et al., 2010).

Despite dynamism, there is still likely to be value in measuring angular deformity. In some circumstances, such as explanatory studies, correcting angular deformity at particular sites, such as the PIPJ, may be important. Furthermore, angular measurements probably ought to remain as secondary outcomes in pragmatic studies, but with an appropriate PROM as the primary outcome measure. The data presented here support the use of active extension angles rather than passive extension angles, as the former are less likely to be influenced by dynamism. Alternatively, composite loss of extension might be used, as has been previously described (Makela et al., 1991).

5.6 Conclusions of study of dynamism

Dynamism is common in preoperative contractures involving the PIPJ. It occurs at both the MCPJ and PIPJ, and still affects MCPJ measurements when hyperextension is not studied. Dynamism is of sufficient magnitude that it could represent a significant source of bias if passive joint angles are studied, especially if observers are unblinded, as is frequently the case in studies of Dupuytren's disease (see Systematic Review and Meta-analysis of Surgical Trials in Dupuytren's Disease).

SUMMARY

This chapter demonstrates that dynamism is encountered in 89% of digits when measuring passive extension deficit angles at either MCPJs or PIPJs. It is frequently comparable in magnitude to widely used definitions of early correction of deformity and late recurrence of contracture. The previous chapter identified that passive angles have been used in other studies, often as unblinded measurements, and with no evidence of consideration of the influence of dynamism (van Rijssen et al., 2006, van Rijssen et al., 2012). The data in those papers should be reassessed in light of the findings of this chapter. This chapter supports the abandonment of passive extension deficit as the primary outcome measure in future studies of Dupuytren's disease.

Alternative outcome measures need to be considered. It is possible that a PROM will be more appropriate than passive extension deficit as the primary outcome measure of future studies. To investigate this, patients' experience

of Dupuytren's disease needs to be better understood. The next chapter will consider patients' goals for the treatment of their Dupuytren's disease, and relate these to one of the candidate PROMs that has been designed for use in Dupuytren's disease specifically.

6 British Patients' Goals for the Treatment of Dupuytren's Disease, and their Relationship to the Unité Rhumatologique des Affections de la Main Scale

6.1 Preface

The previous chapter investigated dynamism as a potential source of bias in passive measurement of extension deficit. Such angular measurements of finger joints have been used as quantitative outcome measures in high quality studies. However, the incidence and magnitude of the effect of dynamism on the measurement of passive extension deficit angles may constitute an important and previously unmeasured source of bias. Furthermore, the correlation between existing patient-reported outcome measures of hand function and angular deformity is poor (Jerosch-Herold et al., 2011). To understand better which of these is more valid, it is essential to appreciate the goals of treatment. These may be cited as 'correcting angular deformity', and this is typically used as the surgeon-centred endpoint of research. However, patient expectation of treatment is poorly described. In particular, surgery may be performed based on surgeon-centred indications relating to the technical challenges of correcting proximal interphalangeal joint (PIPJ) contractures that have been discussed in the introduction. As a result, some guidelines recommend operating on any PIPJ contracture (BSSH, 2008). It is possible that a proportion of patients are operated on prophylactically at a

stage at which they do not experience symptoms, but before their disease reaches a stage at which full correction becomes technically difficult due to the anatomical considerations. It is not known whether this is the case.

An understanding of patients' goals can also be used to appraise existing patient-reported outcome measures, such as the recently developed Dupuytren's disease-specific Unité Rhumatologique des Affections de la Main scale.

6.2 Introduction

The use of patient-reported outcome measures (PROMs) in healthcare has been stimulated by recent UK Department of Health publications (Darzi, 2008, Department of Health, 2010). PROMs may be of particular use in hand surgery, as patients' functional requirements vary, as do their expectations of treatment and definitions of a satisfactory outcome. PROMs used in hand surgery are classified into generic tools that assess global wellbeing (e.g. the EuroQol 5D), domain-specific tools that assess a particular region (e.g. the Disability of the Arm, Shoulder and Hand (DASH) tool (Hudak et al., 1996) or the Patient Evaluation Measure (PEM) (Macey et al., 1995)), and disease-specific tools (Szabo, 2001). Whilst generic and domain-specific tools can be used to assess outcome for upper limb conditions, they have low sensitivity and specificity, and may not detect changes in symptoms or hand function that are relevant to patients. They can also be subject to ceiling or floor effects, which occur when many individuals' scores are close to the upper or the lower scale limits respectively (Szabo, 2001). This limits the scale's ability to detect further improvement or deterioration respectively.

Recently, a Dupuytren's disease-specific tool was developed in France (Beaudreuil et al., 2011). This Unité Rhumatologique des Affections de la Main, or URAM, scale comprises nine items, which are each scored between 0 and 5 depending on the difficulty in performing that particular function (Table 6.1). Nine patients and seven healthcare professionals generated an initial

battery of items that were then reduced to the nine items comprising the final scale, via a process involving a further 85 patients.

However, all were recruited at a single hand surgery centre in Paris, and may have been treated with the same type of intervention (Beaudreuil et al., 2011). It is unclear whether the URAM is broadly applicable and culturally generalisable, and therefore it is unclear whether it is valid for use elsewhere.

This prospective study aimed to identify patients' goals for Dupuytren's disease surgery at a British hand surgery centre, and to compare these to URAM scale items.

Can you…	Without difficulty (0)	With very little difficulty (1)	With some difficulty (2)	With much difficulty (3)	Almost Impossible (4)	Impossible (5)
1. Wash yourself with a flannel, keeping your hand flat?						
2. Wash your face?						
3. Hold a bottle in one hand?						
4. Shake someone's hand?						
5. Stroke something or caress someone?						
6. Clap your hands?						
7. Spread out your fingers?						
8. Lean on your hand?						
9. Pick up small objects with your thumb and index finger?						

Table 6.1: URAM scale

6.3 Methods

Data collection involved the baseline time point of a service evaluation, and took place between September 2011 and April 2013. The inclusion criteria were:

- Patients awaiting fasciectomy or dermofasciectomy at this UK hand surgery centre
- Primary or recurrent Dupuytren's disease
- Available for a preoperative assessment at a time when the candidate was available.

Exclusion criteria were:

- Cognitive impairment preventing informed consent
- Refusal of invitation to participate

Eligible patients were issued with a letter, which explained the project and invited them to participate on a voluntary basis. The candidate assessed all patients who consented to participate. Details of patient demographics and planned procedure were captured.

Patients were asked to specify up to three functional problems they were experiencing and wished their surgery to resolve. Free text answers were recorded, so that participants were not restricted to grading their ability to specify tasks or the severity of particular symptoms. They were not asked to rank these indications. Indications were specified prior to the patients completing any established patient-reported outcome measures as part of

their service evaluation, to avoid them being influenced by the items in existing PROMs.

The candidate then grouped the free text indications, so that different descriptions of the same task were consolidated into 'types' of problem. For example, "difficulty washing my face" and "I catch my eye or nose with my finger when washing my face" were grouped together as "wash self".

The consolidated types of patient problem were mapped against the nine items that comprise the URAM scale. When a patient problem would have been captured by the URAM, a positive result was recorded (+). Where a problem was not captured by the URAM, a negative result was recorded (-). Where a patient problem was interpreted as having some overlap with a URAM item, this was recorded as unclear (+/-).

As well listing the types of indication for their surgery, a question was posed to establish the relative importance patients assigned to issues such as speed of recovery or better long-term outcome of treatment. The question asked them to rank the following characteristics of their ideal treatment:

- Early recovery from surgery
- Better long term outcome
- Surgeon's recommended treatment

The third of these options was provided to gauge whether patients deferred to surgeon opinion or felt unable to prioritise speed of recovery against long-term outcome.

6.4 Results

There were 117 eligible patients. Six declined to participate and one was excluded due to Alzheimer's disease of a severity that precluded informed consent, leaving 110 patients. Eighty of the 110 were awaiting limited fasciectomy, 28 dermofasciectomy, and two aponeurotomy. Their mean age was 68 (range: 34-90). Eighty-four (76%) were men, 69 (63%) were awaiting surgery to their right hand and 67 (61%) were awaiting surgery to their dominant hand. The mean total passive extension deficit across the metacarpophalangeal and proximal interphalangeal joints of the fingers to undergo surgery was 68°. In these measurements, the other joint in the ray was held in passive flexion, thus these were the 'best' passive angles when using the definition from the previous chapter.

In total, patients provided 278 problems, a mean of 2.5 per patient. Fifteen patients provided one problem, 22 gave two problems and 73 specified three problems.

PROBLEM	Number (%) of patients citing problem	Capture in URAM scale	RELEVANT URAM SCALE ITEM
Difficulty washing self	54 (49%)	+	1, 2
Difficulty picking things up (large or small items)	27 (25%)	+/-	9
Finger hooking on things	26 (24%)	-	
Difficulty putting on gloves	25 (23%)	-	
Pain	17 (15%)	-	
Difficulty gripping	16 (15%)	-	
To prevent worsening	12 (11%)	-	
Difficulty putting hands in pockets	9 (8%)	-	
Difficulty placing hand flat	7 (6%)	+/-	8
Difficulty with palmar hold of items	7 (6%)	-	
Difficulty opening bottle tops	7 (6%)	+/-	3

Table 6.2: The most common problems specified by patients

Once consolidated by the authors, there were 43 types of problem, and 94 different combinations of these amongst the 110 patients. The most common problems, their frequencies and capture by the URAM, are shown in Table

6.2. Less common problems included five problems that were specified by four patients each: difficulty using computer keyboard (unclear whether this was captured by the URAM, +/-), difficulty shaking hands (+), difficulty driving (-), difficulty clapping (+), and difficulty dressing oneself (-). Nine further indications were specified by three patients each: difficulty playing piano (+/-), difficulty using cutlery (-), difficulty using the computer other than the keyboard (e.g. the mouse) (-), difficulty with fine grip (+), difficulty writing (+/-), finger knocks things over (-), difficulty playing bowls (-), difficulty playing golf (-), and non-specific difficulty at work (-). Five more indications were common to two patients each: difficulty playing snooker (+), difficulty applying cream to body (+), difficulty cutting fingernails (-), appearance of finger upsetting (-), difficulty stroking (+). There were also 13 indications specified by one patient each: paraesthesia (-), difficulty leaning on hand (+), difficulty placing hand into enclosed spaces (e.g. washing cup) (-), difficulty washing clothes (-), difficulty playing flute (-), difficulty using wheelchair (-), difficulty tying shoelaces (-), difficulty performing housework (-), unable to massage (+), difficulty cooking (-), concerned about finger swelling (-), dropping items (-), and difficulty sleeping (-).

Some problems might not relate directly to extension deficit (for example, difficulty using computer keyboards and difficulty playing piano may relate to reduction of palmar span, and difficulty using cutlery and dropping items may relate to fine motor control rather than loss of extension). Seventeen patients (who had no comorbid painful condition) listed pain in the digit, despite pain

not being considered a relevant symptom in Dupuytren's disease by the team that designed the URAM. Two patients cited the unpleasant appearance of the digit as a problem, a symptom not assessed by the URAM or the DASH, but one which is captured by the PEM. Of the 15 patients who only cited one problem, the problem was difficulty washing oneself for five, concern about future deterioration for four, difficulty putting on gloves for two, and one each for difficulty gripping, difficulty holding items in the palm of the supinated hand (for example when given coins in a shop), difficulty dressing oneself, and difficulty writing.

URAM scale items would have directly assessed ("+") nine of the 43 different types of problem obtained in this study. When frequencies of these nine indications were considered, URAM items would have captured 73 of the 278 problems specified by patients (26%). A further six of the 43 types of problem showed some overlap ("+/-") with URAM items, accounting for a further 51 of the 278 (18%) problems. The remaining 28 indications showed no overlap with URAM items ("-"), and accounted for 154 of the 278 problems (55%).

Long-term outcome and surgeon's recommendation were relatively important to patients, whereas achieving rapid early recovery was of less importance (Table 6.3).

RANK	OPTIONS		
	Early recovery	**Long-term outcome**	**Surgeon's recommendation**
First	27 (24%)	42 (38%)	41 (37%)
Second	35 (32%)	34 (31%)	40 (36%)
Third	48 (44%)	34 (31%)	29 (27%)

Table 6.3: The relative importance of various options to the patients

6.5 Discussion

6.5.1 Present Study Findings

This study demonstrates that preoperative patients have a wide range of goals for the treatment of their Dupuytren's disease. Individuals' expectations were usually unique. Relatively few considered the treatment to be prophylactic; most were expecting improvement in some functional limitation. The URAM scale only captured a minority of these goals. It specifically assesses Dupuytren's disease, and its design used standard, contemporary methodology (Beaudreuil et al., 2011). However, it failed to describe most of the problems that this study's patients experienced. This may be due to the URAM's development, or due to cultural differences between its developers' French patients and this British cohort. The initial item generation phase of the URAM involved input from only nine French patients. It is not clear from the paper whether item generation was conducted before or after treatment. Furthermore, the small French group may not have been representative of patients with Dupuytren's contractures. Reduction of these items was then achieved by removing:

- Items never performed by at least 5% of patients
- Items with a low spread of responses amongst preoperative patients
- Items with low test-retest reliability
- Items with redundancy (defined as a high correlation coefficient with another item)
- Items showing poor factor loading in an exploratory factor analysis

Eliminating items never performed by 5% of their study population of patients might have removed problems that were very important to others. Additionally, removing items because of a small spread of responses obtained when administered to a small group of preoperative patients may not be appropriate. Based on this methodology, a task that was commonly found to be severely limiting by all patients with the condition would be rejected.

It is possible that some of the common goals identified in this study that are not present in the URAM relate to cross cultural differences. For example, difficulty wearing gloves was a common problem for British patients, but may not be relevant if the French patients participating in item generation were based in a warmer climate.

However, the absence of other British patients' goals from the URAM is less easily understood. The authors of the URAM scale state that Dupuytren's disease is painless. However, improving pain was a goal for 15% of the patients in this study, and pain has been previously documented in Dupuytren's disease (Hueston, 1963, von Campe et al., 2012). Furthermore, pain may be an important aspect of postoperative functional impairment. The URAM scale appears unsuitable for the assessment of common complications of treatment of Dupuytren's disease (Crean et al., 2011), which differ from preoperative disabilities. Instead, the URAM scale items correlated most closely with Tubiana stage, a classification of loss of extension (Beaudreuil et al., 2011). As a result, the use of the URAM to assess postoperative outcome of treatment, particularly in the first few weeks when differing treatments may

be expected to result in different levels of pain, may not be appropriate. The developers of the URAM present interpretability data in the original paper (Beaudreuil et al., 2011), but this is performed in a cohort of patients undergoing aponeurotomy. As a less invasive treatment, this has a lower risk of complications than fasciectomy (Crean et al., 2011), at the expense of a greater risk of recurrence (van Rijssen et al., 2012). Whether the URAM's interpretability is acceptable for more invasive treatments, after which complications are likely to be more common, is not known.

The findings of the present study also suggest that the long-term outcome of treatment may be more important than quick recovery to the majority of patients. A larger sample size would have allowed a meaningful sub group analysis, as it may comprise distinct patient types with different preferences. For example, self-employed working age people might prefer quicker recovery at the risk of greater recurrence whereas retired patients might prefer to minimise the likelihood of further treatment for recurrence.

6.5.2 Limitations

There are limitations to the study. Only preoperative goals for treatment were considered in the present study. It is implied that the patients involved in the development of the URAM were preoperative patients (Beaudreuil et al., 2011). Postoperative functional impairment may relate to complications. As discussed, common postoperative complications differ from preoperative impairment, and may include features such as pain and altered sensation, which are not assessed by the URAM (Crean et al., 2011). Preoperative

patients may be unlikely to appreciate the extent or effect of postoperative complications. Studying goals in a mixed cohort of preoperative and postoperative patients might have generated two types of goal: disabilities to be improved by treatment, and complications or adverse effects to be avoided. The sample of patients included some who had already had treatment of Dupuytren's disease in the past. The problems that they reported, and their expectations of treatment, might have been influenced by their previous experiences. However, as many patients with Dupuytren's disease would be expected to undergo more than one procedure in their lifetime, this may increase the generalisability of the results. Another potential issue is that most of the patients studied were awaiting fasciectomy or dermofasciectomy, rather than needle aponeurotomy. Needle aponeurotomy has a quicker recovery (van Rijssen et al., 2006), but exhibits more recurrence (van Rijssen et al., 2012). Patients who specifically sought quick recovery might have been referred elsewhere. However, such a selection bias would not necessarily influence the symptoms reported by the patient.

6.5.3 Relationship to existing literature

Alternative PROMs to the URAM do exist, and have been used to assess outcome of treatment for Dupuytren's contracture (Ball et al., 2013). The most commonly used PROM has been the DASH. Like the URAM, it mainly assesses activity limitations, and so may also fail to capture the problems experienced by many patients. However, it does capture a broad range of impairments affecting the upper limb that are not measured by the URAM. In

particular, symptoms that might arise as complications of treatment (Crean et al., 2011), such as pain and paraesthesia are measured.

The PEM outcome measure assesses impairments and generalised, rather than specific, activity limitations (Macey et al., 1995). It also measures the psychological impact of the appearance of the hand. Therefore, the PEM might better capture the broad range of specific problems that patients experience.

An alternative solution would be to use individualised patient-reported outcome measures, which allow individual patients to specify tasks with which they have difficulty, or symptoms they find troublesome. Most of the patients had unique personal combinations of problems that caused them to seek treatment. A study from Sweden also reported a broad range of functional problems experienced by patients with Dupuytren's contractures and found these self-defined tasks improved significantly following surgery (Engstrand et al., 2009). This individualised approach to measuring outcome may represent a responsive, meaningful, patient-centred and pragmatic endpoint.

Several other individualised tools exist, such as the Measure Your Medical Outcome Profile (MYMOP) (Paterson, 1996) and Canadian Occupational Performance Measure (COPM) (Law et al., 1990). These have been used in other clinical areas, but have not been validated for use in Dupuytren's disease.

Finger joint angle measurements (passive extension deficit, active extension deficit and range of motion) have been used to measure outcome in many

previous studies (Ball et al., 2013). However, angular measurements are not patient-centred outcomes, and the previously unmeasured influence of dynamism has already been studied. The data in this study suggest that some patients experience problems from their Dupuytren's contractures that might not relate directly to loss of extension or reduced range of motion.

6.6 Conclusions of study of patients' goals

Patients had wide-ranging and often individualised goals for the treatment of Dupuytren's disease. These data suggest that the face validity of the URAM needs to be reassessed, and perhaps modified for general use. However, in this study, patients experienced a broad range of problems specified, which may prove hard to capture with any existing rigid scale based on activity limitation.

SUMMARY

This chapter has demonstrated that patients' goals for the treatment of Dupuytren's disease are broad. With 94 different combinations of goals provided by 110 patients, they often have distinct individual expectations of the treatment of Dupuytren's disease. A significant proportion of these goals are not captured by the URAM.

Surgeons may believe that they are treating patients 'prophylactically', before contractures become severe. However, only a very small proportion of patients underwent treatment for Dupuytren's disease prophylactically; nearly all described a functional limitation that they wanted improved by the surgery. In this study, patients prioritised avoiding recurrence in the long term, rather

than rapid functional recovery after treatment. However, this may reflect a selection bias as the cohort was undergoing more invasive surgery (fasciectomy or dermofasciectomy rather than aponeurotomy), which is believed to deliver better long-term outcome.

Further study of the behaviour of the URAM scale in the treatment of Dupuytren's disease is needed, and such work is presented in the next chapter. Given the broad range of goals identified in this chapter, the URAM's comprehensiveness will be examined in the next chapter. Given that the DASH was used in the most contemporary trial of Dupuytren's disease surgery (van Rijssen et al., 2006; van Rijssen et al., 2012), and that the DASH is the most popular PROM for studying Dupuytren's disease (Ball et al., 2013), its validity will also be studied. In this way, the next chapter a balanced consideration of its performance can be made.

7 Validity of the Disabilities of the Arm, Shoulder and Hand tool, the QuickDASH tool, and the Unité Rhumatologique des Affections de la Main scale in Dupuytren's disease

7.1 Preface

Earlier chapters have demonstrated that passive extension deficit, the primary outcome measure in previous randomised trials of the treatment of Dupuytren's disease (van Rijssen et al., 2006, van Rijssen et al., 2012), may be subject to previously unmeasured bias in the form of dynamism. Furthermore, the first Dupuytren's disease-specific patient-reported outcome measure (PROM), the Unité Rhumatologique des Affections de la Main (URAM) scale may still require further assessment of cross-cultural sensitivity and interpretability.

The most popular PROM in Dupuytren's disease, the Disabilities of the Arm, Shoulder and Hand tool (DASH) has been validated in a range of clinical scenarios, but not specifically against modern standards of validity (Mokkink et al., 2010), or in Dupuytren's disease specifically. Furthermore, its validity has been questioned (Packham, 2011), albeit on the basis of it poor correlation with angular deformity (Engstrand et al., 2009, Jerosch-Herold et al., 2011, Zyluk and Jagielski, 2007), which does not constitute a valid assessment of criterion validity according to current standards (Mokkink et al., 2010). Also, the inclusion of items assessing pain has been criticised in Dupuytren's

disease (Beaudreuil et al., 2011). As a result, there is a need to assess the validity of the DASH in Dupuytren's disease, to guide outcome measure selection for future research, but also to inform appraisal of the existing evidence base that has used it, in clinical guideline development, for example. This chapter considers some aspects of the performance of the DASH, its shortened version the QuickDASH, and the URAM.

7.2 Introduction

Several PROMs have been used to evaluate Dupuytren's disease, and the 30-item DASH is the most popular (Ball et al., 2013). However it has been suggested that the DASH may not be valid for use in Dupuytren's disease (Beaudreuil et al., 2011, Packham, 2011), as neither it nor the QuickDASH correlates closely with angular deformity (Budd et al., 2011, Engstrand et al., 2009, Jerosch-Herold et al., 2011, Zyluk and Jagielski, 2007). Furthermore, both include items that assess pain whereas it is claimed that Dupuytren's disease is not painful (Beaudreuil et al., 2011). Other groups suggest that pain may be present (Hueston, 1963, Rodrigues et al., 2014, von Campe et al., 2012), and treatment-related pain may affect postoperative recovery of function.

Much of the data describing the validity and reliability of the DASH was obtained from mixed cohorts involving upper limb conditions widely accepted as painful (Kennedy et al., 2011). Much of the data describing the validity and reliability of the DASH was obtained from mixed cohorts involving upper limb conditions widely accepted as painful (Kennedy et al., 2011). Other than the recent publication of the secondary analysis of a randomised controlled trial (Forget et al., 2014), there is limited data describing the DASH's validity and reliability in Dupuytren's disease specifically.

PROMs have been developed to study Dupuytren's disease specifically, such as the URAM (Beaudreuil et al., 2011) and the Southampton Dupuytren's

scoring system (Mohan et al., 2014). The latter was published since the completion of data collection for this project.

Other PROMs that have been used to assess Dupuytren's disease (Ball et al., 2013), include the Michigan Hand Questionnaire (MHQ) (Chung et al., 1998), the Patient Evaluation Measure (PEM) (Macey et al., 1995), and the QuickDASH (Beaton et al., 2005) . In a study of patients with a range of hand conditions, the DASH took longer to complete than the PEM, but was quicker than the MHQ (Dias et al., 2008). Patients contributing to research, service evaluation or audit might be asked to complete more than one outcome measure. For example, a specific PROM and a generic measure to assess health-related quality of life, such as the EuroQol 5 D (EQ5D) (Herdman et al., 2011), may be required to facilitate cost effectiveness analysis (NICE, 2008). As a result, using PROMs that are quicker for the patient to complete may be more convenient and facilitate higher response rates.

The QuickDASH comprises 11 of the 30 items in the DASH, and should be quicker to complete. However, it has not been used extensively in Dupuytren's disease (Ball et al., 2013).

The URAM is the first Dupuytren's disease-specific PROM to have been developed. It has been subjected to assessments of its validity by the developer (Beaudreuil et al., 2011). It purposefully does not assess domains such as pain. As such, it may prioritise relevance of preoperative symptoms ahead of comprehensiveness for use in the postoperative setting.

Consensus-based standards for the selection of health status measurement instruments (COSMIN) have been developed (Mokkink et al., 2010). These define different aspects of the validity of PROMs.

Content validity assesses whether the items that comprise a PROM are an adequate reflection of what is trying to be measured. It involves assessing the relevance and comprehensiveness of the items in a PROM.

Construct validity examines hypotheses about the PROM. Such hypotheses may relate to its structural validity (internal relationships between items), hypothesis testing (assessing its relationship with other PROMs) and differences between groups (cross-cultural validity). For example, the former can be investigated by studying whether all items in a tool contributing to a summary score reflect the same underlying construct in a multivariate analysis.

Internal consistency, considered part of the reliability of a PROM, is the interrelatedness of the items within a PROM. It assumes that all of the items that contribute to a summary score intend to reflect the same underlying entity, or factor (e.g. impairment of structures in the hand versus restriction of function involving the shoulder), i.e. they are 'unidimensional'. If this assumption is met, internal consistency assesses how closely the items reflect the construct concerned.

Criterion validity tests a PROM against a 'gold standard'. The only accepted methodology for this is the comparison of a shortened PROM against the long version (e.g. the QuickDASH against the DASH).

This cross sectional study assessed aspects of content validity, construct validity and reliability of the DASH and the URAM in Dupuytren's disease, and studied the relationship between the DASH and the QuickDASH (i.e. the criterion validity of the QuickDASH). Responsiveness (the ability to detect change over time) differs from 'validity', in that responsiveness assesses a change score, whereas validity assesses a single time point score. This will be studied in the next chapter based on cohort study data.

7.3 Methods

7.3.1 Patient recruitment and data collection

The data presented in this study were gathered as part of a larger service evaluation.

Patient recruitment took place between September 2011 and April 2013. The inclusion criteria were primary or recurrent Dupuytren's disease, and either:

1. Patients awaiting fasciectomy or dermofasciectomy at one UK hand surgery centre or
2. Patients available for assessment at five UK hand surgery centres 1 year or 5 years (+/- 2 months) after their surgery when the candidate was available.

Exclusion criteria were:

- Cognitive impairment preventing informed consent
- Refusal of invitation to participate

For the first inclusion criterion group, preoperative patients were recruited at the routine preadmission clinic visit prior to surgery. Those who were eligible and consented to participate completed the DASH prior to surgery. These patients were also sent questionnaires for completion by post at 3 weeks, 6 weeks and 1 year after surgery. Patients who were scheduled for surgery to the left and right hand at different times during the study recruitment period were eligible for recruitment twice. This happened on four occasions.

Patients in the second inclusion criterion group were invited to participate with a letter explaining the project and inviting them to participate on a voluntary basis, with a fixed stipend offered to cover travel expenses. The candidate assessed those who consented to participate. The assessment included collection of demographic data and completion of the 30-item DASH questionnaire.

The URAM scale was published during the study period. Patients recruited later in the study (August 2012 onwards) also completed the URAM at assessments.

7.3.2 Angular measurement: total passive extension deficit (TPED)

Patients who completed PROMs whilst a surgeon was present (as opposed to completion by post – which was the case for 3 and 6 week postoperative measurements) had the passive extension deficit of the treated digit assessed by a single examiner. Total passive extension deficit (TPED) was calculated by adding the measured passive extension deficits of the metacarpophalangeal joint and proximal interphalangeal joint while the other joints of the digit were passively flexed. The measurement thus minimised the influence of dynamism (Rodrigues et al., 2014).

7.3.3 Content validity: relevance of pain questions in the DASH

The relevance of items assessing pain was assessed by extracting and analysing responses to question 24 of the DASH (which assesses pain, and is question 9 of the QuickDASH) and question 25 of the DASH (which assesses pain during specific activity) at different time points. It was hypothesised that if

pain items were relevant, they would change significantly through the recovery period.

7.3.4 Construct validity and reliability

How the different items in the DASH related to each other (their internal relationships) for different patients was analysed. This constituted a study of the structural validity of the DASH. When used as instructed by the developer, the DASH generates a single summary 'DASH score', using all of the 30 items (in contrast to other tools such as the Michigan Hand Questionnaire, which generates several summary scores for different areas). For the single DASH score to be valid, all items contributing to the score should measure, or 'reflect', the same underlying entity or 'factor', in this case upper limb function, i.e. the tool should be unidimensional (Mokkink et al., 2010). To evaluate whether the DASH is unidimensional, exploratory factor analysis (EFA) was employed (Mokkink et al., 2010). EFA analyses the relationship between items when completed by different people to identify underlying latent factors that explain the variance; the differences seen between individuals across a population. Some of the relevant concepts involved are defined in Table 1. It was expected that the responses obtained would have a tendency towards low scores and so not fit a normal distribution. This was examined by calculating the kurtosis and skewness for items. If the responses were not normally distributed (defined as kurtosis or skewness >+2 or <-2), then logarithmic transformation of all items was performed (Pallant, 2010) and their distributions then reassessed prior to factor analysis. EFA

may be performed using different statistical methods. In this study, principal axis factoring was used to extract latent factors that were being reflected by the DASH's items, and the number of factors extracted was determined and confirmed by using two different accepted techniques (scree plots and parallel analysis, see Table 1) (Cattell, 1966, Horn, 1965, Patil et al., 2007). If the DASH were unidimensional, then there would only be one factor that could be extracted. Cronbach's alpha was calculated to assess internal consistency. However, this was interpreted with caution if unidimensionality had not been confirmed.

Aspect Studied	Method(s) Used	Concept(s)	Description
Content validity	Cohort study of pain items	Comprehensiveness & Relevance	Content validity considers whether PROM items are relevant to what is being measured, and whether the scale overall is comprehensive.
Distribution of item responses	Kurtosis & skewness	Kurtosis	Defines the sharpness of the peak of a distribution of data
		Skewness	Defines the amount of asymmetry of a distribution of data
Construct validity (structural validity)	Exploratory Factor Analysis	Unidimensionality	All items contributing to a summary score reflect the same underlying factor (some PROMs generate more than one summary score, each describing a different subscale of the PROM)
		Factor	Factor analysis aims to describe the variation in measured items that correlate with each other in terms of fewer unobserved 'factors'
		Principal axis factoring	A form of factor analysis in which factors are extracted based on common variance, rather than total variance.
		Eigenvalue	Describes the amount of the total variance that is explained by a particular factor
		Catell's scree plot	Determines the number of factors to extract. All potential factors are plotted in order of Eigenvalues (see Chart 1). The turning point where the connecting line flattens sharply is the point at which the last

		significant factor has been passed, as further factors represent a flat level of background noise (Catell, 1966).
	Parallel analysis	Determines the number of factors to extract. A second set of Eigenvalues is generated, but which are based entirely on chance. All factors in the real model with Eigenvalues greater than their counterparts in the chance model are significant and are then extracted. Avoids the risk of bias that exists with scree plot interpretation (Horn, 1965).
	Kaiser-Meyer-Olkin statistic	Assesses sampling adequacy. It lies between 0-1, describing the proportion of variance that is common variance. The minimum acceptable level for analysis being 0.6, and >0.9 being described as 'marvellous' (Kaiser, 1974).
	Bartlett's test of sphericity	Assesses whether an identity matrix would result if correlations between the items included were studied, i.e. whether the correlation between all of the DASH items is zero. Some correlation between items is needed for EFA. A significant result is achieved if the data is suitable for EFA.
	Factor loading	The output of EFA. Described how closely an item correlates with a factor.
	Rotation	Presents the factor loadings in a manner that makes interpretation more straightforward, resulting in a pattern matrix. Only possible when more than one factor is extracted. When done, the majority of the items may each show strong loading with one of the factors, a situation described as simple structure.

Reliability	Cronbach's alpha	Internal consistency	Studies the inter-relatedness between items within a scale. However, it is dependent on the scale being unidimensional and reflective
Relationship between DASH and QuickDASH	Bland-Altman plots	Agreement	Studies the relationship between two variables that are expected to correlate. In such circumstances, reporting correlation is common, but may not be appropriate.

Table 7.1: Data handling and relevant statistical concepts

7.3.5 Relationship between the DASH and QuickDASH

The DASH summary score was calculated using the standard formula provided:

DASH=((a/b)-1)*25

Where “a” is the sum of the scores for the responses completed (each response could be scored between one and five), and “b” is the number of responses the patient completed.

The QuickDASH summary score was calculated by extracting the answers to the relevant 11 questions. The score was calculated using the same formula as for the DASH, only with these eleven items.

As the summary scores of the DASH and the QuickDASH are virtually continuous scales (each scored 0-100), and the sample comprised a large number of independent observations, parametric analyses were used to compare them, in keeping with the central limit theorem (Norman, 2010). Pearson’s correlation coefficients were calculated between the total scores for the DASH and the QuickDASH for a) the total sample and b) for different time point subgroups. If the relationship between the QuickDASH and the DASH was not absolute and did not lie on the line of equality (i.e. correlation coefficient was less than 1, the maximum possible correlation coefficient), then agreement was also studied. Agreement was assessed by calculating 95% limits of agreement, using Bland-Altman analysis of the difference between the QuickDASH and the DASH (Bland and Altman, 1986).

7.3.6 Handling of incomplete responses

If more than three of the 30 responses are missing (i.e. fewer than 27/30 responses provided), then the DASH cannot be calculated (Kennedy et al., 2011), and if more than one response of the eleven is missing (i.e. fewer than 10/11 responses provided), then the QuickDASH cannot be calculated (Kennedy et al., 2011). If either was the case, then that questionnaire was excluded from the whole study.

Therefore, some of the included questionnaires still had missing data (up to 3/30 responses missing for the DASH, or 1/11 missing for the QuickDASH). As the study of the relationship between the DASH and QuickDASH used summary scores that can still calculated despite such missing data, this was of no consequence for the analysis of the criterion validity of the QuickDASH. However, in the EFA, individual item data are required. The questionnaires with some missing data were still included for analysis, but with missing responses excluded pairwise.

If required (e.g. for repeated measures ANOVA), then listwise exclusion was used.

Clear guidance for handling incomplete questionnaires was not available for the URAM. However, as the URAM was only included in EFA, which requires individual item data, pairwise exclusion was used to handle missing data.

7.3.7 Sample size

Sample size estimates were based on guidance for designing EFA analyses. Whilst different heuristics exist for this, there is little evidence to support formal estimation of sample size for such analyses (Mundfrom et al., 2005). A suggested sample size of 300-400 for factor loadings of 0.4 was followed, as summarised in a review of EFA methodology (Floyd and Widaman, 1995). A target sample size of 300 was set.

7.4 Results

7.4.1 Demographics

768 DASH questionnaires were received. These described the preoperative and/or postoperative assessment of 527 different procedures. Nine cohort study questionnaires were incomplete to the extent that calculation of a summary score was not possible based on the guidance issued with the DASH and/or the QuickDASH, and they were excluded from all analysis. Thus, 759 DASH questionnaires describing 527 procedures on 523 patients were analysed (Figure 7.1). The 527 procedures comprised 126 needle aponeurotomies, 327 fasciectomies and 74 dermofasciectomies. The mean age at the time of assessment was 68 (range: 34 to 94) years and 403 of the 523 (77%) patients were men. TPED measurements were made at the time of completion of the DASH scores in 522 of the 759 occasions (109 preoperative and 413 postoperative).

330 URAMs were also received. For three, the corresponding DASH was incomplete. Therefore, 327 URAMs with matching DASHs were suitable for inclusion in the EFA of DASH and URAM items. These described the outcome of 284 procedures in 284 patients. The 284 procedures comprised 103 needle aponeurotomies, 144 fasciectomies and 37 dermofasciectomies. The mean age at assessment of this group was 67.5 (range: 35 to 92) years and 227 of the 284 (80.0%) were male. The mean URAM summary score was 7.7/45 (95% CIs: 6.7,8.6). The mean preoperative score was 17.6, the

mean 1-year postoperative score was 5.6 and the mean 5 year postoperative score was 6.5.

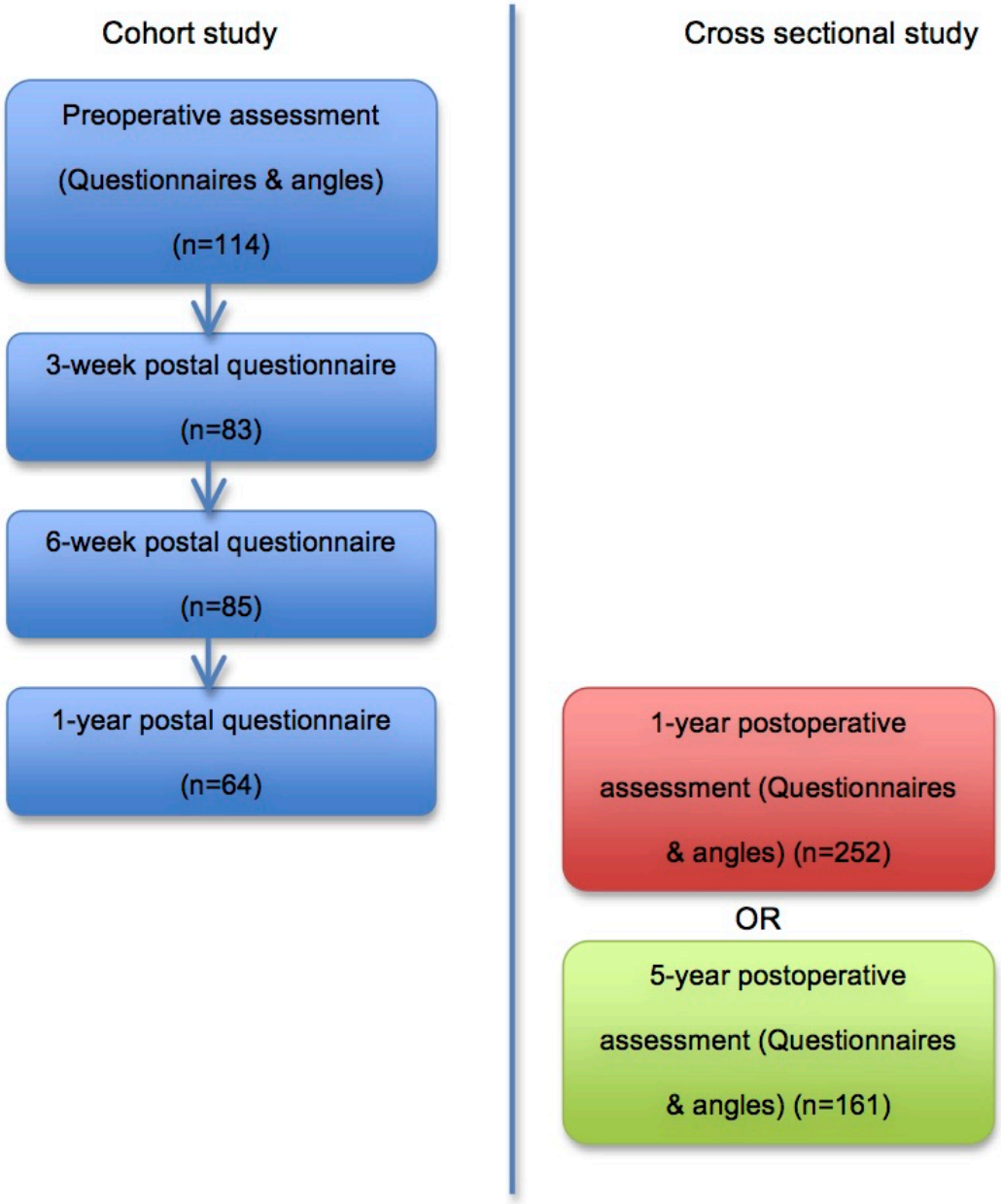

Figure 7.1: Flow diagram demonstrating time points from which complete DASH questionnaires were received

TIME POINT	MEAN DASH (95% CIs)	MEAN QuickDASH (95% CIs)	PEARSON's *r* (95% CIs)
Preoperative	27 (23-31)	28 (24-32)	0.98 (0.97-0.99)
3 week postoperative	37 (33-42)	41 (36-46)	0.97 (0.95-0.98)
6 week postoperative	22 (18-26)	24 (20-28)	0.98 (0.97-0.99)
1 year postoperative	11 (9-13)	12 (10-14)	0.99 (0.98-0.99)
5 years postoperative	11 (9-13)	12 (10-15)	0.97 (0.96-0.98)

Table 7.2: Mean DASH and QuickDASH scores and their correlations, by time point

95% CIs – 95% confidence intervals

7.4.2 Content validity: relevance of pain questions

Question 24 of the DASH (which is question 9 of the QuickDASH) requires participants to rate pain experienced in the arm, shoulder or hand in the preceding week on a scale from 1 (no pain) to 5 (severe pain). This question was completed in 750 of the 759 questionnaires studied. The median score for question 24 was 2/5 ("mild" pain) for the total study. This was also the

case when preoperative responses were studied alone. Sixty-eight patients provided answers to question 24 preoperatively, at 3 weeks and at 6 weeks. When these responses were compared, there was a significant difference between them (p=0.003, repeated measures ANOVA test). Specifically, scores were lower (i.e. less pain) at 6 weeks than at 3 weeks (Tukey's multiple comparison test). Question 25 of the DASH (which is not part of the QuickDASH) rates pain when performing a specific activity. It was completed in 745 of 759 questionnaires. The median score overall was again 2/5, and this was also the case for preoperative responses. The median score was 3/5 for 3-week postoperative responses, falling again to 2/5 in 6 week postoperative responses. Sixty-one patients provided answers to question 25 preoperatively, at 3 weeks and at 6 weeks. Again, when these were compared, there was a significant difference between them (p=0.003, repeated measures ANOVA test). Scores were higher at 3 weeks than preoperatively and were lower at 6 weeks than at 3 weeks (Tukey's multiple comparison test, Table 7.3).

Comparison	Mean first time point (out of 5)	Mean second time point (out of 5)	Difference between means (95% CIs)
Preop vs 3 week	2.12	2.41	-0.29 (-0.59, 0.00)
Preop vs 6 week	2.12	2.03	0.09 (-0.17, 0.35)
3 week vs 6 week	2.41	2.03	**0.38 (0.15, 0.62)**

Table 7.3: Tukey's multiple comparison test of responses to question 24 of the DASH at different time points from patients who completed responses at all time points (n=68)

7.4.3 Construct validity and reliability of DASH

Across the subgroup that also had angular deformity measured, the DASH showed weak correlation with TPED (Pearson's *r* = 0.30, 95% CIs: 0.22 to 0.38). The QuickDASH also correlated weakly with TPED (Pearson's *r* = 0.29, 95% CIs: 0.21 to 0.37).

For the 759 DASH questionnaires in the analysis of the DASH's structural validity, 14 of the 30 were not normally distributed based on the kurtosis and/or skewness of their distributions. Therefore, logarithmic transformation was performed. After logarithmic transformation, 28 of the 30 DASH items had normal distributions. The hypothesis that the correlation matrix of the

DASH items is an identity matrix, with no correlation between variables at all, is investigated in Bartlett's test of sphericity (see Table 7.1). This was highly statistically significant, confirming that there was correlation between variables. 467/471 correlation coefficients between log transformed DASH items were over 0.3. Some correlation between the items is required for EFA, as it studies their inter-relationships. The Kaiser-Meyer-Olkin measure of sampling adequacy (see Table 7.1) was 0.974, confirming the suitability of the data for factor analysis (Kaiser, 1974). Two major factors were present based on all common tests for determining factor numbers (Figure 7.2, confirmed by parallel analysis) (Cattell, 1966, Horn, 1965, Patil et al., 2007); hence the DASH was not unidimensional.

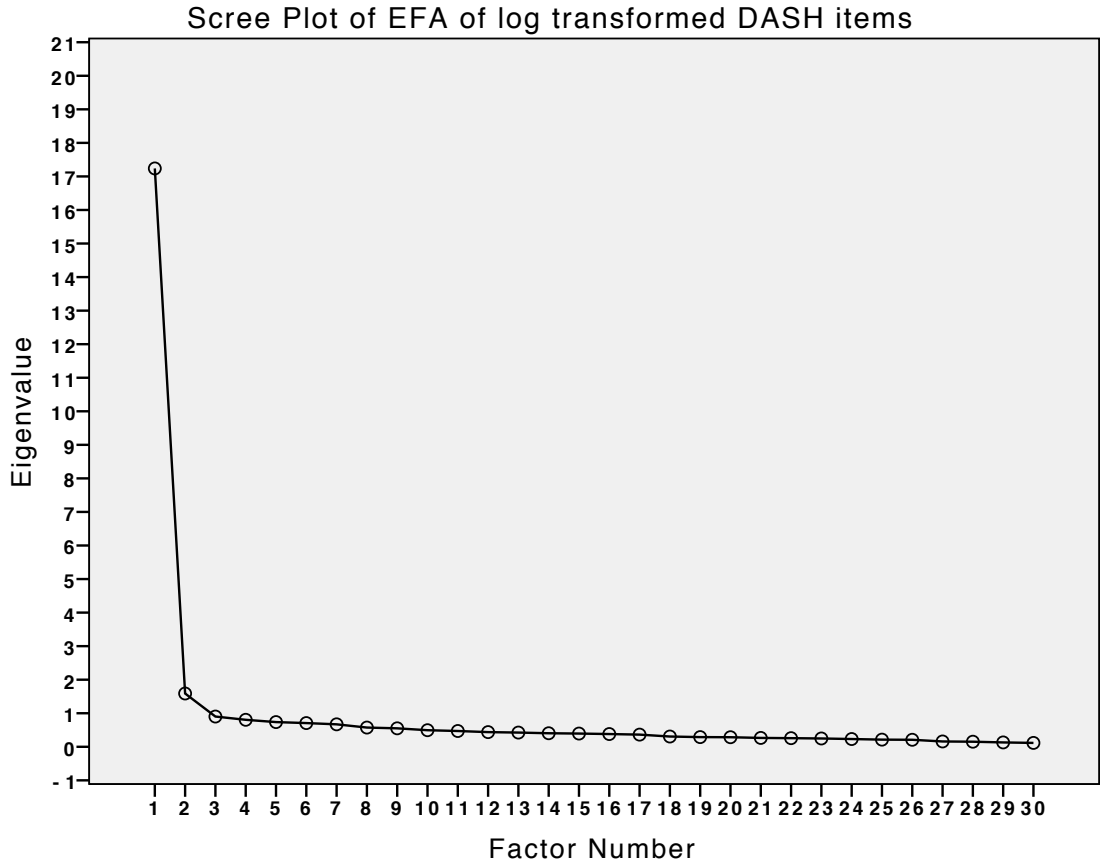

Figure 7.2: Scree plot of exploratory factor analysis for log transformed DASH items

In the scree plot, the line plateaus from the third point onwards, supporting the extraction of two factors.

These factors explained 57.5% and 5.3% of variance respectively. The results of the EFA are shown in Table 7.4. The results are factor loadings, which are the correlation coefficients for each item in the questionnaire with the factor derived in the EFA. Functional limitation items correlated well with Factor 1, whereas impairment and participation items generally correlated with

Factor 2, including pain-related items. The EFA was also rerun using raw, untransformed data, and generated the same pattern of results.

Cronbach's alpha for the DASH was 0.975. However its interpretation was limited by the finding of the DASH potentially not being unidimensional.

Cronbach's alpha for the QuickDASH was 0.933. Both results were consistent with there being redundancy of items within the scales.

Item number	Item question	Loading with factor 1	Loading with factor 2
12	Difficulty changing a light bulb overhead	**0.86**	-0.05
14	Difficulty washing your back	**0.84**	-0.09
7	Difficulty doing heavy household chores	**0.84**	0.05
5	Difficulty pushing open a heavy door	**0.82**	-0.06
13	Difficulty washing or blow drying hair	**0.82**	-0.03
9	Difficulty making a bed	**0.80**	0.03
16	Difficulty using a knife to cut food	**0.80**	0.01
11	Difficulty carrying a heavy object (>10 lbs)	**0.80**	0.04
8	Difficulty gardening	**0.80**	0.08
6	Difficulty placing an object on a shelf above head	**0.78**	-0.04
17	Difficulty with recreational activities that require little effort	**0.78**	-0.05
4	Difficulty preparing a meal	**0.77**	0.07
18	Difficulty with recreational activities in which force is taken through the limb	**0.73**	0.13
21	Difficulty with sexual activities	**0.72**	-0.11
19	Difficulty with recreational activities in which the arm moves freely	**0.71**	0.12
3	Difficulty turning a key	**0.69**	0.03
15	Difficulty putting on a pullover sweater	**0.68**	0.09
20	Difficulty managing transportation needs	**0.68**	0.02
10	Difficulty carrying a shopping bag/briefcase	**0.67**	0.15
2	Difficulty writing	**0.63**	0.02
1	Difficulty opening a tight or new jar	**0.61**	0.15
22	To what extent has your problem interfered with normal social activities?	**0.45**	**0.37**
23	To what extent has your problem interfered with work or other daily activities?	**0.44**	**0.42**

24	Rate your arm, shoulder or hand pain	-0.06	**0.90**
25	Rate your arm, shoulder or hand pain when performing a specific activity	-0.02	**0.87**
28	Rate the stiffness in your arm, shoulder or hand	0.08	**0.72**
27	Rate the weakness in your arm, shoulder or hand	0.17	**0.69**
29	How much difficulty have you had sleeping because of pain in the limb?	0.10	**0.62**
26	Rate the tingling in the arm, shoulder or hand	-0.04	**0.59**
30	Is this true: I feel less capable, confident or useful because of the limb problem?	0.29	**0.51**

Table 7.4: Pattern matrix of exploratory factor analysis of log-transformed DASH items, with two-factor solution and oblimin rotation.

** Loadings assess the correlation between the item concerned and the latent factor extracted in the analysis. They may range from -1 (perfect inverse correlation) through 0 (no correlation) to +1 (perfect correlation).*

N.B. Large factor loadings (>0.3) are shown in bold.

7.4.4 Relationship between the DASH and QuickDASH

Across the entire study, the QuickDASH was higher than the DASH indicating apparently worse upper limb function) (mean difference 1.6 (95%CIs: 1.3 – 1.8), paired t test). However the QuickDASH correlated very well with the DASH, Pearson's *r* was 0.98 (95% CIs: 0.98 – 0.99), as shown in Figure 7.3. Linear regression analysis was performed with the Y-intercept constrained to y=0, as the QuickDASH must equal zero if the DASH equals zero. Runs test confirmed that there was no significant deviation of the residuals from the

model in the linear regression analysis (p=0.228). The slope for the relationship between the two was:

QuickDASH = 1.054 x DASH

Similar correlations were seen in separate preoperative, 3 week, 6 week, 1 year and 5 year follow-up subgroup analyses (Table 7.5).

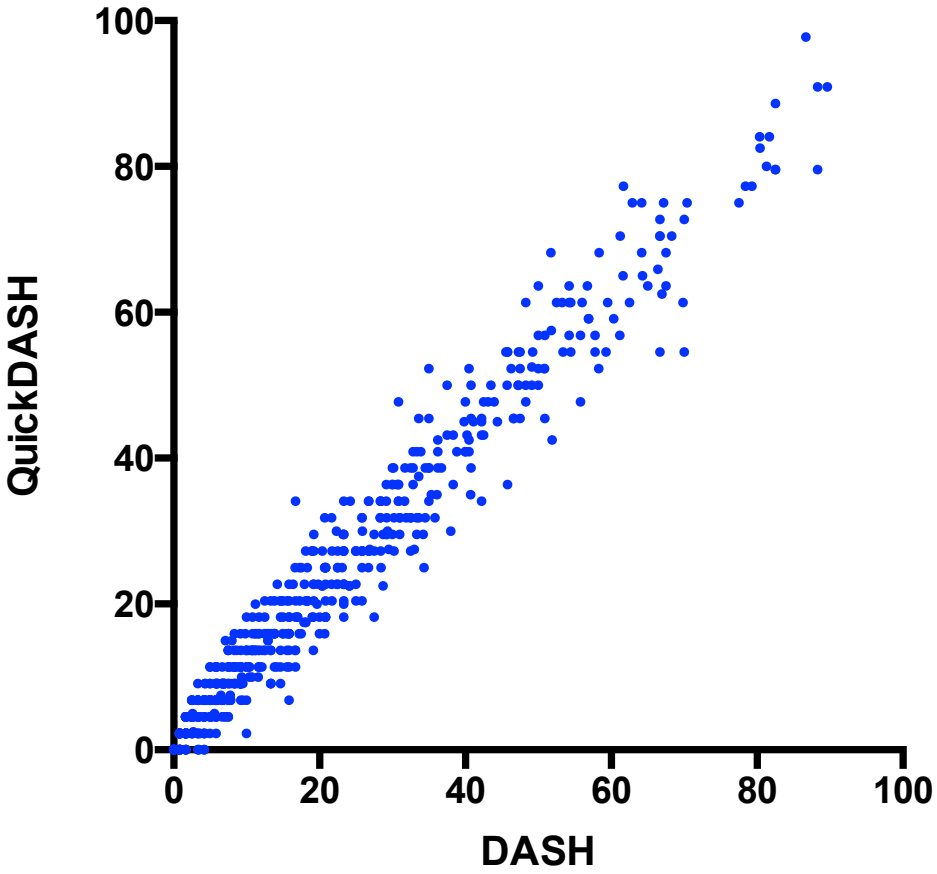

Figure 7.3: Scatterplot of QuickDASH versus DASH (n=759).

TIME POINT	MEAN DASH (95% CIs)	MEAN QuickDASH (95% CIs)	PEARSON's *r* (95% CIs)
Preoperative	27 (23-31)	28 (24-32)	0.98 (0.97-0.99)
3 week postoperative	37 (33-42)	41 (36-46)	0.97 (0.95-0.98)
6 week postoperative	22 (18-26)	24 (20-28)	0.98 (0.97-0.99)
1 year postoperative	11 (9-13)	12 (10-14)	0.99 (0.98-0.99)
5 years postoperative	11 (9-13)	12 (10-15)	0.97 (0.96-0.98)

Table 7.5: Mean DASH and QuickDASH scores and their correlations, by time point

The 95% limits of agreement between the QuickDASH and the DASH were -5.8 to +8.9 (Figure 7.4). As relatively few differences were outside the 95% limits of agreement for mean scores under 30 (those with good upper limb function), further Bland-Altman analyses were performed for scores considered asymptomatic (<15/100), and scores considered symptomatic (>15/100) by the DASH's creators (Kennedy et al., 2011). When the mean of the DASH and QuickDASH for a patient was 15 or less (asymptomatic upper

limb function), the 95% limits of agreement were -3.3 to +5.5 (see Figure 7.5), and when the mean was over 15 (symptomatic upper limb function), they were -7.8 to +12.3 (see Figure 7.6).

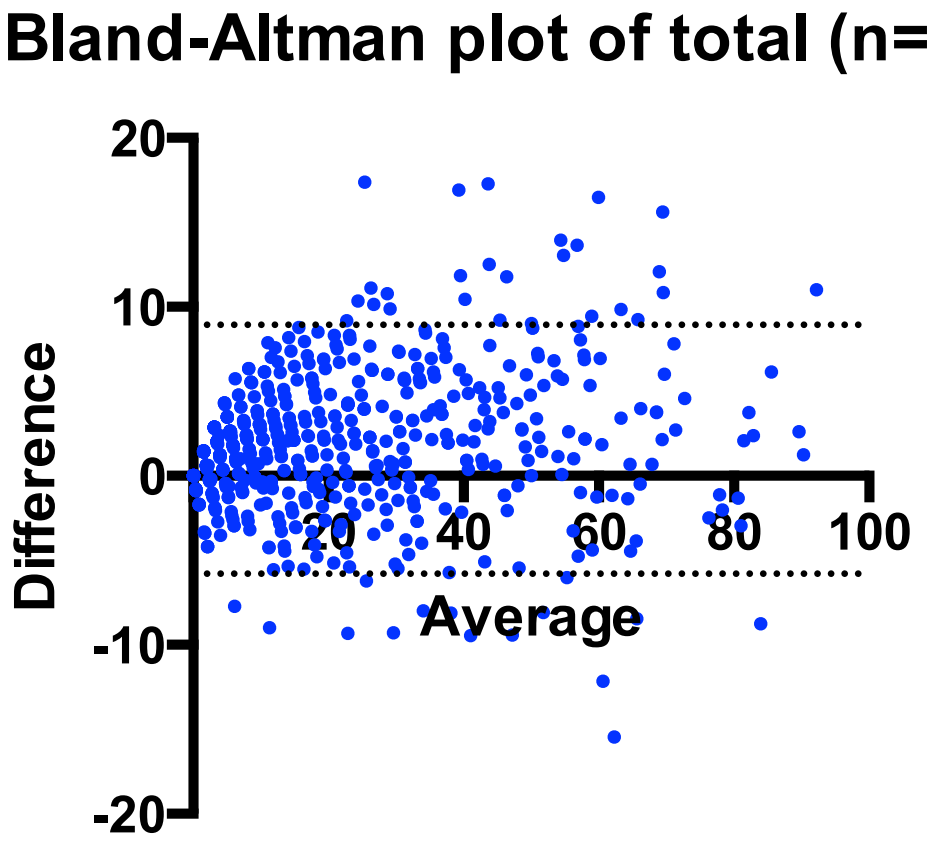

Figure 7.4: Bland-Altman plot of total sample, with 95% limits of agreement shown as dotted lines

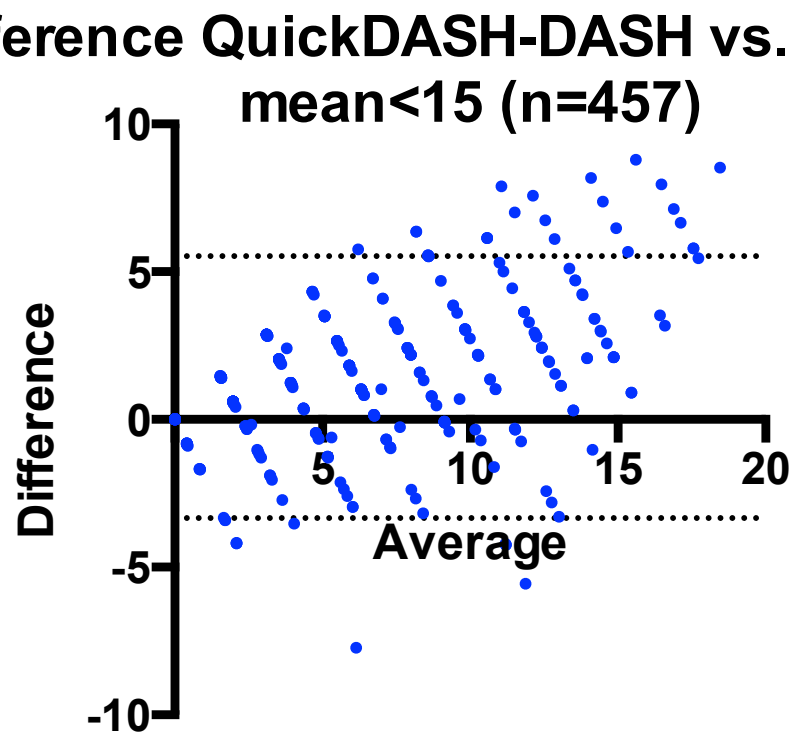

Figure 7.5: Bland-Altman plot of sub group of sample for which the mean of the DASH and the QuickDASH was <15, with 95% limits of agreement shown as dotted lines

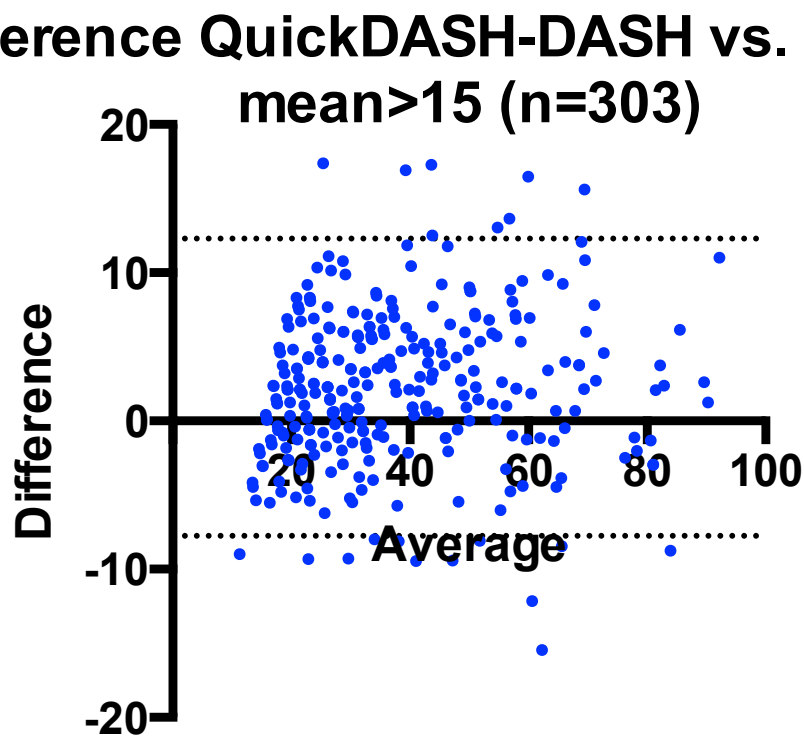

Figure 7.6: Bland-Altman plot of sub group of sample for which the mean of the DASH and the QuickDASH was >15, with 95% limits of agreement shown as dotted lines

7.4.5 Construct validity of URAM

For the 327 matched DASH and URAMs, 4/9 URAM items and 17/30 DASH items were not normally distributed initially, based on the kurtosis or skewness of their distributions. After logarithmic transformation was performed, all URAM items and 23/30 DASH items were normally distributed.

Bartlett's test of sphericity was highly statistically significant, confirming correlation between the 39 log-transformed URAM and DASH items. The Kaiser-Meyer-Olkin measure of sampling adequacy was 0.957.

Overall, five factors had Eigenvalues above 1 (one of the three rules employed to determine the number of factors to extract). However, on examination of the scree plot (Figure 7.7), the curve plateaus markedly after the first three of these factors. This was confirmed by parallel analysis, in which the randomly generated theoretical parallel analysis Eigenvalue for Factor 4 was 1.52, which was larger than the actual Eigenvalue for the equivalent Factor 4 in the EFA (which was 1.20). Similarly, the theoretical parallel analysis Eigenvalue for Factor 5 was 1.47, which was larger than the Factor 5 Eigenvalue in the EFA (which was 1.06). Therefore, for Factor 4 and 5, randomly generated Eigenvalues were larger than the actual Eigenvalues, supporting the conclusion that the Eigenvalues for Factors 4 and 5 being above 1 was due to chance. Therefore, only three factors were extracted.

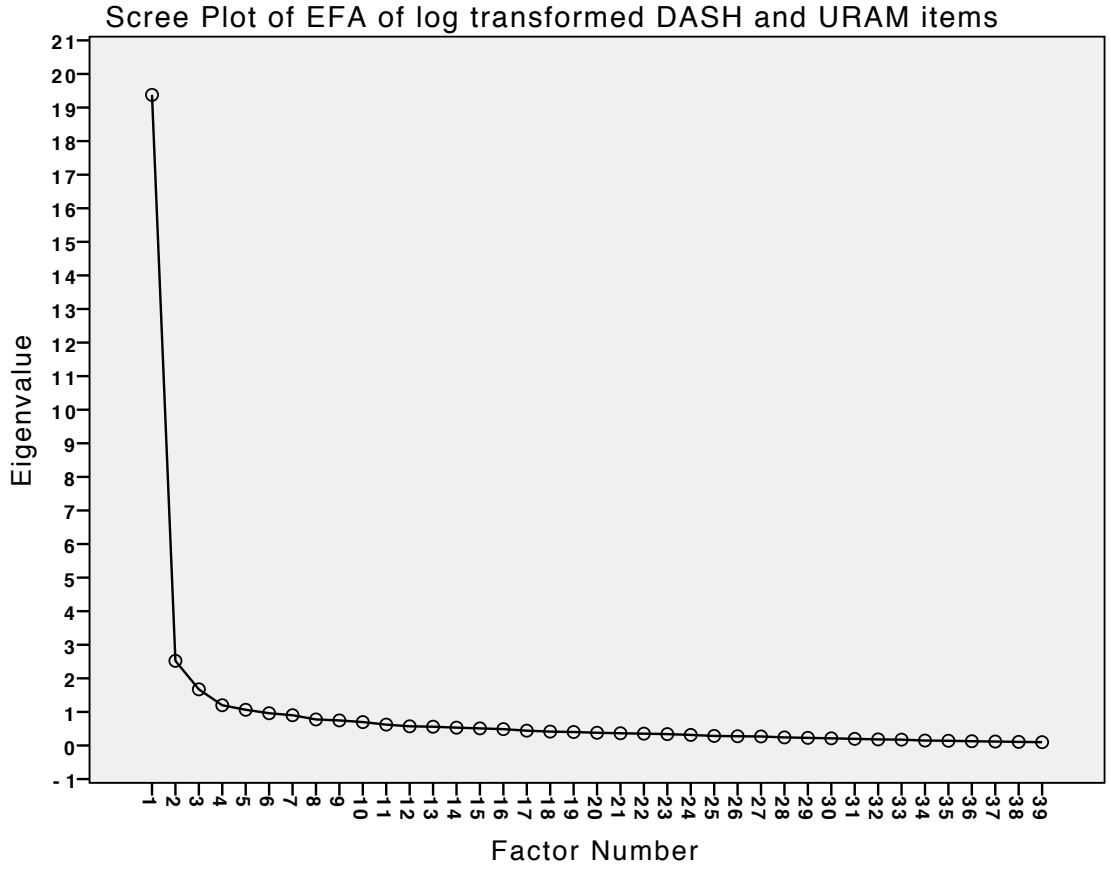

Figure 7.7: Scree plot of exploratory factor analysis for log transformed DASH items

In the scree plot, the line plateaus from the fourth point onwards, supporting the extraction of three factors

The extracted factors accounted for 48.7%, 5.60% and 3.36% of the total variance respectively. The results of the EFA are shown in Table 7.6

Item	Content of item	Loading* with Factor 1	Loading with Factor 2	Loading with Factor 3
URAM 1	Wash self with a flannel, keeping hand flat	-0.006	**0.811**	-0.011
URAM 2	Wash face	0.079	**0.764**	0.003
URAM 3	Hold bottle in one hand	0.305	**0.473**	0.137
URAM 4	Shake someone's hand	0.134	**0.694**	-0.091
URAM 5	Stroke something or caress someone	0.072	**0.833**	-0.085
URAM 6	Clap your hands	-0.021	**0.824**	0.015
URAM 7	Spread out your fingers	-0.055	**0.744**	0.157
URAM 8	Lean on hand	-0.091	**0.747**	0.272
URAM 9	Pick up small objects with thumb and index finger	**0.390**	0.106	0.179
DASH 1	Difficulty opening a tight or new jar	**0.474**	0.020	0.229
DASH 2	Difficulty writing	**0.551**	0.046	0.055
DASH 3	Difficulty turning a key	**0.699**	-0.045	0.054
DASH 4	Difficulty preparing a meal	**0.769**	-0.069	0.068
DASH 5	Difficulty pushing open a heavy door	**0.765**	0.076	-0.126
DASH 6	Difficulty placing an object on a shelf above head	**0.759**	0.012	-0.032
DASH 7	Difficulty doing heavy household chores	**0.835**	0.043	-0.049
DASH 8	Difficulty gardening	**0.800**	0.085	-0.032
DASH 9	Difficulty making a bed	**0.803**	0.067	-0.054
DASH 10	Difficulty carrying a shopping bag/briefcase	**0.712**	0.040	0.079

DASH 11	Difficulty carrying a heavy object (>10 lbs)	**0.769**	0.034	0.030
DASH 12	Difficulty changing a light bulb overhead	**0.842**	-0.037	-0.012
DASH 13	Difficulty washing or blow drying hair	**0.846**	0.005	-0.088
DASH 14	Difficulty washing your back	**0.675**	0.168	-0.135
DASH 15	Difficulty putting on a pullover sweater	**0.747**	-0.049	0.024
DASH 16	Difficulty using a knife to cut food	**0.800**	0.014	-0.014
DASH 17	Difficulty with recreational activities that require little effort	**0.667**	0.040	0.024
DASH 18	Difficulty with recreational activities in which force is taken through the limb	**0.607**	0.003	0.228
DASH 19	Difficulty with recreational activities in which the arm moves freely	**0.566**	0.004	0.240
DASH 20	Difficulty managing transportation needs	**0.682**	-0.067	0.070
DASH 21	Difficulty with sexual activities	**0.394**	-0.027	0.136
DASH 22	To what extent has your problem interfered with normal social activities?	**0.375**	0.145	**0.301**
DASH 23	To what extent has your problem interfered with work or other daily activities?	**0.464**	0.119	**0.308**
DASH 24	Rate your arm, shoulder or hand pain	0.030	0.073	**0.789**
DASH 25	Rate your arm, shoulder or hand pain when performing a specific activity	0.022	0.135	**0.776**
DASH 26	Rate the tingling in the arm, shoulder or hand	0.066	0.057	**0.420**
DASH 27	Rate the weakness in your arm, shoulder or hand	0.213	0.000	**0.642**
DASH 28	Rate the stiffness in your arm, shoulder	0.091	0.149	**0.621**

	or hand			
DASH 29	How much difficulty have you had sleeping because of pain in the limb?	0.323	-0.063	**0.413**
DASH 30	Is this true: I feel less capable, confident or useful because of the limb problem?	0.125	**0.304**	**0.494**

Table 7.6: Pattern matrix of exploratory factor analysis of log-transformed DASH and URAM items, with three-factor solution and oblimin rotation

** Loadings assess the correlation between the item concerned and the latent factor extracted in the analysis. They may range from -1 (perfect inverse correlation) through 0 (no correlation) to +1 (perfect correlation).*

N.B. Large factor loadings (>0.3) are shown in bold.

DASH items loaded on Factor 1 and Factor 3, in a similar pattern to that observed in the earlier EFA of DASH items alone. Eight of the nine URAM items loaded on a distinct factor (Factor 2). Item 9 within the URAM (difficulty picking up small objects with the thumb and index finger loaded separately from the other URAM items, suggesting that the URAM may not be unidimensional either.

7.5 Discussion

7.5.1 Present study findings

This study assessed the content and construct validity and reliability of the DASH and the URAM in Dupuytren's disease, and the QuickDASH's criterion validity against the DASH.

PROM items covering pain have been previously criticised in Dupuytren's disease, but our patients did report preoperative upper limb pain. This may have been due to other comorbid upper limb conditions rather than Dupuytren's disease, but even still, this would affect the overall function of their upper limb. Furthermore, pain levels rose early after surgery, and then decreased after surgery. This is important to capture, as postoperative pain may differ between treatments, and affect early recovery to different extents. These data support the relevance of assessing pain when treating Dupuytren's disease, and therefore suggest that tools that do not measure pain are not comprehensive.

The DASH was not unidimensional, with function items loading better with one construct, and patient perception items loading with another. Some other PROMs are multidimensional; they assess distinct constructs, with the generation of separate subscale scores. Examples include the Michigan Hand Questionnaire, which has distinct subscales for function, pain, work and other constructs (Chung et al., 1998). However, the DASH is designed to generate a single summary score. As a result, its items were not necessarily selected to measure specific distinct constructs. As a result, the different

constructs identified here may not be easily interpreted. Although the items loading with factor 1 appear to be related in that they are function items, alternatively, they may load similarly as they might also all relate to global upper limb, including the shoulder. In contrast, whilst the items that load with factor 2 are related as patient perception items, they might also be answered specifically in terms of the patient's experience of Dupuytren's disease, rather than for the global upper limb, as is the case for items such as 'doing yard work'. These data suggest that the DASH's single summary score may not be appropriate in Dupuytren's disease. Although factor 2 accounted for much less variance than factor 1 in this study, it was still statistically significant based on three separate assessments (scree plot, Eigenvalue and parallel analysis). Nevertheless, as it accounts for such little variance, this finding does not completely preclude the use of the DASH and its summary score in Dupuytren's disease. However, its selection as an endpoint for future studies should be carefully considered.

Interpreting whether agreement is adequate or not is a clinical decision (Bland and Altman, 1986). Given that the minimum detectable change at the 95% confidence level (MDC_{95}) for the DASH is around 13, and that the MDC_{95} of the QuickDASH is around 16 (Kennedy et al., 2011), then the 95% limits of agreement seen here are of similar magnitude to both tools' abilities to detect true change, though these MDCs have not been specifically confirmed in Dupuytren's disease. Therefore, it is likely that the level of agreement seen

here would support the use of the QuickDASH as an alternative to the DASH if either were considered appropriate in Dupuytren's disease.

Our large sample size allowed meaningful subgroup analysis, which demonstrated that close correlation between the two tools was seen at preoperative, early postoperative, and late outcome time points.

The second EFA performed demonstrates that the DASH and the URAM may reflect different underlying constructs. EFA demonstrates the statistical relationship between different items in scales, but cannot explain what the factors extracted actually represent. It is possible that the activity limitation-based items of the DASH might reflect shoulder impairment, that the URAM items might reflect impairment from loss of extension, and that the symptom-related DASH items might reflect more generalised hand function impairment. Even if this were the case, it is possible that the URAM is still not ideal in terms of face validity. As discussed already, pain appears to be a relevant symptom in the early postoperative phase, and the URAM may not be comprehensive enough to reflect this adequately. It may be biased towards loss of extension in isolation. At gross examination, the first eight items of the URAM appear to reflect tasks affected by loss of finger extension. In contrast, item nine (difficulty picking up small objects with the thumb and index finger) does not. Instead, it appears to represent more generalised fine control of hand function. In the developers' own EFA, the URAM was reported as being unidimensional, though the original data is not provided (Beaudreuil et al., 2011). However, this might be an artefact of the analysis itself, which has

been clarified in the present study. In factor analysis, it is recommended that items are selected such that several items load each factor that is to be modelled (Floyd and Widaman, 1995). Failure to do this may weaken the model. In the developers' EFA, if only item nine loaded on a distinct factor, then that factor may not have been adequately modelled. As discussed in the methods, several different statistical methods can be employed in 'factor analysis'. These include principal component analysis as well as the principal axis factoring method employed here. Principal component analysis has advantages and may be mathematically simpler (Pallant, 2010). However, if only few items load on a factor, as is the case here, and was even more so in the developers' EFA (Beaudreuil et al., 2011), the principal component analysis may lead to spurious results (Floyd and Widaman, 1995). For this reason, principal axis factoring was used in the present project. The URAM's developers did not specify which technique was used in their original validation study (Beaudreuil et al., 2011).

However, the results of the EFA might also relate to logistical considerations. In particular, the DASH and URAM questionnaires were administered in their original forms. The DASH's layout was produced by the developer, and maintaining the original layout is a condition of its licence for use (Kennedy et al., 2011). In contrast, the URAM questionnaire used was that in the original paper (Beaudreuil et al., 2011). As a result, the two questionnaires had different styles, layouts, font types and font sizes. These variables may have

influenced the patient's completion of the different questionnaires, leading to them apparently loading separately.

Even if this were the case, it may not be important when considered pragmatically. Standard layouts of the two tools were used here, and it would be reasonable to expect the two tools to still reflect the same single construct in the present study. Furthermore, the fact that the one URAM item whose face validity suggested that it may not reflect the same construct as the other items (URAM item nine appears more in keeping with DASH task items than with the other URAM items), did load on the same factor as DASH task items suggests that the loading pattern is not due to layout and style of questionnaires, but instead due to the performance of the items studied.

7.5.2 Limitations

There are limitations to this study. The sample included some patients who provided more than one measurement. In such circumstances, specific approaches to assessing agreement have been proposed (Bland and Altman, 2007). However, multiple measurements over time (preoperative, 3 weeks postop, 6 weeks postop and 1 year) do not constitute replicate measurements. They are best considered as independent assessments, as the patient's functional status is expected to be different at each time point, due to treatment of disease and progressive recovery.

Logarithmic transformation was used in an attempt to normalise the positive skew and kurtosis encountered. Suggested methods for handling skewed data vary (Ferguson and Cox, 1993, Floyd and Widaman, 1995). However,

similar results were obtained when the EFAs were run using raw, untransformed data.

Here, as in previous studies (Gummesson et al., 2006, Niekel et al., 2009), the QuickDASH and DASH were calculated from a single set of responses, on the DASH questionnaire's proforma. Intra-observer reproducibility, or test-retest reliability, is not observed when using a single set of responses. However, the methodology was the most appropriate to fulfil the specific objective of this study. The objective of this aspect of the study was to determine whether the QuickDASH formula demonstrates acceptable criterion validity with the longer DASH formula for a given set of responses, in keeping with the COSMIN checklist (Mokkink et al., 2010).

Test-retest reliability has been shown to be consistently high in previous studies, as summarised in the user manual for the DASH and the QuickDASH (Kennedy et al., 2011). The present study aims to establish whether for a given set of responses, the QuickDASH formula delivers the same summary score as the DASH formula. If patients had completed the DASH and then the QuickDASH, then there would have been additional error due to intra-observer reproducibility. This would have affected the ability to answer the study question.

7.5.3 Relationship to existing literature

Although the DASH is the most commonly used Patient Reported Outcome Measure (PROM) tool in Dupuytren's research (Ball et al., 2013), Dupuytren's-specific outcome measures are available, for example the URAM scale

(Beaudreuil et al., 2011) and the Southampton Dupuytren's score (Mohan et al., 2014). Exploratory factor analysis was incorporated into the design process of the URAM to ensure that it was unidimensional (Beaudreuil et al., 2011). Preoperative Dupuytren's disease is considered painless, and pain is not assessed in the URAM (Beaudreuil et al., 2011). However, pain has been described as a symptom, particularly related to Dupuytren's nodules (Hueston, 1963, Rodrigues et al., 2014, von Campe et al., 2012). These data demonstrated that pain was relevant, in that it was present preoperatively, increased at 3 weeks postop compared to baseline, returned to the preoperative level by 6 weeks postop, but did not disappear completely. Omitting pain from a tool for use before and after treatment for Dupuytren's disease may mean that the scale is not comprehensive.

The QuickDASH was produced from the DASH using item reduction methodology (Beaton et al., 2005). The two showed good correlation in mixed cohorts of hand surgery patients (Gummesson et al., 2006, Niekel et al., 2009). However, correlation is not appropriate for studying agreement (Bland and Altman, 1986, Bland and Altman, 1990, Schuck, 2004). No study assessing agreement between the DASH and the QuickDASH in Dupuytren's disease specifically, with the technique recommended by Bland and Altman (Bland and Altman, 1990), could be identified. Studying the strength of relationship between the two correlated measures, as has been done elsewhere, may conceal absolute differences in values between them. Such

differences are unmasked when agreement is studied using other techniques, as has been done here (Bland and Altman, 1986, Bland and Altman, 1990).

Poor correlation between angular deformity and the DASH has been previously reported in Dupuytren's disease (Degreef et al., 2009, Engstrand et al., 2009, Jerosch-Herold et al., 2011, Zyluk and Jagielski, 2007). This has led to the suggestion that the DASH may not be valid for use in Dupuytren's disease (Packham, 2011), though the basis for this claim was a series of only seven patient interviews (Pratt and Byrne, 2009). However, such conclusions are dependent on angular deformity being a 'gold standard' for assessing Dupuytren's disease, allowing the assessment of criterion validity. This is not appropriate, as no true 'gold standard' for hand function can exist (Mokkink et al., 2010).

Concerns about the absence of the expected unidimensionality for the DASH have been previously reported. A recent study investigated the structural validity of the DASH as the secondary analysis of a randomised controlled trial of splinting in Dupuytren's disease (Forget et al., 2014). In that article, EFA had been performed on preoperative and early postoperative DASH scores from 153 fasciectomies and dermofasciectomies randomised to receiving postoperative splinting or no postoperative splinting. In contrast, the present study evaluated preoperative, early postoperative and late postoperative outcomes of over 500 procedures, including needle aponeurotomies in addition to fasciectomies and dermofasciectomies. All procedures in the present study were performed under routine clinical conditions. Forget and

colleagues demonstrated that the DASH was not unidimensional in their EFA, though they proceeded to then perform Rasch analysis validation, which requires unidimensionality. The DASH did not perform well in Rasch analysis either (Forget et al., 2014). A range of items in the DASH did not load well with the equivalent Factor 1 in Forget and colleagues' EFA, which are difficult to interpret with clinical logic. In contrast, there was a clear division in our factor analysis between task items and patient symptom items. One other study reported a heterogeneous cohort comprising a range of upper limb conditions (mainly affecting the shoulder) and only a small minority with Dupuytren's disease (Franchignoni et al., 2010). In that study, the DASH items loaded on three distinct factors in EFA, and this was confirmed using other techniques classified as confirmatory factor analysis. Franchignoni and colleagues reported similar results to the present study, with patient perception items 22 to 30 loading separately from the main factor. However, Franchignoni and colleagues found that of the other items, those relating to manual function loaded distinctly from those assessing shoulder functions.

In the present study and in both of the previous studies, the further factors extracted accounted for relatively small proportions of the variance. This might mean that besides Factor 1 in all EFAs, the other factors are relatively insignificant. Indeed, this was the explanation given for proceeding to Rasch analysis in one study (Forget et al., 2014). However, the present study is now the third study to consistently demonstrate a lack of unidimensionality for the DASH. The URAM's apparent lack of unidimensionality when modelled along

with DASH items in the present study is a novel finding, and conflicts with the developers' original EFA (Beaudreuil et al., 2011). However, the present study findings fit with face validity examination of the URAM items. This should be further investigated.

This supports the possibility that Factor 1 in both of the EFAs in the present study is not a reliable indicator of hand function in relation to Dupuytren's disease. Our data further question the structural validity of the DASH in Dupuytren's disease, and by association, the QuickDASH. Due to this, it was not believed necessary or appropriate to subject the DASH to further analyses such as Rasch analysis. Instead, other PROMs that also assess pain may be more appropriate, such as the MHQ or the PEM.

7.5.4 Conclusions of study of validity

This study supports the assessment of pain when studying recovery from Dupuytren's surgery. As a result, the URAM may not be comprehensive for use after treatment. The QuickDASH show acceptable agreement with the full DASH. However, neither the current versions of the DASH nor the URAM may be structurally valid for use in Dupuytren's disease as they were both not unidimensional, and further study of existing PROMs for use in Dupuytren's disease is needed.

SUMMARY

The chapter demonstrates that the DASH is not structurally valid when used in Dupuytren's disease, in keeping with other studies. In particular, all of the items that comprise the DASH do not assess the same latent constructs, so it is not appropriate that they all contribute to a single summary score. As hypothesised, the QuickDASH exhibited acceptable agreement with the DASH and so is also likely to be unsuitable for future use.

The URAM's structural validity was demonstrated as part of its development (Beaudreuil et al., 2011). However, when the factor analysis model was expanded here, the final item in the URAM reflected a different construct from the others, which could be explained in terms of the URAM's face validity. This illustrates one of the limitations of studying validity, but is likely to be relevant when the findings are considered logically.

The URAM does not measure pain, yet here pain is identified as a relevant domain to consider during the early recovery from treatment for Dupuytren's disease.

As a result, in order to provide comprehensive measurement, the URAM needs further study and it is hypothesised from these novel findings that with expansion, a modified URAM might include to subscales: one that reflects symptoms related to the contracture, and one that reflects sequelae and complications of treatment like pain and stiffness.

Neither the DASH nor the URAM are ideal for use in their existing forms. However, based on the above interpretation and its more contemporary design, further work based using the existing URAM scale as the starting point is indicated.

As well as the validity of outcome measures, their interpretability may also affect their suitability for use. Assessments of interpretability for specific outcome measures are also necessary to interpret data from previous studies. Given that both the DASH and the URAM have been used previously (Ball et al., 2013), assessing their interpretability is a logical next step. The following chapters comprise a systematic review of interpretability studies in hand surgery to identify existing interpretability data followed by analysis cohort study data for the DASH and the URAM.

8 Systematic Review of Interpretability of Outcome Measures for Use in Dupuytren's Disease

8.1 Preface

An appropriate measure for future studies of treatments in Dupuytren's disease should be appropriately responsive, but should also have demonstrated interpretability. Not only is this an important parameter of the outcome measure, alongside validity and responsiveness, but also evidence of what constitutes a clinically meaningful change, or minimal important change (MIC), is necessary to design a suitably powered randomised controlled trial. However, as a situational metric, it may vary for different diagnoses, interventions and follow up periods, and can be calculated in distinct ways.

The previous chapter demonstrated that different outcome measures have been used in trials of the treatment of Dupuytren's disease. However, interpreting the methodological quality of these studies and their findings requires evidence of the interpretability of the measures used. As a result, interpretability data is of importance in clinical guideline development as well.

This chapter will review studies of interpretability for outcome measures that could be used to investigate treatments in Dupuytren's disease, and identify targets for future research.

8.2 Introduction

The change in an outcome measure after treatment can be statistically significant without being large enough to be considered worthwhile by patients. Indeed statistical significance following treatment is influenced by variables like the sample size of the study (Dawson et al., 2008). Therefore, the interpretation of clinical data, in either research or clinical practice, is critical.

8.2.1 Terminology

The minimum detectable change (MDC), or smallest detectable change (SDC), is the smallest change that can be reliably detected by an outcome measure. Whilst the MDC describes the responsiveness of an outcome measure, it may not be a clinically important change. Instead, what constitutes a change in a measure that is relevant to a patient should be appreciated (de Vet et al., 2006). The smallest change in an outcome measure that is clinically meaningful, and not simply statistically significant, has been referred to as the 'minimal clinically important difference' (MCID) or 'minimal clinically important change' (MCIC) (Copay et al., 2007). 'Difference' is used to describe the inter-individual difference that would be important (i.e. the difference between a patient who has undergone successful treatment compared to a patient who has not experienced benefit), whereas 'change' describes an important pre-treatment to post-treatment change for an individual (i.e. the smallest change that one individual would perceive as beneficial). As MCIDs and MCICs are patient-centred (Revicki et al., 2008),

the elimination of the first 'C' from MCIC, to 'minimal important change, or MIC, has been proposed (Schunemann and Guyatt, 2005). MICs describe a measure's interpretability, which is distinct from other measures of responsiveness. MICs for specific measures and for particular conditions can be used to inform power calculations during study design, and to ensure detection of clinically meaningful differences between treatments. Thus, having appropriate MICs for the outcome measures being used is important for designing future studies of treatment of Dupuytren's disease.

8.2.2 MICs may be context-specific

MICs can be calculated for objective measures, such as angular deformity or grip strength, as well as for patient-reported outcome measures (PROMs) (Katz et al., 1994, Waljee and Chung, 2012, Witthaut et al., 2011). However, as discussed, MICs may vary between conditions, patient groups, treatment types, outcome measures used, and length of time between treatment and outcome assessment (Ring, 2013). Indeed comorbidities may even influence MICs, despite the same treatment and follow up for the same condition. For example, MICs for carpal tunnel release have been shown to differ depending on whether or not a patient is diabetic (Ozer et al., 2013). Therefore, the context in which an MIC was derived should be considered before it is used for other studies, and ideally MICs that have been calculated specifically for the circumstances concerned should be used. MICs may be calculated by different methods; hence MICs are often reported as estimated ranges, rather than discrete values (Revicki et al., 2008).

8.2.3 Calculating MICs

Methods that can be used to determine MICs include anchor-based techniques and distributional analyses (Copay et al., 2007).

8.2.3.1 Anchor-based methods

These use an external criterion, which is a separate measure, to separate those who are 'meaningfully better' from those who are not. A common external criterion is the Global Rating of Change scale (GRC), which comprises a 15-point scale that ranges from -7 ('worse') to +7 ('better') (Jaeschke et al., 1989). Patients complete the GRC in addition to the outcome measure being studied. 'Meaningful improvement' is defined as +4 to +7 on the GRC scale, and 'stable' as -3 to +3, based on its original use (Jaeschke et al., 1989). As a retrospective anchor, patients complete the GRC after treatment, and must recall whether they have improved from baseline. Prospective anchors also exist. In such cases, the anchor is completed before and after treatment, and the change in the anchor is the external criterion.

Once outcomes have been sub-grouped into improved, stable and deteriorated, MICs can be estimated in different ways. The average change in the improved group is sometimes considered to represent the MIC, and may be compared statistically to the stable group. Alternatively, the difference between the change in the improved group and the change in the stable group may is reported as the MIC (Copay et al., 2007). Alternatively, the ability of the outcome measure to separate 'improved' patients from 'stable' patients

can be assessed with a receiver operating characteristic (ROC) curve. This considers the outcome measure as if it were a binary diagnostic test, and evaluates different cut offs on the quantitative scale for their ability to separate improved from stable outcomes. Using a ROC curve analysis, the MIC is the change in the outcome measure that is most sensitive and specific for identifying meaningful improvement (defined using the anchor measure) (Deyo and Centor, 1986).

The area under the curve (AUC) in ROC curve analysis varies from 0.5 to 1.0. When the AUC is 0.5, then measure being assessed has 50% sensitivity and 50% specificity at all cut offs, i.e. it is no better at determining whether a patient has had a meaningful improvement than a random guess. In contrast, when the AUC is 1.0, then the measure has 100% sensitivity and specificity at all cut-offs. The generation of such AUC data as part of ROC curve analysis provides further meaningful information to guide the selection of outcome measures for future studies.

An example of the anchor-based method is the calculation of the MIC for the treatment of shoulder pain with physiotherapy, as assessed by the shortened Disabilities of the Arm, Shoulder and Hand (QuickDASH) PROM (Mintken et al., 2009). Patients completed the QuickDASH before and 2-4 weeks after treatment. They also completed the GRC at follow up. Patients were grouped according to their GRC score into those who were ‘stable’ (GRC of -3 to +3) and those who were ‘better’ (GRC of +4 or more). None rated the GRC worse than -3. A ROC curve was generated to study the ability of the QuickDASH to

separate ‘stable’ outcomes from ‘better’ outcomes, and the MIC was estimated to correspond to an improvement of 8 points (out of the 100 points in the QuickDASH summary score scale), as this corresponded to point at the upper left hand corner of the curve, which represents the best combination of sensitivity and specificity at identifying meaningful change.

8.2.3.2 Distributional analysis methods

Distributional analyses relate the change in the score of an outcome measure to a parameter of its variability, to generate assessments that are used as the MIC, for example, the effect size (ES), standard error of the mean (SEM), or half of its standard deviation (SD) (Copay et al., 2007). As a result, they use properties derived from the outcome measure’s distribution to estimate the MIC. Each of the above metrics has been considered to represent the MIC (Copay et al., 2007). To do so, patients only complete the study outcome measure, and these ‘MIC’ variables are then calculated without the need for any other data.

An example of the use of distributional analyses is the estimation of the MIC for physiotherapy as a treatment for soft tissue shoulder disorders, assessed with the DASH PROM, at 12 weeks or at the time of discharge from therapy (Beaton et al., 2011). Several distributional analyses were used in this estimation, including 0.2 SD, 0.5 SD and the SEM for their dataset. The use of several different analyses resulted in an estimated range for the MIC DASH of between 3.9 and 15 points (the DASH summary score range is between 0 and 100). This approach may have limitations. Different treatments for

different conditions will have different failure rates, and this rigid approach does not account for this.

8.3 Aims and objectives of the present review

This aim of this chapter was to identify relevant MIC values for the treatment of Dupuytren's disease. The objectives were to identify MICs for surgery for Dupuytren's disease (or for other elective hand surgery applications that could be extrapolated to Dupuytren's disease) for commonly used outcome measures, in particular the Disabilities of the Arm, Shoulder and Hand (DASH) (which is the most popular PROM in Dupuytren's disease research (Ball et al., 2013)), the Unité Rhumatologique des Affections de la Main (URAM), which is the first Dupuytren's disease-specific PROM to have been developed (Beaudreuil et al., 2011), and loss of extension. The methodologies used to calculate such MICs would also be appraised.

8.4 Methods

The methodological design was based on the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement (Moher et al., 2009).

8.4.1 Search strategy

A search strategy was developed to identify articles describing the calculation of the MIC in Dupuytren's disease, or in other aspects of elective hand or wrist surgery. As some measures (for example DASH PROM (Hudak et al., 1996)) can be used throughout the upper limb, the search strategy also included upper limb studies. The AMED (Allied and Complementary Medicine) (1985

to January 2014), Ovid MEDLINE (1946 to January Week 3 2014), and Embase Classic + Embase (1947 to 2014 Week 04) databases were searched in parallel on 29th January 2014. No search limitations were used. The search strategy identified studies with any of the following in their titles or abstracts: 'minimal important change', 'clinically important difference', 'clinically important change', 'minimal important difference', or 'mcid'. The results were then combined using the AND Boolean operator so that the results only included articles that had one or more of the following terms in their titles or abstracts: 'hand', 'finger', 'wrist', or 'upper limb'.

The Online First section of the Journal of Hand Surgery (European Volume) website was also manually searched for the period of November 2013 – May 2014. All clinical articles were reviewed and their reference lists and citing articles screened to identify further studies reporting relevant MICs.

8.4.2 Article Selection

After de-duplication of the electronic search results, the candidate applied a stepwise selection strategy. Only articles describing MICs of relevance to the elective treatment of Dupuytren's disease or of outcome measures that could be used in Dupuytren's disease were included. For example, a study of the MIC for the DASH would be included even if the patient population involved had shoulder pathology, as this MIC might be extrapolated to the use of the DASH in Dupuytren's disease. In contrast, a study of a shoulder-specific measure, such as the Shoulder Pain and Disability Index (SPADI), would be excluded, as this outcome measure cannot be used in Dupuytren's disease.

Similarly, studies of acute trauma were excluded, as preoperative state cannot be measured as the baseline for describing change, unlike chronic conditions that are treated electively, such as Dupuytren's disease. Studies assessing multiple anatomical sites in addition to the upper limb were excluded, for example cerebrovascular accident rehabilitation studies, as they tended to employ outcome measures pertinent to neurology rather than hand surgery (Bindra et al., 2003). The latter selection stages involved screening full text articles. Reference lists from the full text papers that were screened, and articles citing them, were screened to identify further studies for inclusion. A second author verified the classification of search results.

8.4.3 Data Extraction

Data describing the conditions and treatments, outcome measures, follow up time point(s) studied, MICs obtained, and the methodologies used to generate the MIC were extracted for analysis.

8.5 Results

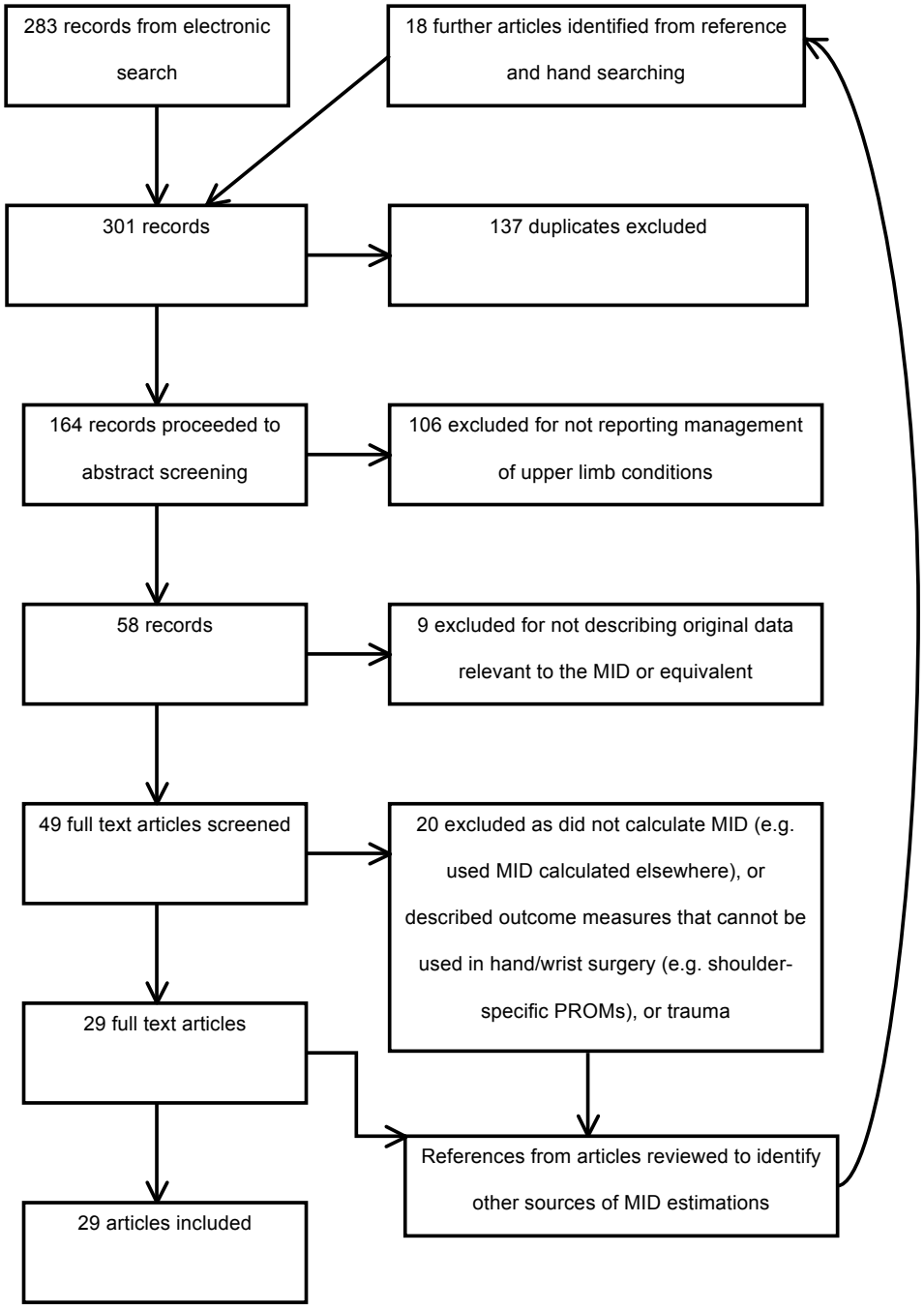

Figure 8.1: PRISMA-style flow diagram of article selection

8.5.1 Search results

The electronic search retrieved 283 records. A further 18 articles were later identified by the searches of reference lists. The selection process retained 29 articles that were analysed (Figure 8.1) (Adams et al., 2012, Amirfeyz et al., 2009, Atroshi et al., 2011, Beaton et al., 2001, Beaton et al., 2011, Beaudreuil et al., 2011, Bessette et al., 1998, Dawson et al., 2008, Franchignoni et al., 2014, Gummesson et al., 2003, Katz et al., 1994, Kim and Jeon, 2013, Kim and Park, 2013, Levine et al., 1993, London et al., 2014, Malay and Chung, 2013, Mintken et al., 2009, Ozer et al., 2013, Ozyurekoglu et al., 2006, Polson et al., 2010, Poltawski and Watson, 2011, Schmitt and Di Fabio, 2004, Shauver and Chung, 2009, Sorensen et al., 2013, Spies-Dorgelo et al., 2006, Tashjian et al., 2009, van der Giesen et al., 2008, Waljee and Chung, 2012, Witthaut et al., 2011). During the manual search, one other MIC was identified in a reference from a review article (Carswell et al., 2004). This was an MIC of 2 points for The Canadian Occupational Performance Measure (COPM), but was in a textbook rather than a peer-reviewed journal, therefore it was not included in the analysis.

8.5.2 Article characteristics

Twenty-eight of the 29 were full-length articles, and the other was a published conference proceeding abstract. Twenty-six were prospective cohort studies, two were secondary analyses from prospective randomised controlled trials, and the remaining article was a retrospective study. Most were published after 2000, with the majority published in the last 5 years (2009-2014).

Terminology varied, with 17 of 29 articles using the term ‘MCID’, and two using ‘MIC’. Two more used ‘minimal important change’, and two used ‘clinically important difference’. The remaining six used no specific terminology to describe their reporting of an analysis consistent with an MIC. These articles were typically only identified from screening lists of references from other papers and from full article screening by the review group, and provided narrative descriptions, for example: “We analyzed all measures of objective hand function variables as continuous variables to determine the cut-off point that corresponds with the presence or absence of patient satisfaction, and created ROC curves for each variable against patient satisfaction scores” (Waljee and Chung, 2012).

Several articles reported more than one MIC. Articles with multiple MICs either studied several conditions, outcome measures, comorbidities, or follow up time points. In total, there were 99 MICs in the 29 included articles (Table 8.1).

Condition	Source	Treatment(s)	Outcome Measure	Measure Scale Min	Score Max	Assessment Time point (days)	MIC	Method used
Carpal Tunnel Syndrome								
	Amirfeyz 2009	CTD	CTQ (Kamath modification)	0	12	42	1.97	B
	Kim & Jeon 2013	CTD	CTQ (total)	1	5	90	0.92	A, B
	Bessette 1998	CTD	CTQ (unweighted modified)	1	5	180	0.74	A
	Bessette 1998	CTD	CTQ (weighted modified)	1	5	180	0.79	A
	Amirfeyz 2009	CTD	CTQ FSS	1	5	42	0.47	B
	Kim & Jeon 2013	CTD	CTQ FSS	1	5	90	0.74	A, B
	Levine 1993	CTD	CTQ FSS	1	5	425	1.16	C
	Ozer 2013	CTD (in diabetics)	CTQ FSS	1	5	90	1.95	B
	Ozer 2013	CTD (in diabetics)	CTQ FSS	1	5	180	2.05	B
	Ozer 2013	CTD (in non-diabetics)	CTQ FSS	1	5	90	1.25	B
	Ozer 2013	CTD (in non-diabetics)	CTQ FSS	1	5	180	1.45	B
	Amirfeyz 2009	CTD	CTQ SSS	1	5	42	0.16	B
	Kim & Jeon 2013	CTD	CTQ SSS	1	5	90	1.14	A, B
	Levine 1993	CTD	CTQ SSS	1	5	425	1.84	C
	Ozer 2013	CTD (in diabetics)	CTQ SSS	1	5	90	1.45	B
	Ozer 2013	CTD (in diabetics)	CTQ SSS	1	5	180	1.55	B

Ozer 2013	CTD (in non diabetics)	CTQ SSS	1	5	90	0.8	B
Ozer 2013	CTD (in non-diabetics)	CTQ SSS	1	5	180	1.6	B
Amirfeyz 2009	CTD	DASH	0	100	42	20.9	B
Bessette 1998	CTD	PCS-12	0	100	180	4.76	A
Bessette 1998	CTD	PCS-36	0	100	180	2.01	A
Bessette 1998	CTD	quality of life rating	0	100	180	7.88	A
Bessette 1998	CTD	SF-36 body pain subscale	0	100	180	8.73	A
Bessette 1998	CTD	SF-36 physical functioning subscale	0	100	180	21.34	A
Shauver 2009	Limited Incision CTD	MHQ function	0	100	270	13	B
Shauver 2009	Limited Incision CTD	MHQ pain	0	100	270	23	B
Shauver 2009	Limited Incision CTD	MHQ work	0	100	270	8	B
Atroshi 2011	Open CTD	6-item CTS symptom scale	1	5	105	0.9	A
Katz 1994	Open Vs Endoscopic CTD	2 point discrimination sensation	n/a	n/a	90	1.6mm	C
Katz 1994	Open Vs Endoscopic CTD	ADL score	0	20	90	3.8	C
Katz 1994	Open Vs Endoscopic CTD	APB strength (MRC)	0	5	90	0.23	C
Katz 1994	Open Vs Endoscopic CTD	CTQ FSS (modified)	0	24	90	4	C
Katz 1994	Open Vs Endoscopic CTD	CTQ SSS (modified)	0	44	90	8.14	C
Katz 1994	Open Vs Endoscopic CTD	grip strength	n/a	n/a	90	5.47	C
Katz 1994	Open Vs Endoscopic CTD	pinch strength	n/a	n/a	90	1.66	C
Katz 1994	Open Vs Endoscopic CTD	Semmes Weinstein sensation	n/a	n/a	90	0.3	C

	Ozyurekoglu 2006	Steroid Injection	CTQ SSS	0	5	21	1.04	B, C
Dupuytren's Disease								
	Witthaut 2011	Collagenase Vs Placebo	range of motion	n/a	n/a	30	13.5°	F
	Beaudreuil 2011	Needle Aponeurotomy	URAM	0	45	30	2.9	C
Elbow Surgery								
	Dawson 2008	Unclear	DASH	0	100	180	≈10	B, C
Idiopathic Ulnar Impaction								
	Kim & Park 2013	Ulnar Shortening Osteotomy	DASH	0	100	365	13.8	B
	Kim & Park 2013	Ulnar Shortening Osteotomy	PRWE	0	100	365	17.3	B
Lateral Epicondyle Tendinopathy								
	Poltawski 2011	Novel Adjunct To Physiotherapy	Patient-rated tennis elbow evaluation	0	100	21	7	B
	Poltawski 2011	Novel Adjunct To Physiotherapy	Patient-rated tennis elbow evaluation	0	100	21	21	B
Mixed								
	Schmitt 2004	Physiotherapy	DASH	0	100	90	17.1	A
	Schmitt 2004	Physiotherapy	PCS-12	0	100	90	7.3	A
	Schmitt 2004	Physiotherapy	PRWE	0	100	90	24	A

	Adams 2012	Unclear	DASH	0	100	14, 28	11	A
	Beaton 2001	Unclear	DASH	0	100	84	15-20	B
	Adams 2012	Unclear	PRWE	0	100	14, 28	14	A
	Adams 2012	Unclear	quickDASH	0	100	14, 28	16	A
	Sorensen 2013	Various	DASH	0	100	14 or 28	10	A
	London 2014	Various	MHQ ADL	0	100	28 or 84	10.1	A, B, D
	London 2014	Various	MHQ aesthetic	0	100	28 or 84	*	N/A
	London 2014	Various	MHQ function	0	100	28 or 84	10.3	A, B, D
	London 2014	Various	MHQ overall	0	100	28 or 84	8.7	A, B, D
	London 2014	Various	MHQ pain	0	100	28 or 84	18.4	A, B, D
	London 2014	Various	MHQ satisfaction	0	100	28 or 84	33.0	A, B, D
	London 2014	Various	MHQ work	0	100	28 or 84	10.0	A, B, D
	Sorensen 2013	Various	PRWE	0	100	14 or 28	14	A
	Sorensen 2013	Various	quickDASH	0	100	14 or 28	14	A
Mixed Hand And Wrist Problems								
	Spies-Dorgelo 2006	Unclear	CTQ SSS	1	5	90	0.23	B
	Spies-Dorgelo 2006	Unclear	Dutch-AIMS2-HFF	0	10	90	0.31	B
Orthopaedic Surgical								
	Gummesson 2003	Unclear	DASH	0	100	180	10	A

Rheumatoid Arthritis Of Hand								
	Van Der Giesen 2008	Operative, Non Op, Both	MHQ total	0	100	90	11.3	A
	Waljee 2012	Silicone MCPJ Arthroplasty	extension lag	n/a	n/a	730	30°	B
	Waljee 2012	Silicone MCPJ Arthroplasty	MCPJ arc of motion	n/a	n/a	730	31°	B
	Shauver 2009 (193)	Silicone MCPJ Arthroplasty	MHQ ADL	0	100	365	11	B
	Shauver 2009 (193)	Silicone MCPJ Arthroplasty	MHQ function	0	100	365	13	B
	Shauver 2009 (193)	Silicone MCPJ Arthroplasty	MHQ pain	0	100	365	3	B
	Waljee 2012	Silicone MCPJ Arthroplasty	ulnar drift	n/a	n/a	730	9°	B
Rotator Cuff Disease								
	Tashjian 2009	Various	pain VAS	0	10	108	1.37	E
Shoulder Disorder								
	Schmitt 2004	Physiotherapy	DASH	0	100	90	10.2	A
	Schmitt 2004	Physiotherapy	PCS-12	0	100	90	6.5	A
Shoulder Pain								
	Mintken 2009	Physiotherapy	NPRS	0	10	14-28	1.1	B
	Mintken 2009	Physiotherapy	quickDASH	0	100	14-28	8	B
Soft Tissue Shoulder Disorder								
	Beaton 2011	Physiotherapy	DASH	0	100	84	3.9 - 15	C, D

Ulnar Nerve Entrapment							
Malay 2013	Simple Decompression	CTQ FSS	1	5	90	0.3	B
Malay 2013	Simple Decompression	CTQ FSS	1	5	180	0.3	B
Malay 2013	Simple Decompression	CTQ FSS	1	5	365	0.4	B
Malay 2013	Simple Decompression	CTQ SSS	1	5	90	0.4	B
Malay 2013	Simple Decompression	CTQ SSS	1	5	180	0.7	B
Malay 2013	Simple Decompression	CTQ SSS	1	5	365	0.7	B
Malay 2013	Simple Decompression	DASH	0	100	90	8	B
Malay 2013	Simple Decompression	DASH	0	100	180	7	B
Malay 2013	Simple Decompression	DASH	0	100	365	7	B
Malay 2013	Simple Decompression	MHQ ADL	0	100	90	6	B
Malay 2013	Simple Decompression	MHQ ADL	0	100	180	8	B
Malay 2013	Simple Decompression	MHQ ADL	0	100	365	6	B
Malay 2013	Simple Decompression	MHQ function	0	100	90	11	B
Malay 2013	Simple Decompression	MHQ function	0	100	180	11	B
Malay 2013	Simple Decompression	MHQ function	0	100	365	13	B
Malay 2013	Simple Decompression	MHQ pain	0	100	90	9	B
Malay 2013	Simple Decompression	MHQ pain	0	100	180	8	B
Malay 2013	Simple Decompression	MHQ pain	0	100	365	13	B

Upper Limb MSK Disorder								
	Franchignoni 2014	Physiotherapy	DASH	0	100	14-35	10.83	B, C, D
	Schmitt 2004	Physiotherapy	DASH	0	100	90	12.6	A
	Franchignoni 2014	Physiotherapy	quickDASH	0	100	14-35	15.91	B, C, D
	Polson 2010	Physiotherapy	quickDASH	0	100	42 or discharge	19	A
Upper Limb MSK Disorder Requiring Physiotherapy								
	Schmitt 2004	physiotherapy	PCS-12	0	100	90	6.8	A

Table 8.1: MICs grouped by condition, sub-grouped by treatment, and ordered by outcome measure and length of follow up

Key to abbreviations:

CTD – carpal tunnel decompression, CTS – carpal tunnel syndrome, ADL – activities of daily living, APB – abductor pollicis brevis, MRC – Medical Research Council, CTQ – carpal tunnel questionnaire, FSS – functional status scale, SSS – symptom severity scale, Dutch-AIMS2-HFF – hand and finger function subscale of the Dutch arthritis impact measurement scales, OA - osteoarthritis, MSK - musculoskeletal, MCPJ – metacarpophalangeal joint, MHQ – Michigan Hand Questionnaire, NPRS – numerical pain rating scale, VAS – visual analogue scale, PCS – physical component summary, PRWE – patient rated wrist evaluation, SF-36 – short form 36, URAM – Unité Rhumatologique des Affections de la Main scale.

** – MIC could not be calculated*

Key to methods used:

A – Change in 'minimally improved group', with comparison to other subgroups presented (anchor-based)

B – ROC curve (anchor-based)

C – Change in 'minimally improved group', without comparison to other subgroups (anchor-based)

D – Distributional data

E – Difference between ‘minimally improved’ and ‘stable’ groups (anchor-based)

F – Regression against anchor (anchor-based)

8.5.3 Outcome measures, conditions and time-points studied

There were only two MICs for Dupuytren's disease specifically, and both were calculated at 30 days' follow up.

One was for range of motion in a randomised controlled trial of collagenase versus placebo. Sub-groups for each treatment arm were not studied individually (Witthaut et al., 2011). The other estimated an MIC for the URAM scale following needle aponeurotomy, and used an unconventional method involving regression of the URAM score against an uncited anchor. No MICs were identified for the most common measure used (the DASH) in Dupuytren's disease, and none were identified for surgical treatments such as fasciectomy or dermofasciectomy.

MIC estimates for the DASH varied widely. The lowest DASH MIC was 3.9/100 for physiotherapy of soft tissue shoulder conditions (Beaton et al., 2011) whereas the highest was 20/100, for a study of mixed treatments of upper limb disorders (Beaton et al., 2001). Both extremes of DASH MIC came from studies with equivalent follow up periods (84 days). Even within a single study, a broad range of estimates for the MIC of the DASH were reported, varying from 3.9 to 15, depending on the method used (Beaton et al., 2011).

Thirty-two of the 99 MICs were calculated in heterogeneous cohorts of patients with different conditions. Heterogeneity was sometimes also seen in studies describing one condition. For example, rotator cuff disease was classified as one condition in the studies in the present review, though may involve different pathologies affecting the rotator cuff (i.e. rotator cuff tear,

subacromial bursitis and adhesive capsulitis (frozen shoulder)). Eighty MICs involved conditions affecting the elbow, forearm, wrist and hand, six described conditions proximal to the elbow, and the other 13 described outcomes for conditions both proximal and distal to the elbow. Sixty-five MICs were based on surgical treatments, 24 on non-operative treatment (most commonly physiotherapy), and eight involved a mixture of operative and non-operative treatments. Treatment modalities were unclear in the others. Twenty MICs were estimated in heterogeneous cohorts that underwent more than one type of treatment. All MICs involved the study of improvement following treatment, with none considering the MIC of later clinical deterioration in chronic conditions, such as recurrence after surgery for Dupuytren's disease.

Overall, MICs had been calculated for 40 different measures (Figure 8.2). Of these the DASH had the most MICs, with 15 different estimations. Eight of the 40 measures were variants of the Boston carpal tunnel questionnaire (Levine et al., 1993).

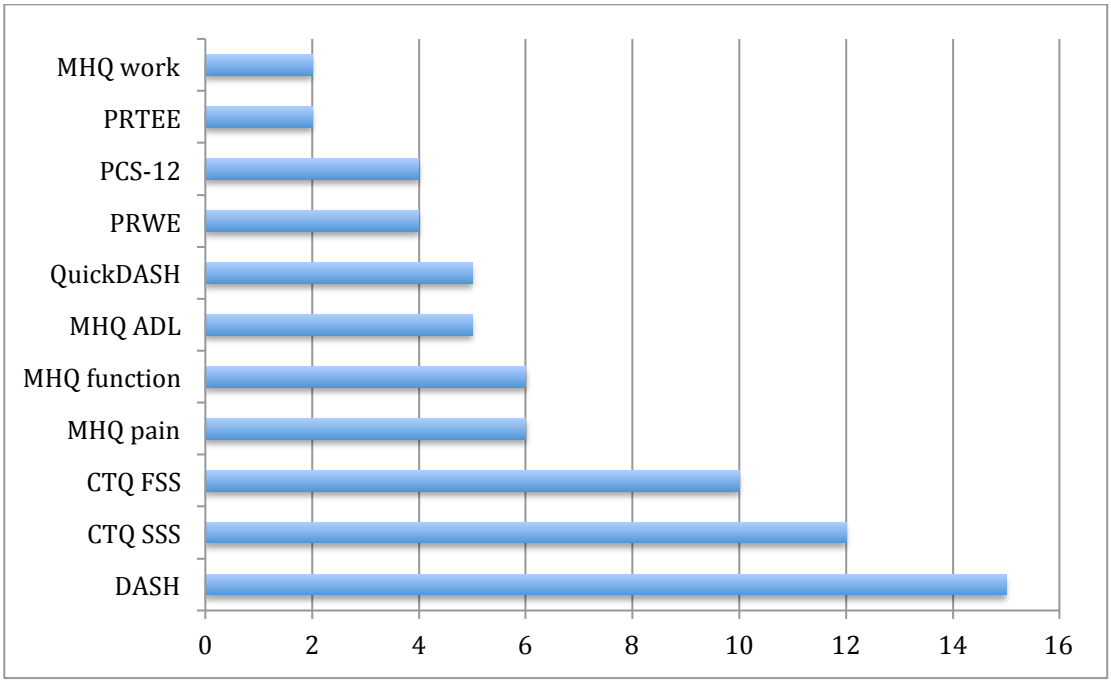

Figure 8.2: Numbers of MICs for different outcome measures

Only measures with more than one MIC are shown in this chart. A further 29 measures had one MIC each, and are shown in Table 8.1.

Key: CTQ – carpal tunnel questionnaire, FSS – functional status scale, SSS – symptom severity scale, MHQ – Michigan Hand Questionnaire, PRTEE – Patient-rated tennis elbow evaluation, PCS – physical component summary, ADL – activities of daily living, DASH – Disabilities of the Arm, Shoulder and Hand.

Amongst other hand surgery treatments, MICs were identified for open carpal tunnel decompression, limited incision carpal tunnel decompression, endoscopic carpal tunnel decompression, steroid injection for carpal tunnel syndrome, ulna-shortening osteotomy for idiopathic ulnar impaction, simple

decompression of ulnar nerve entrapment at the elbow (cubital tunnel syndrome) and silicone metacarpophalangeal joint arthroplasty.

Length of time between treatment and follow up assessment varied from 14 to 730 days, and several MICs were derived from studies involving assessments performed at inconsistent follow up time points.

One study investigated the MIC for the Jebsen Hand Function Test in a heterogeneous cohort (Sears and Chung, 2010). The area under the curve in ROC curve analysis was low (close to 0.5), indicating that the Jebsen Hand Function Test had failed to identify meaningful improvement. As a result, an MIC estimation could not be made.

8.5.4 Methodologies used

All 99 MICs were estimated using at least one anchor. For 41 of them, this involved a prospectively assessed anchor (e.g. several used the change in the satisfaction domain of the Michigan Hand Questionnaire as the anchor), and in the remainder it was a retrospective anchor (i.e. an assessment of improvement obtained during follow-up, after treatment had occurred). The anchors varied, and their origin was often unclear and not referenced.

Fifty-two of 99 MICs were in articles that also presented distributional analyses. For example, the MDC was frequently presented in articles in addition to the MIC. Nine MICs were based on both distributional and anchor-derived estimations.

Sixty-one MICs involved ROC curve analyses, and 49 presented the average changes in subgroups of patients classified as ‘minimally improved’. One MIC

was calculated from a regression of the measure (angular deformity) against the anchor (Witthaut et al., 2011). To do so, patient rated change that had been scored between 0% and 100% was split into four groups, and then this was used as the external criterion against which the change in angle was studied.

8.6 Discussion

8.6.1 Present study findings

This review identified only two MICs specifically for Dupuytren's disease treatment, and none for open surgery (fasciectomy and dermofasciectomy). The two identified may be of limited value. One studied the MIC of range of motion (Witthaut et al., 2011), which is a relatively uncommon objective outcome (Ball et al., 2013). The other used an unusual methodology, which was not used in any other study (Beaudreuil et al., 2011). Both MICs were calculated after 30 days of follow up. It is not clear whether this is the optimal time point to study recovery following collagenase or needle aponeurotomy. It is less likely that recovery from more invasive surgery such as fasciectomy or dermofasciectomy would be complete by this time. If these patients were assessed whilst still recovering, they may be less satisfied with their 'result' than if they were assessed once fully rehabilitated.

Other MICs for a range of clinical scenarios were identified. As the MIC is likely to vary between patient groups, conditions, treatments and time points, these MICs may be best interpreted as an approximation of meaningful improvement after treatment of Dupuytren's disease (Ring, 2013). However extrapolating them in such a fashion is not ideal.

The heterogeneity between the studies included here prevented comparison between MICs from different articles. However, considerable variability in the MIC for a particular measure was seen. For example, MIC estimates for the DASH varied widely. The lowest DASH MIC was 3.9/100 for physiotherapy of

soft tissue shoulder conditions (Beaton et al., 2011) whereas the highest was 20/100, for a study of mixed treatments of upper limb disorders (Beaton et al., 2001). Both extremes of DASH MIC came from studies with equivalent follow up periods (84 days). Even within a single study, a broad range of estimates for the MIC of the DASH were reported, varying from 3.9 to 15, depending on the method used (Beaton et al., 2011). The MICs for the DASH at different time points after decompression of ulnar nerve entrapment were presented in one study, and were similar when assessed using a prospective anchor (8 at 90 days, 7 at 180 days, and 7 at 1 year) (Malay and Chung, 2013). However, such stability may not be seen for all time points in all conditions. This variability emphasises the need for careful selection of an MIC when designing a study, or interpreting clinical outcome data.

We did not identify any studies aiming to define what constitutes a meaningful late deterioration after initially successful treatment (such as recurrence of Dupuytren's disease). Instead, all MICs identified described improvement after treatment, and are not applicable in defining a meaningful deterioration of on-going conditions (de Vet et al., 2006).

There is no consensus on which method should be used to calculate MICs (Bago et al., 2009). It has been suggested that several techniques be employed, and the MIC described as a range of values from these different techniques (Revicki et al., 2008). However, it is unlikely that all methodologies are equally robust. Anchor-based measures are dependent on the external criterion being valid and reliable itself. Whilst there has been support for the

use of retrospective anchors (Guyatt et al., 2002), concerns have been raised about their robustness (Norman et al., 1997), and the decision as to what indicates an "improved" score on scales such as the 15 point GRC is arbitrary (scores of +4 or more are improved) (Jaeschke et al., 1989). Retrospective anchors may reflect the status at the time of assessment, rather than the change in state over time that they are supposed to measure (Garrison and Cook, 2012, Schmitt and Di Fabio, 2005). Most MICs identified in this review were calculated using retrospective anchors, and further studies using prospective anchors may be warranted to confirm them. There may be other problems from some anchor-based methodologies. Although calculating the mean change in the 'minimal improved' group is an accepted technique for deriving the MIC, and was used in several of the articles included in this paper, there may be issues with the value obtained by this technique. If used without further analysis, it does not confirm whether the change in the 'minimally improved' group actually differs from the 'no change' group, and so whether this MIC is able to discriminate meaningful change from other outcomes. To tackle this, several of the studies did test for a statistically significant difference between the 'minimal improved' group and others. Alternatively, the difference between the 'minimally improved' group and the 'no change' group has been considered as the MIC.

Distributional analyses also have problems. They are affected by the variation amongst the subjects in a study, and it has been suggested that due to their nature, they may fail to describe 'meaningful' change at all (Copay et al.,

2007). Few of the MICs in this review involved the use of distributional data. This might reflect the input of clinicians in the design to the studies. However, given that the MIC might be best considered from a series of calculations using different methods (Revicki et al., 2008), perhaps more use of these analyses alongside anchor-based methods should be considered in future.

Besides such issues that are intrinsic to particular methodologies, there were other potential weaknesses in the study designs used to calculate several MICs reviewed here. Some did not have a fixed length of follow up amongst the study cohort, which is unsatisfactory as the MIC may depend on the length of time between the treatment and assessment; patients still recovering may be less satisfied than those who are fully rehabilitated. Also the risks associated with a retrospective anchor, which have already been discussed, might increase with longer follow up. This raises concerns about the robustness of some MICs, and further study to support the estimates of these MICs at specific time points may be required. Additionally, some MICs were calculated for mixed cohorts, comprising patients with different conditions, or patients who had received different treatments for the same condition. MICs generated in this way may be of less value when applied to changes in specific conditions or following particular treatments. They will be influenced by the case-mix of the cohort study, and the resulting 'average' MIC will not be applicable to any particular individual patient, condition or treatment. Some groups that studied multiple conditions or time points reported separate MICs for each clinical situation (Malay et al., 2013), which provides more applicable

data. This illustrates some of the issues already discussed: the MIC for the Carpal Tunnel Questionnaire Symptom Severity Scale (CTQ SSS)

Some articles screened did not present MICs, but instead reported other assessments of responsiveness. Others avoided calculating MICs (Chatterjee and Price, 2009). In one case, ROC curve analysis demonstrated that the Jebsen Hand Function Test was too poor at identifying improvement to allow the estimation of an MIC (Sears and Chung, 2010). This conclusion was based on the Area Under the Curve (AUC) of the ROC curve being lower than the typically acceptable cut-off of 0.75 (Fan et al., 2006).

8.6.2 Limitations

There are some limitations to this review. Steps were taken to ensure that the search strategy was comprehensive, but studies containing relevant data may have been missed. This is a particular risk due to the variation in terminology used between studies. By extending the search to include the screening of reference lists, it is believed that all articles of relevance to Dupuytren's disease have been included.

8.7 Conclusions of Systematic Review of Interpretability

No studies of interpretability of outcome after open surgery for Dupuytren's disease specifically could be identified. Two MICs had been derived in patients with Dupuytren's disease, one for the URAM 30 days after needle aponeurotomy and the other for range of motion in a trial comparing collagenase to placebo. Several MICs did exist for the DASH, but were derived for different conditions and treatments, and showed considerable

variation. Given the nature of MICs, extrapolating them from other circumstances to open surgery for Dupuytren's disease is not appropriate. Further study of interpretability of outcome measures in Dupuytren's disease would inform future study design and guide interpretation of data.

The Consensus-based Standards for the Selection of Health Status Measurement Instruments (COSMIN) collaboration separates validity of single time point measurements, from responsiveness and from interpretability (Mokkink et al., 2010), and the following chapters will consider aspects of the validity and reliability of different candidate outcome measures that might be used in future research. As they have been used in a range of existing studies, the investigations in the following chapters will also inform appraisal of the evidence base, as is required in the development of clinical guidance.

SUMMARY

This chapter demonstrates that whilst MICs have been estimated for use in a range of hand surgery clinical scenarios, few are applicable to surgery for Dupuytren's disease, and even those that are relevant to hand surgery are not suitable for use following open surgery, or at a follow up time point at which recovery after surgery is complete. As a result, there is not currently enough interpretability information available to be able to design an appropriately powered study of different treatment options for Dupuytren's disease. Furthermore, interpretation of findings of existing studies, which have been summarised in the previous systematic review chapter, is also limited by this paucity of evidence. Further examination of what constitutes an important change in outcome measures following treatment of Dupuytren's disease is still required. The next chapter will present cohort study data to investigate the interpretability of the DASH and the URAM.

9 Recovery, Responsiveness And Interpretability Of Patient-Reported Outcome Measures After Surgery For Dupuytren's Disease

9.1 Preface

The preceding chapters have studied aspects of the reliability and validity of three candidate outcome measures for future studies of Dupuytren's disease, namely angular passive extension deficit, the Disabilities of the Arm, Shoulder and Hand (DASH) patient-reported outcome measure (PROM), and the Unité Rhumatologique des Affections de la Main scale (URAM). The results obtained thus far demonstrate issues with the performance of each of these measures. However, they have all been used to study Dupuytren's disease to date, and further investigation of them is warranted to guide appropriate interpretation of papers that have used them.

Responsiveness and interpretability are other parameters of relevance to the performance of outcome measures, besides validity and reliability. Interpretability can also determine meaningful change, but there are few applicable data available for the measures concerned, as discussed in the Systematic Review of Interpretability of Outcome Measures for Use in Dupuytrens' Disease. Furthermore, such analyses are dependent of appropriate timing of follow up assessment, such that recovery is complete. Knowledge of what constitutes a clinically meaningful change is required when

interpreting data for clinical practice, and for guideline development. These parameters will be investigated in this chapter.

9.2 Introduction

Evaluating the outcome of Dupuytren's disease treatment requires an appreciation of what constitutes a clinically important change following treatment and consideration of the timing of assessment in relation to recovery, assuming that the outcome measure being used is appropriate.

The Consensus-based standards for the selection of health status measurement instruments (COSMIN) study defined validity in terms of the appropriateness of a single time point measurement (Mokkink et al., 2010), for example, whether the score achieved on a preoperative hand function questionnaire reflects hand function at that time. The responsiveness is the validity of a change in score based on a change in the construct being measured, which in the case of Dupuytren's disease is often hand function. In contrast, interpretability is concerned with interpreting a change score, often via the minimal important change (MIC) (Mokkink et al., 2010). The MIC is that smallest change in a score following treatment that is considered clinically important by patients. This can be calculated in different ways and is likely to be context-specific, so it may vary between different conditions and treatments (Revicki et al., 2008). A range of MIC estimations exists for hand surgery, and include values for PROMs, but also for objective measures, such as angular deformity or grip strength (Rodrigues et al., 2014). However, in the cited review, the only estimation of an MIC for a PROM in Dupuytren's disease was

for the URAM Dupuytren's disease-specific scale, and was calculated for needle aponeurotomy (Beaudreuil et al., 2011). It is possible that the MIC after open surgery (fasciectomy or dermofasciectomy) may differ from this. The DASH PROM had the most MIC estimations across all hand surgery conditions (Rodrigues et al., 2014), and has been the most widely used PROM in Dupuytren's disease (Ball et al., 2013). However, its MIC in Dupuytren's disease has not been estimated.

Six weeks following treatment has been used as the time point when early outcome of surgery for Dupuytren's disease is measured (van Rijssen et al., 2006). However, other studies suggest that recovery, at least from open surgery such as fasciectomy and dermofasciectomy, may continue for longer than this (Ullah et al., 2009). Measuring outcome too early might underestimate the benefit of treatment in general, or may bias comparisons towards less invasive treatments that might have quicker recovery, such as aponeurotomy.

A range of outcome measures has been used to study Dupuytren's disease (Ball et al., 2013), and can be broadly grouped into generic, domain-specific and disease-specific measures (Szabo, 2001). Recently, Dupuytren's disease-specific measures have been developed (Beaudreuil et al., 2011, Mohan et al., 2014), and the suitability of the most popular outcome measure, the upper limb domain-specific DASH, has been questioned (Packham, 2011).

This present chapter describes a prospective cohort study that investigated recovery time from open surgery for Dupuytren's disease, and the

responsiveness and interpretability of the DASH and URAM PROMs following fasciectomy and dermofasciectomy, and potential explanations for differences in their performance.

9.3 Methods

The data presented in this study were gathered as part of a larger service evaluation comprising a cohort study and a cross sectional study.

Data from the two studies were used for distinct purposes. Data from the cohort study were used in analyses of recovery, responsiveness and interpretability. Data from the cross sectional study were used in addition to data from the cohort study in analysis of construct validity, which in this study was the exploratory factor analysis.

9.3.1 Patient recruitment and data collection

Patient recruitment took place between September 2011 and April 2013. The inclusion criteria were primary or recurrent Dupuytren's disease and either:

- Patients awaiting fasciectomy or dermofasciectomy at one UK hand surgery centre (for the cohort study)
- Patients available for assessment at five UK hand surgery centres 1 year or 5 years (+/- 2 months) after their surgery when the candidate was available (for the cross sectional study).

Exclusion criteria were:

- Cognitive impairment preventing informed consent
- Refusal of invitation to participate

In the cohort study, preoperative patients were recruited at the routine preadmission clinic visit prior to surgery. Those who were eligible and consented to participate completed the DASH prior to surgery. These patients were also sent questionnaires for completion by post at 3 weeks, 6 weeks and 1 year after surgery. Patients who were scheduled for surgery to the left and right hand at different times during the study recruitment period were eligible for recruitment twice. This happened on four occasions. The URAM scale was published during the study period. Patients recruited later in the cohort (August 2012 onwards) also completed the URAM at preoperative, and postal questionnaires at 6 week and 1 year postoperative time points.

After completion of the 1-year follow up period, all patients were posted the Global Rating of Change questionnaire (GRC) (Jaeschke et al., 1989), either with their 1-year hand function questionnaires or separately.

In the cross-sectional study, patients were invited to participate on a voluntary basis with a letter explaining the project and a fixed stipend offered to cover travel expenses. The candidate assessed those who consented to participate. The assessment included collection of demographic data and completion of the DASH and, from August 2012 onwards, the URAM.

9.3.2 Handling of incomplete questionnaires

The DASH remains reliable as long as at least 27/30 items are complete (Kennedy et al., 2011). Therefore, all returned questionnaires meeting this criterion were included.

Two options were considered for handling missing data (when questionnaires were not returned or were returned with more than 3/30 items incomplete). In pairwise exclusion, incomplete data are omitted from specific analyses, but where possible, are included in other analyses. In listwise exclusion, that patient is excluded from all analysis. Pairwise exclusion was the preferred method for handling unreturned or more incomplete questionnaires. If required (e.g. for repeated measures ANOVA), then listwise exclusion was used. For example, if a patient submitted DASH questionnaires for preoperative assessment and 1-year postop state, but did not return the 3-week and 6-week questionnaires, then that patient was still included in interpretability analyses that only required preoperative and 1-year postoperative assessments. However, that patient would be excluded from the study of recovery, as this looks at patients' changes over all time points.

As clear guidance for handling incomplete questionnaires was not available for the URAM, all URAMs with any missing entries were excluded in the same fashion.

9.3.3 Data handling

The DASH summary score was calculated using the standard formula provided:

DASH=[(a/b)-1]*25

Where "a" is the sum of the scores for the responses completed (each response could be scored between one and five), and "b" is the number of responses the patient completed.

The URAM summary score was calculated by adding the responses to all items.

As the PROM summary scores are virtually continuous scales (the DASH is scored 0-100; the URAM 0-45), and the sample comprised a large number of independent observations, parametric analyses were used to compare them, in keeping with the central limit theorem. The central limit theorem demonstrates that datasets of 'large' sample sizes (typically over n=30) comprising independent observations of a continuous behave acceptably when subjected to parametric statistics, even in the absence of a normal distribution (Norman, 2010). An advantage of this approach is that parametric analyses typically retain more power and so pose less risk of a type II error.

9.3.4 Analysis of recovery

Recovery time was analysed by comparing DASH scores at different time points using repeated measures analysis of variance (ANOVA) with Tukey's multiple comparison test.

9.3.5 Analysis of responsiveness and interpretability

Responsiveness was studied by calculating the effect size, defined as the change in score divided by the standard deviation of the baseline (preoperative) scores across the cohort (Kazis et al., 1989). When interpreting the effect sizes, 0.2 was considered small, 0.6 moderate and 1.0 large (Testa, 1987).

Interpretability was studied using receiver operating characteristic (ROC) curves, as this has been the most common method used in hand surgery

(Rodrigues et al., 2014). ROC curves treat an outcome measure (such as the preoperative-postoperative change in DASH or URAM) as a diagnostic test, with the aim of reliably diagnosing clinical improvement from stable results (Deyo and Centor, 1986). To do this, those patients who are 'improved' are separated from those who are 'stable', using an external criterion, or anchor (Revicki et al., 2008). ROC curves then trial different possible cut values (e.g. improvement in DASH of 20/100 compared to 30/100 or 40/100). Each trial cut value will have its own combination of sensitivity and specificity for diagnosing improvement. The MIC can be considered to be Youden's index, which is the point with the highest combination of sensitivity and specificity for identifying improvement (Youden, 1950). To do this, outcomes were sub grouped based on the response to the GRC into those who were 'improved' (GRC+4 to +7), those who were 'stable' (GRC -3 to +3), and those who were 'worse' (GRC -7 to -4), as previously described (Mintken et al., 2009). ROC curves were generated using Prism 6.0 for Mac OS X (GraphPad® Software, La Jolla, USA, 2012).

9.4 Results

9.4.1 Patients and procedures

In the cohort study, 101 patients were recruited in total, of which 44 were sent URAM questionnaires in addition to DASH questionnaires. The demographics of the cohort are shown in Table 9.1. Sixty eight per cent completed all follow up function PROMs.

Detail	DASH	URAM
Age at recruitment (mean years, range)	67, 34-90	66, 38-90
Gender	83/101 male (82.2%)	38/44 male (86.4%)
GRC, preop, 6 week, 1 year PROMs completed*	68/101 (67.3%)	29/44 (65.9%)
Procedures	73 fasciectomies 28 dermofasciectomies	29 fasciectomies 15 dermofasciectomies
Hand treated	61/101 right (60.4%)	25/44 right (56.8%)
Multiple digits treated	27/101 (26.7%)	12/44 (27.3%)
Digits treated	135 digits in 101 cases 80/135 little (59.3%) 39/135 ring (28.9%) 10/135 middle (7.4%) 4/135 index (3.0%) 2/135 thumb (1.5%)	59 digits in 44 cases 34/59 little (57.6%) 15/59 ring (25.4%) 7/59 middle (11.9%) 2/59 index (3.4%) 1/59 thumb (1.7%)

Table 9.1: Patient demographics and procedure details in study

DASH- Disabilities of the Arm, Shoulder and Hand, URAM – Unité Rhumatologique des Affections de la Main, PROM – Patient-reported outcome measure, GRC – Global Rating of Change
** "Complete" DASH questionnaires defined as at least 27/30 items complete, such that a summary score can be calculated. Complete URAM questionnaires defined as all items completed.*

9.4.2 Recovery

Sixty-five patients completed preoperative, 3-week, 6-week and 1-year postoperative DASHs. The scores at different time points are shown in Figure 9.1. The mean DASH summary score was significantly different between time points ($p<0.0001$, repeated measures ANOVA). Tukey's multiple comparisons test demonstrated significant differences between all time points, with the exception of comparing preoperative to 6-week postoperative scores. Function deteriorated between preoperative and 3-week postoperative assessments. Notably, the difference between 6 weeks and 1 year was significant ($p=0.0006$). The developers of the DASH advise that a DASH summary score above 15/100 is consistent with a symptomatic upper limb, and the mean DASH summary score only fell below 15/100 at 1 year. Of note, there was no difference between scores for fasciectomies and scores for dermofasciectomies at any time point, hence results from both procedures were combined for all further analyses.

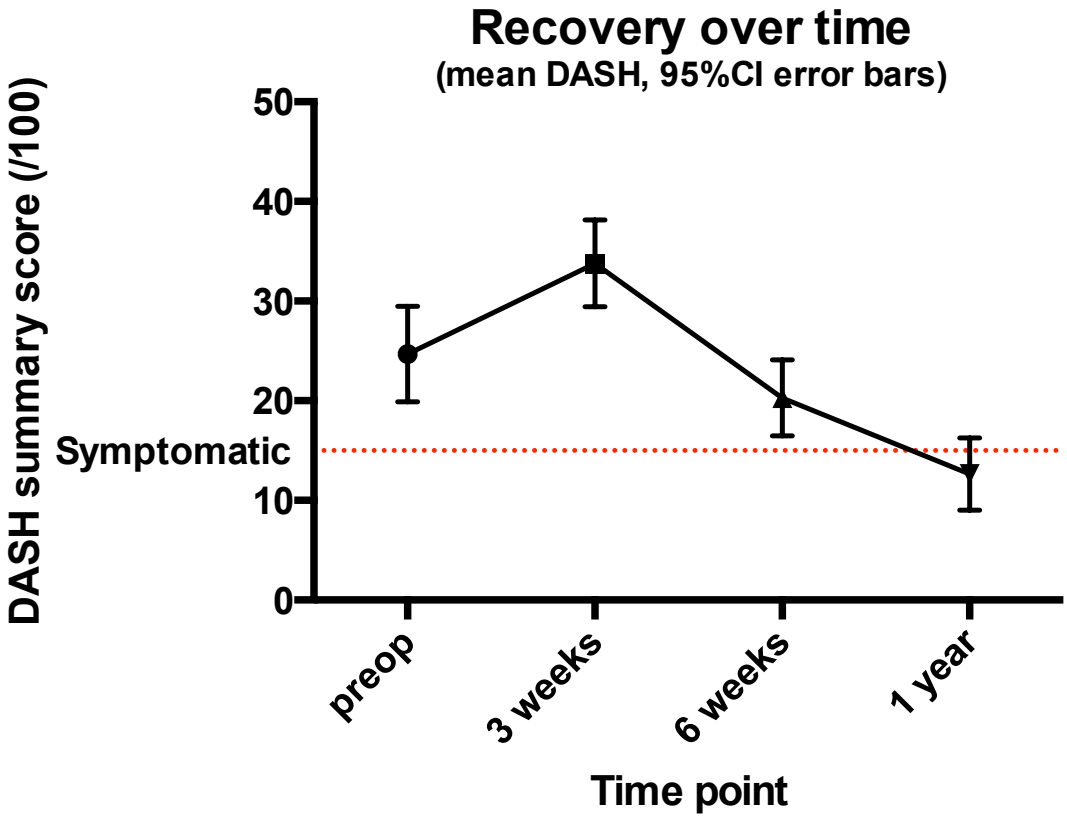

Figure 9.1: Line chart of DASH summary scores from patients who completed all time points (n=65)

95% CI – 95% confidence intervals

9.4.3 Responsiveness and interpretability

As functional state was significantly better at 1 year compared to 6 weeks, responsiveness and interpretability analyses were performed using change between preoperative and 1 year PROMs.

Responses from patients who adequately completed preoperative and 1-year postoperative questionnaires were included in responsiveness analysis. Both the DASH and URAM had significant changes in scores from preoperative

assessment to 1-year postoperative assessment, though the URAM had a larger effect size than the DASH (Table 9.2).

	DASH (n=71)	URAM (n=30)
Preoperative [mean (95% CIs)]	24.5/100 (19.9, 29.0)	17.8/45 (14.5, 21.1)
1-year postoperative [mean (95% CIs)]	12.4/100 (8.9, 16.0)	10.1/45 (6.1, 14.1)
Difference post-preop [mean (95% CIs)], paired t test	12.0/100 (8.2, 15.9) p<0.0001	7.7/45 (3.7, 11.7) p=0.0005
Effect size (mean/SD preop)	0.58	0.87

Table 9.2: Responsiveness of DASH and URAM at 1-year postoperative assessment

The mean GRC in the DASH cohort was +4.3 (95% confidence intervals [95% CIs]: +3.4, +5.2). When sub grouping of DASH outcomes using the GRC was performed, five patients were worse, eleven were stable and 52 were improved. The mean DASH change in the improved subgroup was 13.0/100, and the mean DASH change in the stable subgroup was 10.8/100. The difference between them (2.2/100 [95% CIs: -13.3, 8.9]) was not statistically

significant (p=0.69, unpaired t test). The ROC curve for the DASH is shown in Figure 9.2. The area under the curve (AUC) was 0.51 (95% CIs: 0.33, 0.69), indicating that the DASH could not identify meaningful change defined by the GRC. Consequently, an MIC could not be estimated for the DASH at 1 year after fasciectomy or dermofasciectomy.

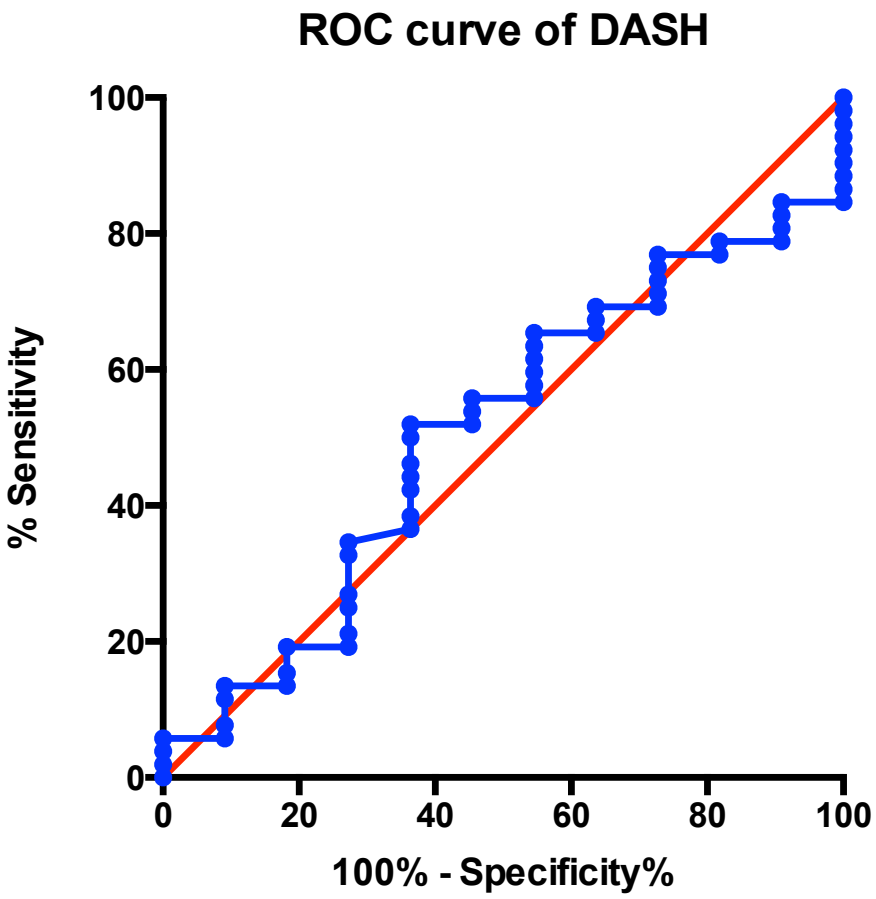

Figure 9.2: ROC curve of DASH's ability to separate 'improved' from 'stable' outcomes, based on the GRC

The red line indicates the line of identity, where sensitivity and specificity are both 50%, and corresponds to an area under the curve of 0.5. The blue points represent trial cut offs for change in DASH that could be used to attempt to separate 'improved' from 'stable' outcomes.

For the cohort that completed the URAM, the mean GRC was +2.9 (95% CIs: +1.2, +4.6). When GRC-based sub grouping was performed, 18 were improved, eight were stable and four were worse. The mean URAM change in the improved subgroup was 11.9/45 (95% CIs: 6.7, 17.0), and the mean URAM change in the stable subgroup was 3.6 (9.3, -2.0). The difference between them (8.3/45, 95% CIs: 0.04, 16.5) was significant (p=0.049, unpaired t test). A change of 8.3/45 on the URAM constitutes an MID (as opposed to an MIC), as it describes the difference between individuals considered clinically important.

The ROC curve determines the MIC, as the change that an individual patient considers significant. For the URAM, it is shown in Figure 9.3. The AUC was 0.74 (95% CIs: 0.55, 0.93). The MIC for the URAM for fasciectomy/dermofasciectomy at one year (defined as Youden's j index) corresponded to an improvement in the URAM of greater than 10.5, which had a sensitivity of 55.6% and a specificity of 87.5%. The likelihood ratio for an improvement in the URAM of 10.5 was 4.4; that is, a patient with a URAM improvement over 10.5 was 4.4 times more likely to be 'improved' than a patient whose URAM had improved by less than 10.5.

Across the cohort studied, the GRC correlated significantly with the 1-year DASH (Pearson's *r*: -0.48, p<0.0001), but did not correlate with the change is DASH (Pearson's *r*: -0.22, p=0.07). It also did not correlate with the preoperative DASH (Pearson's *r*: -0.18, p=0.15). For the URAM, the GRC correlated more closely with the 1-year assessment (Pearson's *r*: -0.68,

p<0.0001), than with the change in the URAM (Pearson's *r*: -0.56, p=0.001). It did not correlate with the preoperative URAM (Pearson's *r*: -0.15, p=0.44).

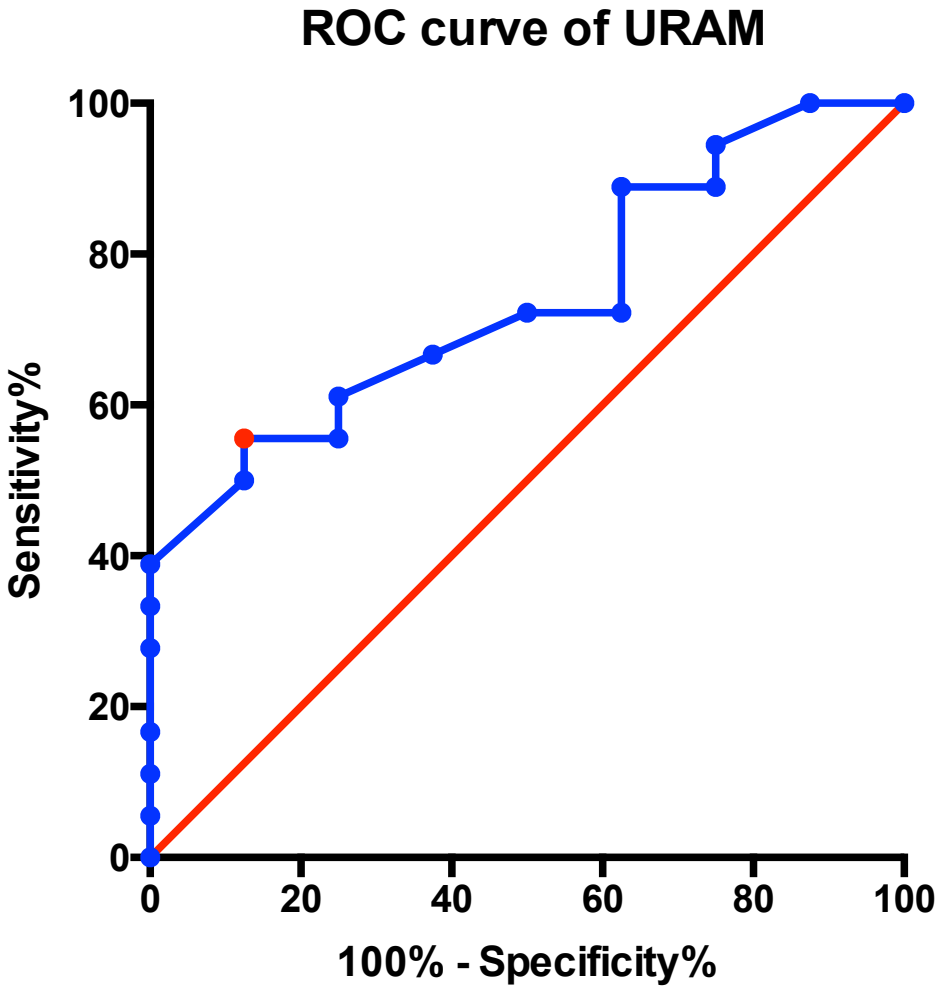

Figure 9.3: ROC curve of URAM's ability to separate 'improved' from 'stable' outcomes, based on the GRC

The red point on the blue line corresponds to the MIC described above. This corresponds to a cut point of an improvement of >10.5 in the URAM.

9.5 Discussion

9.5.1 Present study findings

The timing of early outcome assessment is likely to influence differences between treatment types in Dupuytren's disease. In this study, recovery following fasciectomy or dermofasciectomy took longer than 6 weeks. In contrast, less invasive procedures such as needle aponeurotomy and collagenase, which were not included in this cohort, are likely to recover quicker. In such a situation, functional outcome for more invasive procedures may appear poor, as patients are yet to realise the full benefit of the treatment. To minimise patient burden in the present study, only one further assessment was made, at 1 year after surgery. The statistically significant difference between 6-week outcome and 1 year outcome may also be clinically significant, as the 1-year time point was the only one with a mean summary score below 15/100, the threshold above which a patient's upper limb is considered symptomatic (Chart 1) (Kennedy et al., 2011). This is likely to affect the interpretability of outcome measures as well. Although it might be expected that incomplete recovery at 6 weeks would be accounted for in patients' responses to the anchor item, determining perceived change or satisfaction when patients are still recovering from treatment may not be as meaningful as assessing interpretability once patients have fully recovered. However, it is unclear from these data whether recovery is complete earlier than 1-year, or if it continues after this time point. Given the improvement

between 6 weeks and 1 year, subsequent analyses performed in this study used the 1-year assessment as the follow up time point.

Using 1-year follow up data, the DASH exhibited moderate responsiveness. The URAM's responsiveness was better, albeit in a slightly different and smaller cohort. However, commonly used responsiveness analyses, such as the effect size used here, are only appropriate if all patients studied have undergone a clinically meaningful improvement. If patients underwent a sham procedure, and received no benefit at all (including placebo), then it would be reasonable to expect that an appropriate outcome measure would exhibit no change from before to after the sham treatment, i.e. the effect size would be negligible.

The aim of the interpretability analyses was to separate those who had experienced meaningful improvement from those who had not, and then to calculate the MIC using this dichotomy. Indeed, a notable (16/68) proportion of patients did not experience benefit, or even experienced worsening, as defined by the GRC. The DASH could not identify this meaningful change, and an MIC could not be calculated. In contrast, the URAM showed acceptable interpretability, with an MIC for open surgery of 10.5.

The difference in performance between the DASH and the URAM may relate to the fact that they measure different underlying constructs, as demonstrated in the exploratory factor analysis in the previous chapter. EFA demonstrates the statistical relationship between different items in scales, but cannot describe the nature of the latent factors extracted. It is possible that the

activity limitation-based items of the DASH might reflect shoulder impairment, that most of the URAM items might reflect impairment from loss of extension, and that the symptom-related DASH items might reflect more generalised hand impairment. Thus, the analyses in the present chapter (responsiveness assessed by effect size, and interpretability described by the MIC and MID) in isolation cannot be claimed to demonstrate that the URAM is more appropriate than the DASH for the assessment of hand function in patients with Dupuytren's contractures. However, in conjunction with the studies of validity in the previous chapter, they further question the role of the DASH in future studies of Dupuytren's disease, despite the DASH's previous popularity.

9.5.2 Limitations

As 6 weeks and 1 year were the only time points studied, it is not possible to accurately determine an accurate recovery plateau time from the data here. Indeed, based on studies of other outcome measures (Ullah et al., 2009), it is possible that recovery may plateau close to 3 months after surgery, but the quality of their data is uncertain as it was only presented in graphical form, with no information provided about attrition.

The methodology used to study interpretability has limitations, which may be affected by the long follow up for assessing recovery. Although the GRC was developed for the purpose of anchoring outcomes, it is retrospectively administered, and the use of such retrospective anchors has been criticised (Norman et al., 1997). In particular, the GRC may reflect the status of the hand at the time of assessment, rather than reflecting the change that has

occurred from the preoperative state (Garrison and Cook, 2012, Schmitt and Di Fabio, 2005). In this study, the GRC correlated more closely with the 1-year postoperative state than it did with the change in score. The GRC was chosen here, as most studies of MICs in hand surgery have used retrospective anchors to date (Rodrigues et al., 2014). It is possible that the interpretability of the PROMs studied would be different if a prospective anchor were used. Prospective anchors capture data at baseline and again at follow up, rather than relying on patient recall of the baseline state, which is the case for retrospective anchors like the GRC. Some prospective anchors have been used in hand surgery. In particular, the satisfaction domain of the Michigan Hand Questionnaire has been used as a prospective anchor (Rodrigues et al., 2014). Whether satisfaction is the most appropriate domain to assess as an anchor is not clear, as it may be influenced by patient experience of health care system, rather than the treatment itself, and so results based on satisfaction may not be generalisable to other health care services, even if the treatment itself is similar. Further study of anchors is required. The present study findings should then be confirmed using such methodology.

The methodology for sub-grouping outcomes to construct the ROC charts was based on commonly used methodology from previous studies, with exclusion of patients who experienced deterioration. However, more recent work has led to the suggestion including the entire cohort may improve precision in MIC analyses (Turner et al., 2009). This might be considered in the future work

proposed above, in which interpretability would be studied with a prospective anchor.

The cohort studied may be heterogeneous, as it included both fasciectomies and dermofasciectomies. These data were collected as service evaluation of standard clinical practice, and on several instances, the type of procedure was changed from the preadmission clinic, based on surgeon preference, or for technical reasons during the surgery itself. It is possible that the interpretability of the DASH and the URAM may differ between fasciectomies and dermofasciectomies. However, given that the recovery was not different between procedure types, and that MICs are considered estimates rather than exact values, this is unlikely to be of significance.

9.5.3 Relationship to existing literature

Whilst early recovery from surgery for Dupuytren's disease has been assessed at 6 weeks (van Rijssen et al., 2006), and studies of other treatments have assessed outcome at 30 days (Hurst et al., 2009), the current data support the limited existing data demonstrating that recovery from open surgery takes longer than this (Ullah et al., 2009). Confirmation of the most appropriate time point for studying recovery is required.

Interpretability had been studied in Dupuytren's disease (Beaudreuil et al., 2011, Witthaut et al., 2011). Neither of these studies considered open surgery, though one did generate an MIC estimate of 2.7 for the URAM for needle aponeurotomy (Beaudreuil et al., 2011). The present study generated a considerably larger MIC, which may be due to the different methods used (a

ROC curve of dichotomised outcomes was used here, compared to regression against satisfaction in other studies). However, it is more likely to reflect the differences in recovery between open surgery and aponeurotomy. As open surgery involves a rise in postoperative as demonstrated in previous chapters and a prolonged recovery period, probably with more visits to hospital, it is likely that patients might expect to see greater improvement in their before they begin to consider this more invasive and inconvenient treatment to have been worthwhile.

This is a reasonable suggestion as it is already thought that MICs may vary between treatments (for the same condition) in general (Revicki et al., 2008), Interpretability of outcome in hand surgery has been studied using a range of methods (Rodrigues et al., 2014). Most MICs in this field have used retrospective anchor-based methods, though hand-specific prospective anchoring using the satisfaction domain of the Michigan Hand Questionnaire (MHQ) is also possible, as already discussed (Malay et al., 2013, Shauver and Chung, 2009, Waljee and Chung, 2012). Many of these studies involved assessing the interpretability of different subscales of the MHQ. Whether it is appropriate to use the satisfaction domain of the MHQ, another component of the same tool, as the prospective anchor is not clear.

9.6 Conclusions of study of recovery, responsiveness and interpretability

This study raises further questions regarding the DASH's suitability for use in Dupuytren's disease, in terms of interpretability. However, given the potential issues associated with retrospective anchors, the poor interpretability of the

DASH should be confirmed using a prospective anchor. Measurement of the interpretability of other PROMs for use in Dupuytren's disease is also required to be able to interpret existing research and ensure appropriate outcome measures are used in future studies.

SUMMARY

This chapter demonstrates that recovery from fasciectomy and dermofasciectomy takes longer than the 6 weeks often studied, and measuring outcome at 6 weeks will underestimate the benefit from these procedures.

The DASH shows moderate responsiveness to surgery for Dupuytren's disease, and the URAM shows good responsiveness. However, the DASH was not interpretable in this study. As seen in the earlier Validity chapter, the DASH reflects different underlying constructs to the URAM, which may explain the DASH's poorer performance. Given the earlier chapter's findings of the DASH's validity and this chapter's findings regarding interpretability, the DASH should not be used in future trials in Dupuytren's disease.

The MIC for the URAM after open surgery (10.5/45) was greater than that previously reported for aponeurotomy (2.7/45) (Beaudreuil et al., 2011).

Future work should involve prospective anchors, given that the GRC correlated more closely with final functional state than with change in functional following treatment, and further study of timing of assessment after treatment is required.

Alongside examination of hand function outcome measures themselves as described in the earlier chapters in the thesis, understanding the variables associated with poor functional outcome has been neglected in previous

research, yet understanding which variables are relevant to functional outcome will inform future research design.

The next chapter will investigate the factors associated with poor functional outcome through a cross sectional design.

10 Functional Outcome, Complications and Associated Factors Following Surgery For Dupuytren's Disease

10.1 Preface

The previous chapters have investigated aspects of the performance of outcome measures in Dupuytren's disease. Further work will be required to ensure that the optimal outcome is chosen for the purpose required in future studies, though aspects of poor performance have been demonstrated for the commonly used measures of the DASH, the URAM and passive extension deficit. In addition to the selection of an appropriate measure, study design may need to account for factors that independently influence outcome besides the type of treatment received. Factors associated with recurrence have been studied previously, and collectively have been termed the Dupuytren's 'diathesis'. However, the factors associated with functional outcome are not well described. Identifying and controlling differences in such factors would be important in designing non-randomised comparative studies, to ensure that groups are matched. Confirmation that randomised trial groups are appropriately matched for such factors would also be important. This chapter investigates late outcome of treatment and the factors associated with poor functional state after treatment.

10.2 Introduction

The factors associated with the 'Dupuytren's diathesis' have been studied (Abe et al., 2004, Dias et al., 2013, Gelberman et al., 1980, Hindocha et al., 2006, van Rijssen et al., 2012). However the objective outcomes studied, such as recurrence, provide an incomplete representation of the diverse disability and functional impairment experienced by patients with Dupuytren's disease (Rodrigues et al., 2014). Recurrence and extension are not the only causes of poor outcome after surgery for Dupuytren's disease. For example, complications causing loss of finger flexion may also have serious functional consequences. A recent review has considered the rates of complications reported in the literature following treatment of Dupuytren's disease (Crean et al., 2011). However, factors associated with poor functional outcome and complications of surgery have not been investigated.

The Disabilities of the Arm, Shoulder and Hand (DASH) patient-reported outcome measure (PROM) correlates poorly with angular deformity (Degreef et al., 2009, Engstrand et al., 2009, Jerosch-Herold et al., 2011, Zyluk and Jagielski, 2007), which has led to the conclusion that such PROMs may not be valid in Dupuytren's disease (Packham, 2011). Other measures, such as the Sollerman hand score show a statistically significant, though weak, correlation with angular deformity (Sinha et al., 2002). A new Dupuytren's disease-specific PROM, the Unité Rhumatologique des Affections de la Main (URAM) scale correlates with angular deformity (Beaudreuil et al., 2011), but may not describe other domains of outcome, such as pain (Beaudreuil et al., 2014).

In this study, functional outcome and complications of surgery for Dupuytren's disease were assessed, along with the factors associated with these parameters rather than those associated with recurrence or extension alone.

10.3 Methods

10.3.1 Patient recruitment and data collection

This project was independently approved as a service evaluation at each participating centre prior to commencing data collection at that centre. Information governance, and where required Caldicott Guardian, approval was also obtained locally prior to commencing recruitment. Clinical coding departments at five UK NHS hand surgery centres (Derby, Livingston, Nottingham, Plymouth, Rotherham) identified patients who had undergone aponeurotomy, fasciectomy or dermofasciectomy either 1 year or 5 years earlier. Codes for amputation were not included. Patients living within 20 miles of the centre were invited to attend a locally approved service evaluation. A single surgeon assessed all who consented who could be assessed 1 or 5 years (+/- 2 months) after their surgery. The candidate performed a standardised examination on all patients.

Data captured included patient demographics, known and suggested risk factors for the progression of Dupuytren's disease, complications of surgery, reoperation to the digit since the index procedure, angular deformity, and the DASH PROM.

If more than one digit on a hand had been treated with the same procedure (e.g. fasciectomy to little and ring fingers in a single procedure), then only one digit was assessed. The digit selected in such cases was the one that currently had the worst total active extension deficit.

If different procedures were performed in one operation (e.g. fasciectomy to ring finger and dermofasciectomy to little finger), then both procedures were analysed as separate events for the study of objective outcomes, but the patient was not included in the functional outcome analyses.

If both hands were treated with the same procedure in one operation (which only occurred with aponeurotomy), then only the treated digit on the dominant hand was assessed, and was included in objective and functional outcome analyses. This avoided any patient being recruited to the same subgroup more than once (Sauerland et al., 2003).

Analyses were performed using Prism 6.0 for Mac OS X (GraphPad® Software, La Jolla, USA, 2012) and SPSS® Statistics version 21 (IBM® Software, Redmond, USA, 2012). DASH scores were dichotomised into those above 15 (symptomatic scores) and those below 15 (asymptomatic scores), based on guidance from the developer of the tool (Kennedy et al., 2011).

10.3.2 Objective outcomes

Reoperation (defined as further surgery for recurrence or extension of Dupuytren's disease in the same digit) was assessed by patient recall, and confirmed via hospital records if unclear. The candidate assessed passive extension deficit at MCPJs and PIPJs for all cases. During all measurements, the other joints in the ray being assessed were held in maximum passive flexion, to standardise the effect of dynamism (Rodrigues et al., 2014).

10.3.3 Functional outcome

Proportions of patients with poor functional outcome 1 and 5 years after different types of procedure were compared using Chi square tests.

Binary logistic regression analysis was performed to identify and control for factors associated with poor functional outcome (defined as DASH>15 at 1 year after treatment). The independent variables studied were further ipsilateral Dupuytren's disease surgery since the index procedure ("surgery since"), length of follow up (1 year or 5 years) and eight factors, some of which are part of the traditional Dupuytren's diathesis, and others putative novel factors that might be expected to influence functional outcome:

- Self reported alcohol consumption >28 United Kingdom units per week (1 unit is equivalent to 10 milligrams of ethanol)
- Active smoker
- Self reported positive family history of Dupuytren's disease
- Surgery to the little finger
- Presence of knuckle pads on examination
- Index procedure being revision of previous surgery (defined as previous surgery to the same digit)
- Diabetes mellitus
- Gender

Operation type was entered with aponeurotomy as the constant, so that fasciectomy and dermofasciectomy were compared to it.

The suitability of the data for logistic regression was verified prior to analysis. In particular, the data were examined for the absence of multicollinearity, which occurs when two or more of the independent variables studied correlate with each very strongly. This can affect regression, and in particular can influence the results obtained regarding the independent variables concerned (Pallant, 2010). To do this, tolerance, the amount of variance that cannot be accounted by other variables, was calculated for each variable. If it is low, then the variable may show collinearity with another variable, or multicollinearity with several variables (Pallant, 2010). In keeping with convention, an unacceptable level of tolerance was defined as <0.1.

To control for false discoveries (false positives), the p value threshold considered significant was adjusted using a described method (Benjamini and Hochberg, 1995). As the variables associated with poor functional outcome have not been studied widely, a false discovery rate (Q) of 20% was considered reasonable to minimise type 2 error risk. The variables in the model were ordered by p value and ranked, and the threshold for each variable calculated using the formula (i/m)*Q, where 'i' was the rank of the variable and 'm' was the total number of tests (13 in the analysis of functional outcome). If the p value obtained was smaller than 0.05 and lower than its calculated threshold, then the significant result was considered true.

10.3.4 Adverse outcomes

The adverse outcomes assessed were:

- Cold intolerance (described using an existing scale (Campbell and Kay, 1998))
- Loss of flexion (defined as a fingertip pulp to distal palmar crease distance of over 10mm on active flexion)
- Infection (defined as patient recall of the need for at least one postoperative course of antibiotics that was not prescribed as prophylaxis)
- Complex regional pain syndrome (CRPS) (defined using the modified International Association for the Study of Pain (IASP) criteria (Harden et al., 2007))
- Altered sensation (defined as failure to identify 2/3 tests of two point discrimination at 6 millimetres at distal phalanx level in one or both digital nerve territories on the operated digit)

Proportions of patients with each complication were compared between procedures with Chi square tests. Hierarchical binary logistic regression analyses were performed for each complication in a similar manner as for functional outcome. The independent variables selected for study were ones that might influence the risk of complications. In addition to further ipsilateral surgery for Dupuytren's disease, they were:

- Multiple digit surgery during index procedure
- Gender
- Diabetes mellitus
- Smoking status

- Index procedure being revision of previous surgery (defined as previous surgery to the same digit)

For complications expected to change between 1 and 5 years postoperatively, the time point (1 year versus 5 years) was also studied. These were loss of flexion and cold intolerance (which might improve in the intervening period). For other complications, all assessments were studied together.

10.3.5 Sample size

A sample size with ten outcome events per predictor variable is often quoted for logistic regression analyses. As twelve predictor variables were used here, this would require 120 poor functional outcomes (DASH>15) in our study. However, more recent examination of this rule has suggested that five to nine outcome events per predictor variable may be acceptable (Vittinghoff and McCulloch, 2007). Furthermore, the proportion of patients with poor functional outcome following Dupuytren's disease surgery is not well described. A target of 100 poor functional outcomes was set, and a total target sample size of 400, based on an assumption that approximately 25% would have poor function.

10.4 Results

10.4.1 Patients and procedures

Four hundred and fourteen patients were recruited and assessed between September 2011 and June 2013 across all sites. These 414 patients had

undergone 433 procedures. One had undergone an amputation after the index procedure, and was excluded from the analysis.

All remaining 432 procedures in 413 patients were included in analyses of reoperation and complications, as these were recorded at digit level (Table 10.1). However, function was assessed at patient level. Ten of the 413 patients had undergone aponeurotomy to both hands in a single procedure, and only their dominant hand procedures were assessed. A further nine patients had undergone different procedures to different digits, and so were excluded from analyses of function. Thus, 404 patients were included in analyses of function (Table 10.1).

Nine patients (2%) had different procedures assessed separately at different times in the study period. This comprised seven patients who had undergone fasciectomy and dermofasciectomy to different digits of the same hand 1 year earlier and one patient who had undergone fasciectomy and dermofasciectomy to different digits of the same hand 5 years earlier. The other patient had undergone fasciectomy to one hand and aponeurotomy to the other hand in the same procedure.

	1 YEAR FOLLOW UP		5 YEAR FOLLOW UP	
	Number of procedures in objective analyses	Number of patients in function analyses	Number of procedures in objective analyses	Number of patients in function analyses
Total	270	245	162	159
Aponeurotomy	114	104	20	19
Fasciectomy	126	118	125	124
Dermofasciectomy	30	23	17	16

Table 10.1: Sample sizes studied

The demographics of the 413 patients are shown in Table 10.2. In terms of reoperation choice, 4/11 reoperations after aponeurotomy were further aponeurotomies, and the remaining 7/11 were fasciectomies. After fasciectomy, 1/11 was an aponeurotomy, 5/11 were fasciectomies and the remaining 5/11 were dermofasciectomies. These proportions were significantly different (p=0.028, Chi square test). It was not clear whether these choices were due to patient preference, surgeon preference, or other reasons.

10.4.2 Objective outcomes

The percentage of procedures that that resulted in reoperation was not different between the three procedures in the 1-year postoperative group (p=0.393, Chi square test, see Table 10.3). However, the reoperation rate was significantly greater after aponeurotomy in the 5-year group (p=0.000, Chi square test, see Table 10.3). The reoperation rate after aponeurotomy was significantly higher at 5 years than at 1 year after treatment (p=0.002, Fisher's Exact test, see Table 10.3). There was no significant difference between 1 and 5-year subgroups for fasciectomy or dermofasciectomy (see Table 10.3).

'Poor objective outcome' was estimated to account for patients who may have declined revision surgery or been considered unsuitable for further surgery. This was done by combining those who had undergone reoperation with those who had considerable loss of extension but had not undergone further surgery. The proportion of 'poor objective outcome' cases was significantly

greater 1 year after more invasive procedures (Table 10.3). However, there was no difference between procedures in the 5-year groups.

Demographic or candidate risk factor	
Age	Mean 66, Range 33-89
Male : Female	318 : 95 (77% male)
Hand dominance	371/413 right handed (90%)
Diabetic	61/413 (15%)
Smoker	60/413 (15%)
Self reported weekly alcohol intake (UK units)	Mean 14.7
Previous ipsilateral surgery prior to index operation	103/413 (25%)
Index operation was revision of previously treated digit	85/413 (21%)
Self reported positive family history of Dupuytren's disease	180/413 (44%)
Knuckle pads present	122/413 (30%)
Hand treated	212/413 right (51%)
Digit studied	248 little (60%) 129 ring (31%) 25 middle (6%) 9 index (2%) 2 thumb (0.5%)

Table 10.2: Patient demographics and prevalence of candidate risk factors

Outcome		Aponeurotomy	Fasciectomy	Dermofasciectomy	Chi square test
Reoperation	1 year	5/114 (4.4%)	3/126 (2.4%)	0/30 (0%)	p=0.393
	5 years	6/20 (30.0%)	8/125 (6.4%)	0/17 (0%)	p=0.000
'Poor objective outcome'	1 year	25/114 (21.9%)	48/126 (38.1%)	14/30 (46.7%)	p=0.006
(Reoperation or either MCPJ	5 years	8/20 (40.0%)	61/125 (48.8%)	10/17 (58.8%)	p=0.521
or PIPJ>25^{o})					

Table 10.3: Objective outcomes

MCPJ – metacarpophalangeal joint, PIPJ – proximal interphalangeal joint

10.4.3 Functional outcome

The proportion of patients with symptomatic DASH scores was not significantly different between the three procedures for either 1-year or 5-year postoperative patients (Table 10.4). However different proportions of these patients had undergone further surgery in the 1 or 5 years since, with a significantly higher reoperation rate 5 years after aponeurotomy than after dermofasciectomy.

As the prerequisites were met in terms of tolerance of the variables studied, logistic regression analysis was performed. The omnibus test demonstrates whether the model built by the analysis performs well in terms of 'goodness of fit', i.e. whether the included variables do actually contribute to predicting poor functional outcome. Here, it was statistically significant (p=0.000), demonstrating that this was the case. The results of the logistic regression analysis are shown in Table 10.5. Controlling for confounding variables such as the effect of further surgery since, and of length of follow up, the only other variables studied that showed significant associations with poor function were female gender, diabetes mellitus and previous ipsilateral surgery for Dupuytren's disease. In general, the variables considered part of the Dupuytren's diathesis were not associated with poor functional outcome.

Outcome	Time point	Aponeurotomy	Fasciectomy	Dermofasciectomy	p value
DASH summary score	1 year	9.5 (6.8, 12.2)	10.7 (7.6, 13.8)	14.3 (6.2, 22.5)	0.421*
(mean (95%CIs))	5 years	9.1 (4.7, 13.5)	10.9 (8.3, 13.5)	15.1 (5.5, 24.8)	0.448*
Proportion DASH>15	1 year	19/104 (18.3%)	26/118 (22.0%)	7/23 (30.4%)	0.416†
	5 years	5/19 (26.3%)	34/124 (27.4%)	5/16 (31.3%)	0.940†

Table 10.4: Functional outcomes

**One way ANOVA*

†Chi square test

DASH – Disabilities of the Arm, Shoulder and Hand, 95%CIs – 95% confidence intervals

Independent variable		Adjusted Odds Ratio (OR)	95% confidence intervals of adjusted OR	Rank by p value (i) †	(i/m)*Q p value threshold †	Significance of association (p value)
Gender						
	Female	**3.85**	**2.13-7.14**	**1**	**0.02**	**0.00**
	Male	1				
Previous ipsilateral Dupuytren's surgery						
	Yes	**2.13**	**1.18-3.85**	**2**	**0.03**	**0.01**
	No	1				
Diabetic						
	Yes	**2.07**	**1.10-3.91**	**3**	**0.05**	**0.03**
	No	1				
Smoker						
	Yes	1.67	0.83-3.37	4	0.06	0.15
	No	1				
Little finger surgery						
	No	1.35	0.79-2.27	5	0.08	0.27
	Yes	1				
Follow up length						
	5 years	1.33	0.79-2.27	6	0.09	0.28
	1 year	1				
Knuckle pads						
	Present	1.31	0.76-2.28	7	0.11	0.33
	Absent	1				
Further surgery since						
	Yes	1.60	0.58-4.43	8	0.12	0.36

	No	1				
Age at surgery						
	Under 50	1.52	0.56-4.16	9	0.14	0.41
	50 or over	1				
Procedure was fasciectomy						
	Fasciectomy	1.25	0.68-2.28	10	0.16	0.48
	Aponeurotomy	1				
Procedure was dermofasciectomy						
	Dermofasciectomy	1.21	0.45-3.27	11	0.17	0.70
	Aponeurotomy	1				
Family history of Dupuytren's disease						
	Yes	1.05	0.64-1.74	12	0.34	0.84
	No	1				
Weekly alcohol intake						
	<=28 units	1.01	0.49-2.08	13		0.98
	>28 units	1				

Table 10.5: Logistic regression of function

10.4.4 Complications

The rates of different complications are shown in Table 10.6, grouped by procedure (and length of follow up where relevant). Complications that were hypothesised to improve over time (cold intolerance and loss of flexion) were more common at 1-year follow up compared to 5-year follow up. Infection and altered sensation were observed more frequently after more invasive procedures than after aponeurotomy. At 1-year follow up cold intolerance and loss of flexion were more common after more invasive procedures. There was no difference between procedures at 5-year follow up, though significantly more of the aponeurotomy group had undergone further surgery.

Tolerances for all variables studied in relation to complications were acceptable, and logistic regression analyses were performed for all complications except CRPS, as this was found infrequently. Each of the models for cold intolerance, loss of flexion, altered sensation and infection was significant on omnibus testing. All statistically significant results from the analyses are shown in Table 10.7.

Complication	Time point	Aponeurotomy (total n=134)	Fasciectomy (total n=251)	Dermofasciectomy (total n=47)	p value (Chi square tests)
Reoperation	1 year	5/114 (4%)	3/126 (2%)	0/30 (0%)	0.39
	5 years	6/20 (30%)	8/125 (6%)	0/17 (0%)	**0.002**
Cold intolerance	1 year	11/114 (10%)	39/126 (31%)	19/30 (63%)	**<0.0001**
	5 years	1/20 (5%)	20/126 (16%)	5/17 (29%)	0.13
Flexion loss>10mm	1 year	20/114 (18%)	42/126 (33%)	13/30 (43%)	**0.002**
	5 years	3/20 (15%)	30/125 (24%)	3/17 (18%)	0.60
Altered sensation*		6/134 (4%)	38/251 (15%)	9/47 (19%)	**0.003**
Infection		2/134 (1%)	22/251 (9%)	7/47 (15%)	**0.003**
CRPS		1/134 (1%)	5/251 (2%)	0/47 (0%)	0.42

Table 10.6: Complications

**Defined as absent 2 point discrimination at 6 millimetres in either radial or ulnar digital nerve territories over the pulp of the distal phalanx*

Adverse outcome	Independent variable	Adjusted Odds Ratio (OR)	95% confidence intervals of adjusted OR	Rank by p value (i) †	(i/m)*Q p value threshold †	Significance of association (p value)
Cold intolerance						
	Dermofasciectomy	**14.77**	**5.78-37.74**	**1**	**0.02**	**0.000**
	Aponeurotomy	1				
	Fasciectomy	**4.00**	**1.97-8.12**	**2**	**0.04**	**0.000**
	Aponeurotomy	1				
	Dermofasciectomy	**3.69**	**1.75-7.80**	**3**	**0.06**	**0.001**
	Fasciectomy	1				
	1-year follow up	**2.68**	**1.54-4.67**	**4**	**0.08**	**0.001**
	5-year follow up	1				
	Smoker	**2.66**	**1.44-4.94**	**5**	**0.1**	**0.002**
	Non-smoker	1				
Loss of flexion>10mm						
	Dermofasciectomy	**5.34**	**2.16-13.21**	**1**	**0.02**	**0.000**
	Aponeurotomy	1				
	Fasciectomy	**3.66**	**1.86-7.17**	**2**	**0.04**	**0.000**
	Aponeurotomy	1				

Altered sensation						
	Fasciectomy	**3.09**	**1.21-7.85**	**1**	**0.02**	**0.018**
	Aponeurotomy	1				
	Dermofasciectomy	**3.91**	**1.19-12.80**	**2**	**0.04**	**0.024**
	Aponeurotomy	1				
	Female	**2.11**	**1.10-4.03**	**3**	**0.06**	**0.024**
	Male	1				
Infection						
	Dermofasciectomy	**7.59**	**1.42-43.42**	**1**	**0.02**	**0.018**
	Aponeurotomy	1				
	Fasciectomy	**6.07**	**1.33-27.60**	**2**	**0.04**	**0.020**
	Aponeurotomy	1				
	Revision procedure	**2.36**	**1.03-5.38**	**3**	**0.06**	**0.041**
	Primary procedure	1				

Table 10.7: Significant independent variables in logistic regression analyses of complications

10.5 Discussion

10.5.1 Present study findings

This study confirms that aponeurotomy is associated with a higher reoperation rate than fasciectomy or dermofasciectomy. However, after controlling for some independent variables that might differ between the groups, functional outcome was not significantly different between these three procedures at 1 year and 5 year follow-up. This finding requires confirmation in a study with a larger number of dermofasciectomy and aponeurotomy cases with 5-year follow-up but is important and may be valid. This is as complications that limit function, such as loss of flexion, cold intolerance and altered sensation, may be more frequent following more invasive procedures, which typically had higher complication rates in this study.

The cross-sectional design of our study means that patients' immediate postoperative outcome is not known and limits the interpretation of our data in Table 3. However, other studies have demonstrated reliable rates of initial correction, including for aponeurotomy (Pess et al., 2012).

The choice of recurrence as the primary endpoint for studying treatment in Dupuytren's disease is challenged by the data presented here, which demonstrates the different rates of complications after different treatments. As many of these complications are not associated with recurrence, they will not be captured if recurrence is used as the primary outcome measure. Consequently, recurrence may be a surgeon-centred outcome, but is unlikely to be patient-centred, and it may be of limited value in cost utility analyses.

The variables associated with poorer outcome in this study differ from those identified as contributing to the Dupuytren's diathesis in other studies (Abe et al., 2004, Hindocha et al., 2006, Hueston, 1963). This suggests that those patients whose hand function is worse following surgery may not always be the patients who experience recurrence.

Several variables were associated with poor function here. Patients undergoing revision treatment may not achieve the same degree of improved hand function as those undergoing primary surgery due to an accumulation of iatrogenic insults to the hand, or perhaps due to disease severity. Women reported worse hand function than men, though it is not clear why this is the case. It may be intrinsic to the DASH itself, as similar patterns have been reported with DASH-related measures in other hand conditions (the QuickDASH in carpal tunnel release) (Jenkins et al., 2012). Diabetics might be expected to have greater risk of complications, such as infection and poor healing, and worse rehabilitation as a result. Alternatively, their higher DASH scores may reflect a higher prevalence of comorbid upper limb conditions, such as cheirarthropathy, trigger fingers, and carpal tunnel syndrome (Larkin et al., 2014, Pandey et al., 2013), which may be confounding factors. Newer developments have included the launch of at least two Dupuytren's-specific measures (Beaudreuil et al., 2011, Mohan et al., 2014). However, the DASH has been the most commonly employed measure to date (Ball et al., 2013). Therefore, the data presented here are important to consider when

interpreting the findings of studies regarding functional outcome in Dupuytren's disease.

When the independent variables studied were controlled for, there was no difference in the odds of having poor hand function 5 years after aponeurotomy compared to fasciectomy or dermofasciectomy. This may reflect a balance between a greater risk of recurrence after aponeurotomy (being offset by the less invasive nature of the procedure and perhaps less frequent or less severe complications. However, given the limitations of this study apparent from the different interpretations of the individual findings, a randomised controlled trial with hand function as the primary endpoint is required to confirm this and to facilitate comparison of the relative cost effectiveness of different treatments for Dupuytren's disease.

10.5.2 Limitations

The most important limitation to this study relates to its cross-sectional design. As a result, the preoperative and immediate postoperative states of patients were not known, and may not have been matched. However, steps were taken to improve the reliability of the data presented. The centres that contributed had different preferences for treatment, with some favouring aponeurotomy and others fasciectomy. Those centres that preferred aponeurotomy have considerable experience with it, and so use it as their first line treatment for most cases of Dupuytren's disease. Therefore, it is expected that a large proportion of patients who underwent aponeurotomy at these centres would have received fasciectomy at the others in the study.

However, this cannot be confirmed. Furthermore the logistic regression analyses were performed with the aim of adjusting for differences between groups. Despite this, the comparison between procedure types may not be as robust as would be achieved in a prospective comparative study. Nevertheless, the examination of the factors associated with poor functional outcome is important in its own right. The findings of this study certainly need verification with a prospective study, preferably incorporating randomisation.

Some of the variables studied were self-reported and may not have been accurate: for example, smoking status may have changed since the patient underwent surgery, there may have been recall bias, and there may have been social desirability responses with patients denying or underestimating factors such as excessive alcohol intake or smoking. Studying such variables prospectively would be more reliable and thus allow for more accurate statistical modelling.

Some sub groups within our study were relatively small. However, the rates of complications identified are largely comparable to those reported elsewhere in the literature (Crean et al., 2011). Our findings need to be validated with further large size studies, or even registry-level data.

There are other limitations to our data that might contribute to the findings differing from those in other studies. There may have been selection bias in studies in which patients were invited to participate retrospectively. Also there may be differences in the pre-operative states of the digits treated in different studies, or in patient or surgeon attitudes, either relating to different

international cultural norms or perhaps related to involvement in a trial compared to routine clinical practice. However, given the paucity of literature focussing primarily on functional outcome in Dupuytren's disease, rather than recurrence, it is envisaged that this area merits attention adds value and should influence the design of future research studies.

10.5.3 Relationship to existing literature

Reoperation may be an important clinical and economic endpoint to study, but is a complex variable. In order to undergo further treatment, a patient would have to have recurrent or extended disease that is amenable to further surgery, be offered surgery by a clinician, and consent to the further treatment. Indeed, it was noted during the assessments that some patients described progressive recurrence but had not sought further intervention. This pattern has been previously reported, with 'reoperation rates' lower than 'treatment failure' rates, where reoperation or recurrent contracture is studied (van Rijssen et al., 2012). As a result, reoperation is not a valid surrogate for recurrence. In this study, the proportions of patients undergoing reoperation within 5 years of treatment were higher after aponeurotomy, as might be expected, but were still lower than reported by others (Foucher et al., 2003, van Rijssen et al., 2012). In a randomised controlled trial that compared aponeurotomy to fasciectomy, proportions of patients undergoing reoperation within 5 years were higher, with reoperations rates of 33/52 (63%) and 4/41 (9%) for fasciectomy (van Rijssen et al., 2012).

Abe and colleagues investigated the factors associated with reoperation at a mean follow-up length of 5 years in a small Japanese population (Abe et al., 2004). However, the applicability of findings in this population to other populations is not clear. Additionally, the length of follow-up ranged from 3 to 12 years. As Dupuytren's disease is a slowly progressive condition, patients 3 years following Dupuytren's disease surgery are not comparable to those 12 years after treatment.

Hindocha and colleagues studied the factors associated with recurrence of palpable disease in the operated field (Hindocha et al., 2006). Whilst this is a common definition of recurrence (Becker and Davis, 2010), one may argue that it is less clinically relevant than other endpoints. The reappearance of palpable disease alone does not require treatment, does not necessarily impair function, and has not been shown to predict subsequent deterioration, and as such, does not in itself mandate treatment.

van Rijssen and colleagues studied factors associated with recurrence defined as a progressive angular deformity (van Rijssen et al., 2012). As further treatment might become advisable with deterioration in angular deformity, this may be a more clinically applicable and reliable endpoint than those used in either of the earlier studies. However, it does not describe the patient's hand function or health-related quality of life, which is probably also influenced by factors such as complications.

Most recently, Dias and colleagues investigated factors associated with progressive contracture recurrence in a randomised controlled trial of firebreak

dermofasciectomy versus z-plasty closure of fasciectomy wounds (Dias et al., 2013). The degree of progression that constituted recurrence was not formally defined.

However, not all studies of recurrence support the predictive value of diathesis factors (Gelberman et al., 1980).

Others have investigated the factors associated with poor outcome in the absence of recurrence of disease (Misra et al., 2007), highlighting that 'poor outcome' in Dupuytren's disease is not entirely due to recurrence.

Recurrence has been the focus of much research in Dupuytren's disease (Becker and Davis, 2010). Whilst treating recurrent disease may be challenging, doing so following an aponeurotomy may be more straightforward than after more invasive surgery (van Rijssen and Werker, 2012). Furthermore, it is important to appreciate variables that might be associated with a tendency for Dupuytren's disease to recur following treatment; recurrence alone cannot be used as a surrogate for functional outcome. This is important given that the correlation between angular deformity and loss of function is weak (Engstrand et al., 2009, Jerosch-Herold et al., 2011, Zyluk and Jagielski, 2007). Also an assessment of health-related quality of life should be made to allow cost effectiveness analysis (NICE, 2008). Additionally, these factors might provide a more patient-centred and relevant evaluation of treatment outcome; avoiding recurrence at the expense of an increased risk of complications of surgery may not be desirable. Instead,

considering the impact of both recurrence and complications in affecting function may be more appropriate.

10.6 Conclusions of study of late outcome and complications

Functional state following treatment for Dupuytren's disease remained matched between procedures in this study, after controlling for confounding factors. However, due to the limitations arising from its cross sectional nature, this finding needs verification in a prospective study. However, the factors associated with poor functional outcome were not the same factors associated with recurrence, or the Dupuytren's diathesis. As with previous studies, complications were relatively common, particularly after more invasive surgery. The functional outcome of surgery for Dupuytren's disease may reflect both correction and recurrence of contractures, but also complications.

SUMMARY

The comparison of different procedures in this chapter demonstrates that despite very different recurrence rates, the final functional state that patients achieve after treatment was not different between them. There are limitations to this finding, for example preoperative state was not known and so the groups may not be matched, but further investigation of functional outcome is indicated.

The differences between treatments in terms of complications are considerable, and a logical interpretation is that functional outcome represents the balance of achieving and maintaining correction versus avoiding treatment

sequelae. This might explain the poor correlation between hand function and angular deformity reported earlier in this project, and in previous studies (Engstrand et al., 2009, Jerosch-Herold et al., 2011, Zyluk and Jagielski, 2007).

The factors associated with poor functional state differed from diathesis factors, and hand surgeons need to reassess which patients are at risk of completing rehabilitation with a worse functional outcome in light of this finding. More aggressive hand therapy may be warranted for this subgroup.

The results of this study may inform interpretation of papers in which the DASH was used; the findings should be confirmed for other existing outcome measures and ideally after development of an optimal measure of hand function.

11 Project discussion

11.1 Findings of the chapters in the project

11.1.1 Existing literature

The systematic review of trials in Dupuytren's disease has confirmed that there are very few high quality comparative studies of the treatment of Dupuytren's disease. Even relatively modern studies have significant risks of bias. In the case of those that reported angular deformity as the primary outcome (van Rijssen et al., 2006, van Rijssen et al., 2012), poor blinding of observers posed a significant risk. This potential severity of this risk has been further explored in the present project by assessing the extent of dynamism that might be present. Dynamism is a significant issue that may have been previously underestimated, as it had not been quantified. As has been identified in other reviews (Ball et al., 2013, Becker and Davis, 2010), the outcomes reported have been inconsistent and are not standardised.

A patient-reported outcome measure (PROM) may be a more suitable option. Whilst a range have been used in Dupuytren's disease (Ball et al., 2013), aspects of their validity have not been fully investigated in Dupuytren's disease, and the review of interpretability performed here identified limited data describing their interpretability. There is an urgent need for investigation of candidate outcome measures for future research and for use in clinical practice, some of which has been performed in the present project.

At the same time, the meta-analysis of trials of splinting demonstrated that postoperative splinting, a commonly used component of treatment, may actually impair outcome. This finding was not apparent in any of the meta-analysed trials in isolation; it only became apparent when results were pooled here. It is apparent that further research is required in this area

11.1.2 Methods employed to develop and appraise outcome measures

This project has employed a range of techniques to appraise candidate outcome measures for future use in Dupuytren's disease. Frameworks exist to guide development of PROMs and their validation (Mokkink et al., 2010). However, this project has identified some limitations of contemporary PROM development methodology. In particular, the preoperative impairments experienced in Dupuytren's disease are diverse, and they also differ from the impairments present in the postoperative phase from treatment sequelae. As a result, developing a PROM based on the preoperative experience of Dupuytren's disease may result in a tool that is not comprehensive for use. These issues should be considered in future research design, and are relevant to a range of conditions, not just hand surgery or Dupuytren's disease.

11.1.2.1 Tool Development

Many of the outcome measures used to date in Dupuytren's disease are best considered to be 'legacy measures', which were developed using out-dated methods. For example, surgeons developed the PEM without any clear patient involvement (Macey et al., 1995). More modern tools such as the

URAM involved physician input in the item generation process (Beaudreuil et al., 2011), which it could be argued prevents them from being truly 'patient centred'. This deviation from a truly patient-centred approach may also have occurred in the item reduction stages of development. Modern approaches to item selection include removal of candidate items based on mathematical properties such as a limited spread of responses provided in a pilot study. However, this may render the tool further removed from the 'real world' patient experience of the condition. Certainly, our patients' goals were very diverse (almost unique at the individual level) and difficult to capture with a tool comprising task-based items, as is the case for the URAM. Additionally, the data in the cohort study described here supports the conclusion that the URAM may not be comprehensive in terms of content validity, particularly in relation to postoperative pain.

The developers of the URAM argue that it provides an assessment of the limitation arising from loss of extension (Beaudreuil et al., 2014). However, if loss of extension is not a surrogate for global hand function or patient perception of improvement, then it is of limited practical value, particularly at a healthcare architecture level, where cost utility needs to be assessed to justify investment in treatment. Assessing cost utility in the UK at present typically involves assessment of health-related quality of life (NICE, 2008), though may be performed using global hand function, which might be better assessed with

domain-specific PROMs such as the Patient Evaluation Measure or the Michigan Hand Questionnaire.

Future tool development must adopt a more patient-centred approach, to ensure that they appropriately reflect the lived experience and functional limitation in Dupuytren's disease. To achieve this, the scope of future outcome measures must be clearly defined and should be appropriate. If cost utility of treatment and meaningful clinical improvement is to be achieved, then the items incorporated should describe one or more of global hand function, activity restriction, patient perception or health-related quality of life. To do this, item generation should be based on patient input, and should include the experiences of patients in the early and late postoperative periods, so that the impact of treatment sequelae and complications are incorporated into the measure. Given that domains such as functional impairment from loss of extension are likely to improve following treatment, but that complications such as cold intolerance and stiffness are likely to increase after surgery, it is reasonable to hypothesise that they will load on different constructs. If this is found to be the case, then it is likely that a comprehensive outcome measure for use in Dupuytren's disease will comprise at least two sub scales: one to assess the impact of preoperative symptoms that generally improve after treatment, and one to assess sequelae that arise from treatment.

This approach to tool development may be very apparent from a project such as this. As has been demonstrated in the study of surgical complications in this project, the sequelae of the treatment of Dupuytren's disease are very different to preoperative symptoms. This may not be the case in other conditions where postoperative complications lead to similar functional impairments as the original preoperative symptoms. However, adopting the approach suggested here may provide a more robust model for developing PROMs.

11.1.2.2 Validity and Reliability

Robust evidence for validation processes does not exist. However, international consensus has been achieved through the Delphi process that generated the COnsensus-based Standards for the selection of health status Measurement INstruments (COSMIN) (Mokkink et al., 2010). This venture resulted in a series of 'checklists' against which studies may be benchmarked. It also documented leading international opinion around choice of analyses. Given that most of the PROMs that have been used to study Dupuytren's disease to date were developed prior to COSMIN, it is not surprising that these standards have not always been applied. This project has applied some of these validation techniques to existing legacy measures, such as the DASH, to establish whether they display acceptable validity in spite of having been produced using parochial approaches to PROM development. Typical deviations from COSMIN include inappropriate assessment of criterion

validity, failure to appropriately assess structural validity in relation to reliability, and issues relating to other parameters such as responsiveness and interpretability. For example, COSMIN identifies a range of analyses that are not recommended for studying responsiveness, including all commonly used metrics, such as standardised response mean and effect size. However, COSMIN was not able to establish which alternatives to these analyses are appropriate.

Criterion validity has been studied for a range of PROMs in Dupuytren's disease, with objective angular loss of extension used as the 'gold standard' against which the PROM's performance is benchmarked. Poor correlation with this gold standard has led to criticism of particular outcome measures. However, COSMIN establishes that surrogates such as loss of extension cannot be used as 'gold standards'. The only scenario in which a gold standard can be used is when an abridged version of a PROM is being compared to the original full length PROM, such as when establishing whether the QuickDASH behaves similarly to the full DASH, as studied here.

Using angular deformity as the 'gold standard' against which a PROM's criterion validity is assessed is illogical. If angular deformity is the optimal measurement, then it ought to be used as the outcome measure, and the PROM is redundant. The range and breadth of patient goals identified in this project demonstrates that preoperative functional limitation in Dupuytren's

disease cannot be attributed to loss of extension alone. Furthermore, angular deformity is subject to potential bias in the form of dynamism, which has also been quantified here. Finally, postoperative complications studied in this project demonstrate that postoperative functional impairment is likely to relate to domains other than loss of extension. This is why a suitable alternative to measuring angular deformity is required. To then benchmark candidate outcome measures against the existing flawed measure of angles is not informative.

COSMIN does consider assessing the relationship between a candidate outcome measure and other outcome measures under the umbrella of construct validity. However, this 'hypothesis testing' involves examining specified hypotheses about how a PROM should behave in relation to other outcome measures. As discussed, simply expecting a PROM to correlate well with angular deformity is not a sensible hypothesis to examine. Indeed, if a PROM did correlate strongly with angular deformity, then two alternative conclusions are possible. Firstly, the PROM is of no additional value compared to measuring angles. Secondly, the PROM may be invalid, given that angular deformity is itself a flawed measure. A better outcome measure might not necessarily correlate well with an existing flawed outcome measure. Although a single 'gold standard' is not available to resolve this issue, as already discussed. However, convergent and divergent validity against a

battery of other outcome measures, and careful consideration of the face validity of what matters to patients are central to clarifying this situation.

Another aspect of construct validity is to assess the internal relationships of items that comprise a PROM. Cronbach's alpha is commonly used to summarise the internal consistency of items. However, this assumes that the items contributing to a summary score reflect the same underlying construct. Theoretically, a factor analysis ought to be performed to confirm this prior to measuring Cronbach's alpha. This was done as part of the original validation of the URAM (Beaudreuil et al., 2011), and has been considered for the DASH previously (Forget et al., 2014, Franchignoni et al., 2010), but had not been considered for the DASH when treating Dupuytren's disease in pragmatic clinical practice. The present project identified factor loadings that suggest that the DASH is not unidimensional, nor is the URAM. Furthermore, they appear to reflect different constructs. As already discussed above in outcome measure development, it may be that more than one sub scale is required to assess both the correction of preoperative symptoms and the occurrence of postoperative complications. Two broad approaches to factor analysis may be employed: exploratory factor analysis and confirmatory factor analysis. Given the parochial development of a legacy PROM such as the DASH, its factor structure may not be easily predicted, hence exploratory factor analysis was used in this project. The relevant constructs for assessing outcome in

hand surgery in general, and in Dupuytren's disease may need further exploration before confirmatory factor analyses can be used for validation.

11.1.2.3 Responsiveness and Interpretability

Unlike responsiveness, interpretability is not considered as a measurement property of a PROM by COSMIN, but is acknowledged as an important characteristic (Mokkink et al., 2010). Both the DASH and the URAM demonstrated acceptable responsiveness in this project, as defined by effect size from before to after treatment. However when an external criterion, or anchor, was used to stratify outcomes and separate those that showed improvement from those that did not, the DASH was not interpretable, and an MIC could not be calculated. Although the retrospective anchoring used may not be ideal methodology for studying interpretability compared to prospective anchoring, it is the most widely used anchor found in the review of the literature conducted as part of this project. Besides the DASH in the present project, the systematic review of interpretability found only one other measure is uninterpretable: the Jebsen hand function test. Although assessing interpretability this way has limitations, it may be a more clinically relevant parameter than responsiveness, and an appreciation of relevant MICs is important in trial design and data interpretation as part of guideline development. However, the fact that the retrospective GRC anchor correlated better with postoperative state rather than with change in function, further supports the use of prospective anchors in future studies of interpretability.

According to the systematic review of interpretability, the most commonly used prospective anchor in hand surgery is the satisfaction domain of the Michigan Hand Questionnaire (MHQ), though there are issues with this. It has been used by the MHQ's developer, is frequently used to study the interpretability of other domains of the same tool (other parts of the MHQ), and it is unclear whether satisfaction is an appropriate construct to use as an anchor when assessing hand function. Further investigation of anchors is required.

Studying responsiveness may seem theoretically sound, as the complexity of appraising the performance of an anchor is avoided. However, given that treatment of Dupuytren's disease is not always successful in achieving correction of deformity, and that function might deteriorate from complications, the practicality of assessing responsiveness is a major challenge compared to studying interpretability. Furthermore, PROMs that are not comprehensive, as may be the case with the URAM, may still show desirable responsiveness. Studying interpretability, even given its limitations, may identify poor performance of PROMs that are not comprehensive in a way that is missed by looking at responsiveness alone.

Greater issues regarding interpretability also exist. This project has considered clinically meaningful change following treatment at an individual level. However, whether such a value can be applied to represent the clinical important difference between a bad average outcome from one treatment

compared to a good average outcome from a superior treatment, is not clear. It may be that the MIC is the best of a bad bunch of options, not just for determining meaningful change for an individual, but also for estimating a clinically important difference between individuals as is required for a power calculation for clinical trial comparing these options.

One other issue regarding interpretability was identified in the systematic review of MICs conducted as part of this project. To date, no data exists to describe the MIC for deterioration after treatment. This is particularly important in the context of a progressive condition such as Dupuytren's disease, where recurrence occurs frequently, and may adversely impact on function. Indeed, undergoing multiple procedures was associated with worse functional outcome at five years in this project. These patients may experience a combined effect, with functional impairment arising from both the complications of previous surgery as well as from recurrence of disease. MIC for deterioration could be studied through a cohort study with the use of an appropriate anchor. Given the issue with recall bias that limits the value of retrospective anchors that were confirmed in this project, studies of interpretability of deterioration over the years following treatment should use prospective anchors. This would be achieved by the patient completing the anchor at baseline or preoperative assessment and again at follow up, and the change in the anchor item would then be used to subgroup the outcomes.

11.1.3 Validity, reliability, responsiveness and interpretability of specific outcome measures studied

Investigations of the performance of outcome measures in this thesis have revealed several findings, including limitations of validity of the two PROMs that might be candidates for use in future studies.

11.1.3.1 The Unité Rhumatologique des Affections de la Main scale

As the first Dupuytren's disease-specific PROM to be developed, the URAM has been an important tool to assess in this project. It may not be cross-culturally sensitive for use in the UK, or may not be comprehensive in terms of content validity, as it did not describe some common important goals for patients in this project. The state of validation of the URAM scale is summarised in Table 11.1. Its interpretability was at the threshold of acceptability in this project, based on the area under the receiver operating characteristic (ROC) curve being 0.74, virtually at the lower limit of acceptability of 0.75 (Fan et al., 2006).

In contrast to the more parochial legacy measures used in Dupuytren's disease, modern tools such as the URAM were developed using contemporary methodology (Beaudreuil et al., 2011). Consequently, the URAM would be expected to perform better than other measures in assessments of its performance. Despite this, limitations in its performance have been identified in this project. The methodology used to develop the

URAM (Beaudreuil et al., 2011) may have contributed to the problems with the scale's performance. In the development process, a long list of candidate items was generated to describe the symptoms that patients might experience from Dupuytren's disease, but not necessarily what they might experience from complications of treatment. As discussed, this may limit the comprehensiveness of the URAM for assessing function after treatment, when complications may be relevant to assess.

The candidate item list was narrowed down based on defined criteria, such as eliminating candidate items that were never performed by more than 5% of the study population, or eliminating items if the range of responses provided by the study group were not spread across five or more of the six options provided in the scale (0-5) (Beaudreuil et al., 2011). The overall long list was not provided, nor was a breakdown of which items were excluded on the basis of which criterion so it is only possible to speculate on how this methodology may have affected the final scale. However, there are learning points for the future. While the criteria for excluding candidate items seem sensible, they have reduced its comprehensiveness. If a task was never performed by 5% of the French URAM design cohort, then it will have been eliminated from the scale. Improving pain was cited as a goal by 15% of patients in the current project, and if a similar proportion experienced pain preoperatively in the URAM design group, then it may have been eliminated on the basis of this, or may not have been included given the task-based nature of the items in the

URAM. However, pain was not only seen as a preoperative symptom, but also increased after surgery in our prospective cohort, and so may be important to study. In other clinical conditions, the symptoms and functional limitations of the disease may closely resemble the symptoms and limitations experienced following complications of treatment. For example large joint osteoarthritis may present with pain, and pain might occur after a complicated arthroplasty. In contrast, the treatment of Dupuytren's disease is different. The complications of treatment identified in this project, and in previous work (Crean et al., 2011), may manifest with different symptoms and limitations from the preoperative symptoms. The resulting URAM scale was unidimensional in an exploratory factor analysis conducted by the developers (Beaudreuil et al., 2011). However, in the current project, an exploratory factor analysis was performed using the URAM items and the DASH items. In the current analysis, the URAM was not unidimensional. Instead, the final item in the URAM scale ("difficulty picking up objects with the thumb and index finger"), loaded on a different construct from the other items. It is notable that this item may not directly relate to loss of extension in the same way as the other eight items in the URAM. Furthermore, the thumb and index fingers are typically least affected by Dupuytren's disease. Instead, it loaded with task-based items in the DASH.

The disparity between this result and that obtained by the URAM developers may relate to entering all the DASH and URAM items in the same factor

analysis. Typically each factor or construct generated extracted in a factor analysis should be reflected by several items for the analysis to be reliable (Floyd and Widaman, 1995). Given that only item nine of the URAM loaded on a different factor to items one through eight, this may not have been detected in the factor analysis run by the developers, and only became apparent in this thesis as here constructs may have been better modelled by the incorporation of DASH items that behave similar to item nine of the URAM. Whether this lack of unidimensionality is large enough to be of clinical significance is not clear. The disparity does demonstrate that even contemporary techniques of validation have limitations, and require cautious interpretation.

The exploratory factor analyses extract underlying mathematical constructs that account for proportions of variance seen across the study cohort. However, the analyses cannot explain what exactly the underlying constructs are. Instead, these require logical interpretation. Understanding the constructs reflected by the URAM is potentially straightforward. All nine items in the URAM are activity/participation restriction items, and all were generated from patients and clinicians aiming to capture the preoperative experience of Dupuytren's disease (Beaudreuil et al., 2011). The URAM correlates most closely with loss of extension (Beaudreuil et al., 2011), and items such as difficulty stroking or caressing are more likely to be impaired by loss of extension in the preoperative state, than loss of flexion as a postoperative

complication. Preoperative symptom items should improve with successful treatment, but treatment complication items would deteriorate in cases in which adverse effects occur. As a result, they may load onto different constructs that together provide a more complete reflection of hand function. If they do load differently, then disability or hand function might be better assessed as two separate summary scores. If this is the case, then although the URAM may not be comprehensive as an assessment of disability or hand function in isolation, it might provide a sub scale summary score that describes the former, but a separate sub scale needs to be developed to assess postoperative sequelae as the latter. Indeed, as URAM item nine (which assesses fine pinch grip) loaded separately from the other URAM items when a model was built using DASH and URAM items together, it may constitute the first item for such a sequelae subscale.

ASPECT OF VALIDITY	SUBTYPE	DATA IN PREVIOUS LITERATURE	DATA IN CURRENT PROJECT	COMMENTS
Criterion validity		No	No	Not appropriate
Construct validity	Structural validity	Yes (Beaudreuil et al., 2011)	Yes	Exploratory factor analyses used in both. Unidimensional in developers' study, but not unidimensional in current project when model incorporated DASH items.
	Hypothesis testing	Yes (Beaudreuil et al., 2011)	No	Compared to angular deformity and other PROMs with defined hypotheses
	Cross-cultural sensitivity	No	Yes	Cross-cultural sensitivity to UK patients studied. May not be cross culturally sensitive

Content validity	Comprehensiveness	No	Yes	URAM may not be comprehensive for use based on UK patients' goals. May not be appropriate for comprehensive assessment of postoperative state
	Relevance	No	Yes	All URAM items matched goals provided in this project, therefore, they are likely to be relevant, even if there were not enough items to comprehensively cover goals.

Table 11.1: State of validation of URAM scale

Despite these concerns, the interpretability of the URAM was reasonable in this project. The area under the ROC curve achieved was at the threshold of acceptability. This is despite the study being conducted with an anchor that may be less appropriate than prospective anchors. The interpretability of the URAM should be confirmed with a prospective anchor. The MIC and MID obtained for open surgery in the form of fasciectomy and dermofasciectomy was considerably higher than the MID reported for needle aponeurotomy (Beaudreuil et al., 2011). This may be influenced by the different methods used to estimate the MID. In the current project, the difference between the mean change in the improved group and the mean change in the stable group was considered to be the MID, whereas regression analysis against satisfaction was used in previous literature (Beaudreuil et al., 2011). However, it is more likely to reflect the differences in rehabilitation between the less invasive aponeurotomy procedures and open surgery (fasciectomy or dermofasciectomy). It is likely that a greater improvement in function is needed after open surgery before patients consider the prolonged recovery worthwhile.

Although the MIC and MID of the URAM still requires further investigation, the marked difference between the MID for surgery and that for aponeurotomy may pose an issue in the design of future randomised trials. Even if the URAM is not considered to be the optimal primary outcome measure for such

a trial, similar differences between MICs for aponeurotomy and open surgery are likely to be obtained for other outcome measures, for the reasons discussed. What data would then be used in power calculations for such trials is not clear and will require further consideration.

11.1.3.2 The Disabilities of the Arm, Shoulder and Hand tool

The DASH is the most commonly used PROM in studies of Dupuytren's disease (Ball et al., 2013). It was produced in the 1990s (Hudak et al., 1996), and thus its development and validation were mainly performed prior to current standards such as those specified by COSMIN. The DASH was developed specifically to assess the entire upper limb as a single domain. Item generation for the development of the DASH was based on a literature review, without any direct patient input. Item reduction was then performed in two stages, the first of which involved expert opinion (Hudak et al., 1996). Consequently, it may be argued that it is not patient-centred. Furthermore, it is unclear whether the entire upper limb constitutes a single domain, particularly when using the DASH to study a hand-specific condition such as Dupuytren's disease.

The assumption that the entire upper limb functions as a single domain may not be appropriate: items such as 'difficulty placing an object on a high shelf' would be impaired by shoulder disease, but not necessarily by Dupuytren's disease or other hand conditions. The DASH's user manual includes a review

of papers investigating its construct validity (Kennedy et al., 2011). The methods used in the papers summarised in the DASH's user manual mainly comprised hypothesis testing of the DASH's relationship to other outcome measures using correlation coefficients. They conclude that as hypothesised, the majority of such papers demonstrate that the DASH correlates well with similar measures, and poorly with measures measuring different entities. However, one paper that is not well discussed found the DASH to be so broad that it is responsive to lower limb injuries as well as upper limb conditions (Dowrick et al., 2006).

These issues can be explored further from the findings of the current project. Whilst the project supports the theory that the URAM scale may be too narrow to provide a comprehensive reflection of hand function in Dupuytren's disease, data have been generated that suggest that the DASH may be too broad.

In the exploratory factor analyses in the present project, DASH items loaded on more than one construct, suggesting that the DASH is not unidimensional when used in Dupuytren's disease. Given that the DASH was developed to reflect a single construct, this finding would not be expected, but is in keeping with previous studies (Forget et al., 2014, Franchignoni et al., 2010).

The variance explained by the first extracted construct was much greater than that explained by other factors in both the DASH item model and DASH and

URAM item model. This may mean that the second and third constructs are not particularly clinically relevant. Alternatively, the limited amount of shared variance explained by the second and third constructs here may be because the majority of items (DASH items one through 21 and URAM item nine) loaded on the first construct, so the model is better able to extract the first construct as a result. Assuming that the DASH does reflect two distinct constructs, it may be more appropriate to generate two separate summary scores for subscales based on these different groups of items. However, understanding what these summary scores represent may not be straightforward, as this is dependent on confidently interpreting the factors extracted in the EFA models. For the DASH, this may be more complex than for the URAM. In the EFA of DASH items, the activity-limitation items in the DASH loaded on a distinct factor from the impairment perception-based items. Within the World Health Organisation's International Classification of Functioning, structural impairment, functional limitation and participation restriction are separate tiers (World Health Organisation, 2014). It may be that these three tiers are different domains, and that that the DASH items reflect two of them that both are relevant to the performance of the upper limb, but are separate from each other.

However, this theorem is not supported by the later exploratory factor analysis, in which both DASH and URAM items were studied in a single EFA model. In this later model, the URAM items expected to relate most closely to

loss of extension generally loaded with a third latent factor, distinct from the factors that DASH items loaded onto. The exception was URAM item nine, which loaded alongside DASH activity-limitation items. This suggests that the DASH items may not reflect that functional impairment resulting from loss of finger extension in Dupuytren's disease.

The most appropriate interpretation might be a combination of both of the above. The URAM correlated most closely with loss of extension when it was validated by the developers (Beaudreuil et al., 2011), and previous studies that explored the correlation between the DASH and angular loss of extension, all of which found weak or no correlation (Degreef et al., 2009, Engstrand et al., 2009, Jerosch-Herold et al., 2011, Zyluk and Jagielski, 2007). Given that URAM items one through eight load with one factor and are expected to reflect loss of extension rather than overall hand function, it is possible that DASH items assess the remainder of hand function, along with URAM item nine. When considered together, items 22 through 30 of the DASH resemble the second part of the Patient Evaluation Measure (PEM) tool (Macey et al., 1995), another common hand function legacy PROM. However, URAM item nine loaded alongside the activity and participation items of the DASH (items one through 21), rather than the impairment perception items. This may be due to differences between these groups of items, in keeping with their relative positions within the World Health Organisation's International Classification of Functioning, as already discussed.

ASPECT OF VALIDITY	SUBTYPE	DATA IN PREVIOUS LITERATURE	DATA IN CURRENT PROJECT	COMMENTS
Criterion validity		Yes (Degreef et al., 2009, Engstrand et al., 2009, Jerosch-Herold et al., 2011, Zyluk and Jagielski, 2007)	No	Not appropriate, as no gold standard against which the DASH can be benchmarked
Construct validity	Structural validity	Yes (Franchignoni et al., 2010)	Yes	Exploratory factor analyses used in both. Not unidimensional in either.
	Hypothesis testing	Yes (Kennedy et al., 2011)	No	Compared to angular deformity and other PROMs with defined hypotheses

	Cross-cultural sensitivity	Yes (Kennedy et al., 2011)	No	Developed in English. Cross-cultural sensitivity studied in a variety of settings and languages. Not specifically in Dupuytren's disease, or for UK use.
Content validity	Comprehensiveness	No	No	
	Relevance	No	No	

Table 11.2: State of validation of DASH in Dupuytren's disease

The DASH was not interpretable in this project. This was due to there being a low area under the ROC curve used to estimate the MIC, and there being no significant difference between the mean changes in the 'improved' subgroup and the 'stable' subgroup. There are limitations to all methods of assessing interpretability, which have been discussed already. In particular the use of the GRC retrospective anchor may be problematic. Even still, this methodology has been widely used and the systematic review of MICs conducted only found one scenario in which a recognised outcome measurement was not interpretable, for the Jebsen hand function test. Most other studies of interpretability of relevance to Dupuytren's disease had used the same potentially flawed methodology, and the outcome measures studied had exhibited adequate interpretability to allow the estimation of an MIC or MID. Therefore, the inability of the DASH to identify clinically meaningful change here raises concerns about its on-going use in Dupuytren's disease. Given the close agreement between the DASH and the QuickDASH in this project, the QuickDASH may also not be suitable for use in Dupuytren's disease either.

In contrast to the interpretability results, the DASH exhibited moderate responsiveness. However, when responsiveness is classified in this way, it is not easy to appraise the performance of the measure, and very poor interpretability may exist in the presence of 'good' responsiveness when

studying a cohort of patients in which a significant proportion did not experience clinically meaningful change. The interpretability data from this project reinforce these points.

11.1.3.3 Passive extension deficit

Although the use of 'gold standards' to study the criterion validity of an outcome measure is considered inappropriate in most settings (Mokkink et al., 2010), correlation of PROM scores with angular deformity is widely used to determine the validity in Dupuytren's disease (Beaudreuil et al., 2011, Degreef et al., 2009, Engstrand et al., 2009, Jerosch-Herold et al., 2011, Zyluk and Jagielski, 2007). However, assessing the correlation of a candidate outcome measure with other outcome measures can be considered as a hypothesis test, under the umbrella of construct validity in the framework established by COSMIN. The validation of the URAM involved comparison to a range of other outcomes, including some with which the URAM was expected to show poor correlation (Beaudreuil et al., 2011). The distinction might be best considered by examining the hypotheses made. Even though the URAM's developers made a range of distinct hypotheses for different comparisons, the URAM was still expected to correlate best with angular deformity, and when this pattern was observed, this was considered to represent success. This approach towards extension deficit as a gold standard contradicts the consensus established by COSMIN, and might be considered illogical; if angular deformity is the gold standard, then a PROM that simply correlates

very strongly with angular deformity may not provide any advantage over just measuring joint angles directly. Indeed, this may also be inconsistent with the PROM being truly patient-centred, as direct measurement of angular deformity is a surgeon-centred outcome.

Putting aside the issues of using angular deformity as a gold standard for validating PROMs in Dupuytren's disease, the validity of loss of extension as an outcome in its own right should still be subjected to the same analyses as other outcome measures. The broad range of patient goals identified in the present project suggests that loss of extension alone may not adequately describe the goals and expectations of patients for their planned treatment. Based on the terminology established by COSMIN, loss of extension may not be comprehensive as an outcome measure. It might not even be relevant. The latter may apply when considering specific cut offs for angular deformity to define good and bad outcome or recurrence versus no recurrence. A range of different angular measurements has been studied in the Dupuytren's disease literature (Ball et al., 2013). Study endpoints include correction of angular deformity to 0^{o} to 5^{o} of loss of extension per joint and recurrence defined as increase in the postop contracture by 20^{o} per joint (Hurst et al., 2009). However, the review of interpretability conducted as part of this project failed to identify any evidence to support these cut offs. Not only may they not be evidence-based, they may also be illogical; it is unlikely that correction of a 15^{o} joint contracture to 5^{o} would be considered success, yet correcting a 90^{o}

contracture to 10^{o} would be considered as a failure. Similar logic may be applied to assessing fixed cut offs for recurrence. These particular end points may not be clinically meaningful, and if other possible cut off options also fail to separate meaningful changes and differences, it might be argued that angles are not relevant at all in terms of patient-centred content validity. However, change in angle might be more meaningful. Rather than requiring all joints to achieve correction to zero to five degrees, improving total extension deficit across all joints in a ray by 20 or 30 degrees might describe a meaningful improvement for many patients. This may capture meaningful change for patients with different goals. For example, achieving full correction for a patient who has presented with a mild contracture, but who needs to be able to place the palm flat for work tasks, is as meaningful as allowing the fingers to extend out of the palm to facilitate hand hygiene for a patient who has presented much later with severe contractures and the fingertips close to the palmar skin. This may explain the diverse range of patient goals identified in the present project.

Measurements such as loss of extension are often categorised as ‘objective’, in contrast to ‘subjective’ PROMs. However, there are potential sources of bias that might arise from measuring angles, such that they may not be quite so objective. The effect of dynamism on passive extension deficit angular measurements was considerable in this project, and as it had not been quantified, it may have been underestimated previously. Dynamism was

extremely common in cords crossing the PIPJ, and could influence the classification of treatment success versus failure, or recurrence versus no recurrence in studies. However, the methodology used to study dynamism in the present project aimed to identify the maximum amount of dynamism that could be achieved. This extent is unlikely to be encountered in a trial in which observers are trained to perform measurements. A recent study of Dupuytren's disease assessment found that inter observer reproducibility of angular measurements were better than the dynamism data in the present study would suggest (Broekstra et al., 2015). However, there are numerous limitations to Broekstra and colleagues' study. As an agreement study, bias that might lead to an observer consciously or subconsciously influencing the measurements based on the treatment that a patient had received would not be present. The systematic review of trials conducted in the present project suggests that this assumption is unlikely to be the case in future studies. Unblinded assessment of outcome posed a risk of performance bias in nearly all trials to date. Furthermore, the inter- and intra-observer reproducibility were evaluated using intra-class correlations. This method is flawed (Bland and Altman, 1986), as has been discussed in the present project's discussion on the relationship between the DASH and QuickDASH.

There are strategies that might be adopted to minimise the influence of dynamism on outcome measurement. Assessing active extension deficit may avoid this risk of bias, and is likely to be more clinically relevant than passive

extension deficit, as the active extension deficit measures the angles that the patient achieves himself or herself. Assessment of angles in studies of Dupuytren's disease remains important, particularly in explanatory studies of treatments, and given the complicating effect of differences in finger joint anatomy between the MCPJ and the PIPJ discussed in the introduction. However, it may be more appropriate to use such objective measures as secondary, rather than primary, endpoints in pragmatic trials and assessments of cost effectiveness. The interpretability of angular measurements, and the cut offs used, needs to be studied in more detail.

11.1.4 Choice of outcome measure

The findings of the present project, including the review of interpretability studies applicable to Dupuytren's disease, demonstrate that the existing evidence base is not adequate to make firm recommendations regarding which outcome measures to use for the study of Dupuytren's disease in clinical practice, research or guideline development. Further work is required to determine whether other existing legacy measures perform well in Dupuytren's disease, according to current standards of validity, interpretability and responsiveness.

The results of the present project suggest that there may be different constructs of relevance to Dupuytren's disease, and different outcome measures may reflect distinct underlying constructs. As discussed, the DASH

may not be unidimensional, and face validity examination of the task-based items in it suggest that the main construct that it reflects may not be of relevance to hand function in Dupuytren's disease. In keeping with this, the DASH was not interpretable in the present project. Most of the URAM items may reflect disability related to loss of extension, but the scale does not necessarily assess global hand function.

As a result, even once further validity data have been generated and appraised, the most appropriate choice of primary outcome measure in Dupuytren's disease is likely to depend on intended purpose. In early phase and explanatory studies, such as a study of correcting PIPJ contractures with a novel untested technique, then active extension deficit angles may still be the most appropriate assessment. If this is the case, more careful consideration of meaningful cut offs for endpoints will need to be considered than has been done to date. However, for most applications involving pragmatic studies of clinical outcome and cost utility, it is likely to be more meaningful to assess global hand function than to measure angular deformity, or to assess just the disability associated with loss of extension in isolation. As discussed, this may involve a modification of the URAM scale, with a separate subscale to capture complications, or a completely novel outcome measure.

Further examination of candidate outcome measures for the assessment of hand function in Dupuytren's disease is needed, as the results of the present project do not support the on-going use of the DASH in its current form in Dupuytren's disease. Although there have been learning points derived from this project, such as the potential limitation of studying interpretability using a retrospective anchor, the interpretability of the DASH is so poor that it is improbable that it would exhibit a high level of interpretability even with a prospective anchor. The present project findings suggest that the URAM is a more suitable, though still flawed, outcome measure that could be used.

11.1.5 Outcome and associated factors

In the same manner that continuing to assess passive extension deficit is likely to be inadequate as the primary outcome measure in pragmatic trials, using recurrence as the primary endpoint of similar Dupuytren's disease research is probably not appropriate, given that in the present project, it was not comprehensive in relation to patients' preoperative goals, or to postoperative treatment complications. The latter were common after treatment in this project. As a result, inappropriate prioritisation of recurrence over global hand function might promote the use of more invasive treatments, such as dermofasciectomy, that achieve greater deformity correction and less recurrence, but place patients at risk of poor global hand function outcome due to their relatively greater iatrogenic insult and greater potential for complications that affect hand function.

Factors that are independently associated with poor outcome should be appreciated. This may foster improved clinical care by identifying those at risk of poor outcome at the outset, and would promote high quality research design, conduct, and interpretation in clinical guideline development. This project has demonstrated that the factors associated with a symptomatic upper limb following treatment for Dupuytren's disease are not the same as the factors that comprise the Dupuytren's diathesis. Female gender, previous ipsilateral surgery and diabetes mellitus were associated with poor function (defined here as a DASH score of 15 or more, the cut off considered by the developers' review of the literature to represent the threshold at which the upper limb becomes symptomatically limited (Kennedy et al., 2011)). In contrast, the traditional diathesis factors considered to be associated with disease recurrence include male gender and early age of onset. Some of the present study findings make sense. Repeated operations on the same hand will lead to accumulation of iatrogenic insult, with more scar tissue, and greater risk of complications such as stiffness and altered sensation. Furthermore, diabetics may experience delayed wound healing and a greater risk of infection, which might be expected to affect their functional outcome after surgery.

However, this study investigated associations with poor function, and causality in the relationship could not be studied using this cross sectional study design.

The relationship between some factors and poor functional outcome is not immediately clear: why there was a strong association between female gender and poor functional outcome following surgery for Dupuytren's disease is not immediately apparent. There are other limitations to this aspect of the project. The definition of poor functional outcome was a cut off of 15/100 on the DASH score. As discussed already, the DASH's performance in Dupuytren's disease in the present project was poor, and the interpretability of specific cut offs such as this may not be evidence based. The latent constructs reflected by the DASH differed from the URAM, and may not relate to Dupuytren's disease. As a result, the factors identified in this cross sectional study may be associated with confounding comorbid upper limb conditions. For example, diabetics also have a higher incidence of other upper conditions, such as trigger finger and carpal tunnel syndrome in the hand and adhesive capsulitis in the shoulder. The poor functional outcome seen might be a result of this. One step towards explaining this would be to study a cohort of preoperative Dupuytren's disease patients. Preoperative function and the prevalence of upper limb comorbidities would need to be considered.

Whilst the results of the present project probably do not provide robust evidence of the factors associated with poor functional outcome, they do demonstrate that the traditional diathesis factors cannot be assumed to be the only important predictors of poor outcome in Dupuytren's disease. This is an important finding in its own right, and further work to study the causes of poor

outcome using a more responsive and interpretable measure than the DASH is required. However, in the interim, Dupuytren's diathesis factors cannot be extrapolated to predicting poor treatment outcome beyond recurrence.

11.1.6 Comparing treatment options

The systematic review conducted in the present project confirmed that few randomised studies of surgery for Dupuytren's disease exist, and that methodological weaknesses were encountered in all of them. No randomised trial data exist to describe the investigation of some areas of equipoise in current clinical practice. For example, no randomised trial data were identified that compared surgery to collagenase, or that compared fasciectomy to conventional dermofasciectomy, rather than firebreak skin grafting. The current state of the evidence base limits the establishment of optimal clinical practice, and the development of clinical guidelines. Further comparative data is needed to determine optimal practice for the treatment of Dupuytren's disease.

As already discussed, further investigation of outcome measures is still required before optimal randomised controlled trials could be designed in Dupuytren's disease, both in terms of outcome selection, but also in terms of powering a study to identify an clinically meaningful difference between treatments. However, some learning points can be established based on the systematic review of trials. Greater efforts must be made to ensure that risks

of bias in trials in Dupuytren's disease are minimised. Some risks of bias that were encountered are readily avoidable. For example, adequate randomisation processes and allocation concealment measures should be employed. As blinding may prove difficult in this context, blocked randomisation is best avoided. Patient involvement in study design is important, and may further support the adoption of patient-reported hand function or quality of life as the outcome constructs of greatest importance. Good methodological practice was also identified in the review: where appropriate intraoperative randomisation used (Ullah et al., 2009) that may minimise the risk of performance bias that is otherwise difficult to address in trials of surgery.

The observational study of outcome performed as part of the present project constitutes low quality evidence using the GRADE criteria applied in the systematic review of trials. However, as with the study of factors associated with poor outcome, it does question conventional dogma regarding Dupuytren's disease treatment. Although there are limitations to the data presented, the logistic regression analysis identified no difference in functional outcome at five years between needle aponeurotomy and excisional surgery. This might not be expected from previous studies, such as the trial comparing needle aponeurotomy and fasciectomy included in the systematic review conducted here (van Rijssen et al., 2006, van Rijssen et al., 2012), where a very large difference in recurrence rates was observed between the different

procedures. However, the functional outcome findings can be potentially explained and supported by much of the data in the present project. Given that patient goals did not solely reflect loss of extension, and that complications were much more common after more invasive treatment, the late functional outcome of treatment might be expected to reflect the balance of maintained correction of angular deformity offset against the risk of complications. Although the DASH may not be the optimal outcome measure for studying Dupuytren's disease, the URAM might not be more suitable for this analysis. Given that the URAM correlates closely with angular deformity and may mainly reflect loss of extension, it might fail to identify the impact on global hand function that the present study has alluded to, in the same way that previous studies using recurrence of angular deformity may also have missed the impact of complications. This merits further investigation.

11.2 Qualitative findings from patient interactions

In addition to the data presented in the chapters of this thesis, the candidate identified recurring themes from performing the patient evaluations. The assessments for the cross sectional typically took 20 to 30 minutes per patient, and the cohort study preoperative assessment typically lasted 15 minutes. Both assessments are longer than the typical 10 minutes assigned for an NHS patient assessment in clinical practice.

Formal structuring of the interactions had not been performed, and so detailed analysis was not attempted. However, aspects of relevance to the interpretation of the present project, other literature, and to future study design were considered.

One pattern that was noted was that patients at a centre tended to favour the treatment modality preferred by that centre. For example, at a centre that performed aponeurotomies, the patients often discussed their perceived superiority of that treatment. The advantages cited during informal conversation included rapid recovery, minimum inconvenience and rehabilitation, and a willingness to undergo serial treatments for recurrence. This was particularly noticeable for self-employed individuals who were keen to minimise time off work. In contrast, patients at centres that preferred fasciectomy often identified the high recurrence rate of aponeurotomy as being undesirable.

The above pattern is complex to interpret, and there are several potential explanations, which are not mutually exclusive. Patients who preferred a particular treatment option may have specifically sought referral to a centre with a reputation for that procedure. At least one centre in this study did receive a significant minority of its referrals from outwith its local catchment area, probably for this reason. However, in the cross sectional, patients who resided over 20 miles away from the centre were excluded. This clause was

introduced with the aim of reducing patient inconvenience and fatigue, as some patients were already contributing to other research projects. However, it did mean that patients who had travelled a long distance to receive a particular treatment modality were excluded from this project and so this should not have influenced the findings. This may be an important point to consider in future trial design to ensure that trial groups are comparable. Furthermore, it demonstrates that patients may choose a treatment 'strategy' incorporating the possibility of multiple treatments over a period, rather than choosing a single procedure. This is of relevance to study design and to guideline development, as straightforward head-to-head comparison of treatments may not be appropriate in a pragmatic trial. Given the wide disparity in recurrence and reoperation rates between needle aponeurotomy and fasciectomy observed both in the present project and in previous studies (van Rijssen et al., 2012), it may be more appropriate to study randomise patients to undergoing a series of aponeurotomies over a several year period versus a single more invasive procedure such as fasciectomy.

Alternatively, a form of social desirability response may have been encountered, with the patients influenced by their surgeons and other clinical staff. All cross sectional study patients had undergone treatment at the time of recruitment. Several had been unaware of the nature of Dupuytren's disease, or of its treatment options prior to clinical assessment. Having discussed the diagnosis and options with the clinician at the treating centre,

they may have been influenced by the surgeon's opinions and preferences. This may also apply to the study of patient goals. Although recruitment for that study was performed at the preoperative stage, a proportion of patients had already undergone treatment for Dupuytren's disease previously, and so may have been influenced by their previous clinical counselling and experience. Minimising the potential effect of this in future studies might be achieved by only including patients undergoing primary treatment, and by standardising the clinical counselling provided to study participants.

Given the importance of aspects such as those discussed in this section to study design, a formal semi-structured interview study with patients at preoperative and early and late postoperative stages of treatment may be required to ensure that the data generated are truly patient-centred where appropriate.

11.3 Future work

11.3.1 Outcome measure appraisal

11.3.1.1 Validity of outcome measures

In the present project, aspects of the DASH's validity and interpretability in Dupuytren's disease that had not been previously examined were studied. Overall, it performed poorly. There are further analyses of validity that could also be performed. For example, COSMIN's standards also suggest the use

of confirmatory factor analysis to examine structural validity. However, this was not deemed appropriate for the DASH, given that failings relating to its structural validity and interpretability had already been identified. Confirmatory factor analysis could be applied to the URAM to confirm the factor structure identified in the present project, which differed from that obtained by the developers (Beaudreuil et al., 2011).

Additionally, Rasch analysis has also be used to study validity (Franchignoni et al., 2010). This would constitute a shift to the use of item response theory analyses rather than classical theory analyses such as EFA. Although Rasch analysis is not discussed by COSMIN, the current COSMIN checklist is now at least five years old. Rasch analysis-validated outcome measures have been included in recently-developed standardised outcome sets established by The International Consortium for Health Outcome Measurement (ICHOM) for other health conditions (International Consortium for Health Outcome Measurement, 2014). Whilst the present project was conducted using the criteria set by COSMIN, it is quite possible that the use of Rasch analysis for validating outcome measures will become standard practice in future. As a result, further validation of legacy and novel PROMs should consider employing such methods.

It is likely that the validity data available will be inadequate for other legacy PROMs that could be used in Dupuytren's disease, such as the Patient

Evaluation Measure (PEM) (Macey et al., 1995), or the Michigan Hand Questionnaire (MHQ) (Chung et al., 1998). These PROMs should also be subjected to modern validity analyses in Dupuytren's disease. Based on the findings of the present project, it is reasonable to hypothesise that they will perform better than the DASH and even the URAM. The second part of the PEM comprises items that assess the patient's perceptions of impairment such as pain. As a result it is more likely to be unidimensional than the DASH. The MHQ comprises several subscales to assess distinct constructs separately. As a result, the factor structure of its items is hypothesised to be more appropriate than the DASH or even the URAM, and it may perform more reliably. However, the MHQ is lengthier and harder for patients to understand and to complete (Dias et al., 2008).

11.3.1.2 Responsiveness and Interpretability of outcome measures

Further work is also required to be able to interpret the outcome measures used in Dupuytren's disease. The present project provided new data regarding the interpretability of the DASH and URAM, using a retrospective anchor. Given that the GRC anchor correlated better with postoperative state than with change in score, further studies might include the use of prospective anchors to compare their performance. The systematic review of interpretability performed in the present project identified that a number of studies had been conducted using the satisfaction subscale of the MHQ as the anchor. However, the developer of the MHQ led all these. Not only

should consideration be afforded to the prospective versus retrospective nature of anchors, but also their face validity should be considered. The GRC provides a very general patient perception of improvement. Both this and the change in the satisfaction domain of the MHQ may instead reflect a different construct from using a prospectively assessed measure of health-related quality of life. Furthermore, it is important that future studies clearly distinguish between minimal important differences (MIDs, between individuals) and minimal important changes (MICs, for individual patients). The previous estimate of the MID for the URAM, based on regression against a satisfaction score, constitutes an MIC, but not an MID.

As well as these conventional studies of interpretability, there are other uses of outcome measures that require investigation of their interpretability. As long-term outcome is affected by complications, as studied in the present project, and also by gradual recurrence, the interpretability of late progressive deterioration needs to be studied as well. According to the systematic review of interpretability in the present project, the few MICs and MIDs calculated for Dupuytren's disease so far all relate to the interpretability of improvement early after treatment, and none to the later deterioration. Outcome measures could also be used for other purposes in Dupuytren's disease. In theory, an appropriate PROM could be used to screen patients in primary care to select those at an appropriate stage for treatment in secondary care. This would also require studies of the interpretability of an appropriately validated PROM.

11.3.1.3 Design of novel PROMs

If the above future work is conducted, it may be the case that none of the current legacy PROMs is ideal for use in Dupuytren's disease. In this scenario, it may be necessary to design a novel PROM. The URAM scale was developed relatively recently, and using more modern methodology than other legacy measures such as the DASH and the PEM. However, item generation involved identification of limitations reported by patients with Dupuytren's disease and clinicians who treat Dupuytren's disease. As already discussed in the study of British patients' goals, this may not necessarily generate a long list of items that are relevant or cross culturally sensitive. Additionally, given the factor structure that emerged in the present project when URAM items were combined with DASH items, only item nine within the URAM may assess complications or global hand function. In Dupuytren's disease preoperative impairments, as described in the study of patient goals, differ considerably from the impairments due to complications of treatment, which were also studied in this project. In order to design a PROM that is suitable for use before and again after treatment, its development should incorporate the postoperative experience of treatment in addition to the preoperative symptoms. As already discussed, this may result in two distinct constructs being assessed, resulting in separate subscale scores.

11.3.2 Comparative study design

11.3.2.1 Considerations for future comparative study design

As has been discussed, more work is necessary to identify the best outcome measure for use in studies of Dupuytren's disease. Consequently, discussing the design of comparative studies may be premature. However, there are some points that have arisen from the present project, both directly in experiments and indirectly through the qualitative element of patient interactions by the candidate, which might improve the subsequent design of comparative studies.

As a range of treatments are currently employed, and as clinical equipoise may be encountered between treatment strategies, it is likely that pragmatic trials of treatments would be most informative. However, given the marked differences between treatment modalities such as aponeurotomy and fasciectomy, head-to-head comparison of single treatments may not be appropriate. Indeed, many of the patients assessed following aponeurotomy in the present project accepted that there was greater risk of recurrence, but they accepted this as they prioritised a more rapid recovery from the initial treatment. Therefore, it might be argued that a single fasciectomy ought to be compared to a series of aponeurotomies performed over a period of say five years. This might be akin to comparing a series of steroid injections to control knee osteoarthritis pain versus a single knee replacement operation.

While the above suggestion may more accurately reflect the pragmatic clinical approaches to managing Dupuytren's disease, it might pose methodological challenges for study design. In the present project, the URAM's MIC for surgery was considerably larger than the developers' previously published MIC for needle aponeurotomy. As interpretability metrics are influenced by the context, including treatment type, it is likely that the MIC for aponeurotomy will be significantly lower than the MIC for surgery calculated in the present project. In such circumstances, selecting the correct difference when powering and interpreting comparative studies will be challenging.

In addition to the information regarding outcome measure selection and optimisation for a future comparative study, this thesis has also identified other areas of trial design that will require consideration. Efforts will need to be made to minimise performance bias in a comparative trial. The systematic review of existing trials demonstrated that all previous randomised studies have been risk of bias. In particular, innovative and carefully considered solutions to minimise performance bias are required. This was only rarely achieved in previous studies, where one of the better examples involved intraoperative randomisation (Ullah et al., 2009).

The findings of this thesis demonstrate that considerable work is still required to be able to design and conduct optimal high quality randomised controlled trials in the treatment of Dupuytren's disease. However, it reaffirms the extent

and breadth of the impact of this condition on patients, and the value of investing in further improving treatment for patients with Dupuytren's disease. The particular challenges identified here are also applicable to other aspects of hand surgery, and to other non-terminal conditions. As a result, resolving the issues in Dupuytren's disease discussed in this thesis has the potential to modify practices and raise the standard of clinical research in a range of clinical areas. However, a proposal for a trial that incorporates the novel findings from this thesis will be presented, to summarise the current state of research in this area.

11.3.2.2 Proposed clinical trial protocol

Based on the previous research of the treatment of Dupuytren's disease and the findings of the experiments in this thesis, a trial outline proposal is presented. Based on the findings of the first systematic review presented in this thesis, there are several unresolved uncertainties in Dupuytren's disease treatment. In particular, the age-old issue of disease excision versus disease division remains inadequately investigated. However, in the first instance the differences between different excisional procedures needs to be established, and likewise the differences between different divisional procedures. Once the superior treatment option within each strategy is identified, then comparison of the two strategies could be conducted.

Given that the data presented in this thesis focus largely on excisional strategies, a trial comparing the two main excisional strategies (fasciectomy and dermofasciectomy) will be introduced.

11.3.2.2.1 Objectives of trial

The compare the effectiveness and safety of conventional dermofasciectomy to fasciectomy in terms of hand function, complications and cost utility.

11.3.2.2.2 Design

A patient-centred, pragmatic, multi-centre, randomised controlled trial with attention to minimising risks of bias incorporated into the design. Patient and public involvement will be critical. Given the variation in patient experience of Dupuytren's disease (see Chapter 6, Patients' Goals), more than one patient representative will be involved in the core trial team and named as a co-applicant on funding applications. The core team will also comprise stakeholder representatives including hand therapists, general practitioners and commissioners, a statistician, and a health economist. Given the identification of aesthetic appearance of the hand as a patient goal for treatment, the involvement of a clinical psychologist may be considered. This latter point represents a significant shift in approach to this condition compared to previous research.

11.3.2.2.3 Setting

Recruitment will take place at a minimum of five NHS hand surgery services. These must include orthopaedic hand surgery services, plastic hand surgery services, and fully integrated services (such as Derby, where the Hand Unit comprises both). The surgeons and hand therapists participating will all receive formal training in trial recruitment. The participating centres will include large volume centres and smaller centres, to pragmatically reflect the breadth of current practice in the NHS and to ensure that the external validity of the trial is optimised. All centres will include their routine postoperative management and hand therapy care pathways, and it is noted that this may involve “Any Qualified Provider” arrangements, in keeping with current NHS policy.

11.3.2.2.4 Recruitment

Patients attending elective hand surgery clinics with Dupuytren’s disease as the primary indication for referral will be screened for eligibility in the clinic, provided with written information about the trial and invited to participate. After a minimum two-week interval from initial consultation, patients will be recruited when attending preoperative assessment prior (but close) to their surgery episode. Signed informed consent will be taken at preoperative assessment attendance and baseline measurements performed.

11.3.2.2.5 Inclusion and exclusion criteria

Pragmatic inclusion criteria will be adopted. The target study population are patients with primary or recurrent Dupuytren's disease affecting one or more fingers requiring surgery, and for whom the treating surgeon considers either fasciectomy or dermofasciectomy to be technically possible.

Exclusion criteria will be patients under 18 years old, those with thumb or first web disease requiring treatment and those who are unable to give informed consent. Other variables, such as a diagnosis of diabetes mellitus, will be captured, as this thesis has demonstrated that they are likely to affect outcome. However, these patients will not be excluded, so that the external validity of the trial is optimised to NHS practice.

11.3.2.2.6 Randomisation

Randomisation will involve remote centralised randomisation process to minimise risks of bias identified in previous research regarding both randomisation and allocation concealment (see Chapter 4). Given that blinding is not achievable, standard block randomisation will not be used. Patients will be randomised on the day of surgery immediately prior to anaesthesia.

11.3.2.2.7 Intervention

A pragmatic definition of fasciectomy will be adopted. This will involve excision of Dupuytren's disease responsible for the patient's specific

symptoms followed by direct closure or local flap closure (e.g. z-plasties) at the surgeon's discretion and based on their standard practice. Smaller fasciectomies (such as segmental fasciectomy), where the aim is not to remove all of the disease responsible for a joint contracture, but to remove only a small segment of the disease across a joint, will be excluded.

11.3.2.2.8 Comparator

Conventional dermofasciectomy will be defined pragmatically, but with some criteria to distinguish it from firebreak dermofasciectomy (which has been shown to not affect outcome previously (Ullah et al., 2009)). Skin grafts must cover the span of the joint affected and reach both midaxial lines of the digit in order to replace all of the glabrous skin covering the affected joint.

11.3.2.2.9 Outcome measures

Given the findings of the thesis, the URAM will be adopted as the primary outcome measure of the trial. Although this outcome measure has limitations (see Chapter 7), it is the most contemporary measure of relevance, and its performance has surpassed that of the main alternative (the DASH) throughout this project. However, given its limitations, secondary endpoints will be important. These will include other patient-reported measures to capture the sequelae of surgery, such as a pain visual analogue scale (VAS). Formal recording of complications and satisfaction will also be measured. Finger joint angles will be assessed as a secondary outcome, but given the

effect of dynamism identified here (see Chapter 5), passive extension deficit will not be measured. Instead, active finger extension and flexion will be measured, as this is not subject to dynamism and is more patient-centred. Health service opportunity costs will be captured for cost utility analysis. Return to work or baseline activity data and costing data (e.g. social care costs) will also be captured. Further treatment will be recorded.

11.3.2.2.10 Sample size

A sample size calculation will be performed based on an MID between treatments of 8.3/45 in the URAM for the change score from baseline to one year after randomisation (as this represents the difference between treatments considered relevant, rather than the individual change considered relevant, or MID, see Chapter 9). One year after randomisation in this trial equates to one year after treatment.

11.3.2.2.11 Follow up

Follow up will involve clinic assessments at six weeks, three months, six months and one year after randomisation. This thesis has demonstrated that recovery takes longer than previous appreciated, hence the importance of taking measurements throughout the first year. After this point, annual assessments will be made for five years.

11.3.2.2.12 Blinding

Where possible, objective assessments such as measuring active finger joint angles will be performed by trained health care professionals who are not directly involved in the trial, though it is appreciated that formal blinding will not be possible here.

The above trial proposal incorporates lessons identified from key findings of this project throughout. The quality of future studies in this area is likely to be improved by this project, and areas for future research have been clarified. The key findings of the project will be summarised in the next and final section.

12 Project Summary

- There were few randomised controlled trials in Dupuytren's disease, most of which have methodological flaws such that the quality of evidence available to guide practice is poor.

- Patients have individual preoperative goals that they want improved by surgery, and these vary markedly from patient to patient.

- Not all patient goals matched the impairments that are traditionally considered to be relevant in Dupuytren's disease by clinicians, particularly loss of extension.

- The URAM scale was not fully comprehensive for use in Dupuytren's disease, particularly after treatment was completed.

- Dynamism was common and large in magnitude in Dupuytren's contractures crossing the proximal interphalangeal joint (PIPJ). It might constitute a previously unquantified source of bias in unblinded studies.

- The DASH was not unidimensional when used in Dupuytren's disease, and the constructs that it reflects may not be relevant.

- The URAM appeared unidimensional in Dupuytren's disease, though item nine behaved differently, as illustrated when the factor structure is analysed in conjunction with DASH items.

- Limited interpretability data existed for outcome measures that might be used in trials of Dupuytren's disease treatment, and the minimal important changes (MICs) and minimal important differences (MIDs) that did exist were mainly estimated using retrospective anchors.

- The DASH was not interpretable after open surgery for Dupuytren's disease.

- The URAM showed acceptable interpretability, but with a much larger MIC for open surgery than was previously reported for aponeurotomy.

- The factors associated with poor functional outcome were diabetes, revision surgery and female gender. These differ from traditional diathesis factors.

- After controlling for these factors (diabetes, previous ipsilateral surgery, female gender) and others, there was no significant difference in functional outcome between aponeurotomy and excisional surgery.

- Complication rates were considerably higher after excisional surgery than after aponeurotomy, which is likely to result in permanent functional limitation, and may explain the lack of difference in long-term functional outcome between the procedures, despite their markedly different recurrence rates.

13 Appendices

Appendix 1: Nottingham University Hospitals R&D Approval

Nottingham University Hospitals NHS
NHS Trust

Queen's Medical Centre Campus
Research and Development
E11 Curie Court
Derby Road
Nottingham
NG7 2UH

Direct Dial: 0115 9709049

Fax: 0115 8493295

www.nuh.nhs.uk

6th April 2011

Dear R&D Office,

Dupuytren's disease: a clinical evaluation of the recurrence and complication rates of different surgical procedures.

I can confirm we have reviewed the above project and can confirm that it is considered as a service evaluation.

If any further details are required, please do not hesitate to contact me on the details below or the project lead Dr Tim Davis.

Best regards,

Charlotte Davies
Research and Development Operations Manager
E11 Curie Court
Queens Medical Centre Campus
Derby Road
Nottingham
NG7 2UH

Tel; 0115 9249924 ext 61870
Mobile; 07595284977
Fax; 0115 8493295 internal 3295

E-mail; charlotte.davies@nuh.nhs.uk

For further information about R&D and a complete list of SOPs, processes and NUH sponsorship details please visit our new website; www.nuhrise.org

Appendix 2: Patient Information Letter

BSSH
Research

Mr Jeremy Rodrigues BSc(Hons) MRCS MSc
BSSH Research Fellow
Division of Orthopaedic & Accident Surgery
University of Nottingham
C Floor, West Block
Queen's Medical Centre
Nottingham
NG7 2UH

Dear Sir/Madam,

The UK's hand surgery society (The British Society for Surgery of The Hand (BSSH)) is carrying out a study of surgery for **Dupuytren's disease**.

This is a common condition that leads to the fingers bending up.

We are keen to see how patients get along after standard surgery performed at the moment. This is a "**Service Evaluation**" of what is currently done in the NHS, not a trial of anything new.

By doing so, we hope to be able to be able to improve the treatment of this condition for patients in the future, and to be able to plan studies comparing different treatments in the future.

Taking part is voluntary. You do not have to take part, and your treatment now or in future will not be affected at all if you choose not to take part.

We would like to invite you to take part in the project.

All information collected will be treated confidentially, and will be stored securely.

If you would like any further information, please feel free to contact the surgeon leading the project, Mr Jeremy Rodrigues, or my supervisor Professor Tim RC Davis.

- My postal address is at the top of the page
- Or feel free to e mail me at: Jeremy.rodrigues@nuh.nhs.uk
- Or make telephone contact by calling the Queen's Medical Centre, Nottingham on 0115 924 9924. This is the switchboard number. Once connected to the operator, ask for Professor TRC Davis's secretary in Orthopaedics. She will pass on the message.

Yours faithfully,

Jeremy Rodrigues

Appendix 3: Rotherham Patient Contact Letter Content

Dear ……,

I am writing to you to ask for your help with research into a condition called “Dupuytren’s disease.” It is a common condition and the UK’s national hand surgery organisation, the British Society for Surgery of the Hand (or BSSH) is keen to assess how people fare after having surgery for this.

You may recall that you had an operation for this condition (which causes fingers to bend into the palm) either 1 year ago or 5 years ago.

We would like to invite you to participate in this service evaluation. If you are happy to do so the BSSH will contact you by phone to arrange this. It will involve a visit to a clinic here. Your medical notes will be available in the clinic for the BSSH surgeon to review with you, if required.

You will be reimbursed £5 to help cover transport costs.

If you are happy to be involved, then please complete the enclosed consent form and return it using the stamped addressed envelope provided. The surgeon conducting the research will then contact you.

There are a number of different research projects looking at this condition and you may be approached by other colleagues of mine also. **Please be aware that all of this is entirely voluntary and, if you choose not to take part, your treatment will not be affected in any way.**

It is hoped that the results of this research will help hand surgeons to design and assess new treatments in the future.

Many thanks.

Yours sincerely,

Mr Indranil Chakrabarti
Consultant Hand Surgeon

Appendix 4: Consent form

BSSH

Research

1. I confirm that I have read and understand the letter dated XXXX.

2. I understand that the study is being coordinated by Nottingham University Hospitals NHS Trust, The University of Nottingham and The British Society for Surgery of The Hand, and give permission for a surgeon working on their behalf to be given my details for this study.

3. I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason, without my medical care or legal rights being affected.

4. I understand that sections of any of my medical notes may be looked at by responsible individuals where it is relevant to my taking part in research. I give permission for these individuals to have access to my records.

5. I understand that the data will be stored securely for at least 10 years after which it will be disposed of securely. Regulatory authorities from Research & Development department will have access to identifiable data for monitoring purposes.

6. I agree to take part in the above study.

SIGNATURE: _____

DATE: ____ / ____ / ____

PRINT NAME: _____

TELEPHONE NUMBER: _____
(to arrange clinic appointment via phone)

Appendix 5: Cross Sectional Study Proforma

Cohort 1yr☐ 5yr☐

Assessment date _____

Operation date _____

Operation Hospital & Surgeon _____

Operated Hand R☐ L☐

Operated digits TH☐ IF☐ MF☐ RF☐ LF☐

Operation Needle☐ Blade☐ Limited☐ Fasciec☐ Dermofasc☐
MCP / P1 / PIPJ / P2
mid-axial Y☐ N☐

Graft donor _____

Splint postop Y☐ N☐ details _____

Previous ipsilat surgery Y☐ N☐ details _____

Previous contralat surgery Y☐ N☐ details _____

Hand dominance R☐ L☐

Occupation since op _____

Main Occupation preop _____

Family history Y☐ N☐ details _____

Diabetic Y☐ N☐, type 1☐ type 2☐, meds _____

Steroids Y☐ N☐

Smoker Y☐ N☐

Epilepsy Y☐ N☐ meds _____

Weekly alcohol intake _____

Infection requiring antibiotics Y☐ N☐ details _____

2 point discrim intact 6mm Y☐ N☐ details RDN_____UDN_____

Cold intolerance Y☐ N☐ details _____

CRPS Y☐ N☐ details _____

Surgery since index operation Y☐ N☐ details _____

Passive extension MCPJ _____ PIPJ _____

Active flexion digit _____

Active flexion other digits _____

Operation site: palpable disease Y☐ N☐ Cord & Contracture☐ Cord☐ Nodule☐

Graft recurrence: under graft ☐ lat to graft ☐ prox / distal to graft ☐

Other fingers (same hand) with disease Th☐ Web ☐ IF☐ MF☐ RF☐ LF☐

Other fingers (other hand) with disease Th☐ Web ☐ IF☐ MF☐ RF☐ LF☐

Knuckle pads Y☐ N☐ **Lederhose disease** Y☐ N☐

Appendix 6: Cohort Study Baseline Data Capture Proforma

Name DoB Number

Assessment date_____

Operation date _____

Operation Hospital _____

Surgeon _____

Telephone _____

Operated Hand R L

Operated digits TH IF MF RF LF

Operation Needle Limited Fasciec Dermofasc
MCP / P1 / PIPJ / P2

Previous surgery Y N details _____

Hand dominance R L

Occupation _____

Diabetic Y N type 1 type 2, meds _____

Steroids Y N

Smoker Y N

Epilepsy Y N meds _____

Other PMH _____

Current angles MCPJ _____ PIPJ _____

14 References

Abe Y, Rokkaku T, Ofuchi S, Tokunaga S, Takahashi K, Moriya H. An objective method to evaluate the risk of recurrence and extension of Dupuytren's disease. J Hand Surg Brit Eur. 2004, 29: 427-30.

Adams A, Ketchersid J, Tan WH, Calfee RP, Howard D. Minimal clinically important differences on the DASH, Quick-DASH and PRWE: Level 2 evidence. J Hand Surg Am. 2012, 37(8 supp): 17.

Al-Qattan MM. Factors in the pathogenesis of Dupuytren's contracture. J Hand Surg Am. 2006, 31: 1527-34.

Alioto RJ, Rosier RN, Burton RI, Puzas JE. Comparative effects of growth factors on fibroblasts of Dupuytren's tissue and normal palmar fascia. J Hand Surg Am. 1994, 19: 442-52.

Amirfeyz R, Pentlow A, Foote J, Leslie I. Assessing the clinical significance of change scores following carpal tunnel surgery. Int Orthop. 2009, 33: 181-5.

Armstrong JR, Hurren JS, Logan AM. Dermofasciectomy in the management of Dupuytren's disease. J Bone Joint Surg Br. 2000, 82: 90-4.

Atkins D, Best D, Briss PA et al. Grading quality of evidence and strength of recommendations. BMJ. 2004, 328: 1490.

Atkins RM. Complex regional pain syndrome. J Bone Joint Surg Br. 2003, 85: 1100-6.

Atroshi I, Lyrn PE, Ornstein E, Gummesson C. The six-item CTS symptoms scale and palmar pain scale in carpal tunnel syndrome. J Hand Surg Am. 2011, 36: 788-94.

Badois FJ, Lermusiaux JL, Masse C, Kuntz D. [Non-surgical treatment of Dupuytren disease using needle fasciotomy]. Rev Rhum Engl Ed. 1993, 60: 808-13.

Bago J, Perez-Grueso FJ, Les E, Hernandez P, Pellise F. Minimal important differences of the SRS-22 Patient Questionnaire following surgical treatment of idiopathic scoliosis. European Spine Journal. 2009, 18: 1898-904.

Baird KS, Crossan JF, Ralston SH. Abnormal growth factor and cytokine expression in Dupuytren's contracture. J Clin Pathol. 1993, 46: 425-8.

Ball C, Pratt AL, Nanchahal J. Optimal functional outcome measures for assessing treatment for Dupuytren's disease: a systematic review and recommendations for future practice. BMC Musculoskelet Disord. 2013, 14: 131.

Bayat A, Alansar A, Hajeer AH et al. Genetic susceptibility in Dupuytren's disease: lack of association of a novel transforming growth factor beta(2) polymorphism in Dupuytren's disease. J Hand Surg Brit Eur. 2002, 27: 47-9.

Bayat A, Watson JS, Stanley JK et al. Genetic susceptibility in Dupuytren's disease. TGF-beta1 polymorphisms and Dupuytren's disease. J Bone Joint Surg Br. 2002, 84: 211-5.

Beaton DE, Katz JN, Fossel AH, Wright JG, Tarasuk V, Bombardier C. Measuring the whole or the parts? Validity, reliability, and responsiveness of

the Disabilities of the Arm, Shoulder and Hand outcome measure in different regions of the upper extremity. J Hand Ther. 2001, 14: 128-46.

Beaton DE, Van Eerd D, Smith P et al. Minimal change is sensitive, less specific to recovery: A diagnostic testing approach to interpretability. J Clin Epidemiol. 2011, 64: 487-96.

Beaton DE, Wright JG, Katz JN. Development of the QuickDASH: comparison of three item-reduction approaches. J Bone Joint Surg Am. 2005, 87: 1038-46.

Beaudreuil J, Allard A, Zerkak D et al. Unite Rhumatologique des Affections de la Main (URAM) scale: development and validation of a tool to assess Dupuytren's disease-specific disability. Arthritis Care & Research. 2011, 63: 1448-55.

Beaudreuil J, Orcel P, Bardin T et al. Re: Rodrigues JN, Zhang W, Scammell BE, Davis TRC. What patients want from the treatment of Dupuytren's disease - is the Unite Rhumatologique des Affections de la Main (URAM) scale relevant? J Hand Surg Eur. Epub ahead of print 21 February 2014. DOI: 10.1177/1753193414524689. J Hand Surg Brit Eur. 2014, 39: 673-5.

Becker GW, Davis TR. The outcome of surgical treatments for primary Dupuytren's disease--a systematic review. J Hand Surg Brit Eur. 2010, 35: 623-6.

Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society. Series B (Methodological). 1995, 57: 289-300.

Berndt A, Kosmehl H, Katenkamp D, Tauchmann V. Appearance of the myofibroblastic phenotype in Dupuytren's disease is associated with a fibronectin, laminin, collagen type IV and tenascin extracellular matrix. Pathobiology. 1994, 62: 55-8.

Bessette L, Sangha O, Kuntz KM et al. Comparative responsiveness of generic versus disease-specific and weighted versus unweighted health status measures in carpal tunnel syndrome. Medical care. 1998, 36: 491-502.

Beyermann K, Prommersberger KJ, Jacobs C, Lanz UB. Severe contracture of the proximal interphalangeal joint in Dupuytren's disease: does capsuloligamentous release improve outcome? J Hand Surg Brit Eur. 2004, 29: 240-3.

Bhatia R, Blackshaw G, Barr V, Savage R. Comparative study of "staples versus sutures" in skin closure following Dupuytren's surgery. J Hand Surg Brit Eur. 2002, 27: 53-4.

Bien TH, Burge R. Smoking and drinking: a review of the literature. Int J Addict. 1990 25(12): 1429-54.

Bindra RR, Dias JJ, Heras-Palau C, Amadio PC, Chung KC, Burke FD. Assessing outcome after hand surgery: the current state. J Hand Surg Brit Eur. 2003, 28: 289-94.

Bisson MA, McGrouther DA, Mudera V, Grobbelaar AO. The different characteristics of Dupuytren's disease fibroblasts derived from either nodule or cord: expression of alpha-smooth muscle actin and the response to stimulation by TGF-beta1. J Hand Surg Brit Eur. 2003, 28: 351-6.

Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet. 1986, 1: 307-10.

Bland JM, Altman DG. A note on the use of the intraclass correlation coefficient in the evaluation of agreement between two methods of measurement. Computers in biology and medicine. 1990, 20: 337-40.

Bland JM, Altman DG. Agreement between methods of measurement with multiple observations per individual. Journal of biopharmaceutical statistics. 2007, 17: 571-82.

Boone DC, Azen SP, Lin CM, Spence C, Baron C, Lee L. Reliability of goniometric measurements. Phys Ther. 1978, 58: 1355-60.

Bot SD, Terwee CB, van der Windt DA, Bouter LM, Dekker J, de Vet HC. Clinimetric evaluation of shoulder disability questionnaires: a systematic review of the literature. Ann Rheum Dis. 2004, 63: 335-41.

Brazier JE, Harper R, Jones NM et al. Validating the SF-36 health survey questionnaire: new outcome measure for primary care. BMJ. 1992, 305: 160-4.

Broekstra DC, Lanting R, Werker PM, van den Heuvel ER. Intra- and inter-observer agreement on diagnosis of Dupuytren disease, measurements of severity of contracture, and disease extent. Manual therapy. 2015.

Brouwers M, Kho ME, Browman GP et al. AGREE II: Advancing guideline development, reporting and evaluation in healthcare. Can Med Assoc J. 2010, 182: E839-42.

Brown RA, Prajapati R, McGrouther DA, Yannas IV, Eastwood M. Tensional homeostasis in dermal fibroblasts: mechanical responses to mechanical loading in three-dimensional substrates. J Cell Physiol. 1998, 175: 323-32.

Bruner JM. Incisions for plastic and reconstructive (non-septic) surgery of the hand. Br J Plast Surg. 1951, 4: 48-55.

Bryan AS, Ghorbal MS. The long-term results of closed palmar fasciotomy in the management of Dupuytren's contracture. J Hand Surg Brit Eur. 1988, 13: 254-6.

BSSH. Dupuytren's Disease. London: BSSH, 2008. http://www.bssh.ac.uk/education/guidelines/dd_guidelines.pdf (3rd April 2012 2012).

Budd HR, Larson D, Chojnowski A, Shepstone L. The QuickDASH score: a patient-reported outcome measure for Dupuytren's surgery. J Hand Ther. 2011, 24: 15-20; quiz 1.

Bulstrode NW, Bisson M, Jemec B, Pratt AL, McGrouther DA, Grobbelaar AO. A prospective randomised clinical trial of the intra-operative use of 5-fluorouracil on the outcome of dupuytren's disease. J Hand Surg Brit Eur. 2004, 29: 18-21.

Campbell DA, Kay SP. What is cold intolerance? J Hand Surg Brit Eur. 1998, 23: 3-5.

Carswell A, McColl MA, Baptiste S, Law M, Polatajko H, Pollock N. The Canadian Occupational Performance Measure: a research and clinical

literature review. Canadian Journal of Occupational Therapy - Revue Canadienne d Ergotherapie. 2004, 71: 210-22.

Cattell RB. The Scree Test for the Number of Factors. Multivariate Behavioral Research. 1966, 1: 245-76.

Chammas M, Bousquet P, Renard E, Poirier J-L, Jaffiol C, Allieu Y. Dupuytren's disease, carpal tunnel syndrome, trigger finger, and diabetes mellitus. J Hand Surg Am. 1995, 20: 109-14.

Chatterjee JS, Price PE. Comparative responsiveness of the Michigan Hand Outcomes Questionnaire and the Carpal Tunnel Questionnaire after carpal tunnel release. J Hand Surg Am. 2009, 34: 273-80.

Chignon-Sicard B, Georgiou CA, Fontas E et al. Efficacy of leukocyte- and platelet-rich fibrin in wound healing: a randomized controlled clinical trial. Plast Reconstr Surg. 2012, 130: 819e-29e.

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

Citron N, Hearnden A. Skin tension in the aetiology of Dupuytren's disease; a prospective trial. J Hand Surg Brit Eur. 2003, 28: 528-30.

Citron ND, Nunez V. Recurrence after surgery for Dupuytren's disease: a randomized trial of two skin incisions. J Hand Surg Brit Eur. 2005, 30: 563-6.

Collis EL, Eatock R. Report of an inquiry on Dupuytren’s contraction as a disease of occupation, with special reference to its occurrence among minders of lace machines. . London, Home Office, 1912.

Collis J, Collocott S, Hing W, Kelly E. The effect of night extension orthoses following surgical release of Dupuytren contracture: A single-center, randomized, controlled trial. J Hand Surg Am, 2013: 1285-94e2.

Cooper AP. On Dislocations of the Fingers and Toes - Dislocation from Contraction of the Tendon. Treatise on Dislocations and Fractures of the Joints. London, Longman & Co., 1822: 524-5.

Copay AG, Subach BR, Glassman SD, Polly DW, Jr., Schuler TC. Understanding the minimum clinically important difference: a review of concepts and methods. Spine Journal: Official Journal of the North American Spine Society. 2007, 7: 541-6.

Craft RO, Smith AA, Coakley B, Casey WJ, 3rd, Rebecca AM, Duncan SF. Preliminary Soft-Tissue Distraction vs Checkrein Ligament Release After Fasciectomy in the Treatment of Dupuytren PIP Joint Contracturesa. Plast Reconstr Surg. 2011.

Crean SM, Gerber RA, Le Graverand MP, Boyd DM, Cappelleri JC. The efficacy and safety of fasciectomy and fasciotomy for Dupuytren's contracture in European patients: a structured review of published studies. J Hand Surg Brit Eur. 2011, 36: 396-407.

Critchley EM, Vakil SD, Hayward HW, Owen VM. Dupuytren's disease in epilepsy: result of prolonged administration of anticonvulsants. J Neurol Neurosurg Psychiatry. 1976, 39: 498-503.

Darzi A. High quality care for all: NHS Next Stage Review final report. London: TSO, 2008.

http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_085825 (27th June 2012).

Dawson J, Doll H, Boller I et al. Comparative responsiveness and minimal change for the Oxford Elbow Score following surgery. Quality of Life Research. 2008, 17: 1257-67.

de Vet HC, Terwee CB, Ostelo RW, Beckerman H, Knol DL, Bouter LM. Minimal changes in health status questionnaires: distinction between minimally detectable change and minimally important change. Health & Quality of Life Outcomes. 2006, 4: 54.

Degreef I, Vererfve PB, De Smet L. Effect of severity of Dupuytren contracture on disability. Scand J Plast Reconstr Surg Hand Surg. 2009, 43: 41-2.

Department of Health. Equity and Excellence: Liberating the NHS. London2010.

http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_117353 (26th June 2012).

Deyo RA, Centor RM. Assessing the responsiveness of functional scales to clinical change: an analogy to diagnostic test performance. Journal of Chronic Diseases. 1986, 39: 897-906.

Dias JJ, Braybrooke J. Dupuytren's contracture: an audit of the outcomes of surgery. J Hand Surg Brit Eur. 2006, 31: 514-21.

Dias JJ, Rajan RA, Thompson JR. Which questionnaire is best? The reliability, validity and ease of use of the Patient Evaluation Measure, the Disabilities of

the Arm, Shoulder and Hand and the Michigan Hand Outcome Measure. J Hand Surg Brit Eur. 2008, 33: 9-17.

Dias JJ, Singh HP, Ullah A, Bhowal B, Thompson JR. Patterns of recontracture after surgical correction of Dupuytren disease. J Hand Surg Am. 2013, 38: 1987-93.

Dolmans GH, de Bock GH, Werker PM. Dupuytren diathesis and genetic risk. J Hand Surg Am. 2012, 37: 2106-11.

Dolmans GH, Werker PM, Hennies HC et al. Wnt signaling and Dupuytren's disease. New Engl J Med. 2011, 365: 307-17.

Dowrick AS, Gabbe BJ, Williamson OD, Cameron PA. Outcome instruments for the assessment of the upper extremity following trauma: a review. Injury. 2005, 36: 468-76.

Dowrick AS, Gabbe BJ, Williamson OD, Cameron PA. Does the disabilities of the arm, shoulder and hand (DASH) scoring system only measure disability due to injuries to the upper limb? J Bone Joint Surg Br. 2006, 88: 524-7.

Draviaraj KP, Chakrabarti I. Functional outcome after surgery for Dupuytren's contracture: a prospective study. J Hand Surg Am. 2004, 29: 804-8.

Dupuytren G. Permanent retraction of the fingers produced by an affection of the palmar fascia. . Lancet. 1834, 2: 22.

Early PF. Population studies in Dupuytren's contracture. J Bone Joint Surg Br. 1962, 44B: 602-13.

Ellis B, Bruton A. A study to compare the reliability of composite finger flexion with goniometry for measurement of range of motion in the hand. Clin Rehabil. 2002, 16: 562-70.

Engstrand C, Boren L, Liedberg GM. Evaluation of activity limitation and digital extension in Dupuytren's contracture three months after fasciectomy and hand therapy interventions. J Hand Ther. 2009, 22: 21-6; quiz 7.

Fan J, Upadhye S, Worster A. Understanding receiver operating characteristic (ROC) curves. Cjem. 2006, 8: 19-20.

FDA. US Food and Drug Administration. FDA Approves Xiaflex for Debilitating Hand Condition. 2010. URL: http://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm199736.htm 2010. (18th October 2011).

Ferguson E, Cox T. Exploratory Factor Analysis: A Users' Guide. International Journal of Selection and Assessment. 1993, 1: 84-94.

Fitzpatrick R, Davey C, Buxton MJ, Jones DR. Evaluating patient-based outcome measures for use in clinical trials. Health Technol Assess. 1998, 2: i-iv, 1-74.

Floyd FJ, Widaman KF. Factor Analysis in the Development and Refinement of Clinical Assessment Instruments. Psychological Assessment. 1995, 7: 286-99.

Forget NJ, Jerosch-Herold C, Shepstone L, Higgins J. Psychometric evaluation of the Disabilities of the Arm, Shoulder and Hand (DASH) with

Dupuytren's contracture: validity evidence using Rasch modeling. BMC Musculoskelet Disord. 2014, 15: 361.

Foucher G, Cornil C, Lenoble E. Open palm technique for Dupuytren's disease. A five-year follow-up. Ann Chir Main Memb Super. 1992, 11: 362-6.

Foucher G, Medina J, Navarro R. Percutaneous needle aponeurotomy: complications and results. J Hand Surg Brit Eur. 2003, 28: 427-31.

Franchignoni F, Giordano A, Sartorio F, Vercelli S, Pascariello B, Ferriero G. Suggestions for refinement of the Disabilities of the Arm, Shoulder and Hand Outcome Measure (DASH): a factor analysis and Rasch validation study. Arch Phys Med Rehabil. 2010, 91: 1370-7.

Franchignoni F, Vercelli S, Giordano A, Sartorio F, Bravini E, Ferriero G. Minimal clinically important difference of the disabilities of the arm, shoulder and hand outcome measure (DASH) and its shortened version (QuickDASH). J Orthop Spots Phys Ther. 2014, 44: 30-9.

Gabbiani G, Ryan GB, Majne G. Presence of modified fibroblasts in granulation tissue and their possible role in wound contraction. Experientia. 1971, 27: 549-50.

Garrison C, Cook C. Clinimetrics corner: the Global Rating of Change Score (GRoC) poorly correlates with functional measures and is not temporally stable. The Journal of manual & manipulative therapy. 2012, 20: 178-81.

Gazdzik T, Wasilewski Z. Results of surgical treatment for Dupuytren contracture [Wyniki leczenia operacyjnego przykurczu Dupuytrena]. Chir Narzadow Ruchu Ortop Pol. 1997, 62: 375-9.

Gelberman RH, Amiel D, Rudolph RM, Vance RM. Dupuytren's contracture. An electron microscopic, biochemical, and clinical correlative study. J Bone Joint Surg Am. 1980, 62: 425-32.

Gelberman RH, Panagis JS, Hergenroeder PT, Zakaib GS. Wound complications in the surgical management of Dupuytren's contracture: a comparison of operative incisions. Hand. 1982, 14: 248-54.

Geoghegan JM, Forbes J, Clark DI, Smith C, Hubbard R. Dupuytren's disease risk factors. J Hand Surg Brit Eur. 2004, 29: 423-6.

Georgeu GA, Mayfield S, Logan AM. Lateral digital photography with computer-aided goniometry versus standard goniometry for recording finger joint angles. J Hand Surg Brit Eur. 2002, 27: 184-6.

Gerber R, Perry R, Thompson R, Bainbridge C. Dupuytren's contracture: a retrospective database analysis to assess clinical management and costs in England. BMC Musculoskelet Disord. 2011, 12: 73.

Godtfredsen NS, Lucht H, Prescott E, Sorensen TIA, Gronbaek M. A prospective study linked both alcohol and tobacco to Dupuytren's disease. J Clin Epidemiol. 2004, 57: 858-63.

Goyrand G. Nouvelles recherches sur la retraction permanente des doigts. Memoires de l'Academie Royale de Medicine. 1833, 3: 489-96.

Gudmundsson KG, Arngrimsson R, Sigfusson N, Jonsson T. Increased total mortality and cancer mortality in men with Dupuytren's disease: a 15-year follow-up study. J Clin Epidemiol. 2002, 55: 5-10.

Gummesson C, Atroshi I, Ekdahl C. The disabilities of the arm, shoulder and hand (DASH) outcome questionnaire: longitudinal construct validity and measuring self-rated health change after surgery. BMC Musculoskelet Disord. 2003, 4: 11.

Gummesson C, Ward MM, Atroshi I. The shortened disabilities of the arm, shoulder and hand questionnaire (QuickDASH): validity and reliability based on responses within the full-length DASH. BMC Musculoskelet Disord. 2006, 7: 44.

Guyatt GH, Norman GR, Juniper EF, Griffith LE. A critical look at transition ratings. J Clin Epidemiol. 2002, 55: 900-8.

Hamilton GF, Lachenbruch PA. Reliability of goniometers in assessing finger joint angle. Phys Ther. 1969, 49: 465-9.

Harden RN, Bruehl S, Stanton-Hicks M, Wilson PR. Proposed new diagnostic criteria for complex regional pain syndrome. Pain medicine. 2007, 8: 326-31.

Hart MG, Hooper G. Clinical associations of Dupuytren's disease. Postgrad Med J. 2005, 81: 425-8.

Hazarika EZ, Knight MT, Frazer-Moodie A. The effect of intermittent pneumatic compression on the hand after fasciectomy. Hand. 1979, 11: 309-14.

Health and Social Care Act c. 7. UK, 2012.

Hedges LV. Fitting continuous models to effect size data. Journal of Educational Statistics. 1982, 7: 245-70.

Herdman M, C. G, Lloyd A et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). Qual Life Res. 2011, 20: 1727-36.

Heuston JT. The control of recurrent Dupuytren's contracture by skin replacement. Br J Plast Surg. 1969, 22: 152-6.

Higgins JPT, Green S. Cochrane Handbook for Systematic Reviews of Interventions. The Cochrane Collaboration, 2011. http://www.cochrane-handbook.org (3rd October 2012).

Hindocha S, Stanley JK, Watson S, Bayat A. Dupuytren's diathesis revisited: Evaluation of prognostic indicators for risk of disease recurrence. J Hand Surg Am. 2006, 31: 1626-34.

Hinz B. Formation and function of the myofibroblast during tissue repair. Journal of Investigative Dermatology. 2007, 127: 526-37.

Horn JL. A Rationale and Test for the Number of Factors in Factor Analysis. Psychometrika. 1965, 30: 179-85.

Hospital Episode Statistics. HES Online. Hospital Episode Statistics (HES) : The NHS information centre for health and social care, 2011. http://www.hesonline.nhs.uk/Ease/servlet/ContentServer?siteID=1937 (18th October 2011).

Howard K, Simison AJ, Morris A, Bhalaik V. A prospective randomised trial of absorbable versus non-absorbable sutures for wound closure after fasciectomy for Dupuytren's contracture. J Hand Surg Brit Eur. 2009, 34: 618-20.

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

Hueston JT. Limited fasciectomy for Dupuytren's contracture. Plast Reconstr Surg Transplant Bull. 1961, 27: 569-85.

Hueston JT. Digital Wolfe grafts in recurrent Dupuytren's contracture. Plast Reconstr Surg Transplant Bull. 1962, 29: 342-4.

Hueston JT. The Dupuytren's diathesis. In: Hueston JT (Ed.) Dupuytren's contracture. Edinburgh, E & S Livingstone, 1963: 51-63.

Hueston JT. Principles of management. In: Hueston JT (Ed.) Dupuytren's contracture. London, E &S Livingstone, 1963: 76-94.

Hueston JT. Enzymic fasciotomy. Hand. 1971, 3: 38-40.

Hueston JT. Dupuytren's contracture: selection for surgery. Br J Hosp Med (Lond). 1974, 13: 361.

Hueston JT. Current state of treatment of Dupuytren's disease. Ann Chir Main. 1984, 3: 81-92.

Hueston JT. 'Firebreak' grafts in Dupuytren's contracture. Aust N Z J Surg. 1984, 54: 277-81.

Hurst LC. Dupuytren’s Contracture. In: Wolfe SW, Hotchkiss RN, Pederson WC, Kozin SH (Eds.) Green’s Operative Hand Surgery, 6th ed. Philadelphia, Elsevier Churchill-Livingstone, 2010.

Hurst LC, Badalamente MA, Hentz VR et al. Injectable collagenase clostridium histolyticum for Dupuytren's contracture. N Engl J Med. 2009, 361: 968-79.

International Consortium for Health Outcome Measurement. ICHOM. 2014. https://http://www.ichom.org/ (20th March 2015).

Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. Control Clin Trials. 1989, 10: 407-15.

Jebsen RH, Taylor N, Trieschmann RB, Trotter MJ, Howard LA. An objective and standardized test of hand function. Arch Phys Med Rehabil. 1969, 50: 311-9.

Jenkins PJ, Duckworth AD, Watts AC, McEachan JE. The outcome of carpal tunnel decompression in patients with diabetes mellitus. J Bone Joint Surg Br. 2012, 94: 811-4.

Jerosch-Herold C, Shepstone L, Chojnowski A, Larson D. Severity of contracture and self-reported disability in patients with Dupuytren's contracture referred for surgery. J Hand Ther. 2011, 24: 6-10; quiz 1.

Jerosch-Herold C, Shepstone L, Chojnowski AJ, Larson D. Splinting after contracture release for Dupuytren's contracture (SCoRD): protocol of a pragmatic, multi-centre, randomized controlled trial. BMC Musculoskelet Disord. 2008, 9: 62.

Jerosch-Herold C, Shepstone L, Chojnowski AJ, Larson D, Barrett E, Vaughan SP. Night-time splinting after fasciectomy or dermo-fasciectomy for

Dupuytren's contracture: a pragmatic, multi-centre, randomised controlled trial. BMC Musculoskelet Disord. 2011, 12: 136.

Kaiser HF. An index of factorial simplicity. Psychometrika. 1974, 39: 31-6.

Katz JN, Gelberman RH, Wright EA, Lew RA, Liang MH. Responsiveness of self-reported and objective measures of disease severity in carpal tunnel syndrome. Medical Care. 1994, 32: 1127-33.

Kazis LE, Anderson JJ, Meenan RF. Effect sizes for interpreting changes in health status. Medical Care. 1989, 27: S178-89.

Kemler MA, Houpt P, van der Horst CMAM. A pilot study assessing the effectiveness of postoperative splinting after limited fasciectomy for Dupuytren's disease. J Hand Surg Brit Eur. 2012, 37: 733-7.

Kennedy C, Beaton D, Solway S, McConnell S, Bombardier C. The DASH and QuickDASH Outcome Measure User's Manual, Third ed. Toronto, Institute for Work and Health, 2011.

Ketchum LD, Donahue TK. The injection of nodules of Dupuytren's disease with triamcinolone acetonide. J Hand Surg Am. 2000, 25: 1157-62.

Kim JK, Jeon SH. Minimal clinically important differences in the Carpal Tunnel Questionnaire after carpal tunnel release. J Hand Surg Brit Eur. 2013, 38: 75-9.

Kim JK, Park ES. Comparative responsiveness and minimal clinically important differences for idiopathic ulnar impaction syndrome. Clin Orthop Relat Res. 2013, 471: 1406-11.

Knobloch K, Kuehn M, Vogt PM. Focused extracorporeal shockwave therapy in Dupuytren's disease--a hypothesis. Med Hypotheses. 2011, 76: 635-7.

Larkin ME, Barnie A, Braffett BH et al. Musculoskeletal complications in type 1 diabetes. Diabetes Care. 2014, 37: 1863-9.

Larocerie-Salgado J, Davidson J. Nonoperative treatment of PIPJ flexion contractures associated with Dupuytren's disease. J Hand Surg Brit Eur. 2012, 37: 722-7.

Law M, Baptiste S, McColl M, Opzoomer A, Polatajko H, Pollock N. The Canadian occupational performance measure: an outcome measure for occupational therapy. Can J Occup Ther. 1990, 57: 82-7.

Leclercq C. Clinical Aspects: Clinical Presentation. In: Tubiana R, Leclercq C, Hurst L, Badalamente M, Mackin E (Eds.) Dupuytren's Disease. London, Martin Dunitz, 2000: 79-96.

Leclercq C. Results of surgical treatment: Recurrence. In: Tubiana R, Leclercq C, Hurst L, Badalamente M, Mackin E (Eds.) Dupuytren's DIsease. London, Martin Dunitz, 2000: 239-50.

Leclercq C. Treatment: Fasciotomy. In: Tubiana R, Leclercq C, Hurst L, Badalamente M, Mackin E (Eds.) Dupuytren's Disease. London, Martin Dunitz, 2000: 132-8.

Lefebvre C, Manheimer E, Glanville J. Chapter 6: Searching for studies. In: Higgins, J.P.T., Green, S. (editors). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. http://www.cochrane-handbook.org. 2013).

Levine DW, Simmons BP, Koris MJ et al. A self-administered questionnaire for the assessment of severity of symptoms and functional status in carpal tunnel syndrome. J Bone Joint Surg Am. 1993, 75: 1585-92.

Logan CY, Nusse R. The Wnt signaling pathway in development and disease. Annual Review of Cell & Developmental Biology. 2004, 20: 781-810.

London DA, Stepan JG, Calfee RP. Determining the Michigan Hand Outcomes Questionnaire minimal clinically important difference by means of three methods. Plast Reconstr Surg. 2014, 133: 616-25.

Loos B, Puschkin V, Horch RE. 50 years experience with Dupuytren's contracture in the Erlangen University Hospital--a retrospective analysis of 2919 operated hands from 1956 to 2006. BMC Musculoskelet Disord. 2007, 8: 60.

Luck JV. Dupuytren's Contracture: A New Concept of the Pathogenesis Correlated with Surgical Management. J Bone Joint Surg Am. 1959, 41: 635-64.

Lund M. DUPUYTREN'S CONTRACTURE AND EPILEPSY The clinical connection between Dupuytren's contracture, fibroma plantae, periarthrosis humeri, helodermia, induratio penis plastica and epilepsy, with an attempt at a pathogenetic valuation. Acta Psychiatrica Scandinavica. 1941, 16: 465-92.

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

Makela EA, Jaroma H, Harju A, Anttila S, Vainio J. Dupuytren's contracture: the long-term results after day surgery. J Hand Surg Brit Eur. 1991, 16: 272-4.

Malay S, Chung KC. The minimal clinically important difference after simple decompression for ulnar neuropathy at the elbow. J Hand Surg Am. 2013, 38: 652-9.

MAPI. PROQOLID. 2013 ed. Lyon, France, MAPI Research Trust, 2013.

McCash CR. The Open Palm Technique in Dupuytren's Contracture. Br J Plast Surg. 1964, 17: 271-80.

McFarlane RM. Patterns of the diseased fascia in the fingers in Dupuytren's contracture. Displacement of the neurovascular bundle. Plast Reconstr Surg. 1974, 54: 31-44.

McFarlane RM. The finger. In: McFarlane RM, McGrouther DA, Flint MH (Eds.) Dupuytren's Disease: Biology and Treatment. Edinburgh, Churchill Livingstone, 1990.

McGrouther AD. The palm. In: McFarlane RM, McGrouther DA, Flint MH (Eds.) Dupuytren's Disease: Biology and Treatment. Edinburgh, Churchill Livingstone, 1990.

McGrouther AD. Dupuytren's Contracture. In: Green DP, Pederson WC, Hotchkiss RN, Wolfe SW (Eds.) Green's Operative Hand Surgery, 5th ed. Philadelphia, Elsevier Churchill-Livingstone, 2005, Vol. 1.

McMillan C, Binhammer P. Steroid injection and needle aponeurotomy for Dupuytren contracture: a randomized, controlled study. J Hand Surg Am. 2012, 37: 1307-12.

Messina A, Messina J. The continuous elongation treatment by the TEC device for severe Dupuytren's contracture of the fingers. Plast Reconstr Surg. 1993, 92: 84-90.

Mintken PE, Glynn P, Cleland JA. Psychometric properties of the shortened disabilities of the Arm, Shoulder, and Hand Questionnaire (QuickDASH) and Numeric Pain Rating Scale in patients with shoulder pain. J Shoulder Elbow Surg. 2009, 18: 920-6.

Misra A, Jain A, Ghazanfar R, Johnston T, Nanchahal J. Predicting the outcome of surgery for the proximal interphalangeal joint in Dupuytren's disease. J Hand Surg Am. 2007, 32: 240-5.

Moermans JP. Segmental aponeurectomy in Dupuytren's disease. J Hand Surg Brit Eur. 1991, 16: 243-54.

Mohan A, Vadher J, Ismail H, Warwick D. The Southampton Dupuytren's Scoring Scheme. J Plast Surg Hand Surg. 2014, 48: 28-33.

Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Medicine. 2009, 6: e1000097.

Mokkink LB, Terwee CB, Knol DL et al. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. BMC Med Res Methodol. 2010, 10: 22.

Mundfrom DJ, Shaw DG, Ke TL. Minimum Sample Size Recommendations for Conducting Factor Analyses. International Journal of Testing. 2005, 5: 159-68.

Murrell GA, Francis MJ, Bromley L. Free radicals and Dupuytren's contracture. Br Med J (Clin Res Ed). 1987, 295: 1373-5.

NHS. Alcohol unit calculator. London: NHS Choices, 2013. http://www.nhs.uk/Tools/Pages/Alcohol-unit-calculator.aspx (6th June 2013).

NICE. Interventional procedure overview: Needle fasciotomy for Dupuytren's contracture. London2003. http://guidance.nice.org.uk/IPG43 (5th April 2012).

NICE. Guide to the methods of technology appraisal. London: NICE, 2008. http://www.nice.org.uk/aboutnice/howwework/devnicetech/technologyappraisalprocessguides/?domedia=1&mid=B52851A3-19B9-E0B5-D48284D172BD8459 (26th June 2012).

NICE. Interventional Procedure Guidance (IPG368): Radiation therapy for early Dupuytren's disease. London: NICE, 2011. http://guidance.nice.org.uk/IPG368 (5th April 2012).

NICE. NICE Accreditation. London2011. https://http://www.evidence.nhs.uk/accreditation (20th June 2012).

NICE. NICE Quality Standards. London: NICE, 2013. http://www.nice.org.uk/aboutnice/qualitystandards/qualitystandards.jsp (22nd June 2013).

Niekel MC, Lindenhovius AL, Watson JB, Vranceanu AM, Ring D. Correlation of DASH and QuickDASH with measures of psychological distress. J Hand Surg Am. 2009, 34: 1499-505.

Noble J, Heathcote JG, Cohen H. Diabetes mellitus in the aetiology of Dupuytren's disease. J Bone Joint Surg Br. 1984, 66: 322-5.

Norman G. Likert scales, levels of measurement and the "laws" of statistics. Advances in health sciences education : theory and practice. 2010, 15: 625-32.

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

NRES. Is your project research? : NHS Health Research Authority, 2012. http://www.nres.npsa.nhs.uk/applications/is-your-project-research/ (1st February 2012).

Ozer K, Malay S, Toker S, Chung KC. Minimal clinically important difference of carpal tunnel release in diabetic and nondiabetic patients. Plast Reconstr Surg. 2013, 131: 1279-85.

Ozyurekoglu T, McCabe SJ, Goldsmith LJ, LaJoie AS. The Minimal Clinically Important Difference of the Carpal Tunnel Syndrome Symptom Severity Scale. Journal of Hand Surgery. 2006, 31: 733-8.

Packham T. Clinical commentary in response to: Severity of contracture and self-reported disability in patients with Dupuytren's contracture referred for surgery. J Hand Ther. 2011, 24: 12-4.

Pagnotta A, Specchia N, Greco F. Androgen receptors in Dupuytren's contracture. Journal of Orthopaedic Research. 2002, 20: 163-8.

Pagnotta A, Specchia N, Soccetti A, Manzotti S, Greco F. Responsiveness of Dupuytren's disease fibroblasts to 5 alpha-dihydrotestosterone. J Hand Surg Am. 2003, 28: 1029-34.

Pallant J. Part 3: Preliminary analyses: Manipulating the data. SPSS Survival Manual, 4th ed. Maidenhead, Open University Press, 2010.

Pallant J. Part 4: Statistical techniques to explore relationships between variables: Factor analysis. SPSS Survival Guide, 4th ed. Maidenhead, Open University Press, 2010.

Pallant J. Part 4: Statistical techniques to explore relationships between variables: Logistic Regression. SPSS Survival Guide, 4th ed. Maidenhead, Open University Press, 2010.

Pandey A, Usman K, Reddy H, Gutch M, Jain N, Qidwai S. Prevalence of hand disorders in type 2 diabetes mellitus and its correlation with microvascular complications. Annals of medical and health sciences research. 2013, 3: 349-54.

Paterson C. Measuring outcomes in primary care: a patient generated measure, MYMOP, compared with the SF-36 health survey. BMJ. 1996, 312: 1016-20.

Patil VH, Singh SN, Mishra S, Donovan TD. Parallel Analysis Engine to Aid Determining Number of Factors to Retain [Computer Software]. 2007. http://smishra.faculty.ku.edu/parallelengine.htm (2nd August 2014).

Peimer CA, Blazar P, Coleman S et al. Dupuytren contracture recurrence following treatment with collagenase clostridium histolyticum (CORDLESS study): 3-year data. J Hand Surg Am. 2013, 38: 12-22.

Pess GM, Pess RM, Pess RA. Results of needle aponeurotomy for Dupuytren contracture in over 1,000 fingers. J Hand Surg Am. 2012, 37: 651-6.

Polson K, Reid Duncan D, McNair PJ, Larmer P. Responsiveness, minimal importance difference and minimal detectable change scores of the shortened disability arm shoulder hand (QuickDASH) questionnaire. Manual Therapy. 2010, 15: 404-7.

Poltawski L, Watson T. Measuring clinically important change with the Patient-rated Tennis Elbow Evaluation. Hand Therapy. 2011, 16: 52-7.

Pratt AL, Byrne G. The lived experience of Dupuytren's disease of the hand. J Clin Nurs. 2009, 18: 1793-802.

Qureshi FI, Hornigold R, Spencer JD, Hall SM. Langerhans cells in Dupuytren's contracture. J Hand Surg Brit Eur. 2001, 26: 362-7.

Rayan GM, Parizi M, Tomasek JJ. Pharmacologic regulation of Dupuytren's fibroblast contraction in vitro. J Hand Surg Am. 1996, 21: 1065-70.

Reilly RM, Stern PJ, Goldfarb CA. A retrospective review of the management of Dupuytren's nodules. J Hand Surg Am. 2005, 30: 1014-8.

Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. J Clin Epidemiol. 2008, 61: 102-9.

Ring D. Benchmarking changes in symptom intensity and magnitude of disability. J Hand Surg Am. 2013, 38: 650-1.

Ritchie JF, Venu KM, Pillai K, Yanni DH. Proximal interphalangeal joint release in Dupuytren's disease of the little finger. J Hand Surg Brit Eur. 2004, 29: 15-7.

Rodrigues JN. How is recurrence of Dupuytren's disease defined, and can contractures be assessed more reliably? Department of Trauma & Orthopaedic Surgery, Queen's Medical Centre. Nottingham, University of Edinburgh, 2010, Vol. Master of Sciences in Surgical Sciences.

Rodrigues JN, Mabvuure NT, Nikkhah D, Shariff Z, Davis TR. Minimal important changes and differences in elective hand surgery. J Hand Surg Brit Eur. 2014.

Rodrigues JN, Zhang W, Scammell BE, Davis TR. Dynamism in Dupuytren's contractures. J Hand Surg Brit Eur. 2014; 40(2):166-70.

Rodrigues JN, Zhang W, Scammell BE, Davis TR. What patients want from the treatment of Dupuytren's disease -- is the Unite Rhumatologique des Affections de la Main (URAM) scale relevant? J Hand Surg Brit Eur. 2014.

Ross DC. Epidemiology of Dupuytren's disease. Hand Clin. 1999, 15: 53-62, vi.

Saboeiro AP, Porkorny JJ, Shehadi SI, Virgo KS, Johnson FE. Racial distribution of Dupuytren's disease in Department of Veterans Affairs patients. Plastic & Reconstructive Surgery. 2000, 106: 71-5.

Sauerland S, Lefering R, Bayer-Sandow T, Bruser P, Neugebauer EA. Fingers, hands or patients? The concept of independent observations. J Hand Surg Brit Eur. 2003, 28: 102-5.

Schmitt J, Di Fabio RP. The Validity of Prospective and Retrospective Global Change Criterion Measures. Arch Phys Med Rehabil. 2005, 86: 2270-6.

Schmitt JS, Di Fabio RP. Reliable change and minimum important difference (MID) proportions facilitated group responsiveness comparisons using individual threshold criteria. J Clin Epidemiol. 2004, 57: 1008-18.

Schuck P. Assessing reproducibility for interval data in health-related quality of life questionnaires: which coefficient should be used? Qual Life Res. 2004, 13: 571-86.

Schunemann HJ, Guyatt GH. Commentary--goodbye M(C)ID! Hello MID, where do you come from? Health services research. 2005, 40: 593-7.

Seah MK, Harry L, Davidson D. A three-digit dermofasciectomy for Dupuytren's disease. Hand Surg. 2012, 17: 109-10.

Sears ED, Chung KC. Validity and responsiveness of the Jebsen-Taylor Hand Function Test. J Hand Surg Am. 2010, 35: 30-7.

Seegenschmiedt MH, Olschewski T, Guntrum F. Radiotherapy optimization in early-stage Dupuytren's contracture: first results of a randomized clinical study. Int J Radiat Oncol Biol Phys. 2001, 49: 785-98.

Shauver MJ, Chung KC. The minimal clinically important difference of the Michigan hand outcomes questionnaire. J Hand Surg Am. 2009, 34: 509-14.

Sinha R, Cresswell TR, Mason R, Chakrabarti I. Functional benefit of Dupuytren's surgery. J Hand Surg Brit Eur. 2002, 27: 378-81.

Slullitel M. Our behaviour in Dupuytren's disease [Nuestro comportamiento ante la enfermedad de Dupuytren / Dupuytren´s disease]. Revista de la Asociación Argentina de Ortopedia y Traumatología. 1987, 52: 103-6.

Smith P, Breed C. Central slip attenuation in Dupuytren's contracture: a cause of persistent flexion of the proximal interphalangeal joint. J Hand Surg Am. 1994, 19: 840-3.

Sorensen AA, Howard D, Tan WH, Ketchersid J, Calfee RP. Minimal clinically important differences of 3 patient-rated outcomes instruments. Journal of Hand Surgery. 2013, 38: 641-9.

Spicer J. Logistic regression and discriminant analysis. In: Spicer J (Ed.) Making Sense of Multivariate Data Analysis. Thousand Oaks, California, Sage Publications, 2004: 123-53.

Spies-Dorgelo MN, Terwee CB, Stalman WA, van der Windt DA. Reproducibility and responsiveness of the Symptom Severity Scale and the hand and finger function subscale of the Dutch arthritis impact measurement scales (Dutch-AIMS2-HFF) in primary care patients with wrist or hand problems. Health & Quality of Life Outcomes. 2006, 4: 87.

Srivastava S, Nancarrow JD, Cort DF. Dupuytren's disease in patients from the Indian sub-continent. Report of ten cases. J Hand Surg Brit Eur. 1989, 14: 32-4.

Stanton-Hicks M, Janig W, Hassenbusch S, Haddox JD, Boas R, Wilson P. Reflex sympathetic dystrophy: changing concepts and taxonomy. Pain. 1995, 63: 127-33.

Starkweather KD, Lattuga S, Hurst LC et al. Collagenase in the treatment of Dupuytren's disease: an in vitro study. J Hand Surg Am. 1996, 21: 490-5.

Szabo RM. Outcomes assessment in hand surgery: when are they meaningful? J Hand Surg Am. 2001, 26: 993-1002.

Tashjian RZ, Deloach J, Porucznik CA, Powell AP. Minimal clinically important differences (MCID) and patient acceptable symptomatic state (PASS) for visual analog scales (VAS) measuring pain in patients treated for rotator cuff disease. Journal of Shoulder and Elbow Surgery. 2009, 18: 927-32.

Testa MA. Interpreting quality-of-life clinical trial data for use in the clinical practice of antihypertensive therapy. Journal of hypertension. Supplement : official journal of the International Society of Hypertension. 1987, 5: S9-13.

Tomasek JJ, Vaughan MB, Haaksma CJ. Cellular structure and biology of Dupuytren's disease. Hand Clin. 1999, 15: 21-34.

Tripoli M, Cordova A, Moschella F. Dupuytren's contracture as result of prolonged administration of phenobarbital. European Review for Medical & Pharmacological Sciences. 2011, 15: 299-302.

Tubiana R. Clinical aspects. In: Tubiana R, Leclercq C, Hurst L, Badalamente M, Mackin E (Eds.) Dupuytren's Disease. London, Martin Dunitz, 2000: 79-121.

Tubiana R. Treatment: Fasciectomy. In: Tubiana R, Leclercq C, Hurst L, Badalamente M, Mackin E (Eds.) Dupuytren's Disease. London, Martin Dunitz, 2000: 139-79.

Tubiana R, Thomine J-M, Mackin E. 1.3 Movements of the hand and wrist. In: Tubiana R, Thomine J-M, EJ M (Eds.) Examination of the hand and wrist, 2nd ed. London, Martin Dunitz, 1996.

Turner D, Schunemann HJ, Griffith LE et al. Using the entire cohort in the receiver operating characteristic analysis maximizes precision of the minimal important difference. J Clin Epidemiol. 2009, 62: 374-9.

Ullah AS, Dias JJ, Bhowal B. Does a 'firebreak' full-thickness skin graft prevent recurrence after surgery for Dupuytren's contracture?: a prospective, randomised trial. J Bone Joint Surg Br. 2009, 91: 374-8.

van der Giesen FJ, Nelissen RG, Arendzen JH, de Jong Z, Wolterbeek R, Vliet Vlieland TP. Responsiveness of the Michigan Hand Outcomes Questionnaire--Dutch language version in patients with rheumatoid arthritis. Arch Phys Med Rehabil. 2008, 89: 1121-6.

van Rijssen AL, Gerbrandy FS, Ter Linden H, Klip H, Werker PM. A comparison of the direct outcomes of percutaneous needle fasciotomy and limited fasciectomy for Dupuytren's disease: a 6-week follow-up study. J Hand Surg Am. 2006, 31: 717-25.

van Rijssen AL, ter Linden H, Werker PM. Five-year results of a randomized clinical trial on treatment in Dupuytren's disease: percutaneous needle fasciotomy versus limited fasciectomy. Plast Reconstr Surg. 2012, 129: 469-77.

van Rijssen AL, Werker PM. Percutaneous needle fasciotomy in dupuytren's disease. J Hand Surg Brit Eur. 2006, 31: 498-501.

van Rijssen AL, Werker PM. Percutaneous needle fasciotomy for recurrent Dupuytren disease. J Hand Surg Am. 2012, 37: 1820-3.

Vittinghoff E, McCulloch CE. Relaxing the rule of ten events per variable in logistic and Cox regression. American journal of epidemiology. 2007, 165: 710-8.

von Campe A, Mende K, Omaren H, Meuli-Simmen C. Painful nodules and cords in Dupuytren disease. J Hand Surg Am. 2012, 37: 1313-8.

Waljee JF, Chung KC. Objective functional outcomes and patient satisfaction after silicone metacarpophalangeal arthroplasty for rheumatoid arthritis. J Hand Surg Am. 2012, 37: 47-54.

Ward CM. Oedema of the hand after fasciectomy with or without tourniquet. Hand. 1976, 8: 179-85.

Warwick D, Arner M, Pajardi G et al. Collagenase Clostridium histolyticum in patients with Dupuytren's contracture: results from POINT X, an open-label study of clinical and patient-reported outcomes. J Hand Surg Brit Eur. 2014.

Werker PMN, Pess GM, van Rijssen AL, Denkler K. Correction of Contracture and Recurrence Rates of Dupuytren Contracture Following Invasive Treatment: The Importance of Clear Definitions. J Hand Surg Am. 2012, 37: 2095-105.e7.

Wilburn J, McKenna SP, Perry-Hinsley D, Bayat A. The impact of dupuytren disease on patient activity and quality of life. J Hand Surg Am. 2013, 38: 1209-14.

Witthaut J, Bushmakin AG, Gerber RA, Cappelleri JC, Le Graverand-Gastineau MP. Determining clinically important changes in range of motion in patients with Dupuytren's Contracture: secondary analysis of the randomized,

double-blind, placebo-controlled CORD I study. Clin Drug Investig. 2011, 31: 791-8.

World Health Organisation. International Classification of Functioning, Disability and Health (ICF). Wolrd Health Organisation, 2014. http://www.who.int/classifications/icf/en/ (22nd March 2015).

Yoshida K, Yamada Y, Hara H, Yamanaka K, Inoue H. The surgical treatment for Dupuytren's contracture. 7th Congress of the International Federation of Societies for Surgery of the Hand. 1998: 307-11.

Youden WJ. Index for rating diagnostic tests. Cancer. 1950, 3: 32-5.

Zachariae L. Extensive versus limited fasciectomy for Dupuytren's contracture. Scandinavian Journal of Plastic & Reconstructive Surgery. 1967, 1: 150-3.

Zachariae L. Dupuytren's contracture. How limited should a limited fasciectomy be? Scand J Plast Reconstr Surg. 1969, 3: 145-9.

Zerajic D, Finsen V. Dupuytren's disease in Bosnia and Herzegovina. An epidemiological study. BMC Musculoskelet Disord. 2004, 5: 10.

Zyluk A, Jagielski W. The effect of the severity of the Dupuytren's contracture on the function of the hand before and after surgery. J Hand Surg Brit Eur. 2007, 32: 326-9.