Severity and progression of hand pain and functional difficulty: a prospective cohort study in community-dwelling older adults with hand pain

Elaine Elizabeth Nicholls

A thesis submitted for the degree of Doctor of Philosophy

December 2016

Arthritis Research UK Primary Care Centre

Keele University

Declaration

The datasets used in this thesis form part of a larger program of work investigating the epidemiology of, and treatment for, hand problems in older adults. The research programs that provided the datasets were led by Professors Peter Croft, George Peat, Krysia Dziedzic and Danielle van der Windt and were funded as follows: the Clinical Assessment Study of the Hand (CAS-HA), baseline to 3-year follow-up, Medical Research Council (MRC) Programme Grant (Grant Code: G9900220), 54-month and 6-year follow-ups Arthritis Research UK (ARUK) Programme Grant (Grant Code: 18174); Self-Management in Osteoarthritis of the Hand (The SMOotH study), funded by ARUK, Grant Code: 17958. Support for Science Funding was secured for both studies by the North Staffordshire Primary Care Research Consortium for NHS service support costs.

The funding and baseline data collection for the CAS-HA study had been completed before I enrolled as a PhD student, so I had no involvement in the baseline study design or data collection procedures. However since joining the study team, I was employed as the study co-ordinator at the 18-month follow-up (assisting in the process to gain ethical approval and in the day-to-day running of the study) and the study statistician for all follow-up time points (responsible for testing and checking the accuracy of the study databases after data entry had been completed by administrative staff, and many aspects of analysis). I was also involved with the SMOoTH study as the study statistician throughout its development and implementation, although this data set forms only a small part of the thesis overall.

I can confirm that I have conducted all of the analyses presented in this thesis and have written, and developed, the research project under the guidance of my supervisors, Elaine Thomas and Danielle van der Windt.

Keele University

SUBMISSION OF THESIS FOR A RESEARCH DEGREE

DECLARATION by the candidate for a research degree.

Degree for which thesis being submitted: PhD

Title of thesis: Severity and progression of hand pain and functional difficulty: a prospective cohort study in community-dwelling older adults with hand pain

Date of submission:

Original registration date: 8th January 2008

Name of candidate: Elaine Elizabeth Nicholls

Research Institute: Primary Care and Health Sciences

Name of Lead Supervisor: Dr Elaine Thomas

I certify that:

(a) The thesis being submitted for examination is my own account of my own research
(b) My research has been conducted ethically. Where relevant a letter from the approving body confirming that ethical approval has been given has been bound in the thesis as Appendix 1
(c) The data and results presented are the genuine data and results actually obtained by me during the conduct of the research
(d) Where I have drawn on the work, ideas and results of others this has been appropriately acknowledged in the thesis
(e) Where any collaboration has taken place with one or more other researchers, I have included within an 'Acknowledgments' section in the thesis a clear statement of their contributions, in line with the relevant statement in the Code of Practice.
(f) The greater portion of the work described in the thesis has been undertaken subsequent to my registration for the higher degree for which I am submitting for examination
(g) Where part of the work described in the thesis has previously been incorporated in another thesis submitted by me for a higher degree (if any), this has been identified and acknowledged in the thesis
(h) The thesis submitted is within the required word limit as specified in the Regulations

Total words in submitted thesis (including text and footnotes, but excluding references and appendices) …86,130…..

Signature of candidate ……Date ……

Note

Extract from Code of Practice: If the research degree is set within a broader programme of work involving a group of investigators – particularly if this programme of work predates the candidate's registration – the candidate should provide an explicit statement (in an 'Acknowledgments' section) of the respective roles of the candidate and these other individuals in relevant aspects of the work reported in the thesis. For example, it should make clear, where relevant, the candidate's role in designing the study, developing data collection instruments, collecting primary data, analysing such data, and formulating conclusions from the analysis. Others involved in these aspects of the research should be named, and their contributions relative to that of the candidate should be specified (*this does not apply to the ordinary supervision, only if the supervisor or supervisory team has had greater than usual involvement*).

Abstract

Background

Hand problems are common in older adults causing pain and disruption to daily living. Understanding prognosis of such problems is therefore important to provide information on likely symptom course and to target treatment to those in most need. The aim of this thesis was to investigate prognosis of hand pain and functional difficulty in community-dwelling older adults with hand pain.

Methods

The majority of data analysis was based on a cohort of adults aged 50 years and over reporting hand pain in the last 12-months at baseline (N=623). The Australian/Canadian Hand Osteoarthritis Index (AUSCAN) was the primary measure of hand pain (0-10) and function (0-10), measured at baseline and four follow-up time-points (1.5, 3, 5, and 7.5-years). Random effect models, latent class growth models, parallel process growth models and parallel process growth mixture models were used to model longitudinal trajectories of AUSCAN pain and functional difficulty over time.

Results

Trajectories of hand pain and functional difficulty were shown to be relatively stable for the majority of participants over the 7.5-year follow-up period with an overall mean change per year of 0.05 (95% confidence interval: 0.02, 0.07) and 0.07 (95% confidence interval 0.05, 0.09) points for AUSCAN pain and function respectively. Although combinations of predictors were identified that predicted symptom course, the strongest predictor was the baseline measure for the outcome of interest, with model fit not greatly improved by adding three further predictors to the model e.g. Nagelkerke's pseudo R-square: Hand pain, baseline only 0.64; with additional predictors 0.70; Hand function, 0.80 and 0.83 respectively. A group of participants with hand pain trajectories that differed greatly from

their hand function trajectories was not identified suggesting that changes in hand pain are linked to changes in hand function over time.

Conclusions

Progression of hand symptoms was not inevitable for all participants when assessed over a 7.5 year time-period. Baseline symptom severity may be the single most important predictor to identify those with an unfavourable symptom course and where early onward referral/treatment may be useful. This work remains exploratory however until findings are replicated in an external dataset.

Key words: hand, pain, function, prognosis, older adults

Acknowledgements

I would like to thank my supervisors, Dr Elaine Thomas and Professor Danielle van der Windt for their encouragement, support and guidance throughout this research project – both their patience and wisdom has been invaluable throughout the development of this thesis. I'd also like to thank the funders of the project (listed previously) for their financial support and also all staff at the Arthritis Research UK Primary Care Centre who dedicated their time, and expertise, to design and collect the data for the studies presented in this thesis.

In particular, I'd like to thank my current colleagues: Professor George Peat (for his unending enthusiasm and helpful discussions around the analysis of symptom trajectories), Professor Krysia Dziedic, Dr Michelle Marshall and Dr Helen Myers (for their keen interest to understand and discuss all aspects of hand problems in older adults), Joanne Jordan (for training and guidance on systematic review design), Dr Vicky Strauss and Professor Linda Muthen (for helpful discussions on the workings of Mplus) and also previous colleagues who were pivotal in encouraging me to pursue a research career and undertake this PhD (Professor Mark Haggard, Mary Gannon, Helen Spencer, Vicki Owen and Dr Josie Higson).

Finally, I'd like to thank all my friends who have supported me through all stages of this PhD project, especially those who have joined me for lunch and those who have helped me to relax at the end of the working day. To my family also, for whose support and encouragement I will always be grateful.

Table of contents

List of tables

List of figures

List of boxes

Abbreviations

ACR = American College of Rheumatology

AIC = Akaike's Information Criteria

AIMS2 = Arthritis Impact Measurement Scales 2

AMED = Allied and Complementary Medicine

ANOVA = Analysis of Variance

ARUK = Arthritis Research UK

AUSCAN = Australian/Canadian Hand Osteoarthritis Index

BIC = Bayesian Information Criterion

BL = Baseline

BMI = Body Mass Index

BNI = British Nursing Index

CAS-HA = Clinical Assessment Study of the Hand

CAS-K = Clinical Assessment Study of the Knee

CASP = Critical Appraisal Skills Programme

CFA = Confirmatory Factor Analysis

CFI = Comparative Fit Index

CI = Confidence Interval

CMC1 = First Carpometacarpal Joint

DIP = Distal Interphalangeal Joints

EULAR = European League Against Rheumatism

GARP = Genetics ARthrosis and Progression study

GAT = Hand Grip-ability Test

GBDTM = Group-based Dual Trajectory Model

GEE = Generalised Estimating Equations

GLM = Generalised Linear Model

GM = Growth Models

GMM = Growth Mixture Models

GOGO = Genetics of Generalized Osteoarthritis study

GP = General Practitioner

HADS = Hospital Anxiety and Depression scale

HAQ = Stanford Health Assessment Questionnaire

HMIC = Health Management Information Consortium

ICC = Intra-class Correlation Coefficient

IP = First Interphalangeal Joint

IPQ-R = Illness Perceptions Questionnaire

IQR = Inter-Quartile Range

KL = Kellgren-Lawrence

LASSO = Least Absolute Shrinkage and Selection Operator

LCGM = Latent Class Growth Models

LCM = Latent Curve Models

LGM = Latent Growth Models

LMR-LRT = Lo-Mendell-Rubin Likelihood Ratio Test

LTM = Latent Trajectory Models

MAR = Missing At Random

MCAR = Missing Completely At Random

MCP = Metacarpophalangeal Joints

MCS = Mental Component Score

MDC = Minimum Data Collection

MIC = Minimum Important Change

ML = Maximum Likelihood

MLR = Maximum Likelihood with Robust Standard Errors

MNAR = Missing Not At Random

MRC = Medical Research Council

NHS = National Health Service

NICE = National Institute for Clinical Excellence

NorStOP = North Staffordshire Osteoarthritis Project

NSAID = Non-steroidal Anti-inflammatory Drug

OA = Osteoarthritis

OARSI = Osteoarthritis Research Society International

ONS = Office for National Statistics

OR = Odds Ratio

PB-LRT = Parametric Bootstrapped Likelihood Ratio Test

PCS = Physical Component Score

PIP = Proximal Interphalangeal Joints

PPGM = Parallel Process Growth Model

PPGMM = Parallel Process Growth Mixture Model

PPLCGM = Parallel Process Latent Class Growth Model

PROGRESS = PROGnosis RESearch Strategy

QIC = Quasi-likelihood Independence Model Information Criterion

QUIPS = QUality In Prognosis Studies

RA = rheumatoid arthritis

RC = Research Clinic

REML = Restricted Maximum Likelihood

RMSEA = Root Mean Square Error of Approximation

ROC = Receiver Operating Characteristic

RRR = Relative Risk Ratio

SBIC = Sample-size-adjusted Bayesian Information Criterion

SBSCT = Satorra-Bentler Scaled Chi-square Test

SD = Standard Deviation

SDC = Smallest Detectable Change

SEM = Standard Error of Measurement

SF-12 = Short-form 12

SIGN = Scottish Intercollegiate Guidelines Network

SMOotH = Self-management in Osteoarthritis of the Hand trial

SRM = Standardised Response Mean

SRMR = Standardised Root Mean-square Residual

TLI = Tucker-Lewis Index

TS = Trapezio-scaphoid

UK = United Kingdom

VLMR-LRT = Vuong-Lo-Mendell-Rubin Likelihood Ratio Test

WHO = World Health Organisation

1 Introduction

1.1 Musculoskeletal disorders – background

Musculoskeletal disorders represent a wide range of conditions that affect the joints, bones, muscles and soft tissue of the human body (Parsons et al. 2011). Although a diverse range of conditions (exceeding 200 in number (Parsons et al. 2011)) they are commonly linked by their impact on pain and functional difficulty (Woolf et al. 2003). Musculoskeletal disorders are common in the general population and a common cause of longstanding illness (Woolf et al. 2003).

In the UK, it is estimated that 137 adults per 1000 of the population will report a disabling musculoskeletal condition (Office for National Statistics (ONS) 2009). This equates to just over 3-million adults, given the size of the UK population in 2010 and defining disability to be moderate to severe (Parsons et al. 2011). Patients with such conditions frequently report difficulties with completing tasks of everyday living and participating in work, hobbies and social activities (Hill et al. 2010, Wilkie et al. 2007), thus reducing quality of life for the individual and any family members involved in their care. The cost of such conditions has also been estimated in monetary terms by direct costs (e.g. use of medical services) and indirect costs (e.g. loss of productivity at work). In the UK it is estimated that the annual monetary cost of musculoskeletal disorders is £5.3 billion (Arthritis Research UK 2014).

It is clear that musculoskeletal disorders have major individual, societal and economic impact (Woolf et al. 2003), and are reported as a leading cause of years lived with disability in Western Europe (Vos et al. 2012). This has prompted governments worldwide to set musculoskeletal conditions as a priority for future health-planning and research, with the World Health Organisation (WHO) declaring the ten year period (2000 – 2010) as the 'Bone and Joint' decade – an initiative set up to raise awareness of the growing burden of

musculoskeletal disorders on society. The period for the 'Bone and Joint' decade has since been extended to 2020 highlighting the need for continued high quality research in this area (Bone and Joint decade's Musculoskeletal Portal 2015).

1.2 Hand pain and function in older adults

The hand is a common site for musculoskeletal pain (Urwin et al. 1998). Prevalence estimates for hand pain in the general population range between 5 and 26% depending on how severity and duration of symptoms are defined (Palmer 2003). Unlike the lower limb, the hand is not a weight-bearing joint and (for most people) is not directly involved in mobility. However, its role in achieving satisfactory function is clear, with the ability to grip/pinch objects, touch and feel sensation being vital to the completion of many tasks of everyday living (Bland et al. 2008). Hand problems are therefore the focus of this thesis, and worthy of study, separate from other areas of the body, particularly as they have been less frequently researched in older adults than musculoskeletal problems at other body sites, e.g. the hip or knee (Myers et al. 2007).

The prevalence of hand pain increases with age (Palmer 2003), and for adults aged over 55 and 50 years, respectively, the one-month and one-year period prevalence of hand pain is estimated at 17% and 30% (Dahaghin al. 2005b, Dziedzic et al. 2007). It is anticipated that the percentage of older people in the UK will increase in future years and projections from the Office for National Statistics estimate that by 2030 nearly a quarter of the UK population will be aged 65 years and over (Office for National Statistics (ONS) 2012). The number of older adults with hand problems in the UK can therefore only rise as a consequence.

Older adults represent a particular sector of the population that warrant attention as distinct from the population as a whole. Firstly, older adults are at greater risk of hand pain (Palmer 2003) and risk factors for progression of hand pain may differ from those in their younger years (Gagliese 2009). For example, the role of occupational risk factors may

differ in populations of mainly retired workers and the influence of any co-existing medical conditions may also impact on outcome in older adults. Secondly, the profile of the underlying cause of hand pain varies with age. For example, hand osteoarthritis is the most common cause of hand pain in older adults (National Collaborating Centre for Chronic Conditions 2008), yet this condition is relatively rare for adults aged 18 – 45 years (Haugen et al. 2011).

1.3 Prognosis

Prognosis is defined as the probable course or prediction of the outcome of a health condition over time (Hayden et al. 2010). Providing information to patients and clinicians on prognosis of a health condition is important and musculoskeletal conditions are no exception: in a postal survey of patients consulting for musculoskeletal conditions, 82% reported that information on prognosis would be useful, however only 33% reported discussing prognostic information with their general practitioner (GP) (Mallen et al. 2009).

Prognosis research has been described by the PROGnosis RESearch Strategy (PROGRESS) Partnership[1] as including four key inter-related themes: (1) exploring the course of a condition in the context of current care (Hemingway et al. 2013), (2) identifying specific factors associated with prognosis (Riley et al. 2013), (3) developing, validating, and exploring the impact of using statistical models to predict individual risk of a future outcome (Steyerberg et al. 2013), and (4) using prognostic information to help tailor treatment decisions to individuals with similar characteristics, i.e. stratified medicine (Hingorani et al. 2013). However, to achieve these aims, well designed longitudinal studies are needed.

For musculoskeletal conditions, longitudinal studies with long-term follow-up are desirable so that the prognosis of conditions that are potentially chronic and of a long duration can be tracked over time. In addition, studies with outcome measurements collected at

[1] The PROGRESS Partnership is an initiative funded by the UK Medical Research Council that aims to improve the quality and reporting of prognosis studies: http://progress-partnership.org/

multiple time-points are also needed so that symptom course can be reliably tracked over time. Such studies, however, are lacking for people with hand pain and other common musculoskeletal conditions, especially in primary care and the community (Mallen et al. 2007a). A recent report from the European League Against Rheumatism (EULAR) highlighted that a future research priority for osteoarthritis research is to increase the evidence base on the natural history and progression of osteoarthritis over time (Conaghan et al. 2014). There is therefore a gap in the literature for further studies to provide information on the prognosis of musculoskeletal hand conditions over time.

1.4 Summary and thesis aims

In summary, hand pain is prevalent in the population and commonly affects older adults. It affects the ability to function and reduces quality of life. A greater understanding of the clinical course of hand pain is needed, both to understand the nature and severity of the condition and to inform patients and clinicians of the likely future course and impact of the condition.

The overall aim of this thesis is therefore to investigate the clinical course and prognosis of hand pain and function over time in community-dwelling older adults with hand pain.

Specifically, three key objectives are addressed:

1) To describe the long-term (6-year) trajectories of hand pain and functional difficulty in community-dwelling older adults;

2) To identify baseline predictors of poor prognosis (i.e. worsening of symptoms over time) and to develop potential prognostic models (prediction rules) to identify those at risk of poor outcome;

3) To explore the simultaneous relationship between changes in hand pain and functional difficulty over time.

Although the focus of this thesis is on symptom course over time, rather than diagnosis, it is recognised that multiple causes exist for hand symptoms. The prevalence of such causes will be reported in this thesis, however in a population of older adults, it is recognised that the most likely cause of hand symptoms is osteoarthritis. The data analysed in this thesis therefore represent a population of older adults who either have, or are at risk of, developing hand osteoarthritis over time.

1.5 Thesis overview

Chapter 2 – Factors associated with hand pain and functional difficulty in older adults: a systematic review

In this chapter a systematic search of the literature is presented to summarise current evidence on factors associated with severity and progression of self-reported hand pain and functional difficulty in older adults. The aim of this search is to help identify potential prognostic factors that can then be later tested for their prognostic value in the models that are subsequently presented in Chapters 7 and 8.

Chapter 3 – Recruitment design and data collection procedures

The recruitment design and data collection methods for the Clinical Assessment Study of the Hand (CAS-HA) are presented in this chapter as this study provides the primary source of data analysed in this thesis. In addition, any evidence for selection bias in the cohort is explored.

Chapter 4 – Psychometric properties of the Australian/Canadian Hand Osteoarthritis Index (AUSCAN)

The psychometric properties of the AUSCAN are explored in this chapter and the suitability of this measure to assess hand pain and function is evaluated. This analysis is of particular relevance to this thesis as the AUSCAN is used as the primary measure of hand pain and functional difficulty throughout.

Chapter 5 – Statistical methodology

The statistical methods used in this thesis are described in this chapter and evaluated, both for their strengths and weaknesses, and also for their suitability to address the research questions as stated in Section 1.4.

Chapter 6 – Describing the trajectory of hand pain and functional difficulty in CAS-HA

The AUSCAN (as described in Chapter 4) is applied to the CAS-HA data set in this chapter and its distribution, rates of missing data and course over time are described for the CAS-HA sample as a whole. The analysis in this chapter aims to address Objective 1 as stated in Section 1.4.

Chapter 7– Predicting the course of hand pain and function over time.

The aim of this chapter is to explore whether a set of baseline factors can be identified that predict the course of hand pain and functional difficulty over time (using a pool of baseline variables as identified from the systematic review in Chapter 2). The analysis in this chapter aims to address Objective 2 in Section 1.4.

Chapter 8 – Trajectory subgroups for hand pain and function in CAS-HA

The analysis in this chapter is an extension of that presented in Chapters 6 and 7. The first aim of the chapter is to explore whether subgroups of participants can be identified that have a differing outcome course over time, for hand pain or hand function when analysed as two separate outcomes of interest. The second aim is then to explore whether the baseline predictors, as presented in the thesis so far, differ across the outcome subgroups identified, or predict in combination subgroup membership as defined. The analysis in this chapter aims to address Objectives 1 and 2 in Section 1.4.

Chapter 9 – Joint trajectory modelling

This final results chapter aims to extend the analysis in this thesis by considering the outcomes of hand pain and function jointly, i.e. to explore whether there is a relationship between changes in hand pain over time and changes in hand function. In addition, it is also explored whether groups of participants can be identified that have a differing course of hand pain and hand function over time (e.g. a group of participants with increasing hand pain, but stable hand function). This analysis aims to meet Objective 3 in Section 1.4.

Chapter 10 – Discussion and conclusions

A summary of the key findings from this thesis are presented in this chapter, along with a consideration of the strengths and weaknesses of the analyses as presented. The discussion around strengths and weaknesses will focus solely on issues that are relevant across the thesis as a whole as chapter-specific issues are discussed in separate sections

at the end of each chapter. In this chapter (Chapter 10), the overall implications of these findings for future research are also discussed.

2 Factors associated with hand pain and functional difficulty in older adults: a systematic review

2.1 Introduction

It was highlighted in the previous chapter that prognosis of hand pain and function is important. However, before prognosis can be fully evaluated, potential prognostic factors need to be identified. The aim of this chapter is to present a systematic review that summarises the available evidence on factors associated with progression and severity of hand pain and functional difficulty in community-dwelling older adults. The results of this review are then planned to inform and generate a list of potential prognostic factors that can be tested for their prognostic value in later thesis chapters.

The chapter will be presented in three sections: the first will describe the methodology used to locate and evaluate the quality of the papers in the review (Section 2.2), the second to present the results of the review (Section 2.3) and the third to discuss the findings and their implications for later thesis chapters (Section 2.4).

2.2 Methods

2.2.1 Selection criteria

Publications were included in the review if they had explored factors associated with severity or progression of self-reported hand pain or functional difficulty in older adults selected from the general population[2]. Studies were excluded if they: (1) measured only presence of hand pain or functional difficulty (yes/no), (2) were based in subgroups of the general population (such as those with specific medical complaints, e.g. Parkinson's disease, or specific hand conditions, e.g. rheumatoid arthritis (RA)), (3) were not written in English, (4) were not original research published in a peer-reviewed journal, (5) reported only measures of hand stiffness or numbness, (6) were studies of hand injury, treatment

[2] It was anticipated that the number of prognostic studies in the review would be small so cross-sectional studies were also included

or surgery, (7) were validation studies of questionnaire tools or diagnostic tests (e.g. x-ray or magnetic resonance imaging), or (8) were clinical case studies, case series, or qualitative studies.

2.2.2 Search strategy

NHS Healthcare Databases (search 2.0) was used to search the following databases for relevant review articles: MEDLINE, EMBASE, CINHAHL, BNI (British Nursing Index), AMED (Allied and Complementary Medicine), HMIC (Health Management Information Consortium), PsycINFO (psychology and allied fields). ISI Web of Knowledge was searched to identify any further key articles not included in the main search. The search included any publication in any of the databases prior to January 2011[3].

A maximum of three components were included in the search strategy and were combined using Boolean logic: (hand) AND (pain or function) AND (epidemiological study). Subject Headings (e.g. MESH headings) were used to describe the concepts of "hand", "pain" or "function" if available; otherwise text words were used (Table 2-1). A published search filter was used to focus the search to epidemiological studies in MEDLINE, EMBASE and CINHAL (Scottish Intercollegiate Guidelines Network) (Table 2-2). Titles and abstracts were searched in all databases with the exception of Web of Knowledge – a title search only was completed as an abstract search was not available.

[3] This review has not been updated since January 2011 as the results were needed at an early stage in the thesis to define a list of factors that could be included in the models presented in Chapters 7 and 8. An automatic alert however was created within NHS Healthcare Databases to run the search periodically so that any relevant papers published post January 2011 could be incorporated into the discussion chapter of the thesis. The systematic review has been previously published (Nicholls et al. 2012)

Table 2-1: Literature review search strategy

	Health Care Database							
Search term	MEDLINE[α]	CINHAL[β]	EMBASE[α]	BNI	AMED[β]	HMIC	PsycINFO[β]	Web of Knowledge
Hand								
MESH headings (not exploded[γ])								
Hand	+	+	+	+	+	+		
Hand joints	+							
Hand deformities	+	+						
Hand dermatoses	+							
Hand bones	+							
Carpal tunnel syndrome	+	+	+		+	+		
Dupuytren's contracture	+	+	+			+		
Tenosynovitis	+	+	+		+	+		
De Quervain disease	+							
Hand joint			+					
Hand bone			+					
Hand disease			+					
Hand eczema			+					
Hand endema			+					
Hand malformation			+					
Hand paresthesia			+					
Hand radiography			+					
Hand muscle			+					
Hand (anatomy)							+	
Keywords (title and abstract)								
Quervain*		+	+	+	+	+	+	
Dequervain*		+	+	+	+	+	+	
Carpal tunnel				+			+	
Dupuytren*				+	+		+	
Tenosynovitis				+			+	

Keywords (title only)						
Hand						+
Pain/Function						
MESH headings (exploded$^{\gamma}$)						
Pain	+	+	+	+	+	
Pain Measurement	+	+		+	+	
Activities of daily living	+	+		+	+	
Functional status		+				
Functional assessment		+				
Geriatric functional assessment		+				
Functional assessment inventory		+				
Pain assessment			+			
Musculoskeletal function			+			
Daily life activity			+			
Pain perception					+	
Keywords (title and abstract)						
Problem*	+	+	+	+	+	
Symptom*	+	+	+	+	+	
Function*	+			+	+	
Disabilit*	+	+	+	+	+	
Activit*	+	+	+	+	+	
Keywords (title only)						
Pain						+
Function						+
Epidemiological filter	+	+	+			

Footnote: Search terms within each major section were combined using a logical OR. The major sections were then combined using a logical AND (hand AND pain/function AND epidemiological filter (if used)).
+ = included in the search strategy
* = word truncation
α = Search restricted to Humans and English language articles
β = Search restricted to English language articles
γ = Subject headings can be searched in exploded format. This format generates a larger number of articles as it is searching not only the main heading but on all narrower subject headings relating to that topic as well

Table 2-2: Published epidemiological filters (reproduced from the Scottish Intercollegiate Guidelines Network website – SIGN)

MEDLINE	EMBASE	CINHAL
1. Epidemiologic studies/	1. Clinical study/	1. Prospective studies/
2. Exp case control studies/	2. Case control study	2. Exp case control studies/
3. Exp cohort studies/	3. Family study/	3. Correlational studies/
4. Case control.tw.	4. Longitudinal study/	4. Nonconcurrent prospective studies/
5. (cohort adj (study or studies)).tw.	5. Retrospective study/	5. Cross sectional studies/
6. Cohort analy$.tw.	6. Prospective study/	6. (cohort adj (study or studies)).tw.
7. (Follow up adj (study or studies)).tw.	7. Randomized controlled trials/	7. (observational adj (study or studies)).tw.
8. (observational adj (study or studies)).tw.	8. 6 not 7	8. or/1-7
9. Longitudinal.tw.	9. Cohort analysis/	
10. Retrospective.tw.	10. (Cohort adj (study or studies)).mp.	
11. Cross sectional.tw.	11. (Case control adj (study or studies)).tw.	
12. Cross-sectional studies/	12. (follow up adj (study or studies)).tw.	
13. Or/1-12	13. (observational adj (study or studies)).tw.	
	14. (epidemiologic$ adj (study or studies)).tw.	
	15. (cross sectional adj (study or studies)).tw.	
	16. Or/1-5,8-15	

Footnotes (reproduced from the SIGN website): / after an index term indicates that all subheadings were selected. "exp" before an index term indicates that the term was exploded. .tw. indicates a search for a term in title/abstract .mp. indicates a free text search for a term .$ at the end of a term indicates that this term has been truncated. adj indicates a search for two terms where they appear adjacent to one another

The titles and abstracts of publications generated by the search strategy were screened for possible inclusion in the review. In the first stage, two reviewers (EN & ET[4]) independently reviewed the first 100 abstracts to ensure inclusion and exclusion criteria were appropriately and consistently applied. One reviewer (EN) then continued to review the abstracts of the remaining papers, with a second independent reviewer being consulted where any ambiguity arose. A third reviewer (DvdW[5]) was involved when consensus was not achieved between the first two reviewers. Where inclusion or exclusion could not be determined from the abstract alone, full text articles were obtained and screened using the same consensus process that was applied at the abstract selection stage.

The reference lists of all articles included in the review were hand searched to identify any further relevant publications (EN). In addition, topic or clinical experts within our research centre (DvdW and KD[6]) were asked if any further articles could be identified for inclusion in the review.

2.2.3 Quality assessment

Multiple quality assessment tools exist that can serve as a guide to help reviewers to judge the overall quality and potential for bias in a study (e.g. methodology checklists from the Critical Appraisal Skills Programme (CASP), National Institute for Clinical Excellence (NICE), and the Scottish Intercollegiate Guidelines Network (SIGN)). In this review, study quality was assessed using the QUIPS tool - QUality In Prognosis Studies. (Hayden et al. 2006, Hayden et al. 2013).

The QUIPS tool was used as it was likely to encompass many of the quality assessment items that would be included in other quality instruments. This is noted as the QUIPS tool was developed by reviewing all of the quality items (including those on published checklists and those generated by individual authors) that had been used when assessing

[4] EN = Elaine Nicholls, ET = Elaine Thomas
[5] DvdW = Danielle van der Windt
[6] KD = Krysia Dziedzic

quality for systematic reviews of prognostic studies prior to October 2005. The QUIPS tool was used as it was relevant to the main study design that was of interest to this review, i.e. prognostic study design.

The QUIPS tool includes six major domains each addressing a possible bias that could occur in a prognostic study: study participation, study attrition, prognostic factor measurement, outcome measurement, confounding and analysis. A description of each bias is given for the reviewer to then rate whether the study is at “low”, “medium” or “high” risk of that particular bias.

The authors of the QUIPS tool encourage users to adapt the tool to make it applicable for the review being completed. For this review a minor adaptation was made to the tool to enable it to be applicable to both cross-sectional and longitudinal studies, namely that the domain heading “prognostic factor measurement” was simplified to “factor measurement” (so factors measured concurrently with outcome could also be included). Also, the domain for “study attrition” was automatically labelled as “not applicable” for cross-sectional studies as this could not be assessed for studies with no follow-up data.

The modified QUIPS tool (shown in Appendix 2) was applied to each paper in the review by two independent reviewers (EN & DvdW or EN & ET). Any disagreements in bias ratings were resolved by consensus.

2.2.4 Data extraction

Data extraction was completed for each article and included the following information: author, year of publication, study location, participant inclusion criteria, and measure of hand pain and function. Factors explored for association with hand pain and function were listed, and their strength of association recorded (e.g. odds ratio, mean difference, correlation). When more than one adjusted analysis was presented (from several multiple regression models), data were only extracted for the model with the highest number of

adjusting factors. All data extraction was completed by one reviewer (EN) and was checked by two independent reviewers for completeness and accuracy (DvdW or ET).

A meta-analysis to pool estimates of association was planned if data collection methods and statistical methodology were similar across studies. The meta-analysis would assess the heterogeneity of study results (including a test for homogeneity and computation of I^2) (Higgins al. 2003) and pooling of estimates by random effects modelling if appropriate (Kirkwood et al. 2003). A sensitivity analysis would test stability of associations after excluding any studies of poor quality, i.e. those scoring high risk from bias on any quality assessment domain.

2.3 Results

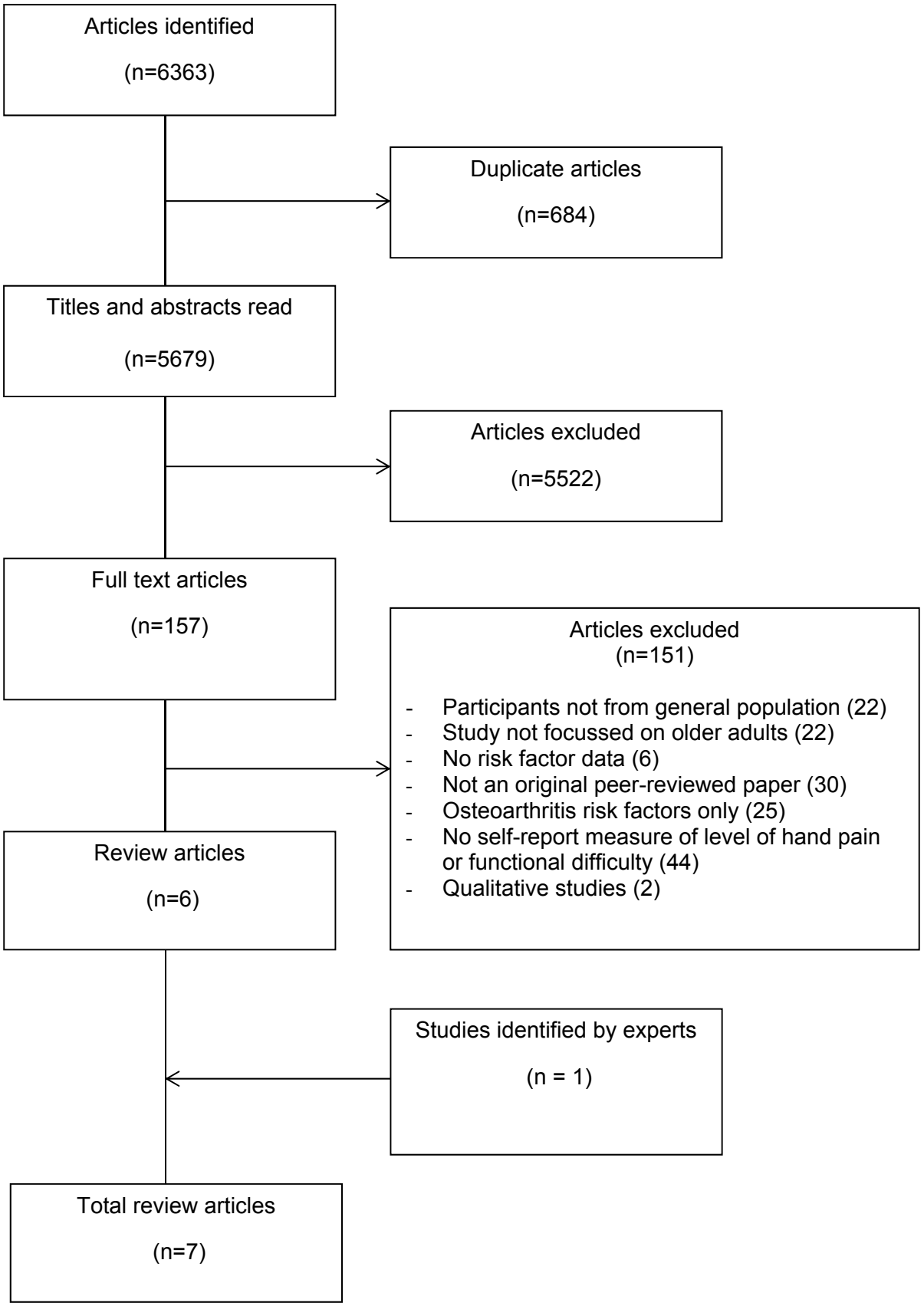
The search strategy identified 6363 citations (MEDLINE 2074; EMBASE 1287; CINHAL 220; BNI 129; AMED 525; HMIC 140; PsychINFO 561; Web of Knowledge 1427). After removal of duplicate citations in more than one database (duplicate citations identified by electronic filters), 5679 citations were considered for inclusion in the review. Screening of article titles and abstracts excluded 5207 articles from the review. Common reasons for exclusion were study samples not selected from the general population (e.g. studies evaluating the effectiveness of surgery or based in a group of patients with a specific clinical condition not directly related to the hand, e.g. stroke) or studies not focussed on older adults.

The remaining 472 abstracts were screened by a second reviewer (ET) and after a consensus meeting 315 were excluded. Articles were mainly removed because they focussed on a particular hand condition requiring specific treatment or specialist care (e.g. RA or carpal tunnel syndrome). Papers on hand osteoarthritis however were kept in at this stage to ensure that no studies using a clinical diagnosis of “hand osteoarthritis” based on “hand pain” were missed and also because this condition is likely to specifically affect the population of interest, i.e. older adults in the community.

A total of 157 full text articles were reviewed and after applying the exclusion criteria, six articles remained in the review (Dahaghin et al. 2005a, Dahaghin et al. 2005b, Hill et al. 2007, Baron et al. 1987, Niu et al. 2003, Marshall et al. 2009). An additional article (Dziedzic et al. 2007b) was identified by a clinical expert at our research centre (KD) giving a total of seven articles in the review[7]. A search of the reference lists of these seven articles did not yield any further articles for inclusion in the review. Further details of article selection are given in Figure 1.

[7] This article was not identified in the original search as the key words were “hand pain” and “hand function” rather than the separate components of “hand” AND “pain”, or “hand” AND “function”

Figure 1: Flow chart of stages of article selection



2.3.1 Description of articles included

Data extracted from all studies were cross-sectional in nature. The articles in the review were based on five independent studies of older adults, which varied in size from 32 to 7983 participants. Response rates varied across studies (16 – 79%).

Three self-reported measures were used to measure hand function: the Arthritis Impact Measurement Scales 2 (AIMS2) hand and finger function subscale (Meenan et al. 1992), the Australian/Canadian Hand Osteoarthritis Index (AUSCAN) hand and finger function subscale (Bellamy et al. 2002a) and the upper limb components of the Stanford Health Assessment Questionnaire (HAQ) (Fries et al. 1982). The AIMS2 and AUSCAN pain subscales (Meenan et al. 1992, Bellamy et al. 2002a) were also used to measure hand pain severity along with the number of painful joints with radiographic OA (Kellgren-Lawrence (KL) grade >=2) (Kellgren et al. 1957). Details regarding the seven articles included in the review are in Table 2-3.

Table 2-3: Summary of study populations, design and outcome measures in the review

Author (year) and Country	Study design	Study population	Response rate	Outcome Hand function	Outcome Hand pain
Dahaghin et al. (2005b) [α] Netherlands	Population-based cohort study	Inhabitants of Ommoord aged 55 years and over	N=7983. Response rate = 78%	Eight questions from HAQ (Fries et al. 1982) -higher score (>=0.5), more functional difficulty	Not included: measured as pain presence not severity
Dahaghin et al. (2005a) [α] Netherlands	Population-based cohort study.	Inhabitants of Ommoord aged 55 years and over	N=3906. Response rate = 35% (only cases with radiographic data at time of analysis are included)	Eight questions from HAQ (Fries et al. 1982) -higher score (>=0.5), more functional difficulty	Not included: measured as pain presence not severity
Hill et al. (2007) [β] UK	Population-based cohort study.	Participants aged 50 years and over, reporting hand pain on a health survey.	N=2113 2-stage survey: Survey 1 response rate = 71% Survey 2 response rate = 79%	Hand and finger function sub-scale of AIMS2 (Meenan et al. 1992) - higher score (>1.5), more functional difficulty	Hand pain sub-scale of AIMS2 (Meenan et al. 1992) - higher score (>3.5), more pain
Baron et al. (1987) Canada	Cohort Study	Tenants of a senior citizens apartment building aged 60 years and over	N=32. Response rate = 16%	Questions from the (HAQ) (Fries et al. 1982) on upper extremity activities - higher score, more functional difficulty	Self-reported data not given
Dziedzic et al. (2007b) [β] UK	Population-based cohort study	Participants aged 50 years and over, reporting hand pain on a health survey	N=2113 2-stage survey: Survey 1 response rate = 71% Survey 2 response rate = 79%	Hand and finger function sub-scale of AIMS2 (Meenan et al. 1992) - higher score, more functional difficulty. Severe functional difficulty = top 25% of observed sub-scale	Hand pain sub-scale of AIMS2 (Meenan et al. 1992) - higher score, more pain. Severe pain = top 25% of observed sub-scale

Niu et al. (2003) United States of America (U.S.A)	Cohort Study (Framingham Osteoarthritis Study)	Participants aged 70 years or over reporting pain, aching or stiffness on most days in any joint	N=976. Response rate unclear	Self-reported data not given	Number of self-reported painful joints with radiographic OA (defined as a joint with KL grade >=2)
Marshall et al. (2009) UK	Cohort study	Participants aged 50 years and over, reporting hand pain on a health survey and attending a hand assessment clinic	N = 623. Response rate to attend clinical assessment = 46%	AUSCAN function subscale (Bellamy et al. 2002a) – higher score, more functional difficulty	AUSCAN pain subscale (Bellamy et al. 2002a) – higher score, more pain

α = Studies by (Dahaghin et al. 2005b) and (Dahaghin et al. 2005a) are derived from the same cohort study, although (Dahaghin al. 2005a) is based on a sub-sample of participants with radiographic data

β = Studies by (Hill et al. 2007) and (Dziedzic et al. 2007b) are derived from the same cohort study

Abbreviations: HAQ = Health Assessment Questionnaire, AIMS2 = Arthritis Impact Measurement Scales 2, KL = Kellgren-Lawrence, AUSCAN = Australian/Canadian Hand Osteoarthritis index

2.3.2 Quality assessment

Complete agreement between reviewers on the quality assessment scores was not obtained initially, but easily resolved. There was no one domain on the quality checklist where disagreements between reviewers were more common. In many instances the reviewer had scored the item as “Low to Moderate”, or “Moderate to High” risk of bias, so the consensus process was to aid a clear decision on category allocation, or to resolve issues where aspects of the study had been overlooked or misinterpreted.

The results of the quality scoring using the QUIPS tool are shown in Table 2-4. As all studies had a cross-sectional design, attrition bias was not scored for any of the studies in the review. Table 2-4 therefore shows the results of the remaining five bias domains only.

Table 2-4: Quality assessment of review articles

	Study participation	Factor measurement[α]	Outcome measurement	Measurement and controlling for confounding variables	Statistical analysis
Dahaghin et al. 2005b	Moderate	Moderate	Moderate	Moderate	Low
Dahaghin et al. 2005a	Moderate	Low	Moderate	Moderate	Low
Hill et al. 2007	Moderate	Low	Low	Moderate	Low
Baron et al. 1987	Moderate	Low	Moderate	Moderate	Moderate
Dziedzic et al. 2007b	Moderate	Low	Low	Not applicable	Low
Niu et al. 2003	High	Low	Low	Moderate	Low
Marshall et al. 2009	Moderate	Low	Low	Moderate	Low

Text in table indicates level of risk from bias
Study attrition not evaluated, as studies were not longitudinal in design.
[α] Includes any factor tested for association with level of hand pain or functional difficulty
Note that Dziedzic et al. (2007b) scored “Not applicable” for “Measurement and controlling for confounding variables”. This study aimed to describe a population of interest rather than look at association, so the issue of confounding was not relevant.

Of the five domains on the quality assessment checklist that were assessed, two were frequently rated as “Moderate” risk of bias: “Study participation” and “Measuring and controlling for confounders”. For “Study participation”, this was mainly due to lack of information describing the target population or insufficient evidence that the sample was representative of the target population; for “Measuring and controlling for confounding”, this was mainly due to lack of explanation of why specific confounding variables were chosen, how they fitted into the conceptual model, and whether confounders were assessed using validated measures. Reasons for deviating from “Low risk” of bias in the remaining domains included use of non-validated assessment methods and presentation of estimates without quantifying their statistical precision (i.e. lack of confidence intervals). Overall, the quality of the papers included in the review was satisfactory.

2.3.3 Main results

The results of the data extraction process are shown in Table 2-5 (detailed format) and Table 2-6 (summary format). The factors tested for association with hand pain and functional difficulty can be broadly categorised under six headings: demographic factors, history of previous health conditions, radiographic/clinical evidence of hand osteoarthritis, illness perceptions, self-reported diagnosis, and performance-based measures of hand function (Table 2-6). Most factors were assessed in a single study with the exception of age, gender and presence of OA which were assessed in 2, 4 and 5 independent studies, respectively. A smaller number of studies assessed associations with self-reported hand pain than with hand function.

Factors significantly associated with limited hand function were older age, female gender, manual occupation, neck or shoulder pain, clinical and radiographic osteoarthritis (although evidence depended on definition of OA), weaker hand strength, hand pain, history of Parkinson’s disease, stroke, diabetes or rheumatoid arthritis, and illness perceptions (namely frustration, impact, and symptom count). Key factors associated with

hand pain severity were age, impact, frustration, patient expectation of a long disease time course and self-reported diagnosis of the cause of the hand problem.

Table 2-5: Factor associated with increasing hand pain and poorer hand function

Factor measurement	Association with hand function		Association with hand pain	
Dahaghin et al. (2005b)[β]	Unadjusted OR (95% CI)	Adjusted[#] OR (95% CI)		
Age 70+ years (c.f. 55-69 years)	6.4 (5.4, 7.6)	4.5 (3.6, 5.6)	α	α
Female Gender	2.8 (2.4, 3.3)	2.2 (1.7, 2.8)	α	α
Manual occupation	2.0 (1.8, 2.3)	1.5 (1.2, 1.8)	α	α
Body Mass index >=30kg/m^2	1.3 (1.0, 1.5)	0.8 (0.6, 1.1)	α	α
Self reported history of:				
Rheumatoid arthritis (RA)	6.3 (4.9, 8.1)	3.3 (2.3, 4.7)	α	α
Osteoarthritis (OA) in any joint	1.6 (1.4, 1.9)	1.1 (0.9, 1.4)	α	α
Diabetes	2.4 (2.0, 3.0)	1.6 (1.1, 2.2)	α	α
Stroke	5.2 (4.1, 6.5)	4.8 (3.4, 6.8)	α	α
Thyroid disease	2.0 (1.7, 2.3)	1.2 (0.9, 1.6)	α	α
Neck & shoulder pain (last month)	2.2 (1.9, 2.5)	1.8 (1.4, 2.2)	α	α
Gout	0.9 (0.4, 2.0)	-	α	α
Hand/wrist fracture last 5 years	1.8 (1.5, 2.1)	0.9 (0.6, 1.3)	α	α

Parkinson's disease	18.4 (10.9, 30.8)	23.8 (11.4, 49.5)	α	α
Hand pain (last month)	2.6 (2.3, 3.1)	2.4 (1.9, 3.0)	α	α
Radiographic OA	2.1 (1.5, 2.9)[δ]	1.4 (0.9, 2.0)[δ]	α	α
Dahaghin et al. (2005a)[β]	Adjusted[μ] OR (95% CI)	Adjusted[λ] OR (95% CI)		
Radiographic hand OA				
KL>=2 in any DIP or IP joint	1.3 (0.9, 1.8)	1.2 (0.8, 1.7)	α	α
KL>=2 in any PIP joint	1.1 (0.8, 1.7)	0.9 (0.6, 1.4)	α	α
KL>=2 in any MCP joint	2.0 (1.3, 3.0)	1.8 (1.2, 2.9)	α	α
KL>=2 at the CMC1 or TS joint	1.3 (1.0, 1.9)	1.2 (0.8, 1.7)	α	α
KL>=2 in two hand joint groups	1.5 (1.1, 2.1)	-	α	α
KL>=3 in two hand joint groups	1.6 (1.1, 2.5)	-	α	α
KL>=4 in two hand joint groups	1.6 (0.9, 2.9)	-	α	α
Number of joints with KL>=2	Borderline significant (data not given)	-	α	α
Number of joints with KL>=2 (Dominant hand only)	1.1 (1.0, 1.2)	-	α	α
KL>=2 in all four hand joint groups	2.7 (1.3, 6.0)	-	α	α

Hill et al. (2007)[γ]	Unadjusted OR (95% CI)	Adjusted[#] OR (95% CI)	Unadjusted OR (95% CI)	Adjusted[#] OR (95% CI)
Age				
60 –69 years (c.f. 50-59)	1.27 (1.03, 1.57)	1.37 (0.98, 1.91)	1.21 (0.98, 1.50)	1.01 (0.80, 1.45)
70 + years (c.f. 50-59)	1.84 (1.49, 2.28)	2.04 (1.44, 2.90)	1.49 (1.20, 1.84)	1.63 (1.20, 2.21)
Female Gender	1.88 (1.57, 2.26)	2.02 (1.50, 2.73)	1.18 (0.99, 1.41)	0.89 (0.68, 1.15)
Self-reported diagnosis				
RA (c.f. OA)	1.06 (0.82, 1.38)	1.24 (0.84, 1.85)	0.92 (0.71, 1.20)	0.90 (0.64, 1.27)
Other (c.f. OA)	0.76 (0.59, 0.97)	1.28 (0.88, 1.89)	0.57 (0.43, 0.73)	0.59 (0.42, 0.83)
Don't know (c.f. OA)	0.47 (0.36, 0.61)	0.92 (0.61, 1.39)	0.38 (0.29, 0.49)	0.53 (0.37, 0.76)
Frustration with hand problem	8.45 (6.85, 10.44)	4.31 (3.17, 5.86)	9.10 (7.36, 11.26)	4.84 (3.70, 6.34)
Illness perception subscales (Moss-Morris, Weinman et al. 2002a)				
Timeline cyclical	P >0.05	-	P >0.05	-
Timeline acute chronic	P >0.05	-	2.51 (2.07, 3.04)	1.41 (1.06, 1.87)
Consequences	1.26 (1.23, 1.29)	1.18 (1.14, 1.23)	1.29 (1.25, 1.32)	1.18 (1.13, 1.22)
Personal control	P >0.05	-	P >0.05	-
Treatment control	P >0.05	-	P >0.05	-
Emotional representations	P >0.05	-	P >0.05	-
Illness coherence	P >0.05	-	P >0.05	-
Psychological attribution	P >0.05	-	P >0.05	-
Identity	5.34 (4.29, 6.64)	2.32 (1.73, 3.12)	-	-
Baron et al. (1987)				
Gender	Females more functional difficulty than males (t=2.35, p=0.026)			α
Hand function index (Smith hand function test (Smith 1973))	Uncorrelated, but no estimates given			α
Hand strength index	R = -0.56 (p=0.001)			α

Hand pain	R = 0.67 (p<0.001)	α
Adduction deformity of the CMC1 joint	R = 0.28 (p=0.057)	α
Tenderness on motion	R = 0.33 (p=0.034)	α
Clinical OA index	R adjusted for gender = 0.37 (p=0.03)	α

Dziedzic et al. (2007b)[γ]	Males		Females		Males		Females	
Age	Mean (SD)	Severe functional difficulties N (%)	Mean (SD)	Severe functional difficulties N (%)	Mean (SD)	Severe functional difficulties N (%)	Mean (SD)	Severe functional difficulties N (%)
50-59 years	1.5 (2.1)	37 (15)	2.3 (2.4)	100 (23)	3.7 (2.4)	47 (19)	3.8 (2.5)	102 (23)
60-69 years	2.0 (2.6)	50 (19)	2.5 (2.3)	109 (25)	4.2 (2.5)	75 (28)	4.1 (2.4)	107 (26)
70-79 years	2.0 (2.6)	40 (20)	3.0 (2.5)	109 (35)	3.9 (2.4)	45 (23)	4.3 (2.4)	86 (28)
80+ years	2.5 (3.0)	13 (27)	4.0 (2.8)	65 (48)	4.2 (2.1)	13 (25)	4.6 (2.7)	49 (38)

Niu et al. (2003)			Males	Females
Number (%) of painful hand joints with radiographic OA				
0	α	α	309 (88)	464 (74)
1	α	α	13 (4)	35 (6)
2	α	α	7 (2)	29 (5)
3	α	α	5 (1)	11 (2)
4	α	α	5 (1)	16 (3)

5+		α		α		12 (3)			70 (11)	
Marshall et al. (2009)										
Radiographic subgroup	Un-adj mean (95% CI)	Adj$^{\infty}$ Mean (95% CI)	Adj$^{\rho}$ Mean (95% CI)	Adj$^{\pm}$ Mean (95% CI)	Adj$^{\Delta}$ Mean (95% CI)	Un-adj mean (95% CI)	Adj$^{\infty}$ Mean (95% CI)	Adj$^{\rho}$ Mean (95% CI)	Adj$^{\pm}$ Mean (95% CI)	Adj$^{\Delta}$ Mean (95% CI)
No OA	8.3 (6.7-9.8)	8.1 (6.6-9.7)	8.7 (7.1-10.3)	8.9 (7.1-10.7)	8.9 (7.1-10.6)	5.4 (4.6-6.2)	5.4 (4.6-6.2)	5.4 (4.5-6.2)	5.7 (4.7-6.6)	5.8 (4.9-6.7)
Finger only OA	8.2 (6.3-10.1)	8.2 (6.4-10.1)	8.3 (6.4-10.2)	8.6 (6.6-10.6)	8.6 (6.6-10.5)	5.7 (4.7-6.8)	5.7 (4.7-6.7)	5.7 (4.7-6.8)	5.9 (4.8-6.9)	6.0 (4.9-7.0)
Thumb only OA	8.6 (6.9-10.3)	8.1 (6.5-9.8)	8.7 (7.0-10.4)	9.0 (7.2-10.8)	8.8 (7.1-10.5)	5.8 (4.9-6.7)	5.7 (4.8-6.6)	5.8 (4.9-6.7)	5.9 (5.0-6.9)	5.9 (5.0-6.8)
Combined thumb and finger OA	10.5 (9.6-11.4)	9.9 (9.0-10.8)	10.3 (9.4-11.2)	10.1 (9.0-11.2)	10.4 (9.4-11.3)	6.5 (6.1-7.0)	6.5 (6.0-7.0)	6.5 (6.1-7.0)	6.4 (5.8-6.9)	6.4 (5.9-6.9)
p-value for overall association	0.018	0.084	0.093	0.601	0.206	0.077	0.095	0.091	0.698	0.573

- term not entered in the model, α = data not available for review, β = studies by (Dahaghin et al. 2005b) and (Dahaghin et al. 2005a) are derived from the same cohort study, although (Dahaghin et al. 2005a) is based on a sub-sample of participants with radiographic data, γ = studies by (Hill et al. 2007) and (Dziedzic et al. 2007b) are derived from the same cohort study, δ = data derived from a smaller subset of cases with radiographic data (N=3906), # = adjusted for all other factors measured, μ = adjusted for age and gender, λ = adjusted for age, gender and all other factors measured, ∞ = adjusted for gender, ρ = adjusted for age, ± = adjusted for number of hand joints with radiographic OA, Δ = adjusted for the presence of moderate to severe radiographic OA. Abbreviations: DIP = Distal interphalangeal joints, IP = First interphalangeal joint, PIP = Proximal interphalangeal joints, MCP = Metacarpophalangeal joints, CMC1 = First carpometacarpal joint, TS = Trapezio-scaphoid joint, SD = Standard deviation, OR = odds ratio, CI = confidence interval, c.f = compared with, p = p-value, R= Pearson's correlation, t= t-statistic

Table 2-6: Summary table of review findings: evidence for factors associated with increasing hand pain and poorer hand function

	Reference for observed association	
Factor	Hand function	Hand pain
Demographic		
Older age	Dahaghin et al. (2005b), Hill et al. (2007)	Hill et al. (2007)
Female gender	Dahaghin et al. (2005b), Hill et al. (2007), Baron et al. (1987)	α
Manual occupation	Dahaghin et al. (2005b)	α
Higher body mass index (BMI)	n.s.	α
Self-reported history of previous health conditions		
Rheumatoid arthritis (RA)	Dahaghin et al. (2005b)	α
Diabetes	Dahaghin et al. (2005b)	α
Stroke	Dahaghin et al. (2005b)	α
Thyroid disease	n.s.	α
Neck and shoulder pain (last month)	Dahaghin et al. (2005b)	α
Gout	n.s.	α
Hand/wrist fracture last 5 years	n.s.	α
Parkinson's disease	Dahaghin et al. (2005b)	α
Hand pain/tenderness	Dahaghin et al. (2005b), Baron et al. (1987)	α
Osteoarthritis		
Radiographic	Dahaghin et al. (2005a) Evidence of association depends on how OA presence	n.s.

	is defined	
Clinical	Baron et al. (1987)	α
Higher scores for Illness perceptions (Moss-Morris, Weinman et al. 2002a)		
Timeline cyclical	n.s.	n.s.
Timeline acute/chronic	n.s.	Hill et al. (2007)
Consequences	Hill et al. (2007)	Hill et al. (2007)
Personal control	n.s.	n.s.
Treatment control	n.s.	n.s.
Emotional representations	n.s.	n.s.
Illness coherence	n.s.	n.s.
Psychological attribution	n.s.	n.s.
Identity	Hill et al. (2007)	α
Frustration	Hill et al. (2007)	Hill et al. (2007)
Self-reported diagnosis (OA, RA, other, don't know)	n.s.	Hill et al. (2007)
Performance-based measures of hand function		
Poorer score Hand function index (Smith hand function test) (Smith 1973)	n.s.	α
Hand strength index	Baron et al. (1987)	α

Table includes studies with factors that show a statistically significant association ($p<0.05$) with more hand pain or poorer function. Studies with descriptive data only are excluded (Dziedzic et al. 2007b, Niu et al. 2003) α= data not available for review, n.s. = tested in at least one study but not statistically significant ($p>=0.05$)

2.3.4 Meta-analysis

A meta-analysis was not conducted to pool estimates across studies because the factors measured in each study differed greatly and the statistical methods used to describe associations were not consistent (e.g. odds ratio vs mean difference vs correlation).

2.4 Discussion

2.4.1 Summary of review findings

In this review, five cross-sectional studies (seven published articles) have been identified that have investigated factors associated with the severity of hand pain and functional difficulty in general population samples of older adults. No longitudinal studies investigating the prognosis of hand problems that meet the criteria for the review have been identified. The factors associated with hand pain and/or function included generic factors (e.g. age, gender), those related to concurrently occurring clinical conditions (e.g. stroke), and hand specific factors (e.g. illness perceptions and radiographic/clinical hand osteoarthritis).

2.4.2 Factors associated with hand pain and functional difficulty

Older age was associated with both worse hand pain and functional ability and was the only factor where a (cross-sectional) “dose-response” (Woodward 1999) relationship was tested, i.e. that the severity of hand pain and functional difficulty progressively worsened with increasing age. These findings are in line with evidence from population and clinical studies that show that the prevalence of hand pain, pain interference and functional difficulty increases with age (Palmer 2003, Thomas et al. 2004a, Jones et al. 2001).

Although female gender was associated with hand pain, such an association was not found for hand function (Hill et al. 2007), despite the prevalence of upper limb musculoskeletal pain being higher for women (Walker-Bone et al. 2003) and female gender being a risk factor for many common hand conditions (Hart et al. 2000, Walker-Bone et al. 2003). This may suggest that a more complex relationship between hand

function and gender may exist which could be explained by other external factors, such as ability to cope and adapt to limited hand function (Myers et al. 2008)

Some of the strongest predictors of hand function found in this review relate to disease history, e.g. history of Parkinson's disease or stroke. Prevalence of such conditions in the population is likely to be low, which may lead to unreliable estimates of association, although it has been shown in several clinical studies that such conditions are related to impaired hand function (e.g. Hunter et al. 2002, Cano-de-la-Cuerda et al. 2010).

Only three out of the seven papers in the review have examined both hand pain and function thus allowing direct comparisons to be made (Hill et al. 2007, Dziedzic et al. 2007b, Marshall et al. 2009). In these studies findings were similar for hand pain and function, however, hand pain, but not hand function, was related to self-reported diagnosis of the cause of the hand problem and to patient expectation of a long disease time course (Hill et al. 2007). This may reflect patients focus on absence of pain as the main sign of recovery from their hand condition or that pain may encourage consultation to receive a clinical diagnosis. It may also be that receiving a medical diagnosis may be associated with more pain perceptions or that some diagnoses reflect more painful conditions.

Of the cross-sectional factors identified in the review, many cannot be modified by treatment (e.g. age, gender and occupation), or relate to disease history that cannot be altered at the point of consultation. Illness perceptions, however, have the potential to be modified and have been identified as important predictors of outcome in studies of primary care consulters with hand pain (Spies-Dorgelo et al. 2008) and back pain (Foster et al. 2008, Macfarlane 2008).

The majority of factors identified by this review are generic factors (e.g. age and gender) or related to clinical conditions not solely affecting the hand (e.g. stroke). Few hand-specific factors were tested; the exceptions were illness perceptions measured in the context of a patients' current hand problem and radiographic/clinical hand osteoarthritis.

The inclusion criteria specified that only studies from the general population would be included in the review, possibly biasing the type of factors tested to those that are generic; hand-specific factors may be more likely to be tested in hand–specific clinical subgroups, e.g. hospital patients with RA.

2.4.3 Strengths and limitations

A major strength of this review was the comprehensive and inclusive search strategy that was developed to minimise the risk of missing key articles. This was achieved by searching in multiple health care databases and tailoring searches to apply directly to the particular databases' indexing method. At each stage in the review, methods were piloted and key decisions on abstract inclusion, quality assessment and data extraction were derived by consensus, improving the quality of the data reported.

Selected databases included conference abstracts and other non-journal articles: HMIC and Web of Knowledge (Centre for Reviews and Dissemination 2009) however, only full journal articles were included in the review. Searching of grey literature or unpublished studies was not undertaken. It might be speculated that as the number of published studies in the review is low, the number of extra studies identified by this method would also be small. Articles written in English were selected electronically and included in the review. Only a small percentage of all articles found in the search were written in other languages (<13%) so it is unlikely that this number would bias the results of the review (Centre for Reviews and Dissemination 2009).

The search was focussed to include only self-report measures of pain and function so studies measuring hand function using clinical tests alone, such as grip strength and timed performance tests were not included. Self-report measures, although potentially prone to recall bias, were chosen as they are frequently used in population-based surveys to fully capture the range of limitations experienced during everyday activities (Jordan et

al. 2009) and it is a self-reported measure of hand pain and functional difficulty that is used throughout this thesis.

2.4.4 Implications

All associations described in this review are observed using cross-sectional data so reflect factors associated with severity of hand pain and functional difficulty at a single point in time. It remains to be tested whether such factors would also predict change in hand pain and functional difficulty over time. In addition, several studies in the review presented unadjusted associations of predictors with outcome, which may not remain clinically important or statistically significant after adjustment for other theoretically plausible variables.

A key aim of this review was to generate a list of variables that could be tested for their potential prognostic value in the CAS-HA study. This has been achieved and provides a starting point to define which factors to test in a prognostic model. However, due to the small number of studies included in the review and the lack of prognostic information, additional literature may be useful to ensure that a broad range of prognostic factors are considered within the thesis e.g. using literature from other sites of pain (e.g. knee pain) or other populations (e.g. participants consulting their general practitioner). This has been considered using the approach below (Section 2.4.5) and is the approach that was used to define a list of potential predictive factors to be tested in this thesis.

2.4.5 Generating a list of potential baseline prognostic factors to test in this thesis

The following sources of data were reviewed and used to generate an initial list of potential prognostic factors for analysis:

(1) The systematic review included in this chapter - although the factors in the review were all tested in cross-sectional data they were still considered as they could have potential prognostic value yet to be determined,

(2) A study of knee pain designed with identical recruitment methods to CAS-HA to recruit participants with knee rather than hand pain – the Clinical Assessment Study of the Knee (CAS-K) (Peat et al. 2004, Mallen et al. 2007b).

(3) A sample of systematic reviews of prognostic factors in body areas other than the hand and in populations other than community-dwelling older adults with hand pain (Hayden et al. 2009, van Dijk et al. 2006, Kuijpers et al. 2004, Mallen et al. 2007b, Tas et al. 2007, Kwok et al. 2013), with systematic reviews chosen as the highest level of evidence in the evidence hierarchy (Greenhalgh 1997).

(4) Individual studies found by EN during the systematic review process that did not fully meet the criteria to be in the review (Spies-Dorgelo et al. 2008, Botha-Scheepers et al. 2009, Bijsterbosch et al. 2011).

The initial list of potential factors was reviewed to identify those factors that were particularly relevant to the CAS-HA population (i.e. older adults with hand pain)[8,9] and that mapped onto concepts that had been measured in CAS-HA (i.e. restricting the list to only those factors that could be explored in this thesis). Each factor was then tested for (a) their prevalence in the CAS-HA data and (b) the percentage of missing data that they had, as those of low or high prevalence (<10% or >90%) or with a high percentage of missing data (>20%) were unlikely to produce reliable estimates in the data. This latter stage was also included to try and reduce the number of factors on the list (to limit the possibility of over-fitting the data) and also because prognostic factors with high rates of missing data

[8] Some factors were clearly pain-site specific (e.g. malalignment for knee pain) or were relevant only to participants in a working population (e.g. detailed questions on current work environment such posture at work and ability to modify work environment) so were excluded from the initial list.
[9] At this stage, the factors were not restricted to those that showed a significant relationship to the severity/progression of the outcome as factors were often derived from populations and pain-sites that differed from CAS-HA so presence or absence of an effect in a different population could not be assumed and evidence was often derived from single studies requiring replication. In addition, factors were not restricted to those that had a strong theoretical justification for the mechanism through which they were prognostic as the aim of the analysis was to explore which factors, or combinations of factors, best predicted participants' outcome trajectories over time, rather than to explain why the factors were predictive per se (Hayden et al. 2010).

were likely to have high rates of missing data if measured again in another study or in clinical practice, which is not ideal (Royston et al. 2009) (see Appendix 3 for excluded items).

The initial list of factors was then presented to a group of clinical experts, including an occupational therapist, physiotherapist, epidemiologist, and radiographer, to be reviewed for clinical relevance for use in a primary care setting[10]. In addition to refining the list of prognostic factors (see Appendix 3 for details), the clinical experts also advised EN that the most clinically relevant model would be obtained if factors were grouped by their method of measurement, i.e. by questionnaire, clinical assessment, or x-ray, so EN could observe how much variability in the data could be explained by relatively cheap and simple methods of assessment (i.e. the questionnaire), compared to those that are more complex and that required specialist trained staff and equipment (i.e. clinical assessment and x-ray). The expert group also aimed to identify (by consensus) any important prognostic factors that had been omitted from the list, but this did not generate any more variables than those on the initial list. The list of factors therefore taken forward to the modelling stage is shown in Table 2-7, grouped by measurement method and, for questionnaire measures only, under the sub-headings of demographic, lifestyle, health, hand condition characteristics and psychological factors. It is these factors that are explored in more detail in Chapters 3, 7 and 8 that follow.

[10] The group were only asked to consider whether the construct represented by each prognostic factor was relevant to primary care, rather than whether the measure (as used in CAS-HA) was feasible to apply in this setting. This was because the purpose of the model (at this stage) was to be a research tool, rather than a clinical tool hence the most reliable form of the prognostic factor was required for modelling rather than considering a simplified version that could be more readily applied in clinical practice (e.g. full questionnaire tools were used (where available) rather than single items for the questionnaire-based measures).

Table 2-7: Potential baseline prognostic factors used to model the trajectory of hand pain and function over time

Concept
Questionnaire
Demographic
Age
Gender
Marital status
Occupation/Social Class
Employment status
Education level
Income
Lifestyle
Alcohol consumption
Smoking status
Social networks
Health
Self-rated health
Co-morbidities
Pain in other body areas
Characteristics of hand condition
Hand pain severity
Hand functional difficulty
Side affected
Time since onset of hand problem
Sudden onset of hand problem
Onset of hand condition following accident or injury to the hand
Physical load on hands during work and leisure
Psychological factors
Anxiety
Depression
Illness perceptions
Long disease time course
Consequences
Personal control
Treatment control

	Illness coherence
	Cyclical time course
	Emotional representation
	Frustration with hand condition
Clinical Assessment	
	Body-mass index
	Hand grip-ability
	Muscle strength
	Severity of hand osteoarthritis
	Carpal tunnel syndrome
	Dupuytren’s contracture
	De Quervain’s tenosynovitis
	Trigger finger
X-ray	
	Severity of hand osteoarthritis

3 Recruitment design and data collection procedures

3.1 Introduction

The aim of this chapter is to describe the recruitment, follow-up and data collection procedures used in the Clinical Assessment Study of the Hand (CAS-HA) as this study provides the majority of the data analysed in this thesis. The description of the CAS-HA study includes information on study design (Section 3.2), data entry and accuracy (Section 3.3), recruitment and follow-up rates (Section 3.4), and an assessment of whether the baseline characteristics of participants at each recruitment and follow-up stage are representative of the broader population from which they are drawn (Section 3.5). This latter assessment is included to explore whether there is any selection bias in the sample.

3.2 Study design

The CAS-HA study is a population-based, prospective observational cohort study of all eligible adults aged 50 years and over registered at two general practices in North Staffordshire. The study forms part of the North Staffordshire Osteoarthritis Project (NorStOP), which is a set of linked cohort studies designed to evaluate the course of musculoskeletal conditions over time (Thomas et al. 2004b). All stages of the study have been given full ethical approval prior to data collection[11]. A protocol for the CAS-HA study has previously been reported (Myers et al. 2007), however, as this is the main data set used in this thesis, for completeness, study details are given below.

3.2.1 Baseline study recruitment

Participants were recruited to the CAS-HA study by postal survey. A Health Survey was mailed to all adults aged 50 years and over registered at the two study general practices

[11] Ethical approval was gained from the North Staffordshire Ethics Committee (baseline and 18-month follow-up), Hereford and Worcester Ethics Committee (3-year and 54-month follow-up) and the West Midlands - Solihull ethics committee (6-year follow-up) Project Reference Numbers: 1430, 05/Q2604/89, 06/Q2801/90 (3-year and 54-month) and 11/WM/0196, respectively))

after exclusion of participants for which survey mailing was inappropriate, e.g. those with severe psychiatric or terminal illness. Participants reporting hand pain or problems (e.g. stiffness or knobbly swellings) in the last 12-months on the General Health Survey and providing written consent to further contact were then sent a Regional Pains Survey to collect further information on their reported hand pain or problem.

Participants reporting hand pain or problems and completing both of the mailed surveys were then sent a letter inviting them to attend a CAS-HA research clinic at a local rheumatology hospital (Haywood Hospital). At the research clinic participants underwent a clinical interview and hand assessment, which included questions on the history of the participants' hand problem, an assessment of key clinical features for a range of common hand conditions, e.g. osteoarthritis, and objective measures of hand function. In addition digital images and x-rays were taken of both hands. The questions/assessments collected on the two surveys and at the research clinic provide measurement of each factor listed in Chapter 2 and were therefore used for further analysis in Chapters 7 and 8 (Table 3-1 and Appendix 4 for further details). Participants consenting to, and attending the research assessment form the baseline sample for the CAS-HA study, and provide the sample used for most analyses presented in this thesis.

Table 3-1: Measurement of potential baseline prognostic factors in the CAS-HA study

Concept	Detail of measurement in CAS-HA
Questionnaire	
Demographic	
Age	Age at baseline: years
Gender	Female, Male
Marital status	Married, Separated, Divorced, Widowed, Cohabiting, Single
Occupation/Social Class	SOC 2000 codes (Office for National Statistics (ONS) 2002): Higher managerial, Higher professional, Lower managerial/professional, Intermediate occupations, Self-employed, Lower supervisory/technical, Semi-routine occupations, Routine occupations
Employment status	Employed, Not working due to ill health, Retired, Unemployed, Housewife, Other
Education level	Age when left school (years); Go from school to full time education or university: Yes, No
Income	Find it a strain to get by from week to week, Have to be careful with money, Able to manage without much difficulty, Quite comfortably off
Lifestyle	
Alcohol consumption	Daily or most days, Once or twice a week, Once or twice a month, Once or twice a year, Never
Smoking status	Never, Previously smoked, Currently smoke
Social networks	Live alone: Yes, No[α]
Health	
Self-rated health	General health: Excellent, Very good, Good, Fair, Poor

	Physical component of the SF-12 (0-100) (Ware et al. 1996)
Co-morbidities	Number of co-morbidities including diabetes, raised blood pressure, eyesight problems, deafness, heart and chest problems
Pain in other body areas	Manchester definition of regional pain (MacFarlane et al. 1996): No other pain, Regional pain, Widespread pain
Characteristics of hand condition	
Hand pain severity	Australian/Canadian Hand Osteoarthritis index (AUSCAN) pain (Bellamy et al. 2002a) (0-20)
	Number of days in the last 12-months with hand pain: less than 7-days, 1-4 weeks, >1-month but <3-months, 3-months or more
Hand functional difficulty	Australian/Canadian Hand Osteoarthritis index (AUSCAN) function (Bellamy et al. 2002a) (0-36)
Side affected	Dominant hand only, Non-dominant hand only, One hand affected but participant ambidextrous, Both hands affected
Time since onset of hand problem	Length of time with a hand problem (years)
Sudden onset of hand problem	Bilateral problem - both hands sudden onset, Bilateral problem – one hand sudden onset, Bilateral problem – neither hand of sudden onset, Unilateral problem – of sudden onset, Unilateral problem – not of sudden onset
Onset of hand condition following accident or injury to the hand	Bilateral problem - both hands onset following accident/injury, Bilateral problem – one hand onset following accident/injury, Bilateral problem – neither hand onset following accident/injury, Unilateral problem – onset following accident/injury, Unilateral problem – onset not following accident/injury
Physical load on hands during work and leisure	Past or present job, hobbies or pastimes involved excessive use of your hands? Yes, No
Psychological factors	

Anxiety	Hospital Anxiety and Depression scale (HADS) (0-21) (Zigmond et al. 1983); Mental component of the SF-12 (0-100) (Ware et al. 1996)
Depression	Hospital Anxiety and Depression scale (HADS) (0-21) (Zigmond et al. 1983) Mental component of the SF-12 (0-100) (Ware et al. 1996)
Illness perceptions	
Long disease time course	Subscale of the Illness perceptions questionnaire (IPQ-R) (6-30) (Moss-Morris et al. 2002)
Consequences	Subscale of the Illness perceptions questionnaire (IPQ-R) (6-30)(Moss-Morris et al. 2002)
Personal control	Subscale of the Illness perceptions questionnaire (IPQ-R) (6-30) (Moss-Morris et al. 2002)
Treatment control	Subscale of the Illness perceptions questionnaire (IPQ-R) (5-25) (Moss-Morris et al. 2002)
Illness coherence	Subscale of the Illness perceptions questionnaire (IPQ-R) (5-25) (Moss-Morris et al. 2002)
Cyclical time course	Subscale of the Illness perceptions questionnaire (IPQ-R) (4-20) (Moss-Morris et al. 2002)
Emotional representation	Subscale of the Illness perceptions questionnaire (IPQ-R) (6-30) (Moss-Morris et al. 2002)
Frustration with hand condition	All days, Most days, Some days, Few days, No days
Clinical Assessment	
Body-mass index	Weight (kgs)/(height (meters))^2
Hand grip-ability	Grip-ability test (seconds) (Dellhag et al. 1995)
Muscle strength	Average of left and right grip strength (lbs); Average of left and right pinch strength (lbs) (Mathiowetz et al. 1984)
Severity of hand osteoarthritis	Meets the American College of Rheumatology (ACR) criteria for hand osteoarthritis: Yes, No (Altman et al. 1990)
Carpal tunnel syndrome	Present in either hand: Yes, No (Palmer et al. 2000)
Dupuytren's contracture	Present in either hand: Yes, No

De Quervain's tenosynovitis	Present in either hand: Yes if positive Finkelstein test, No if negative Finkelstein test (Lister 1978, Simpson 2002)
Trigger finger	Present in either hand: Fingers lock, trigger or catch and need to be released by participant: Yes, No
X-ray	
Severity of hand osteoarthritis	Number of hand joints with Kellgren-Lawrence (Lawrence 1977) x-ray grade >=2 (0-32)

[α] A more detailed measure of social support was included in the CAS-HA study (The Berkman-Syme Social Network Index (SNI) (Berkman, Syme 1979)), but this measure had a large percentage of missing data (21%) so was not used for analysis

3.2.2 Study follow-up

Participants in the CAS-HA baseline sample were mailed follow-up questionnaires at regular 18-month intervals over the six-year period following their baseline assessment. Prior to each follow-up mailing, a “thank you” letter was sent to participants to thank them for their response to the previous survey and to inform them that their next follow-up questionnaire was due. Follow-up questionnaires were only mailed to those who had not withdrawn their consent to be in the study. To ensure that follow-up surveys were not mailed to participants who had recently died, or for whom receiving a survey would be inappropriate, the practice lists were screened by the general practitioners (GPs) before mailing and screened by a member of the health informatics staff to ensure that contact details for participants were up-to-date. If needed, address details were updated using the NHS tracing service.

In addition at the 6-year follow-up, all participants were invited to attend a research assessment clinic alongside completing a postal survey. The research assessment clinic was similar in format to that conducted at baseline.

3.2.3 Reminder mailings

Reminder mailings were included to increase response to the postal surveys. Reminder mailings at the baseline stage were sent 2- and 4-weeks after the initial mailing to any participant not responding to the survey. The 2-week reminder was a postcard asking participants to return the survey; the 4-week reminder included a second copy of the questionnaire.

At follow-up, an extra reminder stage was included in addition to those described above – minimum data collection (MDC). Participants not responding 6-weeks after the initial mailing date for that stage were telephoned (if consent was given) and were asked if they were willing to complete a short survey over the phone. If participants could not be contacted by phone a postal version of the short survey was mailed.

3.3 Data entry and data accuracy

The data from each recruitment and follow-up stage was entered onto an ACCESS data entry database (Microsoft Office) after checking that the questionnaire had been completed by the intended participant[12]. It was also checked that only one unique entry was on the data entry database for each participant who had returned a questionnaire and that all participants who had been logged as returning a questionnaire did indeed have a corresponding entry on the data entry database[13]. The accuracy of the data entry was also checked for 10% of the questionnaires entered (i.e. the "1 in 10" check) and any errors revealed by this process corrected on the database. The full dataset was only used for analysis after it was checked that the error rate was low and that any errors found were not occurring systematically (i.e. that the errors were occurring at random and not always for the same question on the questionnaire). The data were also checked to ensure that responses entered for the questionnaire were plausible given the range of values that each question could take.

The quality of the data was also enhanced by using well-validated and reliable research tools (where possible) (e.g. the SF-12, the HADS and the IPQ-R, see Appendix 4 for further details) to minimise information bias[14] and the primary outcome (the AUSCAN) was only selected after a systematic review had been conducted to identify and review the validity of any existing tools to measure this construct (Dziedzic et al. 2005).

In addition, prior to data collection, the research assessors were trained to use a standardised protocol to collect the measures at the clinical assessments, the reliability of

[12] This was achieved by checking that the name and date of birth given by the person completing the questionnaire matched the information that was on a (separate) mailing database used to log when questionnaires were sent and when they were returned

[13] A multi-stage process was used to resolve any discrepancies between the number of questionnaires logged as being returned and the number of questionnaires entered on the data entry database (e.g. checking whether the questionnaire had been entered under an incorrect study identification number, liaising with the administrative team to check whether the questionnaire had been misfiled, searching the phone log to see if the participant had phoned the research centre giving information to explain why a discrepancy arose)

[14] Information bias can occur due to errors in the assessment of patient outcomes or predictor variables (Grimes et al. 2002)

which was tested in a pilot study (Myers et al. 2011). Clinical equipment (i.e. the JAMAR dynamometer and B&L pinch gauge) were calibrated prior to data collection and the average of three readings taken on each hand to improve the reliability of the reading used in the analysis – an approach recommended in reliability and validity studies of grip and pinch strength measures previously reported (Mathiowetz et al. 1984, Dellhag et al 1995).

3.4 Recruitment rates and loss to follow-up

Figure 2 and Figure 3 show the number of participants included at each stage of recruitment and follow-up. From the participants identified as reporting hand pain or problems in the last year (N=2114), 623 attended the clinical assessment (attendance rate = 29%, Figure 2) and were defined as the baseline sample for the CAS-HA cohort. The most common reasons for not attending the baseline clinical assessment were lack of consent to receive postal mailings, lack of time to attend, or ill health; however, in many instances the reason for non-attendance is unknown.

The response rates to the follow-up surveys were 96, 91, 71, and 66%, respectively, and represent the percentage of follow-up data collected at each time point from those in the original baseline cohort (denominator = 623)[15]. When the follow-up rates were calculated excluding those who were not mailed, as they had either died or had withdrawn from the study (their own decision or a decision made by their GP) response rates ranged from 89 – 97% (Figure 3).

[15] The lower response rate to the 54-month follow-up compared to 3-years was partly explained by the fact that participants were asked to re-consent to the study by positively completing a question on consent on the 3-year questionnaire

Figure 2: Recruitment flow chart for the Clinical Assessment Study of the Hand (CAS-HA)

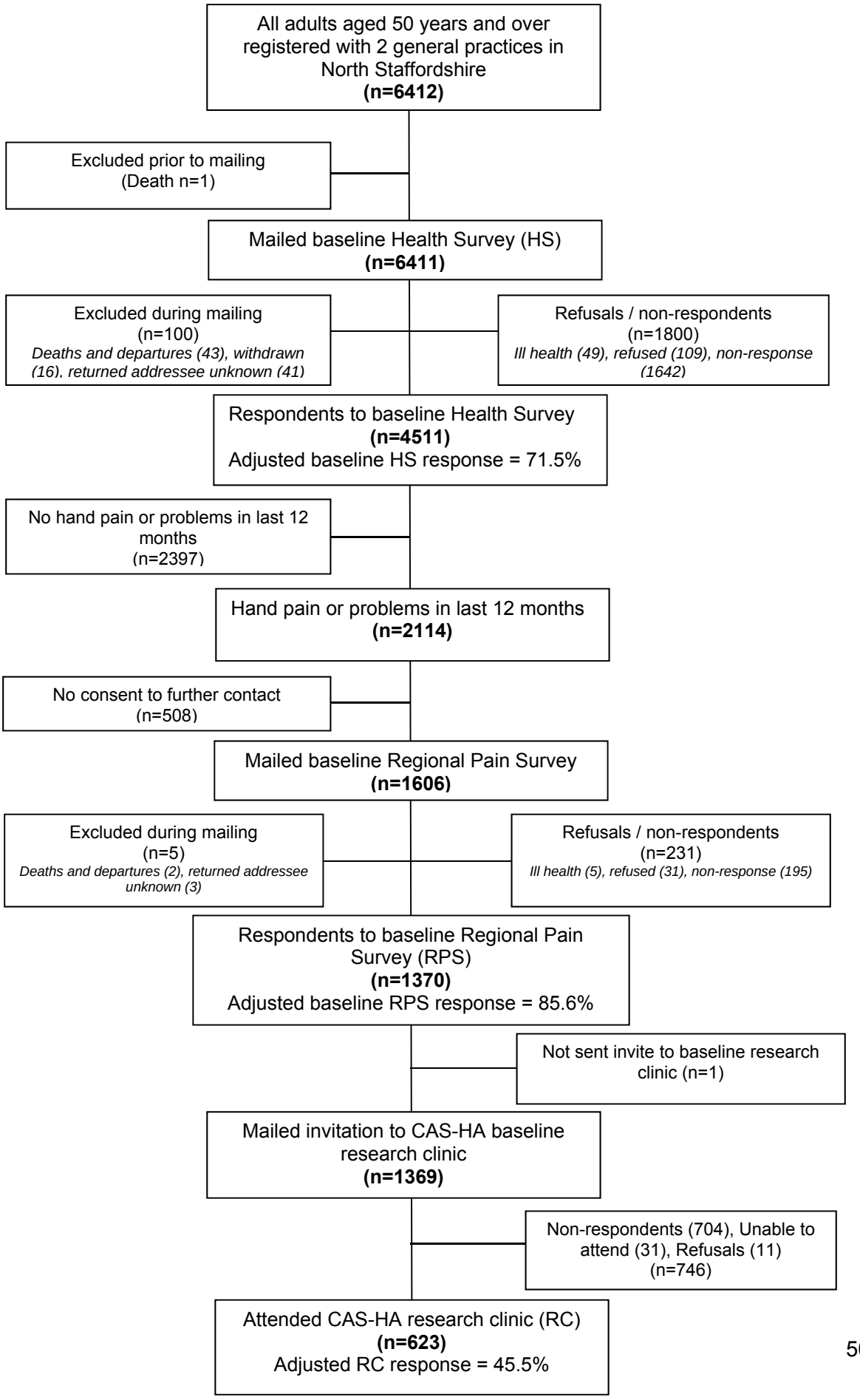
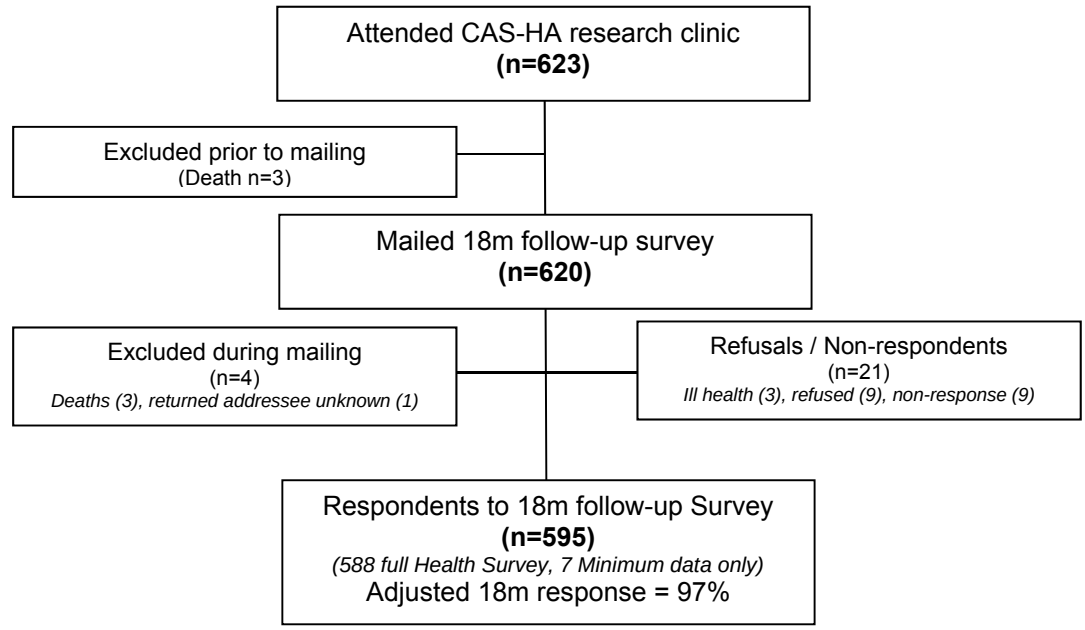


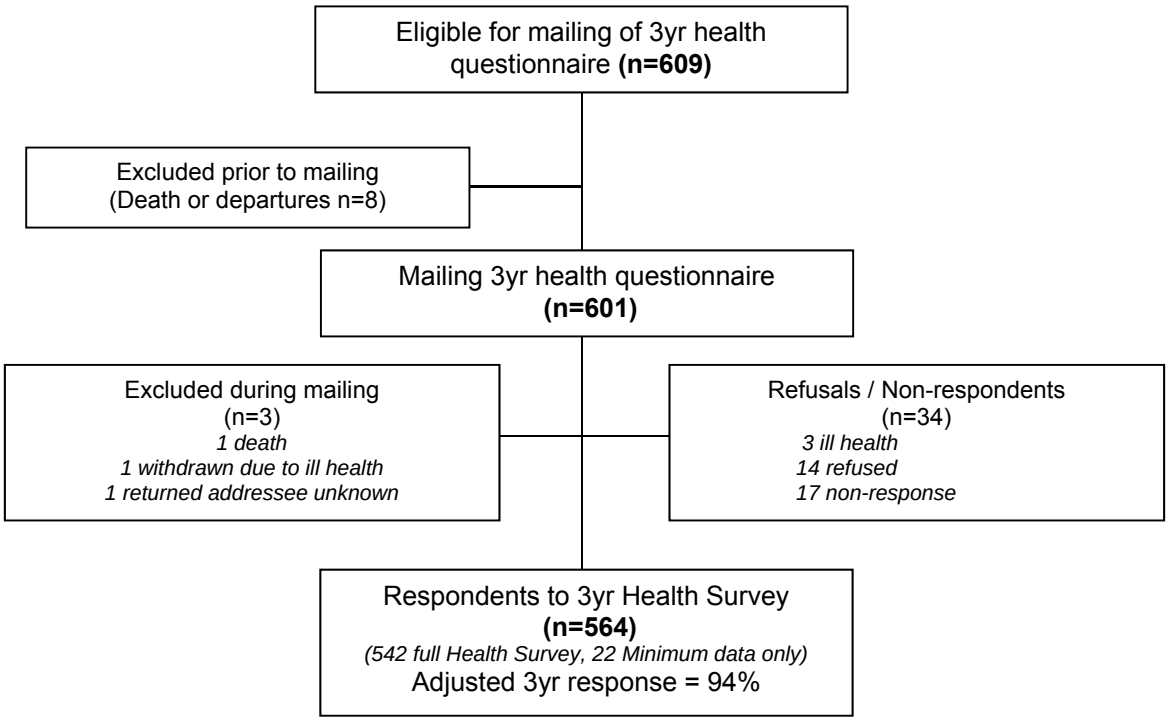
Figure 3: Retention flow charts for the Clinical Assessment Study of the Hand (CAS-HA)

18-month follow-up



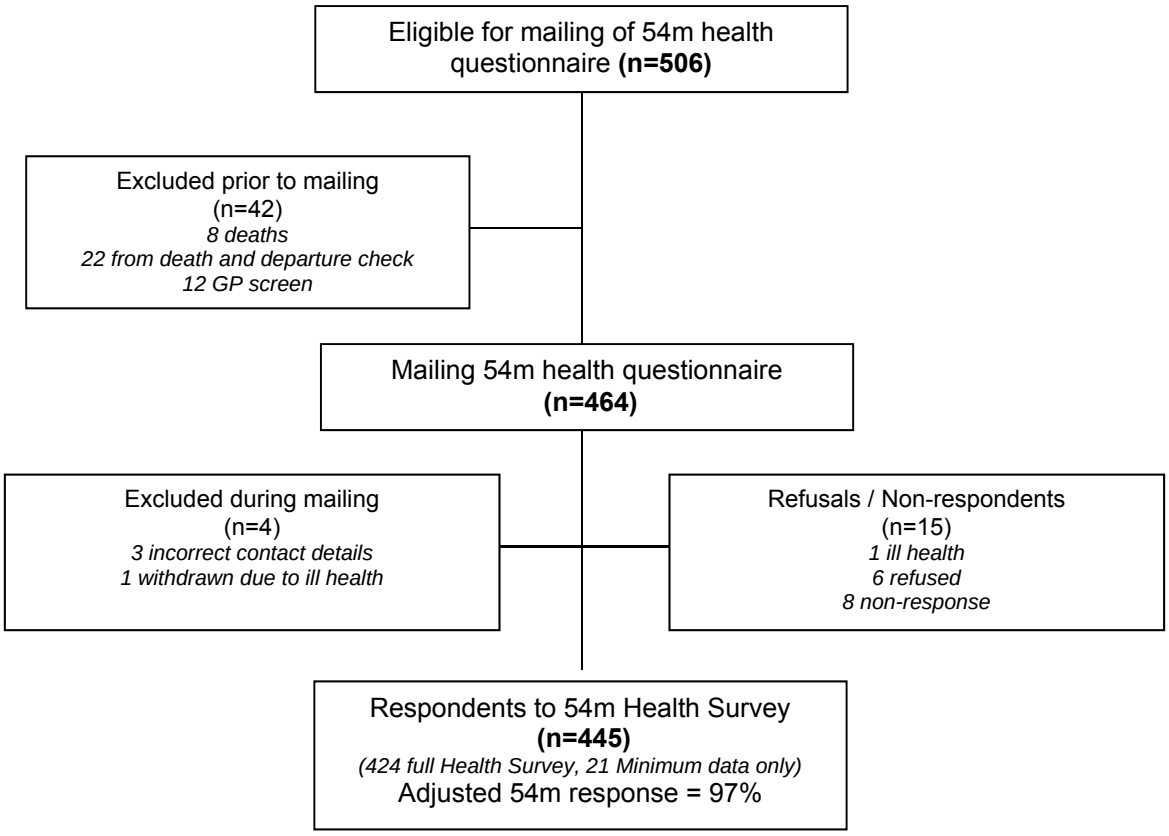
3-year follow-up

Note that eligibility for the 3-year mailing includes non-responders to the 18-month survey (n=9) and five participants who gave consent to the 3-year mailing despite not wanting to complete the 18-month survey



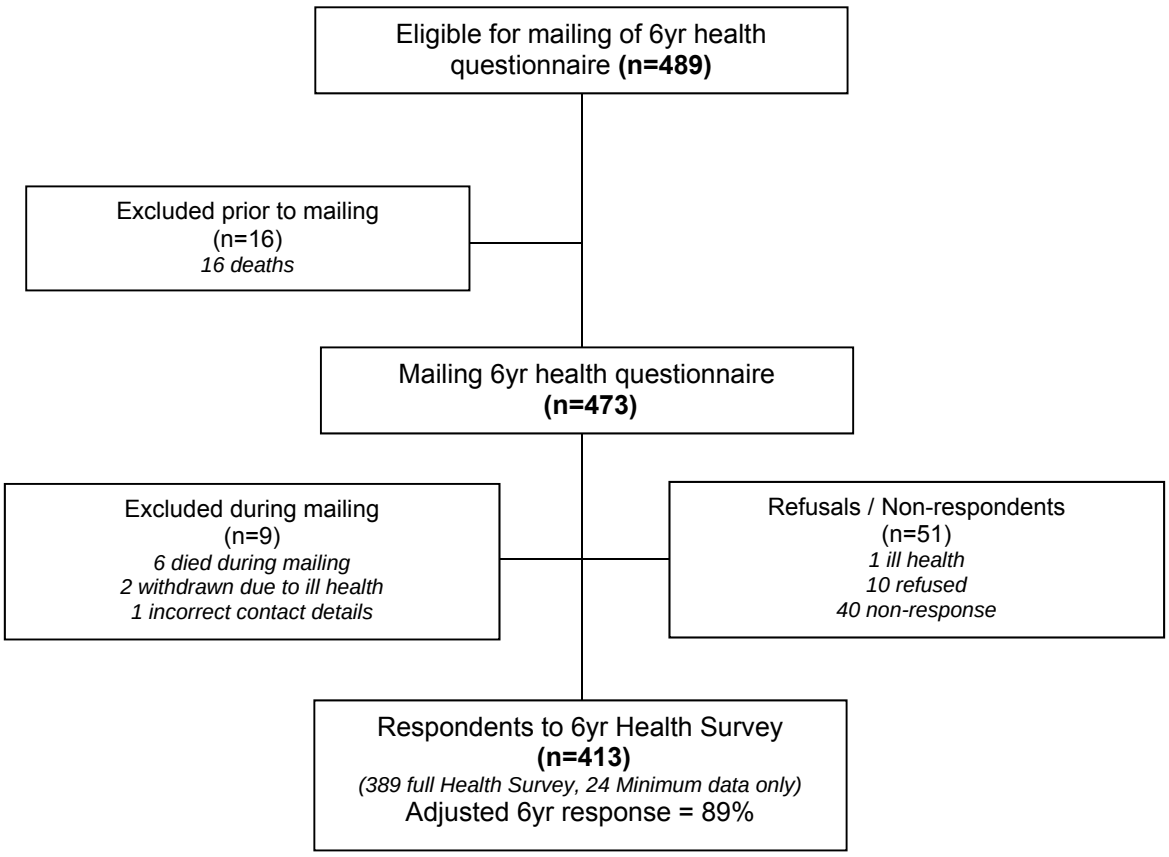
54-month follow-up

Note that eligibility for the 54-month mailing includes only those participants that gave written consent to further contact on the 3-year survey



6-year follow-up

Note that eligibility for the 6-year mailing includes participants who gave written consent on the 3-year questionnaire that was maintained throughout the 54-month follow-up mailing. It also includes 3 participants who phoned the research centre to request to be put back in the mailing sample after receiving the 54-month thank you letter, 5 participants who no longer needed to be excluded by the GP-screen, and 22 participants who had left the practice and were not traced at the 54-month follow-up (the NHS tracing service was not available in the NHS at the time of the 54-month mailing, but was available for the 6-year follow-up).



3.5 Assessing selection bias in CAS-HA

As the CAS-HA study includes multiple stages of recruitment there are multiple opportunities for selection bias to occur, i.e. for there to be a lack of comparability between participant groups studied at each study stage (Grimes et al. 2002). To assess this, the distribution of all variables in Table 3-1 were compared at each study stage (i.e. at each stage of recruitment and follow-up), using numbers and percentages (for categorical data) or means and standard deviations/medians and interquartile ranges (for normally distributed and skewed continuous data respectively) (Table 3-2 and Table 3-3).

The key comparisons in Table 3-2 were between (1) the baseline eligible population and those responding to the Health Survey (column A vs B) and (2) those reporting hand pain or problems in the last 12-months on the baseline Health Survey (column C) and three sub-groups within that a) those that gave consent to further contact (column D), b) responded to the Regional Pains Survey (column E) and c) attended the research clinic, i.e. the baseline CAS-HA sample used in this thesis (column F)[16].

In general, those responding to the Health Survey were largely representative of the baseline eligible population for age, gender and general practice (i.e. for the variables that were available for the entire eligible population). Participants attending the clinical assessment (i.e. the CAS-HA baseline sample) had a tendency to be more likely to be married, of a non-manual social class (i.e. to be in the higher or lower managerial or professional classes), to be quite comfortably off, to have lower alcohol consumption, and have good/excellent general health than those reporting hand pain on the Health Survey questionnaire. They also had slightly worse hand pain severity and functional difficulty than all of those responding to the Regional Pains Survey.

[16] Participants returning the health questionnaire were not compared to participants included at recruitment stages C-F; participants at recruitment stages C - F have been selected (by design) as those with hand pain/problems making comparisons to recruitment stages C-F no longer solely about selection bias

In addition, the effect of loss to follow-up within the CAS-HA cohort is small although there is slight evidence that those remaining in the cohort at the 6-year follow-up have marginally better general health than those responding to the earlier follow-up time-points (Table 3-3).

Table 3-2: Participant characteristics at each stage of recruitment to the CAS-HA study (baseline data collection)

Baseline characteristics	Column A Baseline eligible population N=6411	Column B Responded to Health Survey N=4511	Column C Reported hand pain or problems in last 12 months N=2114	Column D Consented to further contact N=1606	Column E Responded to Regional Pains Survey N=1370	Column F Attended research clinic (i.e. CAS-HA sample) N=623
General Practice A	2752 (43)	1888 (42)	842 (40)	665 (41)	575 (42)	243 (39)
Age (years)[β]	63 (56, 72)	64 (57, 73)	64 (57, 73)	63 (57, 71)	63 (57, 71)	63 (58, 71)
Female gender	3377 (53)	2479 (55)	1300 (62)	967 (60)	835 (61)	385 (62)
Marital status						
Married	-	3136 (70)	1436 (69)	1129 (71)	974 (72)	472 (76)
Separated	-	38 (1)	15 (1)	13 (1)	10 (1)	4 (1)
Divorced	-	286 (6)	153 (7)	119 (8)	98 (7)	38 (6)
Widowed	-	717 (16)	370 (18)	240 (15)	197 (15)	69 (11)
Cohabiting	-	93 (2)	44 (2)	39 (2)	32 (2)	17 (3)
Single	-	185 (4)	75 (4)	55 (3)	50 (4)	19 (3)
Social class						
Higher managerial	-	202 (5)	80 (4)	70 (5)	59 (5)	28 (5)
Higher professional	-	110 (3)	40 (2)	35 (2)	30 (2)	11 (2)
Lower managerial/professional	-	574 (14)	258 (14)	227 (15)	207 (16)	114 (20)
Intermediate occupations	-	438 (11)	209 (11)	172 (12)	156 (12)	82 (14)
Self-employed	-	295 (7)	126 (7)	104 (7)	84 (7)	41 (7)
Lower supervisory/technical	-	253 (6)	115 (6)	95 (6)	76 (6)	33 (6)
Semi-routine	-	1049 (26)	509 (27)	390 (26)	327 (26)	128 (22)
Routine	-	1135 (28)	575 (30)	405 (27)	341 (27)	144 (25)
Employment status						
Employed	-	1366 (32)	563 (28)	483 (31)	403 (31)	165 (28)
Not working due to ill-health	-	274 (6)	188 (9)	138 (9)	113 (9)	56 (9)

Retired	-	2249 (52)	1067 (52)	756 (49)	654 (50)	309 (52)
Unemployed	-	39 (1)	15 (1)	11 (1)	10 (1)	3 (1)
Housewife	-	288 (7)	145 (7)	112 (7)	97 (7)	45 (8)
Other	-	126 (3)	57 (3)	52 (3)	44 (3)	18 (3)
Age when left school (years)[β]	-	15 (14, 16)	15 (14, 16)	15 (15, 16)	15 (15, 16)	15 (15, 16)
Go from school to full time education	-	606 (14)	263 (13)	233 (15)	201 (15)	97 (16)
Income						
Find it a strain to get by from week to week	-	142 (3)	90 (4)	72 (5)	58 (4)	23 (4)
Have to be careful with money	-	1694 (38)	858 (42)	634 (40)	532 (39)	229 (37)
Able to manage without much difficulty	-	1794 (41)	808 (39)	620 (39)	532 (39)	251 (41)
Quite comfortably off	-	781 (18)	311 (15)	253 (16)	227 (17)	111 (18)
Alcohol consumption						
Daily or most days	-	932 (21)	410 (20)	342 (22)	293 (22)	136 (22)
Once or twice a week,	-	1576 (35)	714 (34)	562 (35)	482 (36)	218 (35)
Once or twice a month	-	698 (16)	319 (15)	250 (16)	217 (16)	116 (19)
Once or twice a year	-	717 (16)	367 (18)	267 (17)	219 (16)	97 (16)
Never	-	543 (12)	285 (14)	173 (11)	148 (11)	52 (8)
Smoking status						
Never	-	2153 (48)	1003 (48)	763 (48)	657 (48)	315 (51)
Previously smoked	-	1746 (39)	847 (40)	648 (41)	557 (41)	258 (42)
Currently smoke	-	571 (13)	250 (12)	186 (12)	148 (11)	45 (7)
Lives alone	-	916 (22)	467 (24)	322 (21)	273 (21)	108 (18)
General Health						
Excellent	-	191 (4)	54 (3)	49 (3)	43 (3)	29 (5)
Very good	-	1038 (23)	354 (17)	305 (19)	268 (20)	132 (21)
Good	-	1866 (42)	845 (41)	669 (42)	575 (42)	256 (41)
Fair	-	1113 (25)	642 (31)	438 (28)	365 (27)	158 (26)
Poor	-	256 (6)	192 (9)	130 (8)	104 (8)	44 (7)
Physical component score of the SF-12 (0-100)[β]	-	47 (33, 54)	39 (29, 51)	41 (30, 52)	41 (30, 52)	40 (30, 51)
Number of comorbidities						
0	-	1591 (35)	653 (31)	525 (33)	449 (33)	219 (35)
1	-	1566 (35)	693 (33)	539 (34)	452 (33)	198 (32)

2	-	814 (18)	448 (21)	324 (20)	284 (21)	122 (20)
3	-	342 (8)	190 (9)	129 (8)	113 (8)	55 (9)
4	-	134 (3)	87 (4)	61 (4)	48 (4)	19 (3)
5	-	40 (1)	30 (1)	21 (1)	17 (1)	7 (1)
6	-	24 (1)	13 (1)	7 (1)	7 (1)	3 (1)
Pain in other body areas						
No other pain	-	1506 (33)	387 (18)	278 (17)	228 (17)	101 (16)
Regional pain	-	2476 (55)	1246 (59)	952 (59)	818 (60)	358 (58)
Widespread pain	-	529 (12)	481 (23)	376 (23)	324 (24)	164 (26)
AUSCAN pain (0-20)[β]	-	-	-	-	5.0 (3.0, 9.0)	6.0 (3.0, 9.0)
Number of days in the last 12-months with hand pain						
less than 7-days	-	-	-	-	115 (10)	40 (7)
1-4 weeks	-	-	-	-	146 (13)	66 (12)
>1-mth but <3-mths	-	-	-	-	192 (17)	85 (15)
3-mths or more	-	-	-	-	677 (60)	363 (66)
AUSCAN function (0-36)[β]	-	-	-	-	7.0 (2.0, 15.8)	8.0 (3.0, 17.0)
Side affected						
Dominant hand only	-	-	-	-	-	75 (12)
Non-dominant hand only	-	-	-	-	-	28 (4)
One hand affected but participant ambidextrous	-	-	-	-	-	7 (1)
Both hands affected	-	-	-	-	-	513 (82)
Time since hand problem onset (years)[β]	-	-	-	-	5 (2, 10)	5 (2, 12)
Sudden onset						
Bilateral problem - both hands sudden onset	-	-	-	-	-	78 (13)
Bilateral problem – one hand sudden onset	-	-	-	-	-	34 (5)
Bilateral problem – neither hand of sudden onset	-	-	-	-	-	408 (66)
Unilateral problem – of sudden onset	-	-	-	-	-	38 (6)
Unilateral problem – not of sudden onset	-	-	-	-	-	64 (10)
Onset following accident or injury						

Bilateral problem - both hands onset following accident/injury	-	-	-	-	-	16 (3)
Bilateral problem – one hand onset following accident/injury	-	-	-	-	-	33 (5)
Bilateral problem – neither hand onset following accident/injury	-	-	-	-	-	471 (76)
Unilateral problem – onset following accident/injury	-	-	-	-	-	16 (3)
Unilateral problem – onset not following accident/injury	-	-	-	-	-	86 (14)
Past or present job, hobbies or pastimes involved excessive hand use	-	-	-	-	996 (80)	480 (81)
HADS – Anxiety (0-21)[β]	-	6.0 (3.0, 9.0)	7.0 (4.0, 10.0)	7.0 (4.0, 9.0)	7.0 (4.0, 9.0)	6.0 (4.0, 9.0)
HADS – Depression (0-21)[β]	-	4.0 (2.0, 7.0)	4.0 (2.0, 7.0)	4.0 (2.0, 7.0)	4.0 (2.0, 7.0)	3.8 (2.0, 6.0)
Mental component score of the SF-12 (0-100)[β]	-	54 (44, 58)	52 (39, 58)	53 (41, 58)	53 (42, 58)	54 (44, 59)
Illness perceptions						
Long disease time course (6-30)[β]	-	-	-	-	24 (20, 26)	24 (21, 26)
Consequences (6-30)[β]	-	-	-	-	13 (10, 17)	13 (11, 18)
Personal Control (6-30)[α]	-	-	-	-	17.7 (4.3)	17.9 (4.2)
Treatment Control (5-25)[α]	-	-	-	-	14.4 (3.3)	14.5 (3.3)
Illness coherence (5-25)[β]	-	-	-	-	11 (10, 15)	11 (10, 15)
Cyclical time course (4-20)[β]	-	-	-	-	12 (9, 15)	12 (8, 14)
Emotional representation (6-30)[α]	-	-	-	-	13.5 (4.5)	13.6 (4.6)
Frustration with hand condition in the last month						
All days	-	-	-	-	82 (7)	40 (7)
Most days	-	-	-	-	139 (11)	75 (13)
Some days	-	-	-	-	222 (18)	107 (18)
Few days	-	-	-	-	242 (20)	117 (20)
No days	-	-	-	-	540 (44)	242 (42)
Body-mass index (kg/m^2)[α]	-	-	-	-	-	28.2 (4.8)
Hand grip-ability (GAT) (seconds)[β]	-	-	-	-	-	29 (24, 36)

Grip strength (lbs)[α]	-	-	-	-	-	49.2 (25.9)
Pinch strength (lbs)[α]	-	-	-	-	-	10.4 (4.3)
Meets the ACR criteria for hand OA	-	-	-	-	-	191 (31)
Has carpal tunnel syndrome	-	-	-	-	-	275 (46)
Has Dupuytren's contracture	-	-	-	-	-	165 (26)
Has De Quervain's tenosynovitis	-	-	-	-	-	137 (23)
Has trigger finger	-	-	-	-	-	123 (20)
Number of joints with kellgren-Lawrence x-ray grade >=2[β]	-	-	-	-	-	3 (1,7)

Figures are numbers and percentages unless otherwise stated. α = Mean (standard deviation); β= Median (inter-quartile range)
Abbreviations SF-12 = Short-form 12; AUSCAN = Australian/Canadian Hand Osteoarthritis Index; HADS = Hospital Anxiety and Depression Scale; ACR = American College of Rheumatology; OA = Osteoarthritis

Table 3-3: Participant characteristics at each follow-up stage of the CAS-HA study

Baseline characteristics	Column F Attended research clinic N=623	Column G Responded to 18-month survey N=595	Column H Responded to 3-year survey N=564	Column I Responded to 54-month survey N=445	Column J Responded to 6-year survey N=413
General Practice A	243 (39)	229 (39)	216 (38)	181 (41)	168 (41)
Age (years)[β]	63 (58, 71)	63 (58, 70)	63 (58, 70)	62 (57, 69)	62 (57, 68)
Female gender	385 (62)	368 (62)	349 (62)	279 (63)	255 (62)
Marital status					
Married	472 (76)	454 (77)	427 (76)	341 (77)	318 (77)
Separated	4 (1)	4 (1)	4 (1)	3 (1)	3 (1)
Divorced	38 (6)	36 (6)	35 (6)	28 (6)	30 (7)
Widowed	69 (11)	64 (11)	60 (11)	45 (10)	35 (9)
Cohabiting	17 (3)	15 (3)	16 (3)	10 (2)	11 (3)
Single	19 (3)	19 (3)	19 (3)	15 (3)	14 (3)
Social class					
Higher managerial	28 (5)	28 (5)	27 (5)	22 (5)	22 (6)
Higher professional	11 (2)	10 (2)	11 (2)	7 (2)	7 (2)
Lower managerial/professional	114 (20)	110 (19)	103 (19)	84 (20)	85 (22)
Intermediate occupations	82 (14)	81 (14)	76 (14)	57 (14)	52 (13)
Self-employed	41 (7)	39 (7)	38 (7)	28 (7)	23 (6)
Lower supervisory/technical	33 (6)	33 (6)	31 (6)	27 (6)	24 (6)
Semi-routine	128 (22)	121 (20)	117 (22)	92 (22)	88 (23)
Routine	144 (25)	137 (23)	128 (24)	102 (24)	88 (23)
Employment status					
Employed	165 (28)	160 (28)	150 (28)	131 (31)	125 (32)
Not working due to ill-health	56 (9)	52 (9)	49 (9)	34 (8)	34 (9)
Retired	309 (52)	297 (52)	283 (52)	214 (50)	193 (49)
Unemployed	3 (1)	3 (1)	3 (1)	1 (1)	2 (1)
Housewife	45 (8)	42 (7)	41 (8)	31 (7)	26 (7)
Other	18 (3)	17 (3)	17 (3)	16 (4)	16 (4)

Age when left school (years)[β]	15 (15, 16)	15 (15, 16)	15 (15, 16)	15 (15, 16)	15 (15, 16)
Go from school to full time education	97 (16)	94 (16)	89 (16)	69 (16)	67 (17)
Income					
Find it a strain to get by from week to week	23 (4)	22 (4)	20 (4)	14 (3)	14 (3)
Have to be careful with money	229 (37)	215 (37)	202 (36)	153 (35)	136 (33)
Able to manage without much difficulty	251 (41)	243 (42)	233 (42)	190 (43)	178 (44)
Quite comfortably off	111 (18)	106 (18)	102 (18)	86 (19)	81 (20)
Alcohol consumption					
Daily or most days	136 (22)	131 (22)	127 (23)	103 (23)	97 (24)
Once or twice a week,	218 (35)	207 (35)	191 (34)	149 (34)	144 (35)
Once or twice a month	116 (19)	112 (19)	110 (20)	86 (19)	75 (18)
Once or twice a year	97 (16)	90 (15)	85 (15)	71 (16)	62 (15)
Never	52 (8)	51 (9)	47 (8)	34 (8)	32 (8)
Smoking status					
Never	315 (51)	306 (52)	288 (51)	230 (52)	219 (54)
Previously smoked	258 (42)	245 (42)	232 (41)	180 (41)	165 (40)
Currently smoke	45 (7)	40 (7)	40 (7)	31 (7)	25 (6)
Lives alone	108 (18)	103 (18)	99 (18)	76 (18)	67 (17)
General Health					
Excellent	29 (5)	27 (5)	27 (5)	24 (5)	24 (6)
Very good	132 (21)	128 (22)	122 (22)	105 (24)	103 (25)
Good	256 (41)	246 (42)	233 (42)	186 (42)	172 (42)
Fair	158 (26)	150 (25)	143 (26)	104 (24)	90 (22)
Poor	44 (7)	40 (7)	36 (6)	23 (5)	21 (5)
Physical component score of the SF-12 (0-100)[β]	40 (30, 51)	40 (30, 51)	40 (30, 51)	42 (31, 52)	42 (31, 52)
Number of comorbidities					
0	219 (35)	213 (36)	206 (37)	171 (38)	164 (40)
1	198 (32)	189 (32)	178 (32)	140 (32)	128 (31)
2	122 (20)	117 (20)	111 (20)	78 (18)	75 (18)
3	55 (9)	50 (8)	45 (8)	35 (8)	32 (8)
4	19 (3)	17 (3)	16 (3)	14 (3)	9 (2)
5	7 (1)	7 (1)	6 (1)	5 (1)	4 (1)

6	3 (1)	2 (0)	2 (0)	2 (0)	1 (0)
Pain in other body areas					
No other pain	101 (16)	92 (16)	88 (16)	69 (16)	68 (17)
Regional pain	358 (58)	344 (58)	327 (58)	256 (58)	238 (58)
Widespread pain	164 (26)	159 (27)	149 (26)	120 (27)	107 (26)
AUSCAN pain (0-20)[β]	6.0 (3.0, 9.0)	6.0 (3.0, 9.0)	5.0 (3.0, 9.0)	5.0 (3.0, 9.0)	5.0 (3.0, 9.0)
Number of days in the last 12-months with hand pain					
less than 7-days	40 (7)	40 (8)	39 (8)	30 (8)	27 (7)
1-4 weeks	66 (12)	62 (12)	62 (12)	48 (12)	47 (13)
>1-mth but <3-mths	85 (15)	79 (15)	74 (15)	60 (15)	58 (16)
3-mths or more	363 (66)	346 (66)	328 (65)	259 (65)	235 (64)
AUSCAN function (0-36)[β]	8.0 (3.0, 17.0)	8.0 (3.0, 16.0)	8.0 (3.0, 16.4)	7.0 (2.0, 15.0)	7.0 (2.0, 15.0)
Side affected					
Dominant hand only	75 (12)	73 (12)	67 (12)	59 (13)	56 (14)
Non-dominant hand only	28 (4)	27 (5)	24 (4)	22 (5)	19 (5)
One hand affected but participant ambidextrous	7 (1)	7 (1)	7 (1)	6 (1)	7 (2)
Both hands affected	513 (82)	488 (82)	466 (83)	358 (80)	331 (80)
Time since hand problem onset (years)[β]	5 (2, 12)	5 (2, 12)	5 (2, 12)	5 (2, 11)	5 (2, 11)
Sudden onset					
Bilateral problem - both hands sudden onset	78 (13)	73 (12)	73 (13)	55 (12)	54 (13)
Bilateral problem – one hand sudden onset	34 (5)	31 (5)	27 (5)	21 (5)	20 (5)
Bilateral problem – neither hand of sudden onset	408 (66)	391 (66)	373 (66)	289 (65)	265 (64)
Unilateral problem – of sudden onset	38 (6)	37 (6)	34 (6)	28 (6)	26 (6)
Unilateral problem – not of sudden onset	64 (10)	62 (10)	56 (10)	51 (12)	47 (11)
Onset following accident or injury					
Bilateral problem - both hands onset following accident/injury	16 (3)	16 (3)	14 (3)	10 (2)	11 (3)
Bilateral problem – one hand onset following accident/injury	33 (5)	30 (5)	29 (5)	23 (5)	21 (5)

Bilateral problem – neither hand onset following accident/injury	471 (76)	449 (76)	430 (76)	332 (75)	307 (75)
Unilateral problem – onset following accident/injury	16 (3)	16 (3)	15 (3)	12 (3)	11 (3)
Unilateral problem – onset not following accident/injury	86 (14)	83 (14)	75 (13)	67 (15)	62 (15)
Past or present job, hobbies or pastimes involved excessive hand use	480 (81)	460 (81)	434 (81)	348 (83)	320 (82)
HADS – Anxiety (0-21)[β]	6.0 (4.0, 9.0)	6.0 (4.0, 9.0)	6.0 (4.0, 9.0)	6.0 (4.0, 9.0)	6.0 (3.3, 9.0)
HADS – Depression (0-21)[β]	3.8 (2.0, 6.0)	3.0 (2.0, 6.0)	3.0 (2.0, 6.0)	3.0 (1.0, 6.0)	3.0 (1.0, 6.0)
Mental component score of the SF-12 (0-100)[β]	54 (44, 59)	54 (45, 59)	54 (44, 59)	54 (46, 59)	55 (45, 59)
Illness perceptions					
Long disease time course (6-30)[β]	24 (21, 26)	24 (21, 26)	24 (21, 26)	24 (21, 26)	24 (21, 27)
Consequences (6-30)[β]	13 (11, 18)	13 (11, 18)	13 (11, 18)	13 (10, 17)	13 (10, 17)
Personal Control (6-30)[α]	17.9 (4.2)	17.9 (4.3)	18.0 (4.2)	18.1 (4.3)	18.1 (4.4)
Treatment Control (5-25)[α]	14.5 (3.3)	14.6 (3.3)	14.6 (3.3)	14.7 (3.3)	14.7 (3.4)
Illness coherence (5-25)[β]	11 (10, 15)	11 (10, 15)	11 (10, 15)	11 (10, 15)	11 (10, 15)
Cyclical time course (4-20)[β]	12 (8, 14)	12 (8, 14)	12 (8,14)	12 (8, 14)	12 (8, 14)
Emotional representation (6-30)[α]	13.6 (4.6)	13.6 (4.6)	13.6 (4.6)	13.4 (4.5)	13.3 (4.6)
Frustration with hand condition in the last month					
All days	40 (7)	36 (7)	35 (7)	22 (5)	20 (5)
Most days	75 (13)	71 (13)	71 (14)	60 (15)	51 (13)
Some days	107 (18)	102 (18)	94 (18)	74 (18)	74 (19)
Few days	117 (20)	112 (20)	102 (19)	86 (21)	78 (20)
No days	242 (42)	236 (42)	224 (43)	173 (42)	163 (42)
Body-mass index (kg/m^2)[α]	28.2 (4.8)	28.3 (4.8)	28.2 (4.7)	28.1 (4.5)	28.0 (4.6)
Hand grip-ability (GAT) (seconds)[β]	29 (24, 36)	29 (24, 36)	28 (24, 35)	28 (24, 34)	27 (23, 34)
Grip strength (lbs)[α]	49.2 (25.9)	49.2 (25.9)	49.5 (25.8)	51.0 (25.7)	50.8 (25.6)
Pinch strength (lbs)[α]	10.4 (4.3)	10.4 (4.4)	10.5 (4.4)	10.7 (4.4)	10.7 (4.4)
Meets the ACR criteria for hand OA	191 (31)	181 (31)	172 (31)	134 (30)	127 (31)
Has carpal tunnel syndrome	275 (46)	261 (46)	247 (46)	191 (44)	176 (44)

Has Dupuytren's contracture	165 (26)	157 (26)	150 (27)	110 (25)	102 (25)
Has De Quervain's tenosynovitis	137 (23)	129 (23)	123 (23)	100 (24)	90 (23)
Has trigger finger	123 (20)	118 (20)	112 (20)	77 (17)	76 (18)
Number of joints with kellgren-Lawrence x-ray grade >=2[β]	3 (1,7)	3 (1, 7)	3 (1, 7)	3 (1, 7)	3 (1, 7)

Figures are numbers and percentages unless otherwise stated. α = Mean (standard deviation); β = Median (inter-quartile range)
Abbreviations SF-12 = Short-form 12; AUSCAN = Australian/Canadian Hand Osteoarthritis Index; HADS = Hospital Anxiety and Depression Scale; ACR = American College of Rheumatology; OA = Osteoarthritis

3.6 Summary

The CAS-HA study has been presented in this chapter and it is acknowledged that the study design has both strengths and weaknesses. A discussion of these issues, however, is reserved for Chapter 10, to make the discussion relevant to the results presented in Chapters 6 to 9. The CAS-HA study data is used as a dataset in all chapters of this thesis and is used in the next chapter to explore the reliability, validity and interpretability of the AUSCAN as a measure of hand pain and functional difficulty in the CAS-HA study.

4 Psychometric properties of the Australian/Canadian Hand Osteoarthritis Index (AUSCAN)

4.1 Introduction

The Australian/Canadian Hand Osteoarthritis Index (AUSCAN) (Bellamy et al. 2002a) is the primary measure of hand pain and functional difficulty used in this thesis. The aim of this chapter is therefore to explore the psychometric properties of this measure for use with older adults with hand pain.

The chapter is structured into four sections, the first to describe the AUSCAN measure (Section 4.2), the second to define how the psychometric properties of the AUSCAN are assessed in this chapter (Section 4.3), the third to describe the results of the assessment (Section 4.4) and a final section to summarise the findings and discuss their implications for later thesis chapters (Section 4.5).

The primary source of data used in this chapter is the CAS-HA data, however, for some assessments external datasets are drawn upon, as they are able to provide data that is more appropriate to test specific psychometric properties of the measure. Where such external studies are used, they are referred to in the sections described above and include data from the CAS-HA pilot study (Myers et al. 2011) and the Self-management in Osteoarthritis of the Hand trial (the SMOotH trial) (Dziedzic et al. 2011, Dziedzic et al. 2013). It is therefore assumed that data are from the CAS-HA study unless otherwise stated.

4.2 The Australian/Canadian Hand Osteoarthritis Index (AUSCAN)

The AUSCAN is a self-administered questionnaire used to assess hand pain, stiffness and limitations in hand function in patients with hand osteoarthritis (OA) (Bellamy et al. 2002a). The authors of the tool have previously described its development (Bellamy et al.

2002a) and highlight the importance of developing questionnaires using a multistage approach, including item generation, item reduction, piloting and psychometric evaluation.

Initially 95 potential items were identified for inclusion in the AUSCAN questionnaire which were generated by interviews with hand OA patients, clinicians experienced in their management (rheumatologists, orthopaedic surgeons and physiotherapists), and by review of eight previously published tools purporting to measure hand disability. A multi-stage process was used to select the final 15 items for inclusion in the questionnaire. This included steps to ensure that items selected were important to patients, unambiguous, not gender-specific and relevant to a hand OA population (indicated by selecting only items with a prevalence rate >= 60%).

The AUSCAN questionnaire was originally developed for patients with a clinical diagnosis of hand OA, so it may not be appropriate for use with all CAS-HA participants, as not all participants have a clinical diagnosis of hand OA. In response, the AUSCAN questionnaire was adapted for CAS-HA participants by changing the question reference from ‘hand arthritis’ to ‘hand problem’ and by extending the question time frame from ‘in the last 48 hours’ to ‘in the last week’ to account for the likelihood of milder symptoms in the CAS-HA cohort. However, due to the age of the participants in the sample, many were expected to have radiographic or clinical OA, and results from the clinical assessment indeed showed that 82% had radiographic evidence of hand OA (Kellgren & Lawrence ≥2 in at least one hand joint) and 30% met the ACR clinical criteria for hand OA (Marshall et al. 2009).

Furthermore, for clarity, examples were added to the items ‘carrying a pot with one hand’ and ‘wringing out washcloths’ and some words in the questionnaire were changed to apply more directly to the UK population, e.g. faucet changed to tap. The items included in the AUSCAN questionnaire are shown in Box 1. The adapted version of the AUSCAN tool

was piloted before use; it was well-completed and acceptable to participants (Myers et al. 2011).

Box 1: AUSCAN questionnaire items used in the CAS-HA study

Pain:

Severity of hand pain in the last week when…..

- At rest
- Gripping objects
- Lifting objects
- Turning objects
- Squeezing objects

Stiffness:

Hand stiffness after first waking in the morning

Function:

Degree of difficulty in the last week due to your hand problem…..

- Turning taps
- Turning a round door-knob or handle
- Doing up buttons
- Fastening jewellery
- Opening a new jar
- Carrying a full pot with one hand (e.g. saucepan)
- Peeling vegetables/fruit
- Picking up large heavy objects
- Wringing out wash clothes (e.g. squeezing a wet sponge or flannel)

4.2.1 Scoring of the AUSCAN questionnaire

Each item included in the AUSCAN questionnaire was rated by participants in the CAS-HA study using a 5-point Likert scale with response options of “None” to “Extreme”. After completion of the questionnaire, response options were scored on a scale of 0-4 and, for subscales containing multiple items, added together to make two total scores (one for hand pain and one for hand function). Although the AUSCAN questionnaire includes within it a measure of hand stiffness, the scores for this item were not assessed in this chapter as they were not analysed in later thesis chapters and have potential to be

unreliable as based on a single item only. Hand pain scores ranged from 0 to 20 and the hand function score from 0 to 36, with a higher score representing greater pain or functional difficulty.

The authors of the AUSCAN (Bellamy et al. 2002a) provide guidelines on scoring of missing data and suggest that participants with a single item missing for hand pain, and up to two items missing for hand function, can still have a score computed for the outcome. The authors suggest that missing values (where necessary) be imputed with the mean item response for completed items on the specific subscale. A total score for hand pain or function would therefore be obtained for that participant by summation of both completed and imputed responses.

4.3 Psychometric testing of the AUSCAN questionnaire

The psychometric properties of the AUSCAN were assessed in this chapter using the eight criteria shown in Box 2 (Terwee et al. 2007). The criteria encapsulate the key concepts of reliability; that the questionnaire tool is measuring the concept in a reproducible way, and validity; that the tool is measuring what it intends to measure (Frey 2015). Each concept is described below with an explanation of how it was assessed in this chapter.

Box 2: Psychometric criteria used to determine if the AUSCAN was a reliable and valid measure of hand pain and functional difficulty

- Face and content validity
- Item distribution (missing data rates and floor and ceiling effects)
- Internal consistency
- Reproducibility (test-retest reliability)
- Construct validity
- Criterion validity
- Responsiveness
- Interpretability and definition of minimum important change

List derived from (Terwee et al. 2007) and (Streiner 2003)

4.3.1 Face and content validity

Face and content validity are two terms that are used to describe whether the items included in any questionnaire scale are reasonable (Coolican 2014). The term face validity refers to a subjective opinion that the questionnaire is measuring the desired qualities intended (e.g. when assessed by clinical experts), whereas content validity is a judgement as to whether the questionnaire tool samples all of the important content or domains of a condition (Streiner 2003). In this chapter face and content validity were determined by reviewing evidence in the publication that describes how items in the AUSCAN were initially generated (Bellamy et al. 2002a). In addition, items were assessed to see how suitable they were to measure hand pain and function in the CAS-HA cohort.

4.3.2 Item distribution (missing data rates and floor and ceiling effects)

Missing data rates and floor and ceiling effects were considered in this chapter for each AUSCAN item and their respective subscale scores. High item-level missing data rates could indicate ambiguous or unacceptable items for participants to complete. High floor and ceiling effects, i.e. the percentage of participants at the lowest and highest value on the scale respectively (Bowling 2014), could indicate items or subscale scores that do not fully capture the range of symptoms experienced by CAS-HA participants and where discrimination between participants at the extreme ends of the outcome distribution is poor. When floor and ceiling effects were assessed in this chapter, the guideline by Terwee et al. 2007 was used: that floor or ceiling effects (the proportion of participants scoring either the lowest or highest response option on the scale) greater than 15% indicate inferior content validity and hence limited ability to assess change over time.

4.3.3 Internal consistency

Internal consistency is the degree to which items on a measurement scale measure the same construct (Walsh et al. 1990) and is important to demonstrate, to ensure that items included within the scale are capturing different aspects of a single construct rather than tapping into constructs not intended to be measured by that scale (Streiner 2003). In this

thesis internal consistency was assessed using confirmatory factor analysis, item-total correlations and Cronbach's alpha, as described below.

Confirmatory factor analysis

A 2-factor confirmatory factor analysis (CFA) model was fitted to the hand pain and function AUSCAN items to test whether the items measured two distinct constructs and hence whether it was appropriate to analyse them as two separate subscales[17]. The items were defined in the model to relate only to their hypothesised factor (either hand pain or functional ability), were measured with independent errors and were modelled assuming an ordinal scale of measurement. A correlation term between the latent constructs was also included to represent the overall correlation between hand pain and functional difficulty.

Goodness-of-fit of the confirmatory factor models were assessed using the following fit indices: Tucker-Lewis index (TLI) (Tucker et al. 1973), comparative fit index (CFI) (Bentler 1990), standardised root mean-square residual (SRMR) (Bentler 1995) and the root mean square error of approximation (RMSEA) (Steiger 1990), with good fit indicated if the TLI and CFI were greater than 0.95, and the SRMR and RMSEA were less than 0.05 and 0.06, respectively. In addition, factor loadings were assessed to ensure they were statistically plausible and interpretable, i.e. that all estimates of variance were positive, that correlations were between plus and minus one, and that the standardised regression coefficients were statistically significant and greater than 0.4 (Ferguson et al. 1993).

In addition, if the CFA model was found not to be a good fit to the data, modification indices were used as an exploratory tool to identify any correlations that, if added to the

[17] Confirmatory factor analysis was used rather than exploratory factor analysis as there was a clear hypothesis around which items were planned to measure which construct (i.e. five items to measure hand pain and nine items to measure hand function) (Bellamy et al. 2002a)).

model, would greatly improve model fit[18]. The extra correlations were added to the model sequentially, starting with the correlation that had the potential to improve model fit by the largest amount, and continued until good model fit was achieved using the fit indices as defined above, i.e. model fit was assessed after each additional correlation was added to the model and this process continued until good model fit was achieved. The aim of this analysis was to gain understanding as to why the model did not fit the data well if that was a finding from the CFA.

Item-total correlations and Cronbach's alpha

Item-total correlations and Cronbach's alpha were used to measure the degree of homogeneity of items within each subscale and were reported if the results of the CFA model demonstrated that hand pain and function were two separate subscales[19]. If the item-total correlations were between 0.2 and 0.8 and Cronbach's alpha for each subscale was in the region of 0.7 to 0.95 then internal consistency of the scales would be concluded (Bowling 2014, Terwee et al. 2007). The upper limit was required for these statistics as values greater than this may suggest item redundancy, i.e. items are so highly correlated that the benefit of including two items rather than one is not worthwhile (Streiner 2003).

4.3.4 Reproducibility

Reproducibility is a measure of the degree to which repeated measurements in stable patients provide similar results and includes within it two key concepts: reliability and agreement (de Vet et al. 2006a). Reliability (when discussed in the context of reproducibility) concerns the degree to which patients can be distinguished from each other despite measurement error, whereas agreement concerns absolute measurement

[18] Modification indices give the expected drop in the model chi-square fit statistic if the correlation of interest is added to the model (Muthen et al 2010)

[19] Item-total correlations are defined as the correlations of each individual item with the total subscale score omitting that item. Cronbach's alpha is the average of all possible split-half correlations with a split half correlation defined by splitting the items in a subscale into two parts, and correlating the resulting scales (Streiner 2003).

error and reflects how close scores on repeated measures are (Terwee et al. 2007) (see Appendix 5 for illustration).

Both these concepts were important to consider in this thesis as the AUSCAN is used to measure how symptoms change over time. They were assessed using data from the CAS-HA pilot study only (N=55) because the time-interval between data collection time-points in the main CAS-HA study was too long for an accurate assessment of reproducibility to be conducted (see Appendix 6 for details of the test-retest component of the CAS-HA pilot study). Participants were only included in this analysis if they reported no change in symptoms over time on a global assessment of change question (see footnote for question wording)[20]. These concepts, and how they were assessed in this thesis, are described below:

Reliability

Intra-class correlation coefficients (ICCs) were used to measure the reliability of the AUSCAN hand pain and function measures using the formula defined in Box 3. The ICCs were measured on a scale of zero to one with one representing perfect reliability, i.e. no measurement error. The AUSCAN subscales were considered reliable if the ICC correlations were greater than 0.7 (Terwee et al. 2007).

[20] The wording for this question was "Compared to when you came for your assessment one month ago, how do you think your hand problem has changed?: "Completely recovered", "Much better", "Better", "No change", "Worse", "Much worse"

Box 3: Formulae used to calculate intra-class correlations (ICCs) in this thesis[21]

$$ICC = \frac{\sigma^2_s}{\sigma^2_s + \sigma^2_t + \sigma^2_e}$$

where σ^2_s, σ^2_t, and σ^2_e are estimated from a two-way random effects analysis of variance (ANOVA) model as:

σ^2_s = variability between participants,
σ^2_t = variability between measurement time-points } Measurement error
σ^2_e = residual error

Agreement

Agreement was assessed using the standard error of measurement (SEM). The SEM was calculated using the components of the ICC formula given in Box 3 as $\sqrt{\sigma^2_t + \sigma^2_e}$ (Terwee et al. 2009) to provide an estimate of the within-person standard deviation of a set of AUSCAN scores assuming they were measured repeatedly on a single participant with no symptom change over time (van Kampen et al. 2013). It is therefore expressed in units on the AUSCAN scales and in the context of the range of values that each scale can take (de Vet et al. 2006a).

In addition, Bland and Altman plots were used to explore whether the level of absolute agreement between AUSCAN scores at the two time points (i.e. d = the difference in the

[21] The ICC in Box 3 is referred to as ICC2 (A,1) using standard notation by McGraw and Wong (McGraw et al. 1996). This indicates that the ICCs are in class 2 (so all patients are measured on two occasions) by a single observer. The 'A' refers to agreement to indicate that variability between time points is included as measurement error, along with residual variability. Although the word 'Agreement' is used in this notation De Vet et al (de Vet et al. 2006b) highlight that 'Agreement' in the ICC context is still part of a description of reliability. Other ICC's can be calculated e.g. ICC2 (C,1) where the 'C' stands for 'Consistency' to indicate that measurement error is estimated using residual variability alone (i.e. the denominator in the ICC formula in Box 3 is $\sigma^2_s + \sigma^2_e$. The ICC2(C,1) was not used in this thesis as De Vet et al (de Vet et al. 2006a) support that systematic variability between time points should be included as part of measurement error as it would be considered an error, for example if participants were to systematically rate the AUSCAN one point higher on the second occasion than the first when no change in symptom state had occurred. Other ICCs can also be calculated e.g. when multiple raters provide measurement at each time point, but these ICCs were not relevant to assessing the reproducibility of the AUSCAN so were not considered further.

AUSCAN score between the two time-points) was dependent on the severity of AUSCAN score observed (i.e. m = the mean AUSCAN score between the two time-points)[22] (Bland et al. 1999). It was desirable for such plots to show no pattern in the data and for data points to be equally scattered across the full range of symptom severity. 95% Bland and Altman limits of agreement were also calculated as (m +/- 1.96*(standard deviation of the d score)) and superimposed on the plots to show the upper and lower bounds for which 95% of the difference scores would lie in the data.

4.3.5 Construct validity

Construct validity refers to the extent to which scores on a measurement tool relate to other measures in a manner consistent with theoretically derived hypotheses concerning the concepts being measured (Terwee et al. 2007). Construct validity of the AUSCAN was tested in this chapter using *a priori* hypotheses as stated in Box 4 and Box 5. The hypotheses included tests of convergent validity, i.e. to test whether the AUSCAN was correlated with a set of external variables hypothesised to be related to the AUSCAN, as well as extreme groups/discriminative validity, i.e. to test whether the mean AUSCAN scores differ between known groups expected to differ by their AUSCAN levels (Streiner 2003). The hypotheses were tested using correlation coefficients or mean differences, as appropriate.

[22] Bland and Altman plots were derived by plotting the mean AUSCAN score across the two occasions of measurement against the difference in AUSCAN score across the two measurement occasions

Box 4: Construct validity hypotheses for AUSCAN pain

AUSCAN pain would be positively correlated with:

- Pain severity in the last month (rated on a 0-10 scale)
- Pain subscale of the Arthritis Impact Measurement scales 2 (AIMS2) ((Meenan et al. 1992) and Appendix 7)
- The "Consequences" subscale of the Illness Perceptions Questionnaire – Revised (IPQ-R) (Moss-Morris et al. 2002)

AUSCAN pain scores would be higher if:

- Participants had visited their general practitioner (GP) in the last 12 months
- Participants reported hand pain in the last month using a dichotomous yes/no question

Box 5: Construct validity hypotheses for AUSCAN function

AUSCAN function would be positively correlated with:

- Grip strength (Mathiowetz et al. 1984)
- Pinch strength (Mathiowetz et al. 1984)
- Grip-ability test (GAT) (Dellhag 1995)
- AIMS2 hand and finger function scale (Meenan et al. 1992) and Appendix 7)
- The "Consequences" subscale of the Illness Perceptions Questionnaire – Revised (IPQ-R) (Moss-Morris et al. 2002)

AUSCAN function scores would be higher if:

- Participants had visited their general practitioner (GP) in the last 12 months

4.3.6 Criterion validity

Criterion validity is defined as the correlation of a scale with a 'gold standard' of the measure of the trait or disorder under study (Bowling 2014). A 'gold standard' measure of hand pain and function was not collected in the CAS-HA study so this form of validity testing was not explored further[23].

[23] Measures of grip and pinch strength could have been potential 'gold standards' for hand function. They were not used as a 'gold standard' in this chapter as self-reported hand function is a measure of hand function in daily living, e.g. with the ability to use gadgets to help with daily tasks, so is not a measure of hand strength alone.

4.3.7 Sensitivity to change and responsiveness

Sensitivity to change and responsiveness are two terms often used interchangeably to indicate the ability of a measure to detect change over time when change has occurred. However, it has been highlighted that "sensitivity to change" taps a measure's ability to measure any degree of change, whereas responsiveness assesses ability to measure change that is important to the patient (Streiner 2003). To acknowledge this distinction, sensitivity to change was derived using all participants completing the AUSCAN at baseline and the first follow-up time point, whereas responsiveness was determined by excluding participants reporting no change in their hand problem using a global assessment of change question[24].

Sensitivity to change and responsiveness were measured in the CAS-HA study and the SMOotH trial (Dziedzic et al. 2011, Dziedzic et al. 2013). The SMOotH trial data was used as an extra resource as it provided an opportunity to test sensitivity to change and responsiveness in a different data sample that had a greater potential for change over time to occur, as participants had been randomised to receive potentially effective treatments (see Appendix 8 for details of the SMOotH study). In addition, the time period between the first two measurements in the SMOotH study was much shorter than in CAS-HA (3-months vs 18-months), reducing the potential for recall bias in the global assessment of change question used to select participants for analysis[25].

Cohen's effect size (Cohen 1988) and the Standardised Response Mean (SRM) (Stratford et al. 2005) were used to assess sensitivity to change and responsiveness. Both measures were based on the difference between the mean AUSCAN at baseline and follow-up but varied by the sample used to estimate variability in the data. Cohen's effect

[24] The wording for this question was "Compared to when you came for your assessment x months ago, how do you think your hand problem has changed?: "Completely recovered", "Much better", "Better", "No change", "Worse", "Much worse"

[25] Three versions of the global assessment of change question were included in the SMOotH study which referred to change in (1) 'your hand problem' (2) 'your hand pain' (3) 'your ability to use your hands'. The three versions of the measure were highly correlated so only the global assessment of change question referring to 'your hand problem' was used in the analysis in this chapter to simplify comparison to the CAS-HA study data (see Appendix 9)

size used the standard deviation of the AUSCAN at baseline; the SRM used the standard deviation of the AUSCAN change score (Box 6).

Box 6: Formulae to calculate sensitivity to change and responsiveness

$$\text{Cohen's effect size} = \frac{\bar{x}_{BL} - \bar{x}_{FU}}{\text{SD of BL score}}$$

$$\text{Standardised response mean (SRM)} = \frac{\bar{x}_{BL} - \bar{x}_{FU}}{\text{SD of (BL-FU) change score}}$$

$$\text{Guyatt's responsiveness ratio} = \frac{(\bar{x}_{BL} - \bar{x}_{FU})_{\text{in treated patients}}}{(\text{SD of (BL-FU) change score})_{\text{in control patients}}}$$

Where
BL = baseline, FU = follow-up
$\bar{x}_{BL}$=Mean score at baseline,
$\bar{x}_{FU}$=Mean score at follow-up
SD = standard deviation

Responsiveness was also assessed using Guyatt's responsiveness ratio (Box 6) (Guyatt et al. 1986). This ratio depends on a treatment and control group being identified in the data, which was not possible for participants in the CAS-HA study, and is less relevant for participants in the SMOotH study as no significant treatment effects for the AUSCAN were found in the trial. The formula for Guyatt's responsiveness ratio was therefore adapted in both studies by defining the control group as participants reporting no symptom change over time on the global assessment of change measure and the treated group were those reporting symptom change (improvement or deterioration) on the same measure. This approach was used as it has been considered acceptable in other responsiveness studies where a treatment and control group was not defined by the study design or where the treatment effect was not significant (Crosby et al. 2003, Sim et al. 2006).

Standard benchmarks for Cohen's effect size and the SRM were used to indicate the degree of responsiveness/sensitivity to change that had been achieved: "low": < 0.5; "medium": between 0.5 and 0.79; "high" > 0.8 (Cohen 1988). For Guyatt's responsiveness statistic, a cut-off of greater than one is proposed to indicate a responsive measure as a

value greater than one would indicate that the average change score in improving/deteriorating participants was bigger than the amount of variability observed in stable patients (van der Windt et al. 1998).

4.3.8 Interpretability

Interpretability is defined as the degree to which qualitative meaning can be assigned to quantitative scores (Terwee et al. 2007) and was assessed for the AUSCAN using the concept of minimum important change (MIC). MIC has been defined by Jaeschke et al (1989) as the smallest difference on the score of interest, that is perceived as beneficial by patients, that in the absence of side effects, would change the patient's management (Jaeschke et al. 1989)[26]. This concept adds to the interpretability of the AUSCAN as it enables one to interpret the magnitude of change in the data as being "large" or "small" independently of statistical significance and the sample size used in the study of interest.

Several methods have been proposed to calculate MIC including anchor-based and distribution-based methods. Anchor-based methods use an external criterion, or anchor, to determine what patients or clinicians consider important improvement/deterioration and relate this to change on the outcome measure for which MIC is derived (Wright et al. 2012). Distribution-based approaches, in contrast, are based on the distributional properties of the data or on the standard error of measurement as defined in Section 4.3.4 above (de Vet et al. 2007). Both methods are described below, along with how they were assessed for the AUSCAN in this thesis.

Anchor-based methods

The global assessment of change question (defined in Section 4.3.4) was used as an external anchor to explore whether suitable cut-offs could be identified for change scores

[26] Minimum important change (MIC) is often used interchangeably with minimum important difference (MID), however, De Vet et al 2007 make the distinction that minimum important change relates to within-person change, and minimum important difference to between-person differences. The term MIC is used throughout this thesis as the derivation for MIC uses change between two time-points as the focus of the analysis

on the AUSCAN, baseline to first follow-up, for hand pain and hand function, that best discriminated (1) any worsening in symptoms from those reporting no change, and (2) any improvement in symptoms from those reporting no change. The cut-offs represent values for MIC as they provide “best estimates” for the smallest value on each AUSCAN change score that would need to be observed before symptom change can be concluded. A precursor to this analysis, however, was to check that the correlations between the external anchor and the AUSCAN change scores were greater than 0.5, as relatively high correlations were needed for a reliable estimate of MIC to be obtained (de Vet et al. 2007).

The cut-offs were estimated using Receiver Operating Characteristic (ROC) curves to identify values on the AUSCAN change scores that minimised the quantity: ((1-sensitivity) + (1-specificity)) (de Vet et al. 2007) and minimised the rate of false positive and negative classifications between groups defined by the global assessment of change measure (see Appendix 10 for further details). The analyses were conducted separately for participants in the SMOotH and CAS-HA main studies and for worsening and improvement of symptoms as values for MIC could differ depending on the direction of change analysed in the data.

Distribution-based approaches

Several distribution-based methods have been proposed to define MIC including measures based on statistical significance, effect sizes and the standard error of measurement (SEM) (Crosby et al. 2003). Only the SEM is used in this chapter to estimate MIC as it has the preferred property that it is a characteristic of the measure, so is measured in units of the AUSCAN, rather than being based on specific characteristics of the sample likely to vary from study-to-study, i.e. the size of the sample for statistical significance approaches or the amount of variability in the data for effect size based approaches (Crosby et al. 2003).

Several cut-offs, based on multiples of the SEM, have been suggested in the literature to define MIC, e.g. 1*SEM or 1.96*SEM (Crosby et al. 2003), however some authors recommend that any cut-off used to define MIC should exceed the smallest detectable change (SDC) for the measure (de Vet et al. 2006a). The SDC for a measure is calculated as $1.96*\sqrt{2}*$SEM and represents the smallest within-person change in score that, with $p<0.05$, can be interpreted as 'real change' above measurement error[27] (Terwee et al. 2007). The SEM value derived from the CAS-HA pilot study in Section 4.3.4 was used to give a possible range of values for a distribution-based MIC.

Integrating anchor- and distribution-based approaches

Anchor-based and distribution-based methods can each be used to define MIC, however, each approach is limited. The first, because it does not take into account the measurement error in the outcome tool, and the second, because it does not consider whether the value for MIC is meaningful to patients, i.e. it is exploring minimum detectable change rather than minimum important change. In response, De Vet et al. 2007, proposed an integrated approach to determine minimum important change.

The integrated approach works by plotting participants' AUSCAN change scores for participants reporting "no change" on the global assessment question and calculating two limits on this distribution that are defined as: mean change score ± 1.645 SD of the change score[28]. The limits represent the two change scores that 90% of patients will lie between if they report no symptom change over the specified time period, with the upper limit defining the MIC for improvement, the lower limit, MIC for deterioration (De Vet el al.

[27] The SDC is calculated as the upper limit for a 95% confidence for a paired t-test i.e. 1.96* standard error of the difference score. The standard error of the difference score is equal to $\sqrt{2}*$SEM as it takes into account that there are two measures in the test-re-test study design that are measured with error (Ottenbacher et al. 1988).

[28] The limits of agreement used here are very similar to those proposed by Bland and Altman (Bland et al.1999), however, Bland and Altman suggest using 1.96 as a multiple of the SD of the change score (rather than 1.645) to define the limits that 95% of participants (rather than 90%) would lie between. The difference in approach can be compared to the difference between a 1- and 2-tailed test, with the approach by De Vet et al ensuring that 5% rather than 2.5% of participants have change scores that are above and below the upper and lower limits of interest.

2007). These limits were calculated using data from the CAS-HA pilot and main studies and the SMOotH clinical trial.

4.4 Results

The results in this chapter are presented using sub-headings to represent each assessment of reliability and validity described in Box 2 above.

4.4.1 Face and content validity

The description of the AUSCAN in Section 4.2 demonstrates that the AUSCAN has face validity as questions were developed in consultation with hand OA patients and their treating clinicians thus increasing relevance to this patient group (Bellamy et al. 2002a). Face validity of the questionnaire was also improved for the CAS-HA study by changing the reference from 'hand arthritis' to 'hand problem' and widening the item time frame to 'in the last week', hence making the questions more relevant to the CAS-HA sample.

Content validity was also supported as no further items were added to the questionnaire after patients were asked to generate items or domains they thought were missing from the closed-form version of the AUSCAN (Bellamy et al. 2002a). Also, since its development in 2002, the AUSCAN has been used in several other research studies (e.g. Bijsterbosch et al. 2011 and Stukstette et al. 2013) suggesting that other researchers support that items are clear and relevant for measuring hand pain and functional difficulty.

4.4.2 Item distribution (missing data rates and floor and ceiling effects)

The AUSCAN items had low levels of missing data, less than 6%, in the main CAS-HA cohort at baseline (Table 4-1). This suggests that items could be completed by most participants without difficulty. A hand pain score was derived for 589 participants at baseline (95%) suggesting that information could be obtained on hand pain for most participants; similarly for hand function N=593, 95%. Only a small percentage of participants had any items imputed when the scoring rules were applied (n=9 for both hand pain and function).

The percentage of respondents reporting no difficulty on each item ranged from 21 to 54% (Table 4-1), suggesting that some items were less able to discriminate between participants at the non-severe end of the response distribution. However, after summation of items into subscales, only 11 and 13% of participants had a total score of zero for hand pain and function respectively (Figure 4). This was less than the upper limit, proposed by Terwee et al. 2007, for a scale to demonstrate good discriminative ability, i.e. 15%. The percentage of respondents reporting extreme problems on the AUSCAN items was low, less than 6%, suggesting good discriminative ability at the upper end of the response distribution. Consequently, this resulted in skewed distributions for both AUSCAN pain and function (Figure 4).

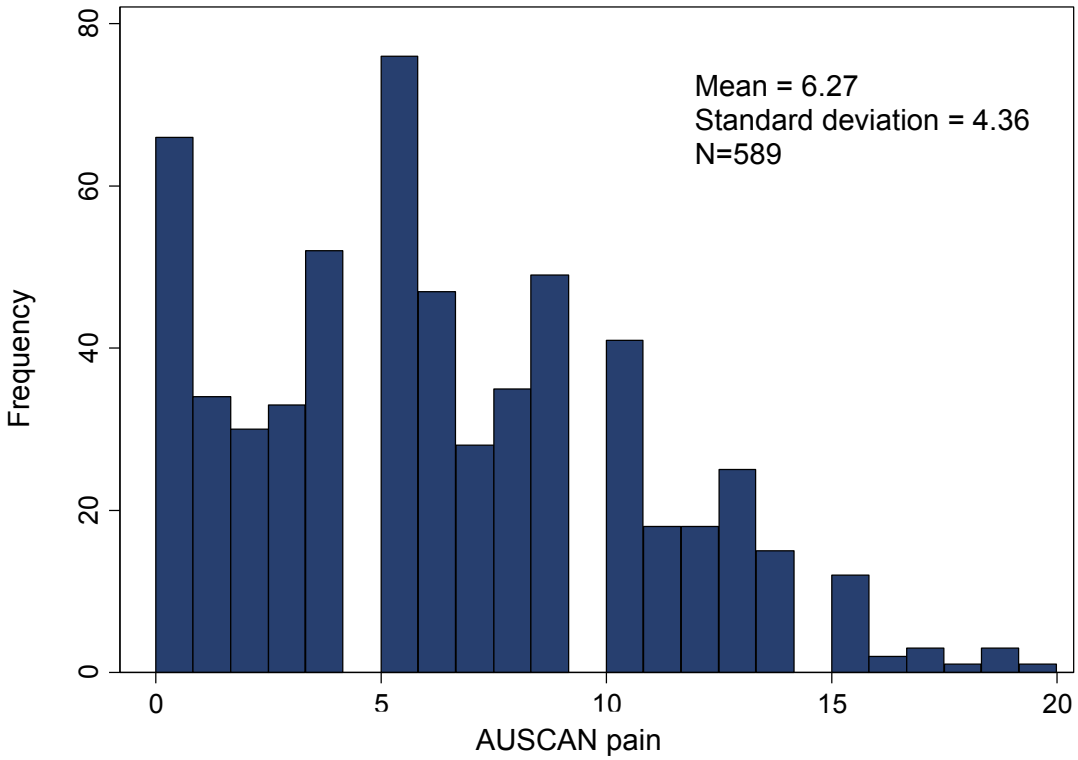
Table 4-1: Item-level AUSCAN data for participants in the CAS-HA baseline survey (N=623)

Subscale/item	% missing	%[α] reporting no problem	%[α] reporting extreme problem	Item-total correlation
Pain subscale				
Hand pain in the last week when…..				
At rest	6	32	1	0.65
Gripping objects	6	21	1	0.86
Lifting objects	6	30	1	0.86
Turning objects	6	27	3	0.85
Squeezing objects	5	21	2	0.87
Function subscale				
Difficulty in the last week with……				
Turning taps	5	54	1	0.83
Turning a round door-knob or handle	5	52	1	0.86
Doing up buttons	5	53	2	0.80
Fastening jewellery	5	40	4	0.83
Opening a new jar	5	25	5	0.82
Carrying a full pot with one hand (e.g. saucepan)	5	30	5	0.86
Peeling vegetables/fruit	5	49	2	0.86
Picking up large heavy objects	5	27	5	0.84
Wringing out wash clothes (e.g. squeezing a wet sponge or flannel)	5	30	4	0.85

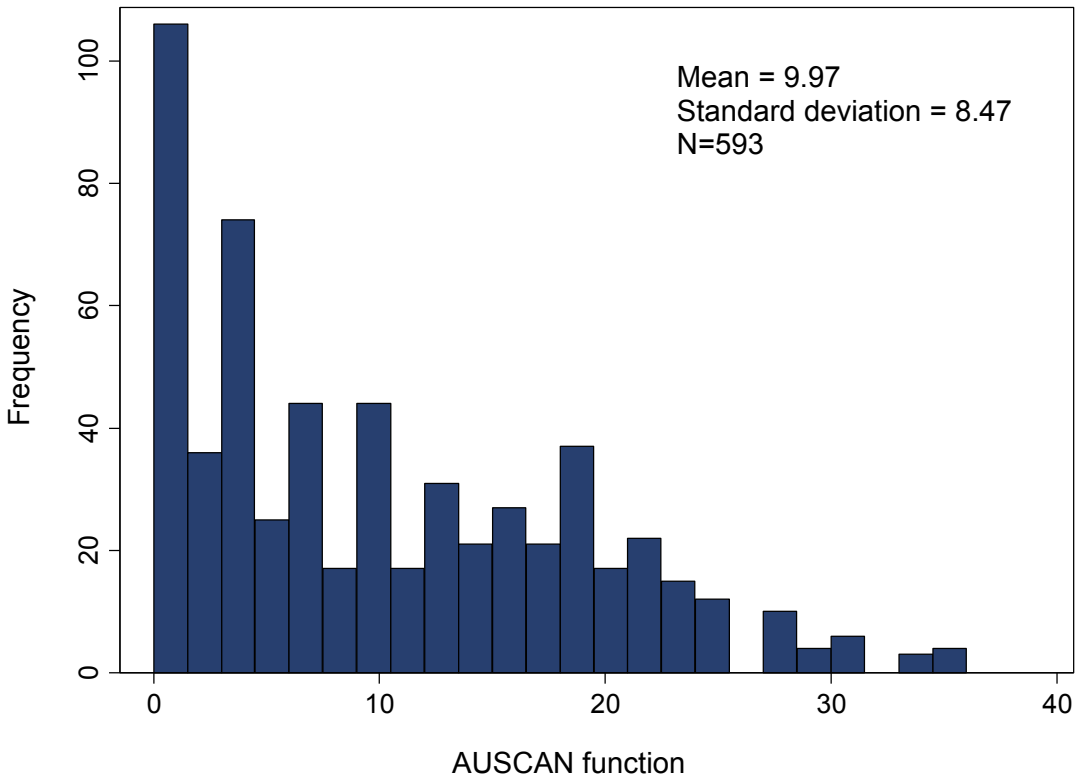
[α] Percentage denominator excludes missing data

Figure 4: Distribution of AUSCAN pain and function subscales for participants in the baseline CAS-HA cohort

a) AUSCAN pain



b) AUSCAN function



4.4.3 Internal consistency

The goodness of fit statistics for the CFA model were TLI = 0.99, CFI = 0.99, SRMR[29] = 0.03 and RMSEA = 0.100 (90% confidence interval = 0.096 to 0.112), which suggested that the model was a reasonable fit to the data; only the RMSEA did not meet the pre-specified criteria for good model fit. All standardised regression coefficients were greater than 0.4 and statistically significant (Table 4-2). Variance estimates in the model were positive and the estimated correlation between hand pain and function was plausible, i.e. 0.82.

Table 4-2: Confirmatory factor analysis of the AUSCAN hand pain and function items and item-total correlations (N=601)

Subscale/item	Standardised regression weight	95% confidence interval	Item-total correlation
Pain subscale			
Hand pain in the last week when…			
At rest	0.73	(0.68, 0.77)	0.65
Gripping objects	0.92	(0.90, 0.93)	0.86
Lifting objects	0.94	(0.93, 0.95)	0.86
Turning objects	0.93	(0.92, 0.94)	0.85
Squeezing objects	0.95	(0.94, 0.97)	0.87
Function subscale			
Difficulty in the last week with……			
Turning taps	0.93	(0.92, 0.95)	0.83
Turning a round door-knob or handle	0.95	(0.94, 0.97)	0.86
Doing up buttons	0.89	(0.87, 0.92)	0.80
Fastening jewellery	0.89	(0.87, 0.91)	0.83
Opening a new jar	0.88	(0.86, 0.90)	0.82
Carrying a full pot with one hand (e.g. saucepan)	0.93	(0.91, 0.94)	0.86
Peeling vegetables/fruit	0.91	(0.90, 0.93)	0.86
Picking up large heavy objects	0.92	(0.91, 0.94)	0.84
Wringing out wash clothes (e.g. squeezing a wet sponge or flannel)	0.91	(0.89, 0.93)	0.85

[29] The SRMR could not be calculated in the presence of missing data when an ordinal outcome was assumed so this statistic is reported on the data set after missing data have been excluded

Modification indices were still explored, however, to see if any correlations could be added to the model that would greatly improve model fit from the reasonable fit that had been originally obtained. Three were identified that were substantially larger than other modification indices in the model and all referred to correlations between items on the hand function subscale: "turning on taps" with "turning a round door-knob or handle", modification index = 147; "Carrying a full pot with one hand" with "Picking up large heavy objects", modification index = 143 and "Doing up buttons" with "Fastening jewellery", modification index = 111; the remaining modification indices ranged from 0 to 35. By adding these three additional correlations to the model a good model fit was achieved: TLI = 0.99, CFI = 0.99, SRMR = 0.02 and RMSEA = 0.06 (90% confidence interval = 0.052 to 0.069).

Item-total correlations were also calculated as a satisfactory CFA had been obtained[30] and all were >0.8, with the exception of "hand pain at rest" at 0.67 (Table 4-2). Good internal consistency was also shown by high Cronbach's alphas for both hand pain (0.93) and hand function (0.96), with both values at the upper end of the limits defined to determine acceptable internal consistency.

4.4.4 Reproducibility

An ICC correlation of 0.88 and 0.87 was obtained for hand pain and hand function respectively and was calculated using data from 41 participants reporting no symptom change in the CAS-HA pilot study. Both ICC values were greater than 0.7 suggesting good reproducibility of the AUSCAN in this data set (Table 4-3). The standard error of measurement was 1.51 and 3.07 for pain and function respectively (Table 4-3) and the Bland and Altman plots showed a broad scatter of data points for hand pain and function (Figure 5), suggesting that the discrepancy between measurement time-points did not

[30] A satisfactory CFA was concluded even though three correlation terms were added to the initial CFA. This was because all of the additional correlations were added within the hand function items and they did not cross over into items on the hand pain scale

depend on participants' average level of hand pain or functional difficulty[31]. The standard deviation of the AUSCAN scores were also similar between the pilot and main CAS-HA studies suggesting that reliability estimates in the CAS-HA pilot study can generalise to the main data set (AUSCAN pain SD_{pilot} = 4.7, SD_{main} = 4.4; AUSCAN function SD_{pilot} = 9.2, SD_{main} = 8.5).

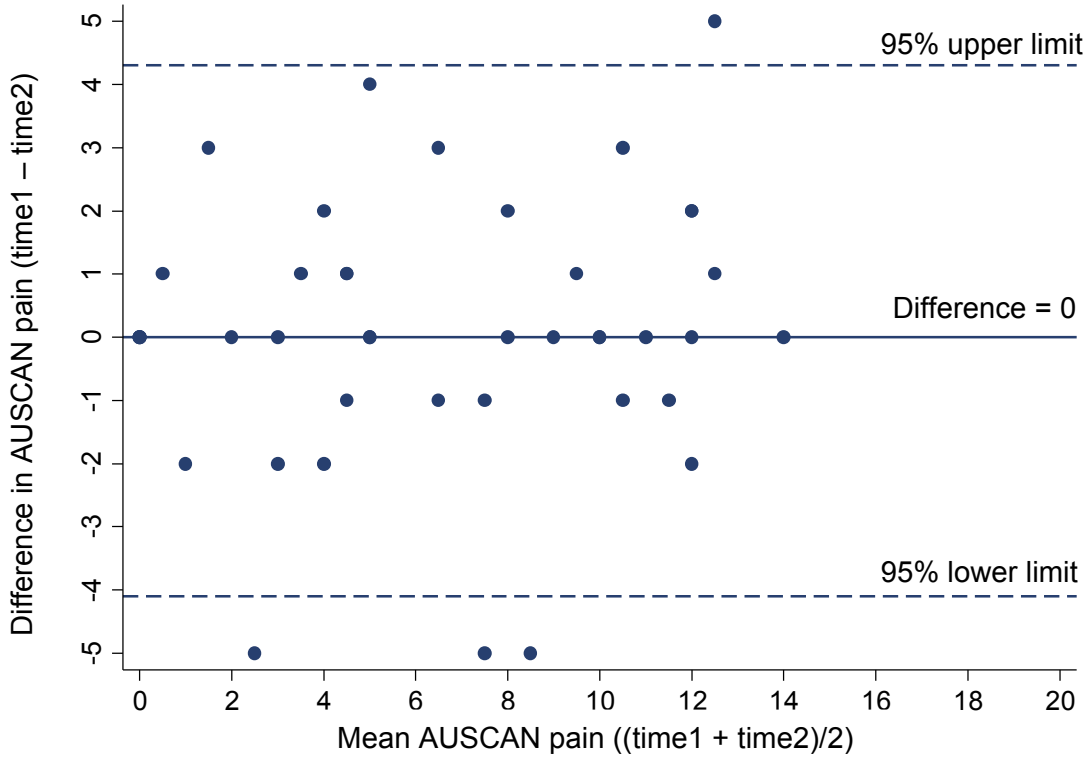
Table 4-3: Variance components, standard error of measurement and intra-class correlation (ICC) of the AUSCAN hand pain and function measures (N=41)

	AUSCAN pain	AUSCAN function
Variance between time points (σ^t)	-0.05	-0.11
Variance between subjects (σ^s)	16.19	64.57
Error variance (σ^e)	2.32	9.55
Standard error of measurement (SEM) (for agreement) = $\sqrt{(\sigma^t + \sigma^e)}$	1.51	3.07
ICC2 (A,1) = $\sigma^s / \sigma^s + \sigma^t + \sigma^e$	0.88	0.87
ICC2 95% confidence interval	(0.78, 0.93)	(0.77, 0.93)

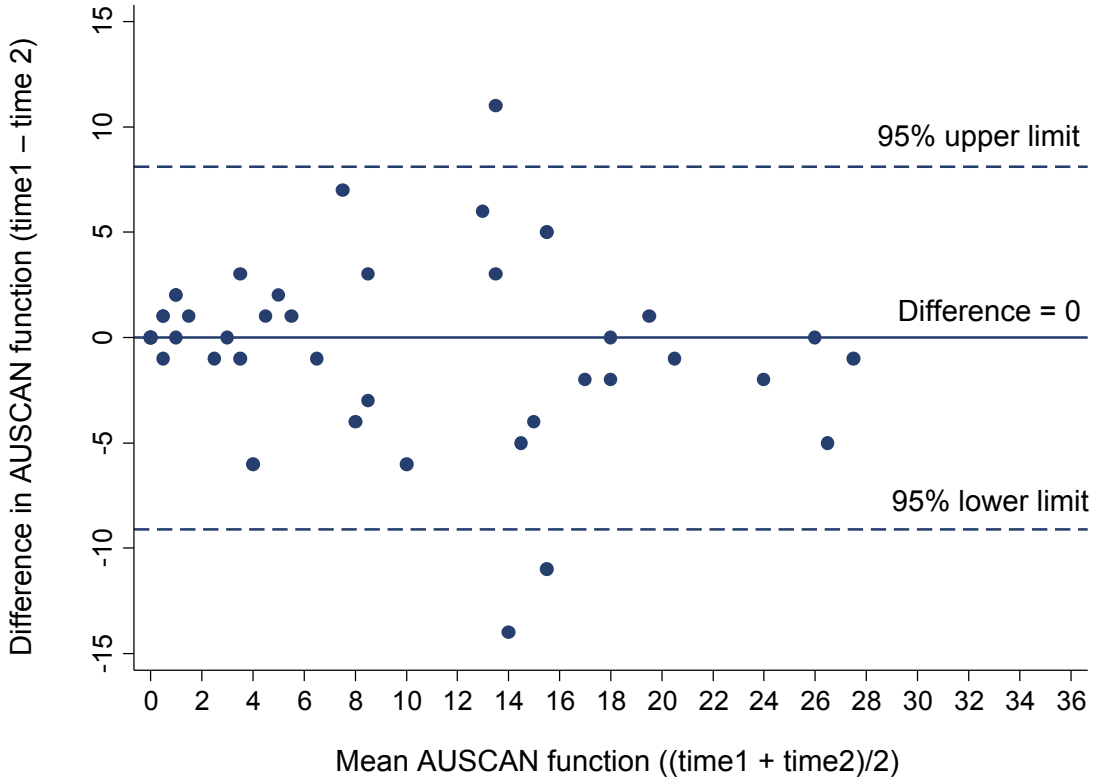
[31] It is noted that the discrepancy between the two time points is larger around a mean difference of fifteen for AUSCAN function, but the reliability of this finding is difficult to confirm as the sample size is small (N=41).

Figure 5: Bland and Altman plots for AUSCAN pain and function

a) AUSCAN pain



b) AUSCAN function



4.4.5 Construct validity

The AUSCAN pain and function subscales were positively correlated with all outcomes they were hypothesised to be related to; correlations ranged from 0.49 to 0.82 (Table 4-4). The AUSCAN hand function measure was more highly correlated with a second self-reported measure of hand function, i.e. the AIMS2 hand and finger function subscale, than the objective measures of grip and pinch strength and the grip-ability test. The average level of AUSCAN pain and function was higher for participants that reported a GP consultation in the previous 12-months on the survey questionnaire. These findings support the construct validity of the AUSCAN as a measure of hand pain and functional ability.

Table 4-4: Construct validity tests for the AUSCAN hand pain and function measures (maximum N= 593)

	Spearman's rank Correlation	Mean (SD)	Median (IQR)
AUSCAN pain			
Pain severity in the last month	0.61	-	-
AIMS2 pain subscale	0.76	-	-
Consequences subscale of the IPQ-R	0.58	-	-
GP consultation for a hand problem in the last year			
Yes (n =150)	-	8.3 (4.9)	9.0 (5.0, 12.0)
No (n = 431)	-	5.5 (3.9)	5.0 (3.0, 8.0)
Hand pain in the last month			
Yes (n = 414)	-	7.4 (4.2)	7.0 (4.8, 10.0)
No (n =135)	-	3.1 (3.1)	3.0 (0.0, 5.0)
AUSCAN function			
Grip strength	-0.54		
Pinch strength	-0.51		
Grip-ability test	0.49		
AIMS2 hand and finger function subscale	0.82		
Consequences subscale of the IPQ-R	0.58		
GP consultation for a hand problem in the last year			
Yes (n =149)		13.8 (9.7)	14.0 (5.0, 21.5)
No (n = 434)		8.6 (7.5)	7.0 (2.0, 14.0)

All correlations and subgroup comparisons were significant at $p<0.001$. AIMS2 = Arthritis impact Measurement Scales 2. IPQ-R = Revised illness perceptions questionnaire, SD = Standard deviation, IQR = Inter-quartile range

4.4.6 Sensitivity to change and responsiveness

The proportion of participants reporting symptom deterioration was greater in the CAS-HA study than the SMOotH trial (42% vs 17%). In contrast, the proportion of participants reporting symptom improvement was higher in the SMOotH trial than the CAS-HA main study (32% vs 16%), with study differences potentially explained by involvement of a treatment intervention or by the length of study follow-up. Despite such differences in the rate of symptom change over time, both studies show low-medium responsiveness for both hand pain and function when assessed using Cohen's effect size and the standardised response mean; a finding that occurs even when participants are selected out to be those reporting symptom improvement or deterioration on the global assessment of change question[32] (Table 4-5). Guyatt's responsiveness ratio provides similar estimates of responsiveness to those using the two other methods, i.e. Cohen's effect size and the standardised response mean, with values below the cut-off of one for responsiveness.

[32] The SRM is possibly preferred over the effect size for this analysis as the change score for the AUSCAN is more likely to be normally distributed than the absolute measure at baseline, which is skewed (Streiner 2003). It has the disadvantage, however, that it requires data to be present on two occasions, but this is less relevant here, as the majority of participants have also completed the AUSCAN at the 18-month follow-up

Table 4-5: Sensitivity to change and responsiveness statistics applied to the main CAS-HA cohort and SMOotH trial data

	Sensitivity to change	Responsiveness	
Data sample	All participants	Participants reporting symptom improvement[α]	Participants reporting symptom deterioration[β]
CAS-HA main study			
AUSCAN pain			
N	546	88	228
Cohen's effect size	-0.06	0.62	-0.38
Standardised response mean	-0.07	0.62	-0.45
Guyatt's responsiveness ratio	Not calculated	0.68	-0.46
AUSCAN function			
N	550	90	230
Cohen's effect size	-0.05	0.37	-0.27
Standardised response mean	-0.07	0.47	-0.41
Guyatt's responsiveness ratio	Not calculated	0.45	-0.41
SMOotH clinical trial			
AUSCAN pain			
N	225	72	39
Cohen's effect size	0.04	0.38	-0.26
Standardised response mean	0.05	0.37	-0.44
Guyatt's responsiveness ratio	Not calculated	0.42	-0.36
AUSCAN function			
N	226	73	39
Cohen's effect size	-0.08	0.21	-0.29
Standardised response mean	-0.12	0.26	-0.55
Guyatt's responsiveness ratio	Not calculated	0.33	-0.53

α Defined as "better", "much better" or "completely recovered" on the global assessment of change question

β Defined as "worse" or "much worse" on the global assessment of change question

4.4.7 Interpretability

Anchor-based approach to determining minimum important change

Table 4-6 shows response frequencies for the global assessment of change question and highlights that the proportion of participants reporting extreme changes, i.e. that they are 'completely better', or 'much worse' on the global assessment of change question, is low (<10%).

Table 4-6: Distribution of the global assessment of change question

Response	CAS-HA main study N(%)	SMOotH clinical trial N(%)
Completely recovered	9 (2)	0 (0)
Much better	36 (6)	17 (8)
Better	47 (8)	57 (25)
No change	246 (42)	115 (50)
Worse	233 (40)	36 (16)
Much worse	15 (3)	3 (1)

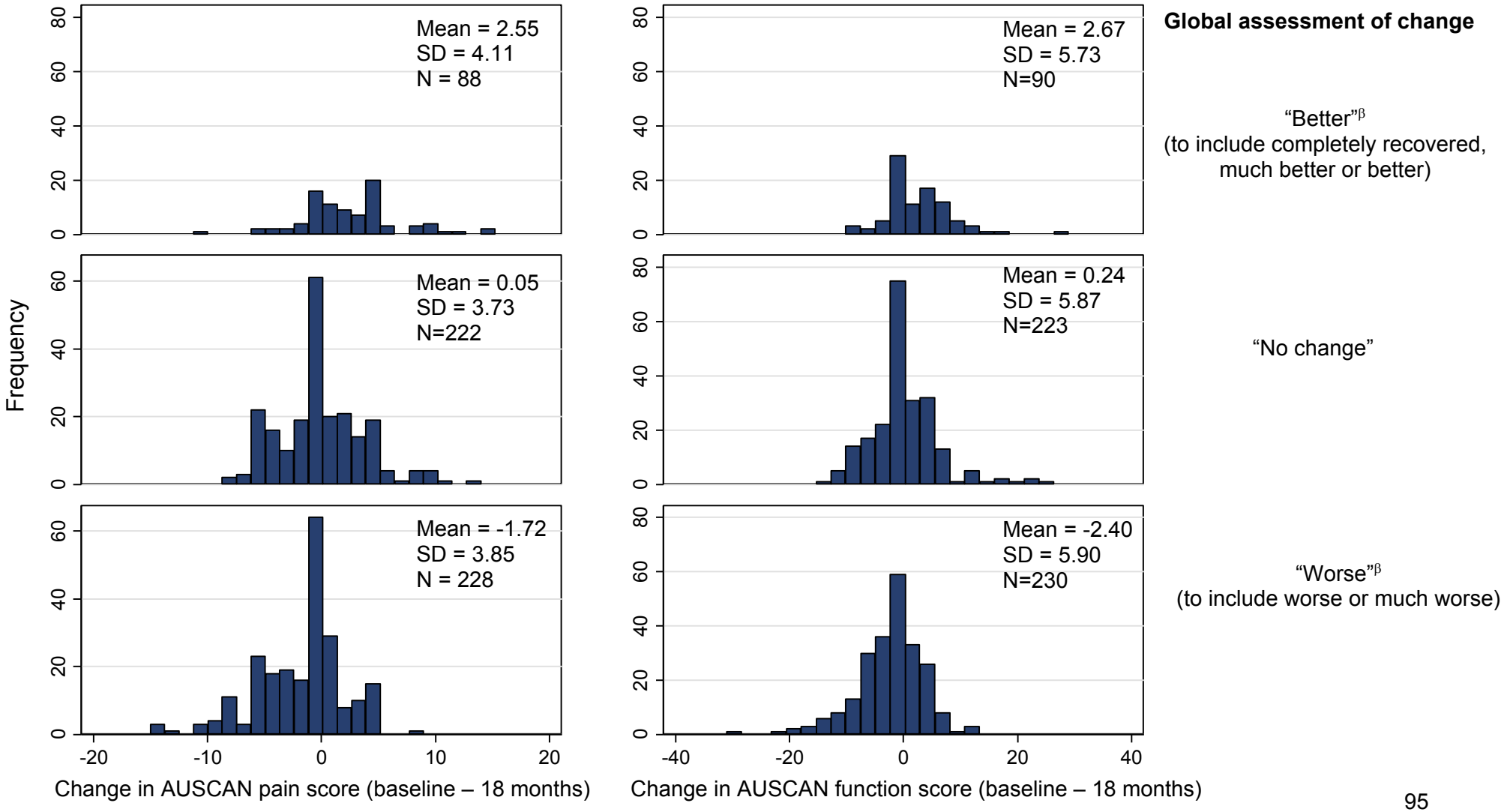
The Spearman's rank correlation between the global assessment of change question and the AUSCAN change score was low and the degree of distribution overlap when the AUSCAN change scores were plotted by response to the global assessment of change question was high (see Table 4-7, Figure 6 and Figure 7). This was not ideal as a high correlation between these two variables was needed for the global assessment of change question to be used as an external anchor to define MIC. In addition, the area-under-the-ROC-curve statistics were all < 0.7 suggesting lower rates of sensitivity and specificity than desirable (Table 4-8). Using an anchor-based approach to define MIC in these datasets was therefore not feasible so this analysis was not taken further in this thesis.

Table 4-7: Correlation between the global assessment of change question and the AUSCAN pain and function change scores

Data sample	CAS-HA N=543	SMOotH N=224
Change in AUSCAN outcome		
Pain (baseline – first follow-up)	-0.34	-0.25
Function (baseline – first follow-up)	-0.30	-0.25

Figures are Spearman's rank correlations
Global assessment of change question measured using six response categories

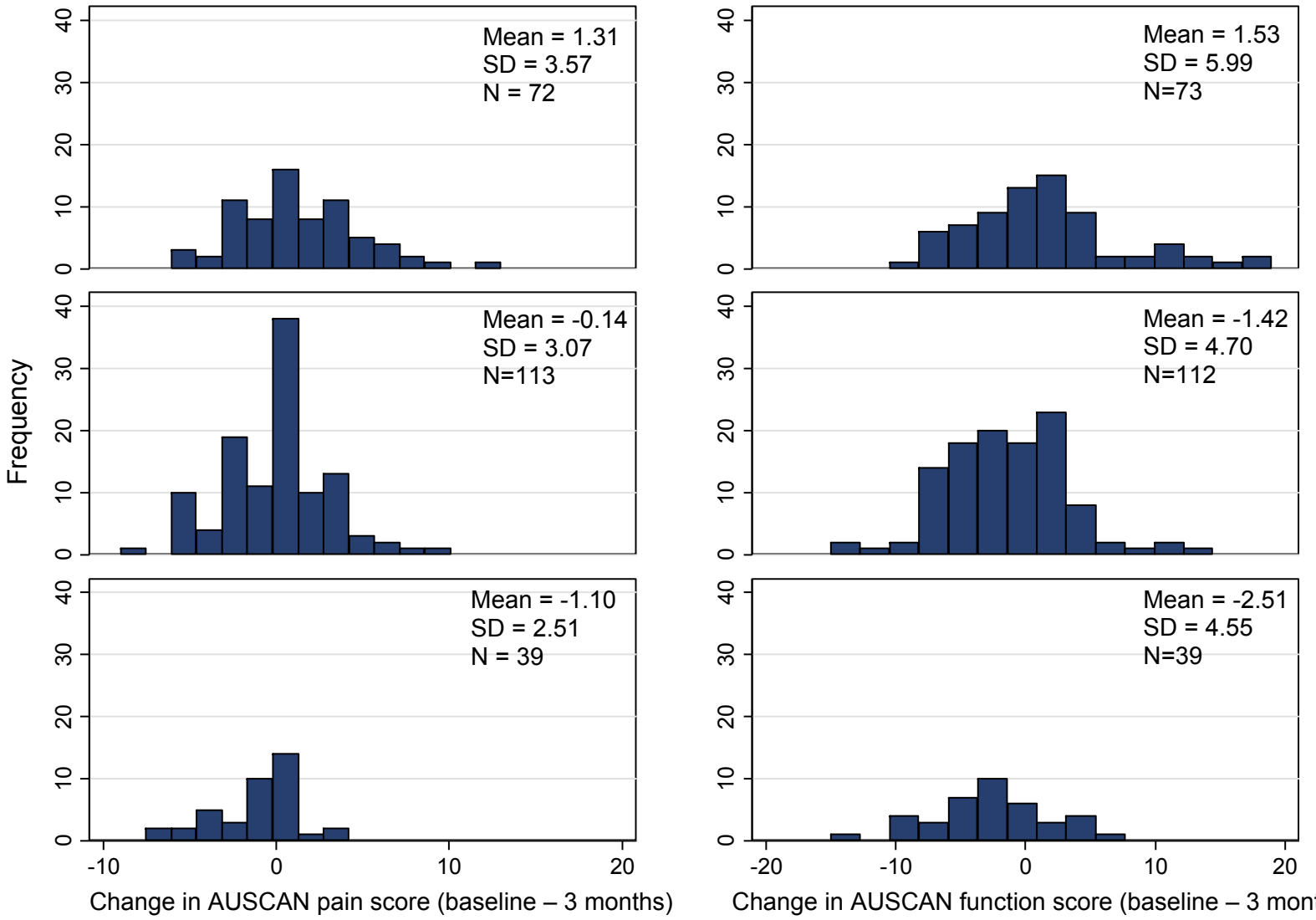
Figure 6: AUSCAN pain and function change scores by global assessment of change question (CAS-HA study)[α]



α = Overall, the mean (standard deviation) of the change score was -0.27 (4.1) for hand pain and -0.43 (6.2) for hand function

β = Category merging undertaken due to small N. SD = Standard deviation

Figure 7: AUSCAN pain and function change scores by global assessment of change question (SMOotH clinical trial)[α]



Global assessment of change

"Better"[β]
(to include completely recovered, much better or better)

"No change"

"Worse"[β]
(to include worse or much worse)

α = Overall, the mean (standard deviation) of the change score was 0.16 (3.3) for hand pain and -0.64 (5.3) for hand function
β = Category merging undertaken due to small N. SD = Standard deviation

Table 4-8: Area under the ROC curves (95% confidence intervals)

Data sample	CAS-HA N=543		SMOotH N=224	
	Symptom improvement$^{\alpha}$	Symptom deterioration$^{\beta}$	Symptom improvement$^{\alpha}$	Symptom deterioration$^{\beta}$
AUSCAN pain	0.69 (0.62, 0.75)	0.61 (0.56, 0.66)	0.61 (0.52, 0.69)	0.58 (0.49, 0.68)
AUSCAN function	0.64 (0.57, 0.70)	0.61 (0.56, 0.66)	0.64 (0.55, 0.72)	0.56 (0.45, 0.66)

α Symptom improvement includes those that are 'completely better', 'much better' 'better' on the global assessment of change question

β Symptom deterioration includes those that are 'worse' or 'much worse' on the global question

Distribution–based approaches

A wide range of possible values for MIC were calculated using the distribution-based method and the SEM estimates reported in Section 4.4.4 (Table 4-9). The maximum value for minimum important change was 4.19 for hand pain and 8.51 for hand function with changes larger than these amounts being greater than the smallest detectable change in the data.

Table 4-9: Minimum important change using distribution-based methods

Data sample	CAS-HA pilot study	
	AUSCAN pain N=41	AUSCAN function N=41
1*SEM	1.51	3.07
1.96*SEM	2.96	6.02
SDC = $1.96*\sqrt{2}*$SEM	4.19	8.51

SEM = standard error of measurement, SDC = smallest detectable change

Integrated anchor- and distribution-based approaches

The integrated approach to calculating MIC suggested that MIC for improvement ranged from 3.64 to 6.19 for AUSCAN pain and from 6.3 to 9.9 for AUSCAN function across the CAS-HA and SMOotH studies (Table 4-10). As the mean change score for hand pain and function is around zero in both data sets, MIC for deterioration is similar to that observed

for improvement, with the exception of AUSCAN function measured in the SMOotH clinical trial.

Table 4-10: Limits of agreement defined using the integrated anchor and distribution-based approach

	N[α]	Mean change score[α]	SD of change score[α]	MIC for deterioration[α,β]	MIC for improvement[α,γ]
AUSCAN pain					
CAS-HA pilot study	41	0.10	2.15	-3.44	3.64
CAS-HA main study	222	0.05	3.73	-6.09	6.19
SMOotH clinical trial	113	-0.14	3.07	-5.19	4.91
AUSCAN function					
CAS-HA pilot study	41	-0.49	4.37	-7.68	6.70
CAS-HA main study	223	0.24	5.87	-9.42	9.90
SMOotH clinical trial	112	-1.42	4.70	-9.15	6.31

α = Defined only for participants reporting no symptom change on the global assessment of change question, β = Mean change score - 1.645 SD of the change score, γ = Mean change score + 1.645 SD of the change score. SD = standard deviation, MIC = Minimum important change

4.5 Discussion

4.5.1 Summary of the key findings

In this chapter, data have been presented that support, in the majority, that the AUSCAN is a valid and reliable measure of hand pain and functional difficulty when adapted for use in community-dwelling older adults with hand pain. Specifically, support has been shown for the face and content validity of the questionnaire, the internal consistency of the two subscale scores for pain and function and their overall construct validity and reliability. In contrast, weaker evidence has been shown for the measures' reproducibility when assessed using the SEM, and its responsiveness, cumulating in relatively large values for MIC being estimated in the data. This finding, however, should be viewed in light of other

studies that have explored the psychometric properties of the AUSCAN and in light of some of the key limitations of the data used for this analysis (see Section 4.5.2 and 4.5.3 below for details).

4.5.2 Comparison with the literature

There are several other studies that have explored the psychometric properties of the AUSCAN, both in populations similar to CAS-HA (Dziedzic et al. 2007a) and in other clinical populations such as those with RA (Massy-Westropp et al. 2004), hand OA (Bellamy et al. 2002b, Allen et al. 2006b, Slatkowsky-Christensen et al. 2005, MacDermid et al. 2007, Wittoek et al. 2009, Moe et al. 2010, Moon et al. 2012), and in community-based samples including people with and without hand pain (Allen et al. 2007, Arreguin Reyes et al. 2012). Despite differences in study inclusion criteria, language, culture and versions of the AUSCAN used[33], all studies support that the AUSCAN has face and content validity although the magnitude of any floor and ceiling effects for the measure can depend on the population of interest (Dziedzic et al. 2007a, Massy-Westropp et al. 2004). These studies also show that the AUSCAN is internally consistent and has good construct validity across all studies, which is consistent with the findings in this study[34].

The repeatability of the AUSCAN has been assessed in a smaller number of studies administering the AUSCAN on two occasions[35] (Bellamy et al. 2002b, Massy-Westropp et al. 2004, Dziedzic et al. 2007a, Moon et al. 2012, Moe et al. 2010, Arreguin Reyes et al. 2012). All studies showed that the AUSCAN was a reliable measure of hand pain and function, with the exception of one study by Moon et al (2012) who found lower ICC values of 0.46 for pain and 0.67 for function. In the CAS-HA data, the AUSCAN had ICC

[33] All studies used either the VAS or the Likert version of the AUSCAN as reported in Bellamy et al (Bellamy et al. 2002a) except for Dziedzic et al 2007a that used the adapted version of the AUSCAN as reported in this chapter

[34] In the study by Allen et al. 2007 the factor structure of the AUSCAN was explored by ethnic origin, Caucasian or African-American. The factor model was only supported in the Caucasian dataset, but, as 99% of participants in CAS-HA report their ethnic origin as 'White' this study gives less concern that the factor structure is not supported in the CAS-HA data

[35] The study by Arreguin Reyes et al. 2012 is not included in this summary as the ICC was only calculated for a total AUSCAN score, rather than for the separate subscales of hand pain and function

values > 0.7 so findings in this chapter are in line with the majority of evidence from other studies. Of these studies, only two reported Bland and Altman limits of agreement (Massy-Westropp et al. 2004, Moe et al. 2010), with only one reporting upper limits for hand pain and function separately, i.e. 1.06 and 0.80 for hand pain and function respectively when measured on a 0-4 scale (Moe et al. 2010). These limits equate to 27% and 20% of the response scales, which are similar to the Bland and Altman upper limits derived in this chapter: 22% and 23% for hand pain and function, respectively.

Evidence for the responsiveness of the AUSCAN differs greatly depending on study setting and on how patients with known symptom change are defined for analysis. For example, two studies have shown that the AUSCAN has “medium” to “high” responsiveness to change before and after non-steroidal anti-inflammatory (NSAID) treatment (Bellamy et al. 2002b, Moon et al. 2012), whereas only “low” to “medium” responsiveness was shown in a study where CRx-102 was the effective treatment compared to placebo for hand OA (Haugen et al. 2009). In this thesis, responsiveness was assessed using a global assessment of change question, which has its limitations that are discussed further in Section 4.5.3. Nevertheless, this analysis still contributes further to the evidence base that highlights that the responsiveness of the AUSCAN is dependent on the setting used to test it.

Across studies providing evidence on the psychometric properties of the AUSCAN there is limited information on a value for MIC for the hand pain and function subscales. One study, using data collected across several countries, reported MIC values of 1.49 and 1.25 for hand pain and function when measured on a 0-20 and 0-36 scale, respectively (Bellamy 2007). Although these figures are roughly in line with evidence presented by Allen et al. (2006a), who report that a one unit change in AUSCAN function relates to clinically important changes[36] of around 1.3kgs for grip strength and 0.2 kg for pinch

[36] Clinically important changes were assessed by the typical forces need to open various containers rather than as self-report by patients

strength, the evidence from these two studies greatly differs from the analysis in this chapter using the integrated method to calculate MIC. The evidence surrounding a single value to recommend for MIC for hand pain and function therefore remains inconclusive.

4.5.3 Strengths and limitations

A key strength of this analysis is that the majority of analyses are based on relatively large sample sizes, with the sample size for the CAS-HA study exceeding most used in other studies exploring the psychometric properties of the AUSCAN. It is acknowledged, however, that both the repeatability analysis, and the distribution method for calculating MIC were limited as they could only be tested using data from the CAS-HA pilot study, hence the sample size was small, though of similar size to other studies published on the repeatability of the AUSCAN (sample sizes ranged from 17 to 51 participants). The analysis in this chapter is strengthened by including data on sensitivity to change/responsiveness, and by attempting to define a value for MIC, as fewer studies have explored these specific psychometric properties of the measure. However, these analyses have their limitations as discussed further below.

Limitations of the responsiveness analysis

A global assessment of change question was used in the responsiveness analysis to define groups of participants with known symptom change over time. This measure was used as it would be difficult, and potentially unreliable, to define a cut-off on an external measure for "improving" and "deteriorating" symptoms. In addition, the global assessment of change question may encourage patients to think about change over the whole time course rather than just at two selected time-points and may be able to incorporate if participants' views of symptom severity change over time (response shift)[37] (Streiner 2003)

[37] Response shift could occur in the CAS-HA study if a participant were to experience an episode of severe pain between the baseline and 18-month follow-up. The participants' recalled baseline

The global assessment of change measure, however, could be prone to recall bias, as current events are easier to recall than those in the past (Wright et al. 2012). This is particularly demonstrated in a study by Kamper et al. (2010), which showed that for patients with musculoskeletal disorders, the global assessment of change measure was more highly correlated with outcomes collected at follow-up than at baseline. This is especially an issue in the CAS-HA study as participants are recalling symptom change over a long time-period (18-months), and in both studies, CAS-HA and SMOotH, there is evidence that AUSCAN pain and function at the first follow-up time period is more highly correlated with the global assessment of change measure than at baseline (see Appendix 11). It is therefore not known whether the groups defined as improving or deteriorating over time did indeed experience true symptom change so this may explain why the AUSCAN does not show high levels of responsiveness in this analysis.

To explore the limitation of using a global assessment of change question for this analysis, the responsiveness of three additional measures of hand pain or function were assessed in the CAS-HA data. The aim of this analysis was to explore if the responsiveness of these measures were also low, as this would provide potential evidence that it was the framework used to test responsiveness that was problematic, rather than the AUSCAN *per se*. The results of this analysis showed low responsiveness for all measures when assessed in participants reporting deterioration in symptoms, yet medium to good responsiveness in participants reporting symptom improvement (Appendix 7 and Appendix 12). The results of this test were therefore inconclusive. The results also did not depend on the statistic that was used to assess responsiveness, i.e. Cohen's effect size, SRM or Guyatt's responsiveness statistic, so it remains unclear whether the low responsiveness of the AUSCAN is due to bias in the global assessment of change question.

---

level of pain severity at follow-up may then be less severe than previously rated as the participant is now more aware of the full range of symptom severity that could be experienced.

A further reason why the AUSCAN shows low-moderate responsiveness in the CAS-HA and SMOotH studies could be because the actual of amount of change over time is small even in participants reporting symptom improvement or deterioration over time. This is in part supported in the data as the mean rate of change is small in the improving and deteriorating groups relative to the scale that the measures are calculated on (Figure 6, Figure 7, Appendix 13, and Appendix 14). There is also a large degree of overlap in the distribution of change scores in the AUSCAN and the additional measures of hand pain and function between participants reporting no change and deterioration in symptoms over time. This highlights a key limitation of any assessment of responsiveness: that it is highly dependent on the setting and population used to test it and that any measure will be more responsive to detect larger, than smaller, changes over time (Streiner 2003), and potentially explains why the AUSCAN is shown to be responsive in other treatment studies when there is a greater potential for symptom change to occur and therefore be detected by the AUSCAN. The data in this thesis may therefore not be very suitable to assess questions around the responsiveness of the AUSCAN as the analysis is limited by the difficulty of defining data subsets where “true” symptom change has occurred over time.

Limitations of the analysis to calculate MIC

As discussed above, the limitations associated with the global assessment of change measure also apply when the global assessment of change question is used in the integrated method to calculate MIC. Also, as the distribution-based approach is based on a small sample size it could be that these results are unreliable, however it is noted that results from this approach are similar to those using the integrated approached that was not based on a small sample size. It may therefore simply be that hand pain and function need to change by a large amount before any impact is observed in daily living, i.e. by around 25% on the scale score for both hand pain and function. However this finding is

not in line with the study by Bellamy et al (2007) who report MIC values that are considerably lower than those presented in this chapter.

This highlights a key criticism of the MIC approach, that it has the potential to generate a wide range of possible values across different studies and across the different methods used to calculate it, e.g. distribution vs anchor-based approaches, or when using a different measure as an external anchor (Terwee et al. 2010). It is therefore recommended that values for minimum important change only be implemented when multiple analyses, often in separate studies, converge to a single value. This requirement therefore means that evidence on MIC in this chapter is tentative and requires further replication in future studies.

A further limitation of the MIC analysis presented is that it does not take into account the relative improvement in hand pain and function that has occurred over time, for example, a 1 point change may be considered large for a participant with a baseline score of 2, but small for a baseline score for 20, so a 50% versus 5% change for an identical value of MIC. It would be possible to explore whether a minimum cut-off for percentage change could be defined, but this was not taken forward given the lack of an appropriate anchor to use in the anchor-based method to calculate MIC.

Limitations of the distribution of the AUSCAN

The skewed distributions of the AUSCAN in the CAS-HA sample could potentially be a limitation to monitoring symptom change over time to address the research questions in Chapter 1 as a sizeable group of participants have AUSCAN scores at baseline of zero so have no ability to improve at the first follow-up time point. This may, or may not, be a limitation however, as it is plausible that participants genuinely have no pain or functional difficulty in the week prior to baseline, as participants were only recruited into the study if they had hand pain or problems in the last 12-months, rather than if they had problems on the day they completed the baseline questionnaire.

4.6 Summary

In this chapter, the AUSCAN was shown to be a reliable and valid measure of hand pain and functional difficulty. It is acknowledged, however, that evidence for the responsiveness of the measure was not fully demonstrated in the CAS-HA data hence this is a limitation to be considered when interpreting the analyses reported in Chapters 6 to 9. This limitation however is made acknowledging that the framework to test responsiveness in the CAS-HA study was not ideal and that longer-term symptom change may be more easily detected in the data if this is larger than the initial change observed between the first two study time-points. A value for MIC has not been determined in the data for hand pain and function so further studies are needed before a single value can be recommended. Alternative methods to judge the magnitude of change in hand pain and function over time are therefore needed and are explored in Chapter 6 after the statistical methods used to model the course of hand pain and function over time have been presented in the next chapter (Chapter 5).

5 Statistical methodology

The aim of this chapter is to describe the statistical methods that are used in this thesis to model the longitudinal data collected from the CAS-HA study, namely the key outcomes of hand pain and function[38]. The methods chosen are those suitable to model longitudinal data, i.e. methods that can appropriately account for the lack of independence in data collected from single participants at multiple time points (Kahn 2011). The methods are described in relation to a continuous outcome measure only, as this is the form of measurement used for hand pain and functional difficulty in this thesis.

The methods used will be discussed in the following sections:

- Generalised Estimating Equations (GEE) (Section 5.1)
- Growth models (GM) and parallel process GM (Section 5.2)
- Latent class growth models (LCGM) (Section 5.3)
- Growth mixture models (GMM) and parallel process GMM (Section 5.4)

A final section will be included to discuss issues that are generic to all of the statistical models above, namely, the computer software used to fit the models of interest (Section 5.5.1), issues of sample size and power (Section 5.5.2) and how to handle missing data (Section 5.5.3).

5.1 Generalised Estimating Equations (GEE)

5.1.1 The basic concept

GEE models were initially developed by Liang and Kung-Lee (Liang et al 1986) and are referred to in the literature as “population average” or “marginal” models as their aim is to estimate the *mean* outcome observed for participants based on any covariates included in

[38] The aim of this chapter is to focus on only those methods that have been used in this thesis to address the research questions in Chapter 1 rather than to give a full and comprehensive guide to all possible approaches to analysing longitudinal data.

the model (Ballinger 2004). They work by extending the standard generalised linear model (GLM) to data that lack independence between time points by treating any correlation between the data points as a "nuisance" parameter to be adjusted out of the data prior to estimating effects of interest (Hu et al. 1998a, Goldstein et al. 2004).

The standard GLM is shown in Box 7 and illustrates that, after an appropriate adjustment for the correlation in the data, GEE models can be applied to a wide range of outcome types. The model illustrates that the expectation of the outcome (the mean) is related to a linear combination of the predictors of interest, but this is achieved via a link function, g, to transform the data when they do not follow a normal distribution. If the data follow a normal distribution, the GEE model can be simplified to: $E(y_{it}) = X_{it}\beta$ with $y \sim N(\,)$ and is fitted to the data after standard errors have been adjusted for the correlation between the data points.

Box 7: The generalised linear model (StataCorp 2013a))

The generalised linear model is defined for an outcome y and covariates x as:

$$g(E(y_{it})) = X_{it}\beta \qquad y \sim F \; with \; parameters \; \theta_{it}$$

where g( ) is the link function, F is the distributional family, $X_{it}\beta$ is the linear predictor of model covariates and i = 1,…..,m and t = 1,……,n_i , where i indexes the individual and t the repeated observations per individual (n_i is the number of repeated observations for individual i)

It was shown, however, in Chapter 4, that both AUSCAN pain and function have skewed distributions. An alternative distributional assumption may therefore be needed, with the gamma distribution being one possible option to model data that are skewed (Azuero et al. 2010). With a gamma distribution, several link functions are possible to relate the mean

to the linear predictor – the identity link (which is equal to 1), log, power or reciprocal links (StataCorp 2013a). The choice of link function is selected as the one that provides best fit to the data (Collett 1999).

For a GEE model to be fitted to the data, along with the outcome distribution and the link function, a working correlation structure for the data also needs to be specified. The working correlation structure represents the within-group correlation in the data (this is the correlation between time points for longitudinal data) and can take on one of several potential forms including: independence, exchangeable, m-dependent, auto-regressive and unstructured (see Box 8 (Twisk 2003)). The working correlation structure is often chosen as the one that parsimoniously reproduces the observed correlation matrix (Twisk 2003), however the GEE model has been shown to be robust to misspecification of the correlation matrix when robust standard errors are used. The choice of the "correct" correlation structure is therefore less crucial as only small gains in model efficiency are achieved when a "correct" structure is chosen[39] (Ballinger 2004).

[39] This is because the model estimation process is iterative in nature, so the initial estimates from the correlation matrix are overwritten during the model fitting process (Ballinger 2004)

Box 8: Working correlation structures for a GEE model

The examples below relate to the CAS-HA study i.e. a study with 5 data collection time points labelled t1 to t5 where ρ = correlation

Independent structure

	t1	t2	t3	t4	t5
t1	-	0	0	0	0
t2	0	-	0	0	0
t3	0	0	-	0	0
t4	0	0	0	-	0
t5	0	0	0	0	-

Correlation between time points assumed to be zero

Exchangeable structure/Compound symmetry

	t1	t2	t3	t4	t5
t1	-	ρ	ρ	ρ	ρ
t2	ρ	-	ρ	ρ	ρ
t3	ρ	ρ	-	ρ	ρ
t4	ρ	ρ	ρ	-	ρ
t5	ρ	ρ	ρ	ρ	-

Common correlation assumed between time points

m-dependent structure (e.g. m=2)

	t1	t2	t3	t4	t5
t1	-	ρ_1	ρ_2	0	0
t2	ρ_1	-	ρ_1	ρ_2	0
t3	ρ_2	ρ_1	-	ρ_1	ρ_2
t4	0	ρ_2	ρ_1	-	ρ_1
t5	0	0	ρ_2	ρ_1	-

Common correlation assumed but dependent on the time distance between data points

Autoregressive structure

	t1	t2	t3	t4	t5
t1	-	ρ^1	ρ^2	ρ^3	ρ^4
t2	ρ^1	-	ρ^1	ρ^2	ρ^3
t3	ρ^2	ρ^1	-	ρ^1	ρ^2
t4	ρ^3	ρ^2	ρ^1	-	ρ^1
t5	ρ^4	ρ^3	ρ^2	ρ^1	-

Correlation t measures apart is ρ to the power of t e.g. 2 measures apart is ρ squared

Unstructured

	t1	t2	t3	t4	t5
t1	-	ρ_1	ρ_2	ρ_3	ρ_4
t2	ρ_1	-	ρ_5	ρ_6	ρ_7
t3	ρ_2	ρ_5	-	ρ_8	ρ_9
t4	ρ_3	ρ_6	ρ_8	-	ρ_{10}
t5	ρ_4	ρ_7	ρ_9	ρ_{10}	-

All correlations between time points assumed to be different

5.1.2 Model estimation and model assumptions

Parameter estimates and their associated standard errors are estimated using quasi-maximum likelihood techniques (Twisk 2003) (for further details see (Liang et al. 1986) and (Heyde 1997)). Briefly, this is an iterative model fitting procedure that is used when there is insufficient information to enable a likelihood function to be constructed, and then maximised, for parameter estimates to be obtained using maximum likelihood (Jiang 2004). A key assumption of the GEE model is that the variance is related to the mean via a known function (Ghisletta et al. 2004), e.g. relating the mean and variance in a normal or gamma distribution as shown in Box 9.

Box 9: Relationship between the mean and variance for a normal and gamma distribution

Distribution	Mean	Variance
Normal $(\mu,\ \sigma^2)$	μ	$E[(X-\ \mu)^2]$
Gamma *(a, b)*[40]	$\frac{a}{b}$	$\frac{a}{b^2}$

In addition, it is assumed that there is a linear relationship between the dependent variable and the linear predictor (via an appropriate link function) (Ghisletta et al 2004) and that participants are sampled at random from a larger population of interest (i.e. that the data are independent when collected from different participants (Kuchibhatla et al. 2003)).

5.1.3 Model interpretation

The parameter estimates from the GEE model are interpreted as the degree to which the *mean* response, given the link function, would change for a one-unit increase in a covariate across the population (Ballinger 2004), i.e. their interpretation is at the

[40] For the gamma distribution, a is the shape parameter and b is the inverse scale parameter to govern the shape of the gamma distribution derived

population rather than at the individual level. In this thesis, the statistical significance of the β-values in the model are estimated using the Wald chi-square statistic, which is distributed as a chi-square test statistic with degrees of freedom equal to the number of parameters being tested (Ballinger 2004)

5.1.4 Coding of time

Time (t) can be modelled in several ways within the GEE framework: as a linear variable (where it is assumed that the change in response is equal when measured for equally-spaced time points), as a non-linear variable (using higher order polynomials (e.g. t^2, t^3) or as a categorical variable when the assumption of linear growth may not be tenable e.g. when rates of change between earlier follow-up time points are hypothesised to be quicker than between later data collection points). In the CAS-HA study, the highest order of polynomial that can be fitted to the data is cubic as the study only has five data collection time points. The most appropriate order of polynomial (linear, quadratic, cubic) will therefore be tested by assessing the statistical significance of each higher order polynomial and including only those that are statistically significant ($p<0.05$).

5.1.5 Model predictors

Model predictors can be included in the GEE models either as continuous or categorical variables and can be measured as time-invariant or time-varying. Time-invariant predictors are those that have a constant value over time (either by definition, or because they are only measured at a single time point), whereas time-varying predictors are those whose value can (potentially) change between time-points (Curran et al. 2010). Interactions between predictors are possible (e.g. to explore whether trajectory shape differs between population subgroups). In this thesis, GEE models are only used to explore the trajectory of the outcome over time so a further description of model prediction is reserved for the growth models described in Section 5.2.

5.1.6 Model fit

GEE model fit cannot be assessed using information-based model fit statistics (e.g. Akaike's information criteria (AIC) (Akaike 1974)) as these statistics can only be calculated for models estimated using full information, rather than quasi, maximum likelihood techniques. However, to address this, an adjusted version of the AIC has been developed, the Quasi-likelihood Independence model information Criterion (QIC) (Cui 2007), that can be used to compare the fit of two competing models of interest. Models with a lower score on the QIC would represent an improved model fit.

In addition, model residuals (i.e. the difference between the observed and predicted values (Everitt et al. 2005)) can be used to assess GEE model fit, with a plot of the residuals against fitted values at each time point aiming to show residuals that do not cluster around an individual value and that have a random pattern that does not systematically change in each time-period. The residuals can also be used to assess the data for any potential outliers that could influence model estimates derived (Ballinger 2004).

5.1.7 Reflections on the GEE model

The GEE model that has been presented so far provides a valid method to analyse the longitudinal data in this thesis and are useful when marginal effects at the population level are of interest (Kuchibhatla et al. 2003). However, the GEE model treats the correlation between time points as a "nuisance parameter" so its ability to provide information on how trajectories differ between participants at the individual level is limited (Kuchibhatla et al. 2003). For this, growth models are needed as described in Section 5.2.

5.2 Growth models

5.2.1 The basic concept

Growth models are referred to by a variety of the names in the literature (e.g. random effect models, random coefficient models, mixed models, multi-level models, hierarchical

models, latent growth models (LGM), latent trajectory models (LTM), and latent curve models (LCM)) (Byrne et al. 2003, Berrington et al. 2006, Curran et al. 2003); with the name used often dependent on the framework used to model the data[41]. For simplicity they will be referred to as growth models throughout this thesis.

Growth models have a similar aim to GEE models in that they both aim to explore the relationship between a set of predictor variables and a longitudinal outcome of interest. Their main difference, however, is the way in which they account for the lack of independence in data when collected at multiple time-points (Hu et al. 1998a). In growth models, the lack of independence between time-points is addressed by estimating an average trajectory curve for the population as a whole, but then, in addition, estimating the amount of individual variation that has occurred around this average curve (Twisk 2003).

In the context of a linear growth model, a single average trajectory curve would be estimated for the sample as a whole by a mean intercept (initial level) and mean slope (rate of growth or decline) (Andruff et al. 2009). These terms in the model are then referred to as the fixed effects as they represent a single value that is identical (i.e. fixed) for all participants in the sample (Curran et al. 2010). In addition to the fixed effects, information on the probability distribution around each fixed effect would also be estimated as the variance of the individual trajectories around the group mean intercept and slope (Curran et al. 2010). These variance estimates would then be referred to as the random effects of the model and can be specified to relate to the fixed intercept term, or to the fixed intercept and slope in the model[42]. It is when the fixed and random effects are taken together that the growth curve modelling approach is described (Curran et al. 2010)

[41] Growth models can be fitted in a multi-level or structural equation modelling framework (see Section 5.2.2 for details) When a structural equation modelling framework is used, common model terms are LGM, LTM or LCM, to highlight that growth is modelled using latent variables, (Byrne et al. 2003). The remaining names are more commonly used when a multi-level framework for analysis is used

[42] It would be inefficient to fit separate regression lines for each individual in the sample (each with their own intercept and slope) as this would use too many degrees of freedom. Estimating a small number of variance parameters is preferred

5.2.2 Frameworks for growth curve modelling

Growth models can be fitted using either a multi-level or structural equation modelling approach (Hox et al. 2005) as described below:

Multi-level approach

Multi-level modelling is often used to analyse data that have a hierarchical or nested data structure, e.g. children nested within classes at school, or patients nested within general practitioners (Curran et al. 2010). It can therefore be used to model longitudinal data as data collected at each time point can be considered to be nested within individuals (Curran et al. 2010).

The multi-level formulae for a linear growth model with a random intercept and slope is shown in Box 10 and illustrates that the model is multi-level (i.e. is split into two parts, level 1 and 2). The level 1 part of the model represents the fixed effects, and shows that the outcome y (measured for an individual i, at time point t) is represented as a function of an intercept term (α_i) multiplied by a constant term equal to one ($\lambda_{\alpha t}$), a slope term (β_i) multiplied by the time of measurement ($\lambda_{\beta t}$) and an error term (ε_{it}). The level 2 part of the model is used to define the random effects, with the level 1 intercept further defined to include a mean intercept (μ_α), and a random effect term ($\zeta_{\alpha i}$) that represents the random departure from the mean intercept that is specific for each individual. A similar format is used for the slope term, with μ_β representing the overall mean slope and $\zeta_{\beta i}$ representing the random effect for the slope (Hox et al 2005)[43]. The equivalent model is also illustrated pictorially in Figure 8.

[43] The level 1 part of the model is referred to as the "within-person" regression as it represents individual change in the outcome over time. In contrast, the level 2 part of the model is the "between person" model as it focusses on inter-individual differences in change in the outcome (Byrne et al. 2003)

Box 10: Linear growth model with random intercept and slope – the multi-level framework (Bollen et al. 2006)

Level 1 – basic model

$$y_{it} = \alpha_i \lambda_{\alpha t} + \beta_i \lambda_{\beta t} + \varepsilon_{it}$$

where α_i = intercept, β_i = linear component, $\lambda_{\alpha t}$ = constant equal to one, $\lambda_{\beta t}$ = time of assessment, ε = residual error

Level 2 - incorporating the random effects

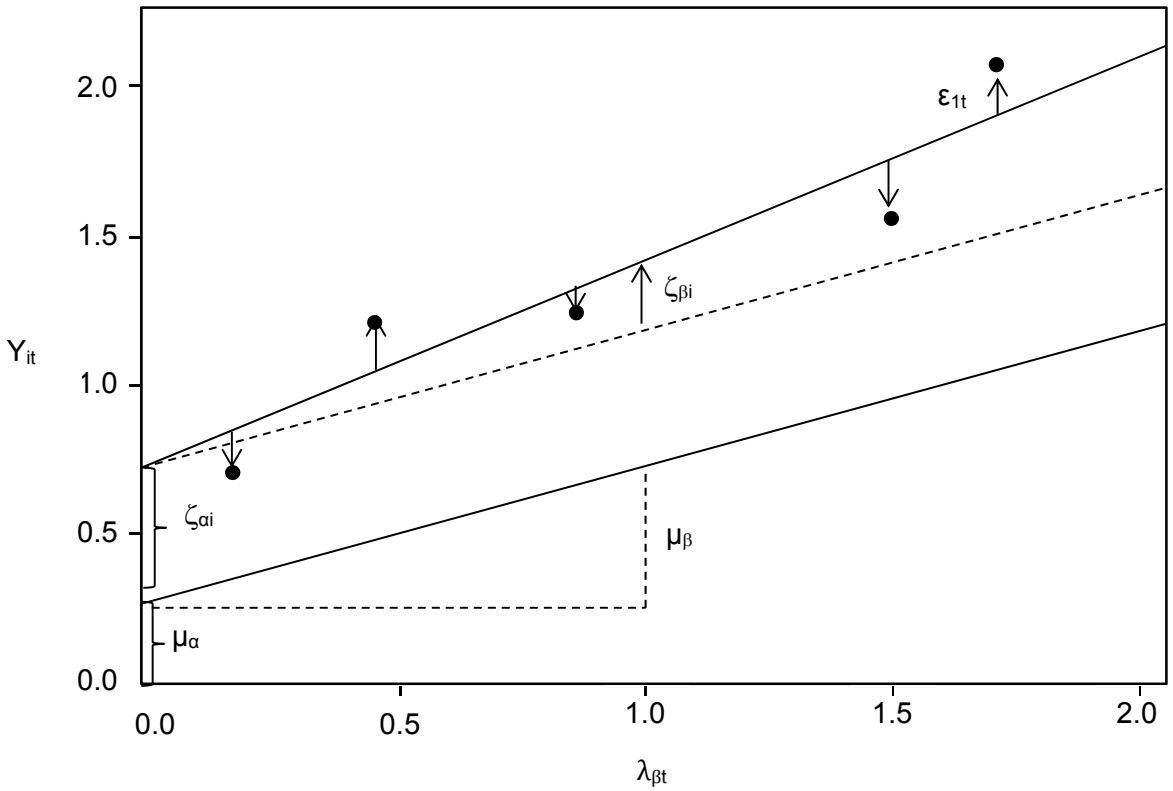
$$\alpha_i = \mu_\alpha + \zeta_{\alpha i}$$

$$\beta_i = \mu_\beta + \zeta_{\beta i}$$

The variances of $\zeta_{\alpha i}$ and $\zeta_{\beta i}$ and their covariance are represented by: $\sum_\zeta = \begin{bmatrix} \sigma_\alpha^2 & \sigma_{\alpha\beta}^2 \\ \sigma_{\alpha\beta}^2 & \sigma_\beta^2 \end{bmatrix}$

and the variance of ε_{it} is σ_ε^2

Figure 8: Linear growth model with random intercept and slope



Reproduced from (StataCorp 2013a)[44]

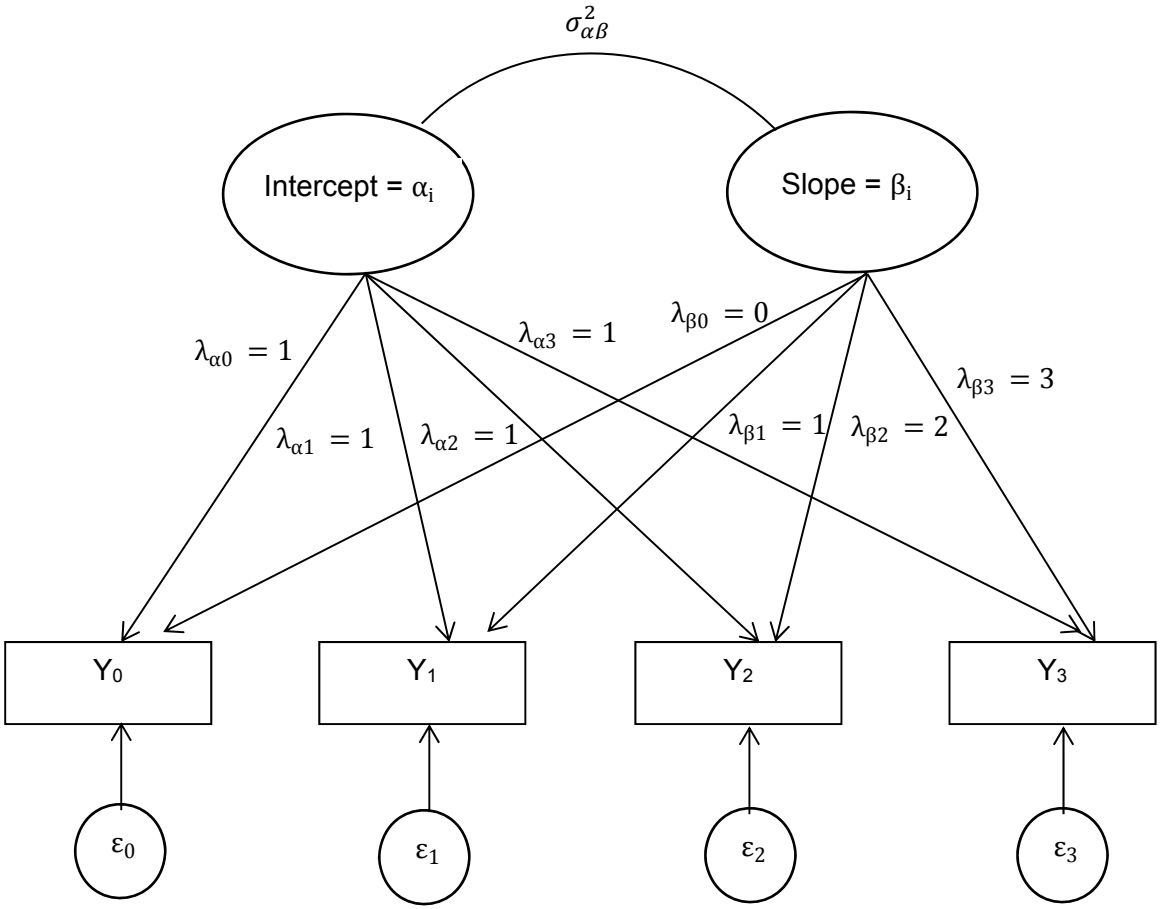
Structural Equation Modelling approach

Structural Equation Modelling is a framework for modelling the mean and covariance structure of a data set and includes within it the option of modelling both measured and unmeasured (latent[45]) variables (Byrne 2001). It lends itself to the modelling of longitudinal data as the random effects in the model can be considered as unobserved continuous latent variables that describe the underlying, unobserved growth trajectory that is marked (or indicated) by the observed repeated measures (Jung et al. 2008, Curran et al. 2003). The structural equation modelling approach to modelling a linear growth model with a random intercept and slope is shown in Figure 9 using notation comparable to that given in Box 10.

[44] This figure was reproduced by permission of the publisher and authors, S. Rabe-Hesketh and A. Skrondal, "Multilevel and Longitudinal Modeling Using Stata, 2nd Edition" (College Station, TX: Stata Press, 2008), fig. 4.4, copyright 2008 by StataCorp LP. (StataCorp 2013a)

[45] A latent variable is one that is not directly observed, but inferred from other measured variables (Byrne 2001)

Figure 9: Linear growth model fitted in a structural equation modelling framework for a longitudinal study with four time points of interest (Hox et al. 2005)



Footnote: Circles (or ellipses) represent unobserved factors (latent variables), squares (or rectangles) represent observed variables, single-headed arrows represent the impact of one variable on another, double-headed arrows represent co-variances or correlations between pairs of variables (Byrne & Crombie, 2003) and the notation in this diagram follows that given in Box 10. In this diagram, the mean of the latent intercept and slope pooled over all participants in the sample represent the model fixed effects; the random effects are characterised by the variance of each latent factor (Curran & Hussong, 2003). Intercept parameters are fixed to 1 to represent that the intercept is constant over time, whereas parameters for the slope term are fixed to 0,1,2,3 to represent linear growth for four equally spaced time-points.

Although the SEM and multi-level model frameworks use different approaches to modelling the longitudinal data[46], they produce identical results when the mean and covariance structure of the latent variables in a SEM analysis are specified to correspond to the fixed and random effects in the multi-level analysis (Wang et al. 2007, Curran et al. 2010, Hox et al. 2005). The main difference between the frameworks lies in the ease with which they can be extended to address more complex research questions of interest (Hox et al. 2005), for example the multi-level approach can be more easily extended to data that have a complex hierarchical structure (e.g. when time points are nested within participants, who are nested within general practices and geographical regions) (Curran et al. 2003), whereas the SEM approach is useful for models that include a combination of categorical and continuous latent variables (e.g. latent class growth curve analysis) (Wang et al. 2007).

In this thesis, the multi-level approach will therefore be used to fit the single process growth models (to gain experience of using the multilevel modelling commands in the STATA software), whereas, the SEM approach will be used for all other models (to gain experience of using the Mplus software, and also for ease of application).

5.2.3 Model estimation

Unlike the GEE models described in Section 5.1, growth models can be estimated using (full information) maximum likelihood (ML) techniques (either ML or restricted ML (REML)). ML will be used in this thesis as it has been shown in large samples that the difference between ML and REML is negligible[47] (Rabe-Hesketh et al. 2008). In addition, as the outcomes in this thesis are skewed (see graphs in Chapter 4) the models are

[46] Multi-level modelling is often considered a uni-variate approach as "time" is measured as a single predictor variable in the model (Hox et al. 2005), whereas SEM uses a multi-variate approach as "time" is not modelled by a single variable but by constraining the factor loadings for the latent variables to represent the particular growth curve of interest (Wang et al. 2007)

[47] Both REML and ML each have advantages: REML tends to produce more reliable estimates of variance, whereas ML tends to produce more reliable estimates of the parameters (Twisk 2003). ML was chosen over REML, however, as it allows model fit to be compared using a likelihood ratio test for nested model comparisons which cannot be done when REML is used (Rabe-Hesketh et al. 2008) – see Section 5.2.7 for details of the likelihood ratio test

estimated using robust standard errors (i.e. MLR – maximum likelihood with robust standard errors) to account for the skewed distribution of the outcome of interest[48] (Muthen et al. 2010).

5.2.4 Coding of time

Coding of time in the growth models is identical to that described for the GEE model in Section 5.1, with both models able to handle time points with equal and non-equally spaced intervals, e.g. time could be coded as 0, 1, 2 and 3 for a study measured at yearly intervals over 3-years, or as 0, 0.5, 1, and 3 for a study measured at baseline, 6-months, 1- and 3-years respectively (Andruff et al. 2009). In this thesis, for the growth models, the most appropriate coding for time will be initially informed by plotting the mean outcome data at each time point, as prior to analysis, no evidence exists as to the most appropriate trajectory shape to model, an approach recommended by Curran et al. (2010).

If higher order polynomials are added to a growth model, they can be added as a fixed effect only (by adding terms to the level 1 of the model only) or as a fixed and random effect (to explore whether including such additional random effects improves model fit). This is illustrated in Box 11 where a quadratic term is added to the model as both a fixed and random effect (the equivalent formulation in a SEM framework is shown in Appendix 15).

[48] Transforming the distribution of the outcomes to approximate a normal distribution (e.g. by using a log or square root transformation) could also be an option to deal with a skewed outcome. It is not used in this chapter as AUSCAN function in particular has a highly skewed distribution so it is unlikely that a suitable transformation could be found. Interpreting effect estimates on a revised scale is also more complex than if the original units of the scale are preserved

Box 11: Formulae for a quadratic growth model with a random intercept and slope (Bollen et al. 2006)

Level 1 – the basic model

$$y_{it} = \alpha_i \lambda_{\alpha t} + \beta_{1i}\lambda_{\beta t} + \beta_{2i}\lambda_{\beta t}^{2} + \varepsilon_{it}$$

where α_i = intercept, β_{1i} = linear component, β_{2i} = quadratic component, $\lambda_{\alpha t}$ = constant equal to one, $\lambda_{\beta t}$ = time of assessment, ε = residual error

Level 2 – incorporating the random effects

$$\alpha_i = \mu_\alpha + \zeta_{\alpha i}$$

$$\beta_{1i} = \mu_{\beta 1} + \zeta_{\beta 1 i}$$

$$\beta_{2i} = \mu_{\beta 2} + \zeta_{\beta 2 i}$$

5.2.5 Model predictors

Growth models (like GEE models) can include within them predictors that are time-invariant or time-varying. Only predictors measured at baseline are considered in this thesis so are assumed to be time-invariant. Such time-invariant predictors can be added to a growth model either as a predictor of the model intercept or as a predictor of the model slope (or both) as illustrated in Box 12 (and Appendix 16). The predictors can be measured on a continuous scale or as a categorical variable (Andruff et al. 2009), (with dummy variables needed to model the latter) as no assumptions are made about the distribution of the predictors of interest (Curran et al. 2003).

Box 12: Formulae for a linear growth model with a random intercept and slope and including a time invariant predictor of the model intercept and slope (Bollen et al. 2006, Hox et al. 2005)

Level 1 – the basic model

$$y_{it} = \alpha_i \lambda_{\alpha t} + \beta_i \lambda_{\beta t} + \varepsilon_{it}$$

Where α_i = intercept, β_i = linear component, $\lambda_{\alpha t}$ = constant equal to one, $\lambda_{\beta t}$ = time of assessment, ε = residual error

Level 2 – incorporating the random effects and including a time invariant predictor

$$\alpha_i = \mu_\alpha + \gamma_1 Z_i + \zeta_{\alpha i}$$

$$\beta_i = \mu_\beta + \gamma_2 Z_i + \zeta_{\beta i}$$

Where γ_1 and γ_2 are the effects of the time-invariant predictor (Z_i) on the intercept and slope

5.2.6 Model interpretation

The terms γ_1 and γ_2 in the above models are interpreted as the change in the average intercept and average slope respectively for a one-point change in the model predictor and are equivalent to fitting a model with the predictor in it and an interaction with time. To aid interpretation of the model, some authors have suggested that centering variables prior to analysis is useful, i.e. subtracting the mean from the independent variable prior to analysis. The key benefit of this is that the model intercept can then be interpreted as the predicted value for a subject with average values for each independent variable in the analysis, which is often more relevant for variables where a value of zero is implausible

(e.g. age in the CAS-HA study would never be 0), but also this is thought to improve model convergence when more complex models are fitted (Rasbash et al. 2012).

5.2.7 Selection of model predictors

Model selection techniques established for linear regression can be used to select important predictors in growth models, i.e. backward selection, forward selection, or backwards/forwards stepwise selection (Altman 1999), however, their application becomes more complex for models that contain both fixed and random effects due to the increased number of combinations of parameters that can be included/excluded from the model. They also require a careful model fitting process that ensures that the number of participants included in each of the models compared is identical (Singer et al. 2003) and hence cannot simply be automated.

A key issue when applying forward/backward model selection techniques is how to determine whether a predictor of interest significantly improves the predictive ability of the model. The likelihood ratio test is commonly used for this, i.e. a test that compares the log-likelihoods between a model with, and without a particular predictor of interest in it. However, for this test to be valid, data are required to follow a normal distribution. As the outcomes modelled in this thesis do not follow a normal distribution an alternative approach is needed. In this thesis the Satorra-Bentler Scaled Chi-square Test (SBSCT) is used as it compares the likelihoods between competing models but includes within it a scaling factor to account for the skewed nature of the outcome of interest[49]. As the SBSCT test is not automatically calculated in the software used, i.e. Mplus, it was calculated "by hand" using the steps described in Box 13.

[49] Two other options were considered alongside the SBSCT test: bootstrapping the model standard errors (rather than estimating them with robust standard errors) or fitting a generalized mixed model with a gamma link function to account for skewness in the data (this works in a similar way to a zero-inflated poisson model in that a model is fitted to predict whether people have hand pain (yes/no) and then a second model is fitted to predict the outcome distribution for those with hand pain i.e. a two part growth model with a preponderance of zeros). The first option was not used as a likelihood ratio test could not be calculated on the data and the second option, although elegant was highly complex to fit to the data and was not needed as the Satorra-Bentler Scaled chi-square test was an acceptable solution to the problem of interest.

Box 13: Calculation of the Satorra-Bentler Scaled Chi-square test

Step 1: Model fitting

Fit the two models for comparison using maximum likelihood estimation with robust standard errors (MLR). Record the following values:

L_0 = Log-likelihood of the model without the predictor in it

L_1 = Log-likelihood of the model with the predictor in it

C_0 = the correction factor associated with the model without the predictor in it

C_1 = the correction factor associated with the model with the predictor in it

P_0 = number of model parameters for the model without the predictor in it

P_1 = number of model parameters for the model with the predictor in it

Step 2: Calculating the Satorra-Bentler Scaled Chi-square test (SBSCT)

$SBSCT = -2*(L_0 - L_1)/cd$

where $cd = (P_0 * C_0 - P_1*C_1)/(P_0 - P_1)$

Step 3: Compare the SBSCT to a chi-square distribution with degrees of freedom equal to $P_0 - P_1$ to calculate the p-value for the test

5.2.8 Model fit

Although the SBSCT test described in Section 5.2.7 is a useful tool in determining whether a predictor adds significantly to a model, it does not give any information as to whether the model is a good fit to the data (i.e. a predictor may significantly improve model fit, but this does not necessarily mean that the model fits the data well). For this, goodness of fit indices, developed in the context of SEM, are useful.

In the SEM context, a wide range of goodness-of-fit indices exist to test how well the model of interest reproduces the observed data (Kenny 2014) and have been developed in response to the problem that the chi-square test is not a perfect measure of model fit (Byrne 2001). The chi-square test is sensitive to sample size[50] and includes a potentially unrealistic null hypothesis that the covariance structure in the raw data will be equal to that implied by the model (Hu et al. 1998b). Choice therefore needs to be made as to which fit indices to report (amongst the range of indices that have been developed and presented in the literature) as there is currently no single index that is agreed to be a definitive measure of model goodness-of-fit.

Hu et al. (1998b) highlight that goodness-of-fit indices can be grouped into families of fit indices and Muthen (2008a) recommend that at least one fit index from each family is reported (rather than reporting multiple measures from the same family). It is only when these indices are considered together can judgements be made as to whether the model fits the data well (Curran et al. 2003). To this end, only a subset of all possible fit indices can be calculated in the Mplus software (from the full range of fit indices shown in Hu et al. (1998b)) so these are the goodness of fit indices that are reported in this thesis (see Table 5-1 for details).

[50] this means that even small departures of the model to the observed data will be statistically significant if the sample size is large

Table 5-1: Goodness of fit indices used to assess goodness of fit in this thesis

Fit index	Literature guidelines on criteria to use to indicate "good" model fit
Incremental fit indices	
Tucker-Lewis index (TLI) (Tucker et al. 1973)	Ranges from 0 to 1 Values > 0.95 indicate good model fit (Byrne 2001)
Comparative fit index (CFI) (Bentler 1990)	Ranges from 0 to 1 Values > 0.95 indicate good model fit (Byrne 2001)
Absolute fit indices	
Standardised root mean-square residual (SRMR) (Bentler 1995)	Ranges from 0 to 1 Values < 0.05 indicate good model fit (Hooper et al. 2008)
Root mean square error of approximation (RMSEA) (Steiger 1990)	Values lower than 0.06 indicate good model fit. A 95% confidence interval can be calculated for this measure. A model with good fit would have a lower limit for the 95% confidence interval close to 0 and an upper limit less than 0.08 (Hooper et al. 2008)
Akaike's information criterion (AIC) (Akaike 1974)	A lower value indicates improved model fit when two competing models are compared
Bayesian information criterion (BIC) (Schwarz 1978)	A lower value indicates improved model fit when two competing models are compared
Sample size adjusted BIC (SBIC) (Sclove 1987)	A lower value indicates improved model fit when two competing models are compared

Footnote: the TLI and CFI are referred to as incremental fit indices as they compare the model of interest to a baseline model i.e. an independence model where no correlation between the observed variables is assumed. The remaining fit indices are absolute measures of fit as they are not compared to specific model of interest in the data (Hooper et al. 2008)

Goodness of fit will also be considered in the context of the multi-level framework by estimating and plotting the model residuals (ε_{it}) to assess how close they are to zero[51]. In addition, the magnitude of the variance around the random effect estimates is also considered as if small this would suggest that the model is fitting well and explaining a high degree of variability in the data[52].

5.2.9 Model assumptions

When fitting a growth model to a data set of interest key assumptions are made, both about the data, and also about the form of the trajectory for analysis. The assumptions can be categorised into four main areas: sampling assumptions, normality assumptions, linearity assumptions, and assumptions for measurement error (see Appendix 17). The assumptions are detailed, so Singer et al. (2003) have suggested that a practical approach be used, largely based on graphical methods, to explore whether model assumptions are satisfied in the growth curve context. This approach is described below and is adopted in this thesis to test whether the assumptions of the growth models are satisfied when the models are fitted to the data.

1) Plot a set of predicted growth curves for a random sample of participants to assess whether the assumption on trajectory form is tenable (e.g. that the trajectory is linear);
2) Plot predicted estimates of the growth parameters (i.e. α_i and β_i for the linear model shown in Box 10) against each time-invariant predictor in the model to check that this relationship is linear;

[51] Model residuals are estimated using empirical Bayes estimation as this estimation method is more precise than ordinary least squares

[52] It was considered whether an R-square value could be calculated for the multi-level model as a measure of model goodness-of-fit (similar to that used in linear regression). Although this measure can be calculated for a multi-level model, the addition of some predictors can decrease the value of the R-square, rather than increase it, due to the multiple variance components involved. This leads to negative values of R-square that are not appropriate (Singer et al. 2003). R-square was therefore not used as a measure of model fit in this thesis.

3) Plot a histogram of the model residuals and random effects (e.g. ε_{it} , $\zeta_{\alpha i}$, $\zeta_{\beta i}$ for the linear model shown in Box 10) to assess their magnitude and ensure that they follow a normal distribution;
4) Plot model residuals and random effects (e.g. ε_{it} , $\zeta_{\alpha i}$, $\zeta_{\beta i}$ for the linear model shown in Box 10) against:
   (a) each predictor in the model (with ε_{it} plotted against time as this predictor is at level 1 in the model, and $\zeta_{\alpha i}$, $\zeta_{\beta i}$ plotted against the time-invariant predictors as they are included in the level 2 part of the model)
   (b) study identification number (study ID)

   For model assumptions to be met, no relationships should exist in the plots as derived. In addition, plots in 4a can also be used to test the homogeneity of variance assumption for any categorical model predictors (i.e. the variance in ε_{it} , $\zeta_{\alpha i}$, $\zeta_{\beta i}$ is roughly equal at each level of the predictor of interest) and plots in 4b can be used to explore whether there are any participants in the data set whose predicted values are excessively large, or small, relative to other participants in the data set;

5) Plots 2 – 4 above can also be used to identify if there are any values on the predictor variables that are outliers in the data that may require checking and correcting in the data.

In addition, the co-variances (ε_{it} $\varepsilon_{it'}$), (ε_{it} β_i), and (ε_{it} α_i) will be inspected (with additional random effect terms included if the model is non-linear) to ensure that they are close to zero. Assumptions around the representativeness of the sample have also previously been explored in Chapter 3 where it was shown that the CAS-HA sample is largely representative of the broader population of participants with hand pain from which it has been drawn.

5.2.10 Variance/co-variance structure of the model random effects

The growth models, as defined so far, have each assumed an unstructured covariance matrix for the random effects, i.e., that the variances and co-variances of the random effects are each estimated separately in the model. It is possible, however, for alternative variance/co-variance matrices to be assumed for the random effects that simplify the model assumptions and may potentially aid model convergence when models are complex (see Table 5-2 for details). In this thesis, the alternative variance/co-variance matrices are explored via a set of sensitivity analyses to check that overall model conclusions do not change when the differing structures as given in Table 5-2 are applied.

Table 5-2: Alternative variance/co-variance matrices that can be used for the random effects in a growth model (StataCorp 2013a)

Variance-covariance matrix	Assumptions implied
Unstructured	All variances and co-variances to be distinctly estimated
Independent	One variance parameter per random effect, all co-variances 0
Exchangeable	Equal variances for random effects, and one common pairwise co-variance
Identity	Equal variances for the random effects, all co-variances 0.

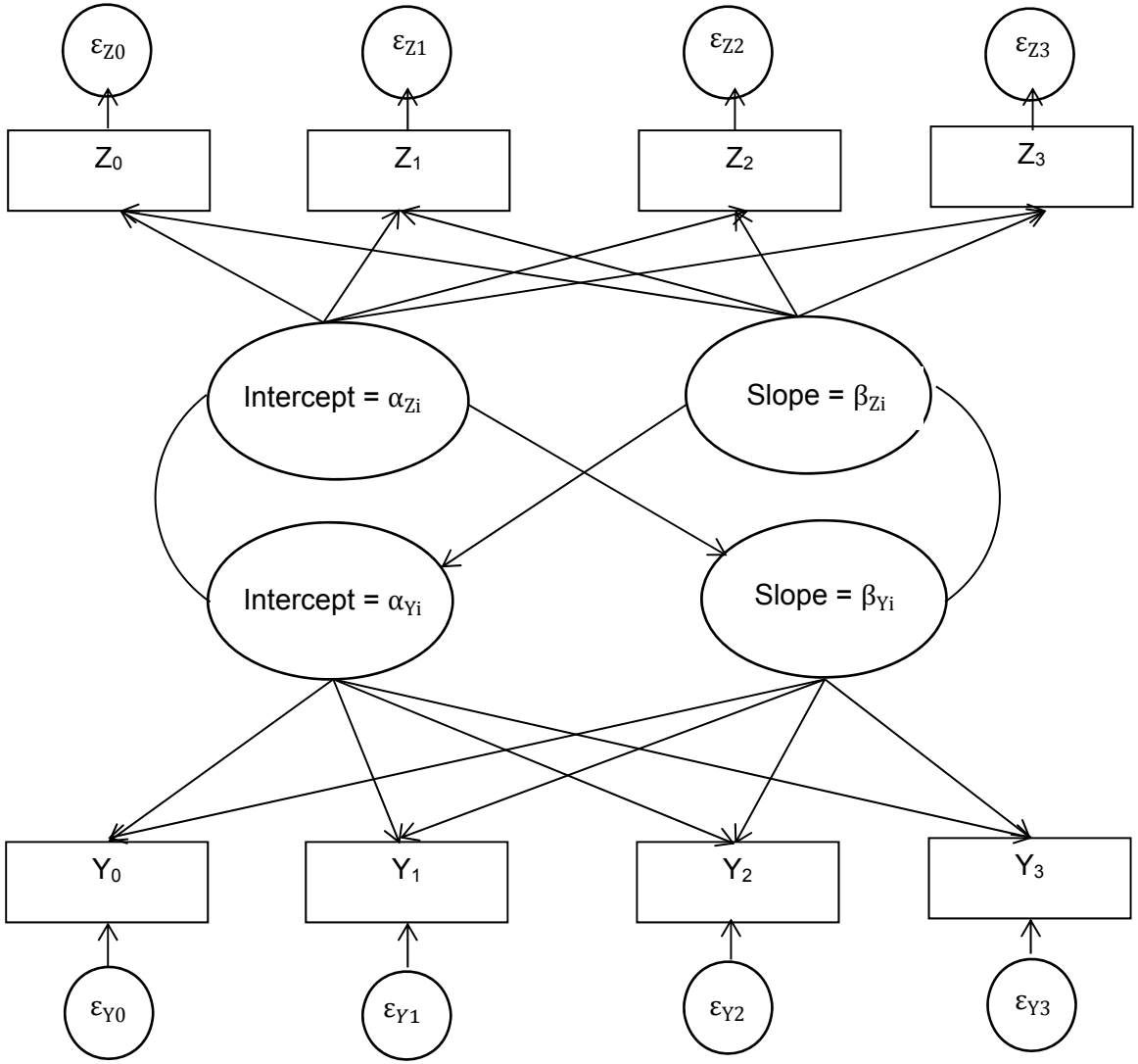
5.2.11 Simultaneous modelling of two growth processes

The models described so far in this chapter have each been concerned with modelling the trajectory of a single outcome over time. Models exist, however, that can be used to simultaneously model the trajectory of two separate outcomes over time (e.g. the trajectories of hand pain and function over time). The models are an extension of the growth models described in Sections 5.2.1 - 5.2.10 and are referred to as parallel process

growth models in the literature[53]. The aim of such models is to test whether the random intercepts and random slopes for an outcome are correlated with the random intercepts and slopes for a second outcome (Payne et al. 2014). Each outcome has its own growth process (akin to that in Figure 9), and is related to the other outcome via a set of additional correlations and predictive relationships added to the model (see the SEM representation of the model given in Figure 10 (Muthen et al. 2010)). The correlations added to the model include those between the intercept for outcomes one and two and between the slopes for outcomes one and two. Alongside this, predictive relationships are also added between the intercept for outcome one and the slope for outcome two, and the intercept for outcome two and the slope for outcome one.

[53] This method was preferred over fitting the second outcome as a time-varying predictor in the growth model as it enables a trajectory process for the second outcome to be modelled. This would not be achieved if it were fitted as a time-varying covariate and would force a choice to be made as to which outcome was the dependent variable and which the independent covariate (Curran et al. 2003)

Figure 10: Parallel process linear growth model fitted in a structural equation modelling framework for a longitudinal study with two outcomes (Y and Z) each measured at four time points of interest (Muthen et al. 2010)



Footnote: The notation in this diagram is identical to that shown in Figure 9 for the linear growth model, however, an additional subscript is included to indicate the outcome of interest that the trajectory parameters refer to, with Y used to indicate outcome one and Z used to indicate outcome two.

The model shown in Figure 10 can be extended to include growth processes for one or both outcomes that are non-linear. This is achieved by adding additional latent variables to the model (such as those shown in Appendix 15) to represent quadratic growth over time. When such models are fitted to the data, the fit indices described in Section 5.2.8 can be used to determine whether the model is a good fit to the data.

5.2.12 Reflections on the growth models

Both growth models and GEE models can be used to explore the relationship between predictors of interest and an outcome trajectory over time. For an outcome measured on a continuous scale, the two approaches lead to equivalent results when an exchangeable correlation structure is specified in the GEE model and a random intercept growth model is fitted to the data (Twisk 2003). It is only when the outcome is measured on a non-continuous scale (e.g. binary, count or ordinal) that noticeable differences between the approaches occur (Twisk 2003). The models have differing assumptions about missing data that are explored more fully in Section 5.5.3.

A key advantage of the growth model approach (over GEEs) is that it provides information not only on the mean trajectory over time, but also on the amount of variation between individuals to indicate how similar trajectories are for participants in a sample (Kahn 2011). It also simultaneously provides information on the correlation between the random intercept and slope terms so it can be explored whether the degree of change over time depends on participants' initial baseline score when entering the study (Byrne et al. 2003). That said though, growth models have been reported to be more sensitive to the misspecification of the covariance structure in the data, so if marginal effects only are of interest, then the GEE approach is preferred (Goldstein et al. 2002).

Although growth models provide a flexible approach to analysing longitudinal data, they do, however, rest on the key assumptions that individuals are sampled from a single population, that a single trajectory curve is sufficient to describe outcome changes over

time, and that there is a uniform influence of any predictors on the variance and growth parameters within the population (Nagin et al. 2010, Jung et al. 2008). This assumption may not always be true, and in some studies unobserved sub-populations may exist, that have differing trajectory shapes over time to the population as a whole (Curran et al. 2003). To respond to this, two additional techniques have been developed in the context of growth models to test the specific assumption of whether a single growth curve is valid to describe the trajectories of interest in a population of interest. The two techniques are latent class growth models (LCGM) and growth mixture models (GMM). They are described in more detail in Sections 5.3 and 5.4 respectively.

5.3 Latent class growth models (LCGM)

5.3.1 The basic concept

Latent class growth models (LCGM) are (semi-parametric) statistical models that can be used to uncover distinct subgroups of participants that have similar outcome trajectories over time within a sample (Andruff et al. 2009). A key feature of LCGM is that prior to modelling, subgroup membership is unknown (unobserved) and is derived, based on data, by grouping trajectories so that outcome trajectories within a group are more similar to each other than to outcome trajectories between groups (Jung et al. 2008).

In LCGM it is not assumed that all participants are sampled from a single observed population and that a single growth curve (with a single estimate of growth parameters) is adequate to describe the trajectory of an outcome over time (as is the case for growth models) (Jung et al. 2008). Instead it is hypothesised that two or more subgroups exist within the data that could (potentially) have differing outcome trajectories over time (Nagin et al. 2010). This approach therefore reduces the potential for important sub-groups of participants to be masked in the data that may not be uncovered by growth curve modelling (Andruff et al. 2009) (e.g. if the trajectories in a 2-group linear model were increasing in the first group and decreasing in the second group, a 1-group model would

mask this by fitting the average of these two lines to the data, i.e. a flat line with no indication that groups who are changing over time existed in the data).

A key assumption of LCGM is that a (small) number of subgroups exist that can be used to approximate the continuous distribution of trajectories of unknown shape that exist in the population (Nagin et al. 2010), which is in contrast to the growth models described in Section 5.2. There, subgroups of interest are known prior to analysis, and are defined using measured variables in the dataset (e.g. gender). LCGM also enables participants to be split into groups so that the baseline characteristics of each group can be described (this is a test of the validity of the groups as if baseline characteristics do not differ between them then the usefulness of the groups is questionable) (Nagin et al. 2010).

5.3.2 Model formulae

LCGM can be expressed by extending the notation for the growth models in Section 5.2 and is achieved by adding an additional subscript K to indicate that separate growth models are fitted for each latent group (assuming that the number of latent groups range from 1 to K) (Box 14). The notation in Box 14 illustrates a key assumption of LCGM - that is, that the within-group variances and co-variances of the random effects (i.e. the random intercept $\zeta_{\alpha Ki}$ and slope $\zeta_{\beta K i}$ for the linear model) are assumed to be equal to zero [54] (Andruff et al. 2009, Jung et al. 2008).

This is an important assumption as it illustrates that in LCGM, between-participant trajectory differences are expressed solely through latent group membership rather than via random effects as used in the standard growth model (Andruff et al. 2009) – an assumption that implies that knowledge of latent group membership is sufficient to describe individual differences in the outcome trajectories over time (Andruff et al. 2009). This assumption also implies homogeneous within-group growth curves that do not vary between participants in a group (Jung et al. 2008). LCGM can also be expressed in an

[54] This is identical to fixing the parameters of the trajectory curve to be equal for all participants within a latent group. Trajectory parameters between latent groups can differ.

SEM framework (Figure 11), which illustrates how growth models can be extended to include an unobserved categorical latent variable that represents group membership and predicts the outcome trajectory over time.

Box 14: Formulae for linear LCGM (Wang et al. 2007)

Level 1 – the fixed effects model

$$y_{Kit} = \alpha_{Ki} + \beta_{Ki}\lambda_{\beta Kt} + \varepsilon_{Kit}$$

where α_i = intercept, β_i = linear component, $\lambda_{\beta t}$ = time of assessment, k=latent group number, ε = residual error

Level 2 - the random effects model

$$\alpha_{Ki} = \mu_{\alpha K} + \zeta_{\alpha Ki}$$

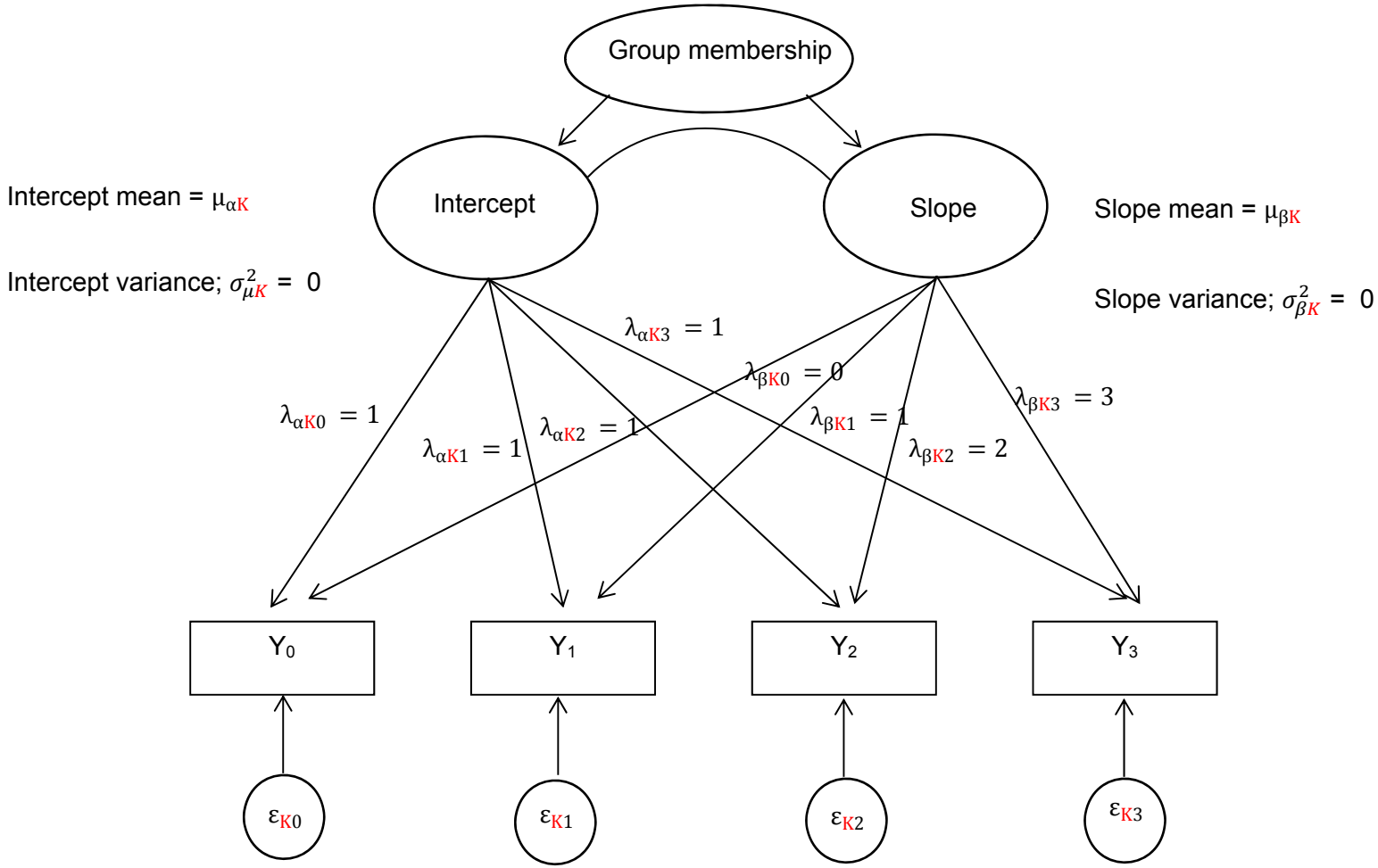
$$\beta_{Ki} = \mu_{\beta K} + \zeta_{\beta Ki}$$

For LCGM, the variances of $\zeta_{\alpha Ki}$ and $\zeta_{\beta Ki}$ ($\sigma^2_{\alpha K}$ and $\sigma^2_{\beta K}$) are set equal to 0 so the level 2 model can be simplified to:

$$\alpha_{Ki} = \mu_{\alpha K}$$

$$\beta_{Ki} = \mu_{\beta K}$$

Figure 11: Linear LCGM fitted in a structural equation modelling framework for a study with data collected at four time points (Wang et al. 2007)



5.3.3 Model assumptions

All of the assumptions described for the growth models in Section 5.2.9 apply to LCGM, however the assumption regarding normally distributed random effects no longer applies. This is because the variance of the random effects is constrained to be zero in LCGM. In particular LCGM are fitted using robust standard errors to account for any skewness in the distribution of the outcome of interest (Muthen et al. 2010).

5.3.4 Model fit

The CFI, TFL, RMSEA and SRMR fit indices described in Section 5.2.8 for growth models cannot be calculated for LCGM when the number of latent groups is greater than one. This is because there is no single covariance matrix for the data to be fitted to and because models with a varying number of groups are not truly nested, i.e. one model cannot be defined by constraining parameter(s) in the other model to be zero (Wang et al. 2007). Model fit for LCGM is therefore conducted using less formal methods (Curran et al. 2003) as described below:

Information-based criteria (i.e. the AIC, BIC, and SABIC)

The AIC, BIC and SABIC (as defined in Section 5.2.8) are based on the model log-likelihood so can be applied to LCGM (Wang et al. 2007, Raftery 1995). They can be used to compare the fit of two competing LCGM, with the preferred model being the one with the smallest value on the information criterion of interest (Wang et al. 2007). Both the AIC and BIC balance model complexity with model goodness of fit (Nagin et al. 2010), however the BIC has a greater penalty for model complexity, so would suggest that a model with a fewer number of groups were optimal than if the AIC were used (IBM 2011), The sample-size adjusted BIC is a modification of the BIC that has been proposed to improve the performance of the BIC for models that either have a large number of parameters or are analysed in a small data sample (Tofighi et al. 2008)

Posterior probabilities

Posterior probabilities represent the likelihood that each participant (with its respective outcome trajectory over time) belongs to each latent group included in the model (Andruff et al. 2009). For each participant, high posterior probabilities are desirable for a single latent group, with low probabilities for the remaining latent groups. This is illustrated (for fictitious data) in Table 5-3, with all participants in the sample having posterior probabilities greater than 0.9 for a single latent group, suggesting that the model is a good fit to the data (Jung et al. 2008). From the posterior probabilities, two further indices can be calculated that can be used to assess model fit: group membership probabilities and average posterior probabilities. They are described below:

Table 5-3: Posterior probabilities for 5 participants in an illustrative fictitious dataset

	Posterior probability			
Participant	Latent group 1	Latent group 2	Latent group 3	Group membership
1	0.02	0.06	**0.92**	3
2	**0.95**	0.01	0.04	1
3	**0.90**	0.05	0.05	1
4	0.03	**0.97**	0.01	2
5	0.04	0.04	**0.92**	3

Group membership probabilities

Group membership can be assigned for each participant using a maximum-probability rule that allocates participants to the group that corresponds to their highest posterior probability (Andruff et al. 2009). For example, in Table 5-3, participant 1 is allocated to group membership 3 as this is the latent group where the highest posterior probability (0.92) is achieved for this participant. The probability of belonging to each group can then

be calculated (using a frequency table of group membership). It has been suggested for a model to fit the data well, the number of participants in each latent group should not fall below 1% (Jung et al. 2008) - 5% (Andruff et al. 2009) of the total sample as small groups may have occurred by chance and not be replicated in other populations.

Average posterior probabilities

Average posterior probabilities can also be calculated by averaging the maximum posterior probabilities for participants in each group of interest. For example, in Table 5-3, the average posterior probabilities are 0.93 ((0.95 +0.90)/2), 0.97, 0.92 for groups 1, 2 and 3 respectively. The average posterior probabilities give an indication of the internal reliability of each group (Andruff et al. 2009). For the model to fit the data well, average posterior probabilities greater than 0.8 are desirable, indicating that the model can discriminate between similar and dissimilar trajectories over time, although some authors consider that a more relaxed threshold of greater than 0.7 is appropriate for this criterion (Andruff et al. 2009, Nagin et al. 2010).

Entropy

Entropy is a standardised index, ranging from 0 to 1, that can be used to measure overall classification accuracy and determine how accurately participants can be classed into one, and only one, latent group (Wang et al. 2007). It is calculated by averaging the posterior probabilities after individuals have been assigned to their most likely group (Nagin et al. 2010). Although a high entropy value indicates better classification accuracy (Jung et al. 2008), no formal criteria exist to indicate how close to one entropy needs to be for good model fit to be assumed (Jung et al. 2008). Several authors suggest entropy values greater than 0.80 are required for classification accuracy to be demonstrated (Wang et al. 2007), so this will be used as a guideline in this thesis.

Trajectory plots

Trajectory plots are a useful graphical tool to assess model fit and are derived by plotting the predicted outcome (from LCGM) for each latent group at each time point. The predicted trajectories are calculated from the regression coefficients estimated in LCGM using techniques common to standard (linear) regression (Andruff et al. 2009). A plot of the predicted values and their associated confidence intervals can be used to reveal whether the model is a good fit to the data, with specific consideration given to any overlap in the confidence intervals for the predicted trajectories (overlapping confidence intervals may suggest that the model is over-fitted and that a model with a fewer number of groups would be preferable) (Andruff et al. 2009), and also whether the predicted trajectories are a good approximation to the raw trajectories within each latent group[55]. This would be an indication of good model fit.

Model convergence and local solutions

Model convergence, although not a direct method to assess model fit, is an indicator of how well a model fits the data of interest; it indicates whether a reliable model solution has been obtained. Poor model convergence often occurs in LCGM due to the complexity of the model to be fitted to the data (Jung et al. 2008) and is more likely to occur as models increase in their complexity (e.g. models with higher order polynomials and/or a large number of latent groups).

One particular issue in such models is that of local solutions, which occur when the modelling algorithm converges to parameter estimates that are not associated with the true maximum log likelihood for the model[56] (Jung et al. 2008). To avoid reporting local

[55] Multi-group SEM can be used to test whether the regression parameters (e.g. the intercept and slope) are equal across the latent groups (Andruff et al. 2009), however this approach will not be used in this thesis; assessment of trajectory overlap will therefore be based on visual inspection only.

[56] A local solution occurs if the algorithm to fit LCGM converges to a maximum value that is only true for a small part of the likelihood estimation curve; it is not necessarily the maximum for the likelihood estimation curve as a whole i.e. it is not a global solution (Jung et al. 2008).

solutions, running the model with multiple starting values is recommended[57], to ensure that, irrespective of the random starting point used, the algorithm converges to the same final stage maximum log likelihood solution. As a general guideline, if the *largest*[58] log likelihood is replicated for two or (preferably) more random starting values, a global, rather than local, solution can be assumed and the model with the largest log likelihood reported as a valid solution (Muthen et al 2010).

5.3.5 Number of latent groups to include in LCGM

For LCGM to be fitted to a data set of interest the number of latent groups needs to be specified [59] and, unless a clear hypothesis exists around the number of latent groups to model, the optimum number of latent groups for the data will need to be identified (Andruff et al. 2009). It has been emphasised by several authors (Jung et al. 2008, Curran et al. 2003) that selection of the optimum number of latent groups cannot be based solely on statistical indices of goodness of fit (such as those described in Section 5.3.4), but should consider (with equal importance) other factors such as model parsimony, clinical interpretability and theoretical knowledge of the outcome of interest (Mora et al. 2009). A range of factors need to be considered as there is no single, commonly agreed, index that can be used in isolation to identify the number of latent groups that are optimum (Wang et al. 2007, Jung et al. 2008).

For statistical based indices, models with varying numbers of latent groups can be compared to find the model that is the best fit to the data. This can be achieved by comparing (between models) the indices described in Section 5.3.4. However, in addition, likelihood ratio based tests have also been developed to specifically compare model fit between models with varying numbers of latent groups, as described below.

---

[57] In this thesis, 500 random starts were used in the analysis, as this was a large value, but not one that was too computationally burdensome (Muthen et al. 2010).

[58] A global solution cannot be concluded if the log-likelihood value that is replicated in two or more model solutions is not the largest log-likelihood derived (Muthen et al. 2010)

[59] The initial choice for the number of latent groups is encouraged to be guided by previous research and clinical knowledge if available (Andruff et al. 2009)

Likelihood-ratio based comparative goodness of fit tests for LCGM

Likelihood-ratio based statistical tests have been proposed to test whether a model with K-1 groups is a better fit to the data than a model with K groups, i.e. to test the null hypothesis that the data are generated from a model with K-1 groups (Muthen et al. 2010). A non-significant likelihood ratio test will therefore identify the model with the optimum number of groups (Jung et al. 2008). Multiple likelihood ratio based tests have been proposed as the standard likelihood ratio test (used for growth models) does not follow a chi-square distribution when models with K-1 and K groups are compared (Wang et al. 2007) and the models are not nested (Curran et al. 2010). The tests include:

- the Lo-Mendell-Rubin likelihood ratio test (LMR-LRT)
- the Adjusted Lo-Mendell-Rubin likelihood ratio test (Adjusted LMR-LRT)
- the Vuong-Lo-Mendell-Rubin likelihood ratio test (VLMR-LRT)
- the Parametric Bootstrapped likelihood ratio test (PB-LRT)

Although the unadjusted and adjusted LMR-LRT are frequently used to guide the selection of the optimum number of latent groups they have been criticised as being sensitive to sample size[60] (Wang et al. 2007). The PB-LRT has therefore increased in popularity as it is less sensitive to sample size and has been shown to perform well in simulation studies (Nylund et al. 2007). However, a limitation of this measure is that it is computationally demanding so burdensome to apply in practice. As a compromise, it has been suggested that the LMR-LRT be used initially to guide the analysis to a small set of near-optimum models and then the PB-LRT be used in the final stages to determine the optimum number of groups for analysis (Jung et al. 2008).

[60] The optimal number of latent groups will be larger if the sample size is large; larger sample sizes tend to give larger test statistics (Wang et al. 2007)

5.3.6 Polynomial form to be included in LCGM

In addition to the number of latent groups, the form of the polynomial (e.g. linear, quadratic) needs to be specified for LCGM to be fitted to a dataset of interest. Opinion varies as to the optimum approach to achieve this, in particular, on how to combine the search for the optimum polynomial form with the search for the optimum number of latent groups. For example, Curran et al. (2003) recommend that the optimum polynomial form is initially derived for the sample as whole[61] and then held constant while the optimum number of latent groups are explored. Alternatively, Andruff et al. (2009) suggest a more data driven approach, where backwards deletion (based on statistical significance[62]) and model fit comparisons are used at each modelling stage to find the optimum polynomial.

Although the approach by Andruff et al. (2009) is more flexible, as it allows the polynomial form to differ between models with varying numbers of groups and between latent groups within a single model, it runs the risk of being tailored too much to the data set of interest, so may not replicate in future data samples (Wang et al. 2007). The approach by Curran el al. (2003) was therefore used in this thesis.

5.3.7 Model interpretability and replication

After statistical criteria have been used to guide the choice of the optimum number of groups, the usefulness of the groups derived needs to be evaluated. This can be achieved by considering whether the groups differ in terms of their (clinical) history, future outcomes, response to treatment, or relationship to trajectories for other outcomes or behaviours (Nagin et al. 2010). If the derived groups do not differ with respect to these external variables, it can be concluded that the model is not a useful representation of the data. Further to this, model replication in a new dataset is also useful to support the true existence of the trajectory groups in the data (Nagin et al. 2010).

---

[61] The optimum polynomial would be tested by adding higher order polynomials to the model to see if they were statistically significant and improved model fit

[62] The Satorra-Bentler Scaled chi-square test can be used for this purpose if the outcome is skewed; models are nested if only the degree of polynomial that is varied between them rather than the number of groups they contain (Muthen et al. 2009).

5.4 Growth mixture models

5.4.1 The basic concept

Growth mixture models (GMM) are similar to LCGM as both models aim to identify trajectory groups with similar outcome trajectories over time and are fitted using finite mixture models[63] (Nagin et al. 2010). The models differ, however, in the assumptions they make concerning within-group trajectories. In LCGM it is assumed that all within-group trajectories are homogeneous, whereas in GMM this assumption is relaxed, allowing within-group trajectories to vary between participants (Jung et al. 2008). GMM therefore works by fitting separate growth models to each trajectory subgroup in the model (Nagin et al. 2010).

5.4.2 Model formulae

GMM can be expressed using the formulae in Box 14 (shown above for LCGM) and are derived by relaxing the LCGM assumption that the within group variance and co-variance of the random effects is zero, i.e. in GMM these parameters are freely estimated (Wang et al. 2007). GMM are therefore a combination of the growth models described in Section 5.2 and LCGM as it includes a categorical latent variable to define the unobserved subgroups of interest alongside a set of continuous latent variables (random effects) to model individual variability of the trajectories within each latent group (Wang et al. 2007).

5.4.3 Model assumptions

Model assumptions for GMM are similar to LCGM, however normality assumptions regarding the random intercept and slope terms are assumed only to be true within each group, rather than for the population as a whole.

[63] Finite mixture models are a class of statistical model to model data where distinct sub-populations are thought to exist but sub-group membership is unknown prior to analysis i.e. they analyse data from a mixture of two or more groups or populations (Nagin et al. 2010)

5.4.4 Reflections on GMM

A key advantage of GMM over LCGM is that it represents the longitudinal data more realistically by assuming a degree of variability around the growth parameters (Wang et al. 2007). It also has the potential to be used to explore whether the reason why outcome trajectories are not sufficiently explained by the predictors and random effects in a growth model is due to the existence of unobserved subgroups within the population of interest (Nagin et al. 2010, Wang et al. 2007). That said though, a key difficulty with such models is their increased complexity and computational burden that can lead to convergence problems and unstable solutions when GMM are fitted to real life data (Jung et al. 2008).

To address the issue of lack of convergence in GMM, several strategies have been proposed to modify GMM, so that convergence to a global model solution can be achieved either by modifying the starting values in the estimation algorithm or by constraining the variances estimated in the model (see Box 15). In addition, to reduce computation burden, it has been suggested that LCGM be used to explore and define the optimum number of groups and polynomial form in the data and that GMM are then only used in a confirmatory manner, to test whether model fit improves by relaxing the assumption that the within-group variability is zero, i.e. the key assumption of LCGM (Wang et al. 2007). If model fit is then improved for GMM, group membership from this model can then be reported, albeit acknowledging that group membership is not as clearly defined as if the LCGM assumptions were completely satisfied. Although practical, this approach has the potential drawback that when random effects are added to LCGM the optimum number of groups required to model the data may be reduced (as adding the random effects is allowing for more within-group variability in the individual-level trajectories) (Nagin et al. 2010). This needs to be carefully explored when selecting a model that is optimum.

Box 15: Strategies to promote model convergence for GMM

Starting values

- Base the starting values for the EM algorithm (i.e. the algorithm that is used to calculate the maximum likelihood estimates) on parameter estimates from LCGM fitted prior to GMM (rather than using random starting values) (Jung et al. 2008)

Constraining variances of the growth parameters

- Constrain any (implausible) negative estimates of variance to 0 if this is an appropriate assumption from visual inspection of the data (Jung et al. 2008)
- Constrain the variance of the random effects to be equal (but not 0) across groups of interest (Jung et al. 2008)
- Use LCGM trajectory plots to assess whether there are any latent groups that can be simplified by constraining a parameter/variance estimate to be zero (Wang et al. 2007)

5.4.5 Parallel process growth mixture models (PPGMM)

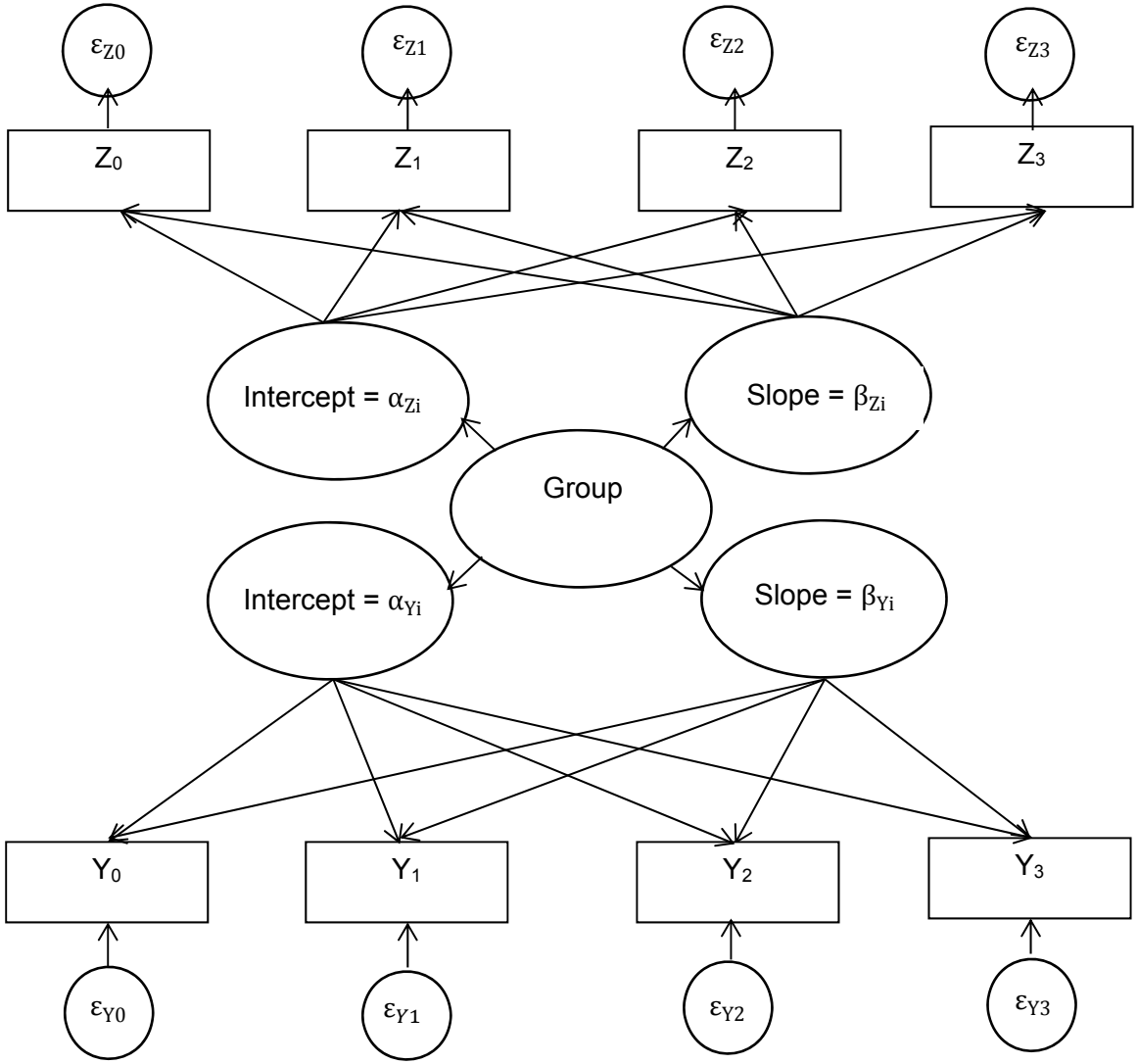
The GMM described in this chapter can be extended so that latent group membership is defined not only by the trajectory of a single outcome of interest but also by the trajectories of two (or more) outcomes assessed simultaneously (e.g. hand pain and function)[64]. This model is referred to as a parallel process growth mixture model (PPGMM) and is illustrated in the SEM framework in Figure 12. This model enables research questions to be addressed such as whether groups of participants exist that have a

[64] The PPGMM is not the same as a group-based dual trajectory model (GBDTM) (the GBDTM simultaneously defines a LCGM for the two outcomes of interest, along with a set of probabilities that relate the groups of the two LCGM models together e.g. probabilities express relationships such as *"if a participant is in group one for outcome one, what is the probability that they will also be in group one for the second outcome"*) (Xie et al. 2010). The GBTM approach was not used in this thesis as the group sizes for the first outcome were not large enough to then support being further split into groups that were defined by the LCGM for the second outcome

characteristic trajectory for outcome one (e.g. hand pain) that is then simultaneously accompanied by a characteristic trajectory for outcome two (e.g. hand function). For example, a PPGMM was fitted in a study of substance abuse and conduct problems to reveal one group of participants with both high substance abuse and high conduct problems throughout the study alongside a separate group of participants (group 2) with increasing substance abuse over time that was not matched by an increase in conduct problems (Wu et al. 2010).

As the PPGMM is an extension of the GMM model, the estimation algorithm to fit the model and the techniques used to assess goodness of fit are identical to that used for GMM. In addition, PPGMM can be simplified to form a parallel process latent class growth model (PPLCGM) if needed by restricting the within-group variance on the growth factors to be zero; an approach quite often needed, as such models are complex and therefore convergence problems are often encountered.

Figure 12: Parallel process linear growth mixture model fitted in a structural equation modelling framework for a longitudinal study with two outcomes (Y and Z) each measured at four time points of interest (adapted from (Wu et al. 2010))



Footnote: The notation in this diagram is identical to that shown in Figure 9 for the linear growth model, however, an additional categorical latent variable has been added to the model along with subscripts to indicate the outcome of interest, with Y used to indicate outcome one and Z used to indicate outcome two.

5.5 Generic issues relating to the application of the models in this chapter

5.5.1 Computer software and computation

The models described in this chapter can be fitted using a range of commercially available software packages (e.g. STATA (StataCorp 2013b), AMOS (Arbuckle 2006), SPSS (IBM Corp 2013), Mplus (Muthen et al. 2010), and Latent Gold (Vermunt et al. 2013)). In this thesis, STATA version 13.0 is used when models are fitted in a multi-level framework and Mplus version 6.0 is used when a SEM framework is adopted.

5.5.2 Sample size calculations for longitudinal data

A sample size calculation for the CAS-HA study was included in the recruitment protocol and was based around the ability to detect a relative risk of deterioration of at least 1.6 at the 18-month follow-up between those with and without a baseline risk factor of interest (e.g. presence of radiographic OA) with 80% power and alpha of 0.05. This required 500 participants at baseline, however as more participants were willing to attend the clinical assessment than planned, a total of 623 participants were recruited (Myers et al. 2007).

The sample size calculation was therefore defined prior to planning the full longitudinal analysis (using all time-points over the 6-year follow-up) as only the baseline and 18-month follow-up studies were initially funded. Although formulae exist to estimate required sample size for longitudinal studies (such as those given in (Diggle et al. 2002)) more general guidelines are used in this study, partly because the sample size calculation was done prior to knowing the number of study time points, but also because it would be difficult to define one single analysis as primary to base a sample size calculation upon as the analysis is largely exploratory. As a general guide, Andruff et al. (2009) suggest that sample sizes between 300 and 500 are sufficient in the context of LCGM, and Byrne et al. (2003) suggest sample sizes > 200 per time point are adequate for such analysis; limits that are both satisfied in the CAS-HA study.

In light of the absence of a specific sample size calculation, a particular focus of the analysis in this thesis is on the width of the confidence interval to explore how reliable the model estimates are. The results are also commented on in light of the estimates given for minimum clinically important change to ensure that the sample size is not so large that clinically meaningless results are presented (Andruff et al. 2009) and especially that higher order polynomials are not included if they are statistically significant yet have little influence on the overall trajectory curve obtained (Wang et al. 2007).

5.5.3 Missing data

Missing data occurs in most longitudinal studies and happens when participants either drop-out of the study or do not respond at one or more time points (Twisk 2003). As both types of missing data could potentially occur in the CAS-HA study, the rates of each type are reported. When considering the models in this chapter, it is noted that they have differing assumptions around missing data, with GEE models assuming that missing data are missing completely at random (MCAR), and all other methods assuming data are missing at random (MAR) (for an explanation of the terms MCAR and MAR see Appendix 18)[65]. Although this is the case, and it could be argued that growth models are preferred as their missing data assumptions are less stringent, this becomes less of an issue when the outcome is continuous, as in practical terms, the differences between the two models are small (Twisk 2003).

All models described in this chapter can be estimated in the presence of missing data which means that the data are included in the analysis if they have data present for at least one time point[66]. This is preferred over running the analysis on participants only with data at all time-points (as this represents a loss of information), but also preferred over

[65] The GEE model has the MCAR assumption for the calculation of the working correlation structure to be unbiased (Ballinger 2004). The growth models have a more relaxed assumption around missing data as maximum likelihood estimates are asymptotically unbiased when missing data are MAR (Nagin et al. 2010))

[66] Models are estimated by summing the individual contributions of each participant such that participants with a large number of data points are weighted more heavily than those with a smaller number of data points (Curran et al. 2010)

using multiple imputation to obtain model estimates when data are missing, this latter decision being made based on findings from a recent study which provided inconclusive evidence as to whether using multiple imputation for growth models led to more accurate results than when growth models without multiple imputation were employed (Twisk et al. 2013).

5.6 Discussion

In this chapter the statistical methods used to address the thesis objectives have been described. Specifically, GEE models were discussed as they are used in Chapter 6 to describe the overall trajectory of hand pain and function over time, along with growth models, that are extensively used in Chapter 7 to identify key baseline predictors of the trajectories of hand pain and function over time. LCGM and GMM were discussed as they are used in Chapter 8 to explore whether distinct subgroups of participants can be identified with differing trajectories of hand pain and functional difficulty over time. Parallel process growth models and parallel process GMM are also discussed as they are used in Chapter 9 to simultaneously model joint trajectories of hand pain and functional difficulty over time.

Whilst considering the range of longitudinal approaches available, several methods were not included, either because the method was limited, or because it addressed a different research question to that in the thesis. For example, repeated measures analysis of variance (ANOVA) was considered, but not used, despite a continuous outcome and time invariant predictors, as this method requires a complete dataset with no missing values and normally distributed data. It also does not use the covariance among the repeated measures to increase the efficiency of the parameter estimates (Diggle et al. 2002, Ballinger 2004)[67]. In addition, examples of techniques that could have been employed to

[67] Other methods that were considered but found to be limited were (1) summary statistics (i.e. where the longitudinal data is combined to produce a single per-person measure), limited as it represents a loss of information in the data and no strong rationale existed for the most appropriate summary measure to use (e.g. average across time points, maximum value across time points)

address different research questions include adding time-varying predictors to the growth models, or incorporating an external (distal) variable (measured at the 6-year follow-up) into the LCGM/GMM to test whether the trajectory groupings obtained predict an external outcome of interest (Wang et al. 2007). A discussion around how such additional statistical techniques could be used to extend the research questions in this thesis is given in Chapter 10 therefore they are not discussed further here. The next four chapters, however, are used to describe the results from fitting the models described in this chapter to address the research questions as stated previously in Chapter 1.

(Matthews 2005) (2) Cluster analysis to group participants into groups based on their repeated measures data – limited as participants are unrealistically allocated to be in or out of a group rather than to have a probability of being in a group as is used in LCGM (Nagin et al. 2010). (3) Latent profile analysis, i.e. latent class analysis with continuous indicators – limited as it does not take into account the time-ordering of the repeated measures as collected (Flaherty et al. 2012).

6 Describing the trajectory of hand pain and functional difficulty in CAS-HA

6.1 Introduction

The overall purpose of this chapter is to describe the longitudinal trajectory of hand pain and function for participants in the CAS-HA study. More specifically, three key objectives will be addressed that will form the three main sections in this chapter:

Objective 1: To describe the distribution of hand pain and function at each time point and their overall trajectories over time (Sections 6.3 and 6.4)

Objective 2: To explore missing data patterns for the AUSCAN and further characterise participants who are lost to follow-up (Section 6.5)

Objective 3: To compare levels of hand pain and function in CAS-HA to population normative data for the AUSCAN at each time point, and then, at an individual level, to compare rates of change in these measures to pre-existing treatment responder criteria (Section 6.6)

The first two objectives are included to gain an understanding of the AUSCAN measures that are modelled in later thesis chapters, e.g. to inform the choice of trajectory shape to model in the data and to assess key modelling assumptions. The third objective is included to support interpretation of the results and consider further what value on the AUSCAN can be considered a "large" change over time in the CAS-HA population, and is included as it was not possible to derive a reliable MIC value using the more established methods applied in Chapter 4[68].

[68] As the analysis in this chapter uses descriptive statistics (i.e. data plots and summary statistics), or analysis techniques previously described in Chapter 5, a separate methods section is not given in this chapter. This chapter is therefore a results chapter only, with findings presented separately for hand pain and function

Prior to presenting the results in this chapter, two additional comments are made regarding the scoring of the AUSCAN and coding of time in the analysis (Sections 6.2.1 and 6.2.2, respectively). These issues are applicable to all sections of analysis and the remaining thesis chapters. A final section will also be included to summarise and discuss the overall findings in this chapter (Section 6.7).

6.2 Data scoring and coding prior to analysis

6.2.1 The AUSCAN

Prior to analysis, the AUSCAN pain and function subscales were re-coded on a scale of 0 – 10[69] so that the magnitude of change over time could be directly compared between hand pain and function outcomes and to simplify comparisons to population normative data that have been published on a 0-10 scale (Bellamy et al. 2011). It is this scaling of the AUSCAN that is used throughout the remainder of the analysis presented in this thesis.

6.2.2 Time

In Chapter 3, the CAS-HA study was described as having regular 18-month intervals over a 6-year period (i.e. 0, 18, 36, 54, and 72-months), however, in practice, this timescale was not achieved for the latter two time points due to a delay in gaining ethical approval and logistical issues when conducting the 6-year assessments. The coding for time used throughout this thesis therefore reflects what happened in practice, rather than what was planned, and is based on the mean number of months since baseline (i.e. 0, 18, 36, 63, and 89-months)[70] converted into "time in years" (with time in years used to improve the interpretability of the estimates and to reflect that many CAS-HA participants are likely to have chronic problems that change over a long time-period). The final coding of time used

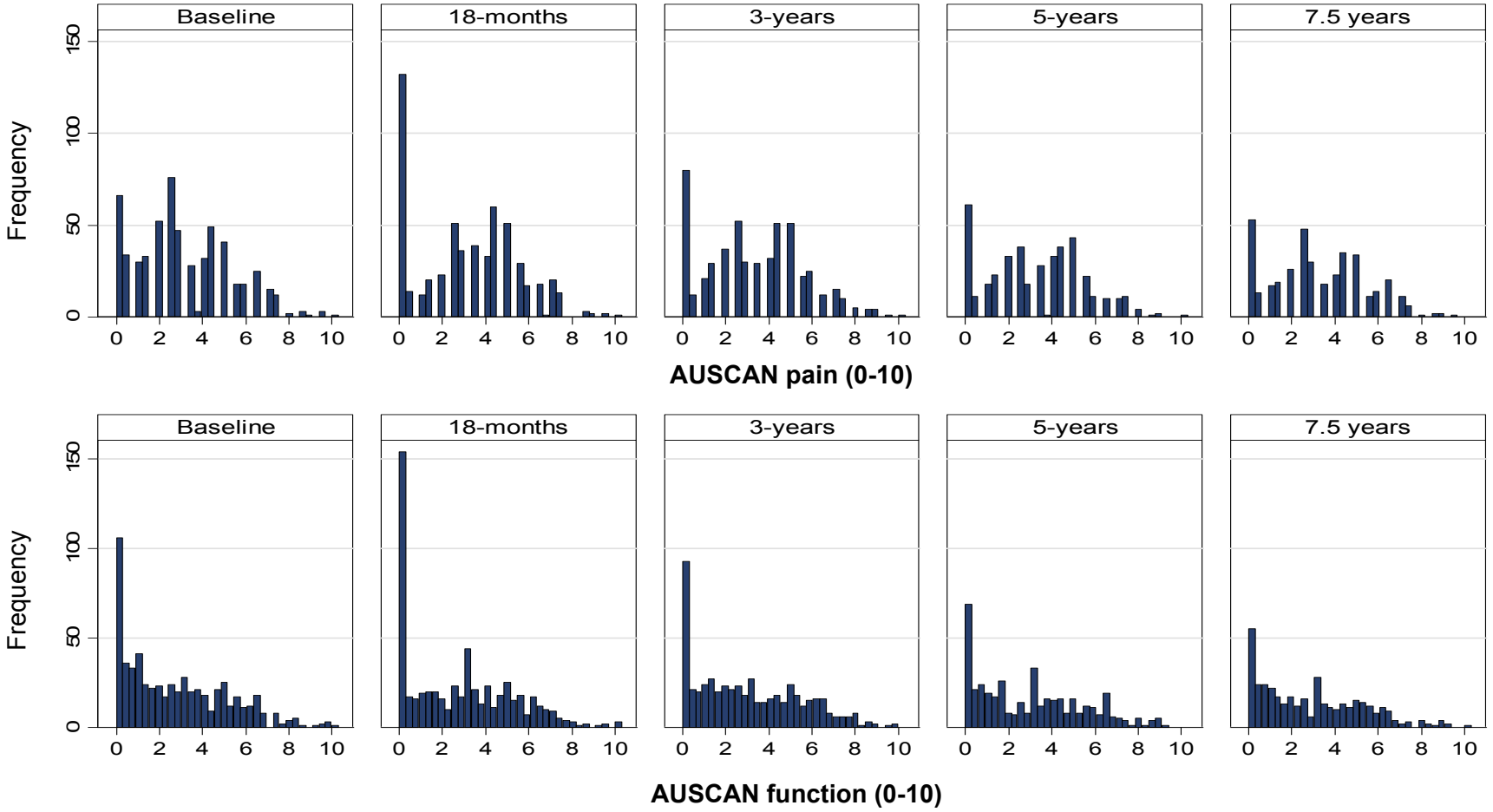
[69] The AUSCAN scale was originally published on a 0-20 scale for hand pain and on a 0-36 scale for hand function. To convert them to a 0-10 scale, hand pain was divided by 2 (i.e. 20/10) and hand function by 3.6 (i.e. 36/10). Zero values remained as zero on the converted scale.
[70] The mean (SD) of the length of follow-up at each time point was: 18-months, 18.2 (0.64); 36-months, 36.2 (2.0); 54-months, 63.4 (1.9); 72-months, 89.0 (1.6).

in the analysis is 0, 1.5, 3, 5.25 and 7.417 years, though for shorthand in the text the latter two time points are referred to as the 5 and 7.5 year follow-up, respectively.

6.3 Outcome distribution

Figure 13 shows that AUSCAN pain and function have distributions that are positively skewed at each analysis time point (i.e. not just at baseline as shown in Chapter 4), which needs to be accounted for in the analyses in this thesis. Almost all distributions have a mode of zero (i.e. no hand pain or problems), except for AUSCAN pain at baseline. If the modal group was excluded from the histogram, the two outcomes would have differing distributions: AUSCAN pain is (approximately) normally distributed whereas AUSCAN function remains positively skewed. The percentage of participants with a score of zero was greater at the 18-month follow-up than for all other time-points – a finding that occurs for both hand pain and function.

Figure 13: AUSCAN hand pain and function at each study time-point in CAS-HA



6.4 Trajectory of hand pain and function over time

The mean trajectory of hand pain and function was relatively stable over the 7.5-year follow-up time period with narrow 95% confidence intervals around the mean estimates (Figure 14 and Figure 15). There was some evidence, however, that the mean AUSCAN scores for hand pain and function increased slightly between the 18-month and 3-year follow-up time points (i.e. increase in pain and functional limitations), but this increase is small, and not observed when only those participants with complete data are included in the plot. The observed trajectories are similar when medians and interquartile ranges are considered taking the skew in the data into account (see Appendix 19).

Figure 14: Mean and 95% confidence intervals for AUSCAN pain, superimposed on individual trajectory curves

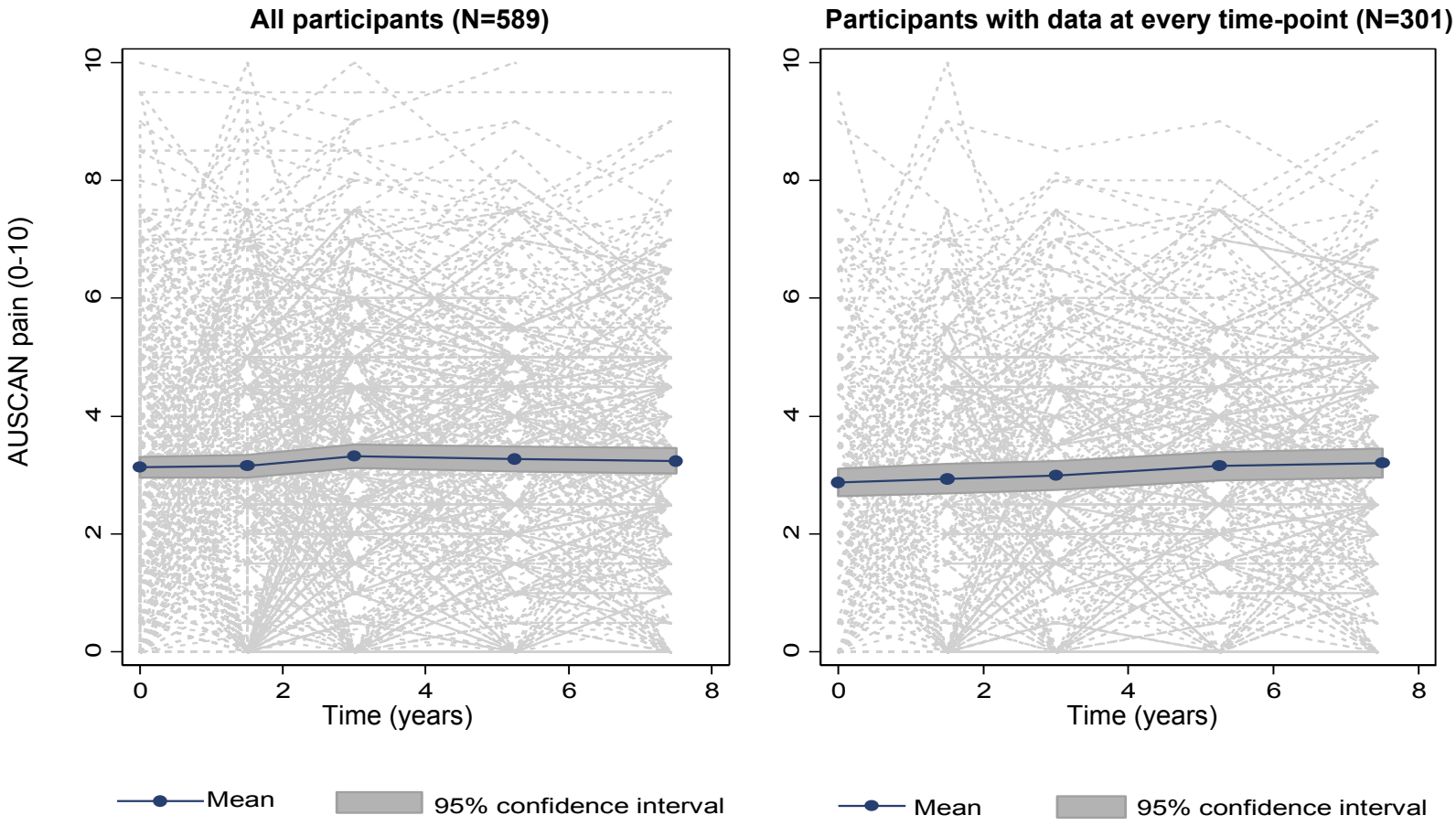
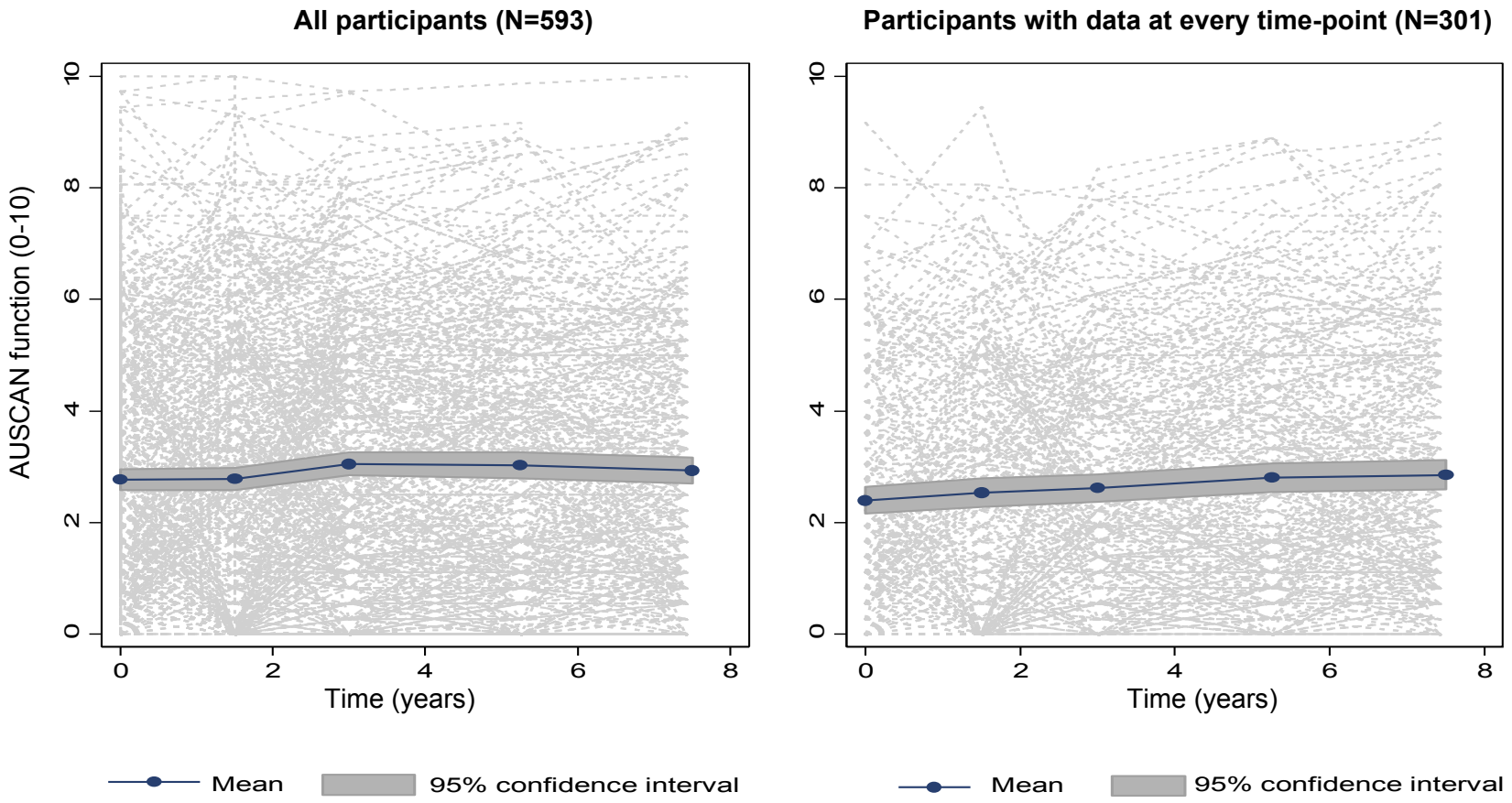


Figure 15: Mean and 95% confidence intervals for AUSCAN function, superimposed on individual trajectory curves



When the mean trajectory of hand pain was described using GEE and growth models, a linear model was found to be optimal for hand pain for both analysis techniques, with the linear model including a random intercept and slope in the growth model (Table 6-1, Table 6-2 and Appendix 20). Adding quadratic terms to each linear model, either as a fixed or random effect term for the growth models, did not significantly improve model fit. In both the GEE and growth models, the per-year rate of change was 0.05 AUSCAN points on a 0-10 scale (95% confidence interval: 0.02, 0.07) (Table 6-1 and Table 6-2).

In contrast, the quadratic model term was statistically significant for hand function when added to the linear model, although there was no significant evidence that fitting the quadratic term as a random effect gave a better model fit than fitting it as a fixed effect only, i.e. the Satorra-Bentler Scaled Chi-square Test (SBSCT) was non-significant when comparing the quadratic model with and without the random effect for the quadratic term in it (SBSCT = 4.58 (d.f. = 3); p = 0.21) (Table 6-3). The per-year rate of change was therefore more complex to interpret from these models as inclusion of the quadratic term resulted in rates of change that differed depending on the time-point of interest, e.g. the per year rate of change in the first year post-baseline was estimated as 0.14, but the equivalent figure between 6- and 7-years post baseline was only 0.01 (Table 6-4). There was no evidence that adding a cubic term to the hand function model improved model fit.

The estimated mean trajectory curves were virtually identical between the GEE and growth models for both outcomes and the GEE model results also did not differ greatly depending on whether a normal or gamma distribution was assumed for the data[71,72] (Table 6-1).

[71] Changing the link function in the GEE model also did not alter the overall model conclusions (i.e. the statistical significance of the time term in the model), however, as expected, the precise values of the parameter estimates differed as the choice of link function influenced the scaling of the outcome (data not shown).

[72] In the growth models, a separate estimate of the residual at each time point was retained in the model as it was observed that the variability in the residuals at the 18-month follow-up time point was larger than for all other time-points. The growth model results did not differ if an unstructured, independent or identity variance/covariance matrix was assumed for the random effects in the hand

pain model, however the model using the exchangeable correlation structure would not converge. For hand function, the unstructured and independent variance/covariance matrices gave similar results, however the quadratic term was not significant when the exchangeable and identity structures were used. Given the results from the unstructured model presented in Table 6-2 and Table 6-3, the assumptions of the exchangeable and identity matrices seem unrealistic, i.e. that the variance of the random effects are equal, so the results from the unstructured model only are presented

Table 6-1: GEE model results for hand pain and function

	Representation of time in the analysis		
	Linear Estimate (95% CI)	Quadratic Estimate (95% CI)	Cubic Estimate (95% CI)
Hand pain			
Normal distribution assumed			
Constant (α)	3.13 (2.97, 3.30)	3.09 (2.92, 3.25)	N/A
Time (β_1)	0.05 (0.02, 0.07)	0.10 (0.03, 0.18)	N/A
Time squared (β_2)		-0.01 (-0.02, 0.00) p =0.090	N/A
Gamma distribution assumed with an identity link function			
Constant (α)	3.13 (2.97,3.29)	3.09 (2.92, 3.26)	N/A
Time (β_1)	0.05 (0.02, 0.07)	0.10 (0.03, 0.18)	N/A
Time squared (β_2)		-0.01 (-0.02, 0.00) p = 0.108	N/A
Hand function			
Normal distribution assumed			
Constant (α)	2.79 (2.61, 2.97)	2.73 (2.54, 2.91)	2.75 (2.56, 2.94)
Time (β_1)	0.07 (0.05, 0.09)	0.15 (0.09, 0.21)	0.06 (-0.07, 0.19)
Time squared (β_2)		-0.01 (-0.02, -0.00) p = 0.006	0.02 (-0.02, 0.07)
Time cubed (β_3)			-0.00 (-0.01, 0.00) p = 0.146

Gamma distribution assumed with an identity link function			
Constant (α)	2.78 (2.60, 2.96)	2.73 (2.55, 2.91)	2.75 (2.57, 2.94)
Time (β_1)	0.08 (0.06, 0.10)	0.14 (0.08, 0.20)	0.04 (-0.09, 0.18)
Time squared (β_2)		-0.01 (-0.02, -0.00) p =0.020	0.03 (-0.02, 0.08)
Time cubed (β_3)			-0.00 (-0.01, 0.00) p = 0.080

Model estimates were generated using an exchangeable correlation structure and robust standard errors. The assumption of a common correlation at each time point was considered reasonable from inspection of the correlation matrix for AUSCAN hand pain and function (Appendix 21). The model results did not differ when an unstructured correlation structure was assumed in the data. Using the independence structure modified the findings (i.e. time in the linear model became non-significant) however this model was disregarded as the outcomes have previously been shown to be correlated across time-points (data from the independence structure model not shown).

Table 6-2: Growth models for AUSCAN pain

	Model 1 Linear: Random intercept	**Model 2** Linear: Random intercept and slope	**Model 3** Quadratic[β]: Random intercept and slope; fixed quadratic	**Model 4** Quadratic[β]: Random intercept, slope; random quadratic term
Fixed part				
Intercept (α)	3.14 (2.98, 3.31)	3.14 (2.97, 3.30)	3.09 (2.92, 3.26)	3.09 (2.92, 3.26)
Time (β_1)	0.05 (0.02, 0.07)	0.05 (0.02, 0.07)	0.10 (0.03, 0.18)	0.11 (0.04, 0.18)
Time squared (β_2)	N/A	N/A	-0.01 (-0.02, 0.00)	-0.01 (-0.02, 0.00)
Random part				
Variance				
Intercept (σ^2_{α})	3.30 (2.91, 3.69)	3.26 (2.79, 3.73)	3.27 (2.81, 3.74)	2.98 (2.32, 3.64)
Slope ($\sigma^2_{\beta 1}$)	N/A	0.02 (0.00, 0.03)	0.02 (0.01, 0.03)	0.06 (-0.09, 0.22)
Quadratic ($\sigma^2_{\beta 2}$)	N/A	N/A	N/A	0.00[α] (-0.00, 0.00)
Covariance[γ]				
Intercept and slope ($\sigma^2_{\alpha\beta 1}$)	N/A	-0.02 (-0.08, 0.04)	-0.02 (-0.08, 0.04)	0.13 (-0.14, 0.41)
Intercept and quadratic ($\sigma^2_{\alpha\beta 2}$)	N/A	N/A	N/A	-0.02 (-0.05, 0.01)
Slope and quadratic ($\sigma^2_{\beta 1\beta 2}$)	N/A	N/A	N/A	-0.01 (-0.03, 0.01)
Residual – variance ($\sigma^2_{\varepsilon t}$)				
Baseline	1.92 (1.61, 2.23)	1.69 (1.35, 2.03)	1.68 (1.35, 2.02)	1.79 (1.23, 2.34)
18-months	2.42 (1.98, 2.87)	2.40 (1.96, 2.84)	2.40 (1.95, 2.84)	2.35 (1.89, 2.81)

3-years	1.62 (1.34, 1.90)	1.65 (1.37, 1.94)	1.65 (1.37, 1.92)	1.47 (1.16, 1.78)
5-years	1.54 (1.26, 1.82)	1.43 (1.16, 1.70)	1.43 (1.16, 1.70)	1.44 (1.12, 1.76)
7.5-years	1.78 (1.44, 2.12)	1.36 (0.96, 1.75)	1.35 (0.95, 1.74)	1.05 (0.19, 1.91)

SBSCT

Model 1 vs Model 2	χ^2 = 14.33 (d.f. = 2); p<0.001		
Model 2 vs Model 3		χ^2 = 2.91 (d.f = 1); p=0.09	
Model 2 vs Model 4			χ^2 = 9.01 (d.f = 4); p=0.06

Unless otherwise stated, figures are parameter estimates and 95% confidence intervals (based on robust standard errors) in brackets. α = The estimate is the same as the upper or lower confidence interval limit due to rounding of model estimates, β = The quadratic model includes within it a term for time (t) and a term for time-squared (t^2) $^{\gamma}$ = co-variances are given in the table rather than correlations, so it is not appropriate to test whether the correlation is significant by assessing whether the 95% confidence interval for the covariance spans zero. p = p-value, N/A = not applicable, d.f. = degrees of freedom, SBSCT = Satorra-Bentler Scaled Chi-square Test

Table 6-3: Growth models for AUSCAN function

	Model 1 Linear: Random intercept	**Model 2** Linear: Random intercept and slope	**Model 3** Quadratic[β]: Random intercept and slope; fixed quadratic	**Model 4** Quadratic[β]: Random intercept, slope; random quadratic term
Fixed part				
Intercept (α)	2.81 (2.62, 2.99)	2.80 (2.62, 2.98)	2.74 (2.55, 2.92)	2.74 (2.55, 2.92)
Time (β_1)	0.07 (0.05, 0.09)	0.08 (0.05, 0.10)	0.15 (0.09, 0.21)	0.15 (0.09, 0.22)
Time squared (β_2)	N/A	N/A	-0.01 (-0.02, -0.00)	-0.01 (-0.02, -0.00)
Random part				
Variance				
Intercept (σ^2_{α})	4.74 (4.24, 5.23)	4.51 (3.96, 5.06)	4.54 (3.99, 5.09)	4.51 (3.75, 5.26)
Slope ($\sigma^2_{\beta 1}$)	N/A	0.02 (0.00, 0.03)	0.02 (0.00, 0.03)	0.09 (-0.03, 0.22)
Quadratic ($\sigma^2_{\beta 2}$)	N/A	N/A	N/A	0.00[α] (-0.00, 0.00)
Covariance[γ]				
Intercept and slope ($\sigma^2_{\alpha\beta 1}$)	N/A	0.02 (-0.04, 0.08)	0.02 (-0.04, 0.08)	0.04 (-0.23, 0.31)
Intercept and quadratic ($\sigma^2_{\alpha\beta 2}$)	N/A	N/A	N/A	-0.01 (-0.04, 0.02)
Slope and quadratic ($\sigma^2_{\beta 1\beta 2}$)	N/A	N/A	N/A	-0.01 (-0.03, 0.01)
Residual – variance ($\sigma^2_{\varepsilon t}$)				
Baseline	1.31 (1.04, 1.59)	1.14 (0.82, 1.46)	1.13 (0.81, 1.45)	0.99 (0.47, 1.50)
18-months	1.88 (1.48, 2.27)	1.83 (1.45, 2.22)	1.84 (1.45, 2.23)	1.86 (1.47, 2.26)

3-years	0.94 (0.72, 1.16)	0.96 (0.74, 1.18)	0.95 (0.74, 1.16)	0.85 (0.61, 1.10)
5-years	1.09 (0.84, 1.34)	0.99 (0.74,1.25)	0.99 (0.74, 1.24)	0.94 (0.69, 1.19)
7.5-years	1.36 (1.06, 1.66)	1.05 (0.71, 1.40)	1.03 (0.69, 1.38)	1.01 (0.34,1.69)

SBSCT

Model 1 vs Model 2	χ^2 = 14.59 (d.f. = 2); p < 0.001			
Model 2 vs Model 3		χ^2 = 7.84 (d.f. = 1); p = 0.01		
Model 2 vs Model 4			χ^2 = 10.91 (d.f. = 4); p = 0.03	
Model 3 vs Model 4				χ^2 = 4.58 (d.f. = 3); p = 0.21$^{\delta}$

Unless otherwise stated, figures are parameter estimates and 95% confidence intervals (based on robust standard errors) in brackets. α = The estimate is the same as the upper or lower confidence interval limit due to rounding of model estimates, β = the quadratic model includes within it a term for time (t) and a term for time-squared (t^2), γ = Co-variances are given in the table rather than correlations, so it is not appropriate to test whether the correlation is significant by assessing whether the 95% confidence interval for the covariance spans zero, δ = this p-value should (in theory) be divided by two to reflect a one-sided test as a variance estimate should never be negative (Rasbash et al. 2012). p = p-value, N/A = not applicable, SBSCT = Satorra-Bentler Scaled Chi-square Test

Table 6-4: Per-year rates of change over time for hand function

Time in years	Mean predicted value[α]	Per year rate of change
0	2.73	
1	2.87	0.14
2	2.99	0.12
3	3.08	0.09
4	3.15	0.07
5	3.21	0.05
6	3.23	0.03
7	3.24	0.01

α Mean predicted value = 2.73 + 0.15*time - 0.01*time*time. Coefficients are taken from the quadratic GEE model however results are virtually identical when the growth models are used.

Although it is shown that the mean trajectories for AUSCAN pain and function are relatively stable over the time-period of the study, the growth models also highlight that there is significant individual variation around the fixed model estimates. This suggests that there is individual variability in the trajectories of hand pain and function over time, which is highlighted by plotting individual trajectories over time (Figure 16)[73] and generating histograms of the standard deviation of participants' AUSCAN scores over time (Figure 17)[74]. If trajectories are stable over time then their standard deviation would be 0, which they are not for AUSCAN pain and function. In addition, the low correlation between the random intercept and slope in the hand pain and function models show that the rate of change over time is not dependent on a participants' baseline starting value, i.e. the correlation between the random intercept and slope in Model 2 for hand pain was -0.08 (95% confidence interval -0.29, 0.14) and in Model 3 for hand function was 0.07 (95% confidence interval -0.18, 0.33).

[73] Some participants showed quite varied AUSCAN scores over time (e.g. participant 30176), whereas others showed more stable patterns over time (e.g. participant 30384) (Figure 16)
[74] If participants have data at all five time-points the unit plotted would be the standard deviation of the AUSCAN at those five time-points

Figure 16: Individual trajectory plots of AUSCAN hand pain and function over time for a sample of 24 participants in CAS-HA

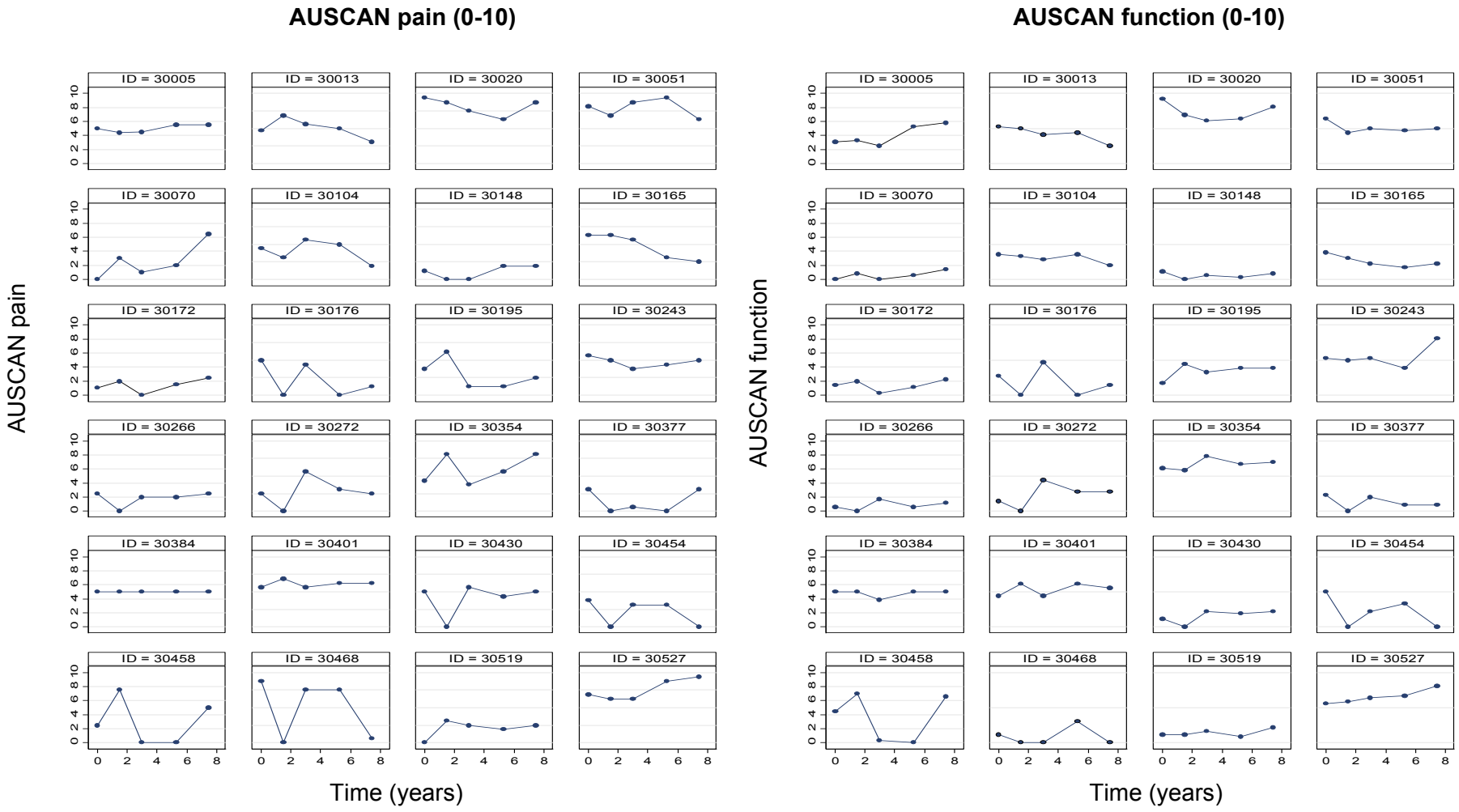
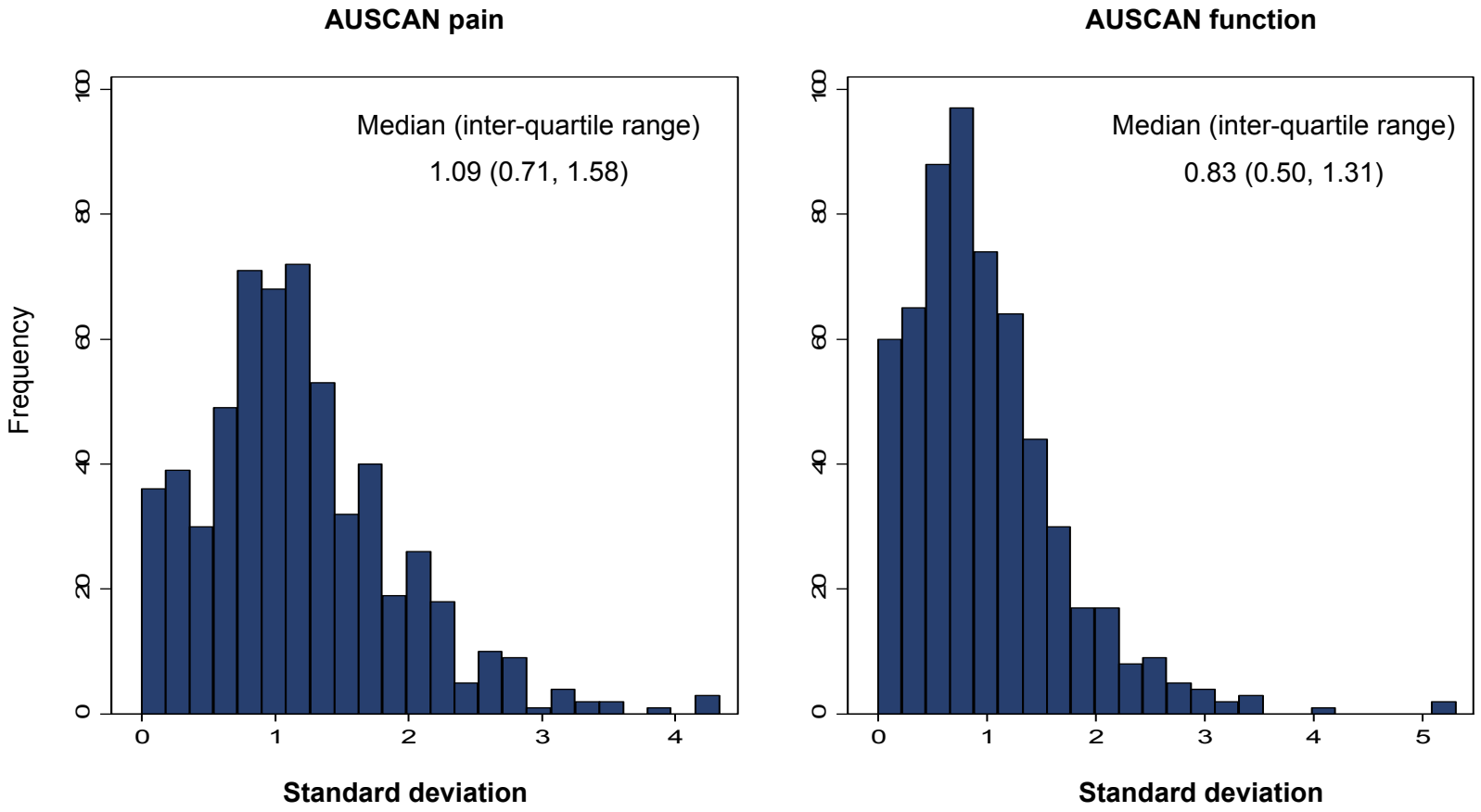


Figure 17: Distribution of individual standard deviations calculated across time for each person in the CAS-HA sample



6.5 AUSCAN missing data patterns and loss to follow-up

6.5.1 Missing data patterns

Missing data rates for AUSCAN pain at each time point are 5, 7, 16, 33 and 38% respectively (Table 6-5 – denominator N = 623) and are nearly identical to those for AUSCAN function: 5, 8, 14, 33 and 38% (Appendix 22). AUSCAN pain data was present at all time-points for 50% of participants and at 3 or more time points (i.e. for more than half of the time-points measured) for 85% of participants. The most common pattern of missing data occurred for participants who were in the study up to the three year follow-up but then did not complete the 5 and 7.5 year follow-up (N=79). Similar missing data patterns were found for AUSCAN function (Appendix 22).

Table 6-5: Missing data patterns for AUSCAN pain

	Missing data pattern (X=data present)					Number of time points with non-missing data
	Baseline	18-months	3-years	5-years	7.5-years	
	N=589	N=577	N=523	N=417	N=384	
N (%)						
311 (50)	X	X	X	X	X	5
79 (13)	X	X	X			3
59 (9)	X	X	X	X		4
51 (8)	X	X				2
32 (5)	X	X	X		X	4
29 (5)	X					1
19 (3)		X	X	X	X	4
10 (2)	X	X		X	X	4
7 (1)	X		X	X	X	4
6 (1)		X	X			2
3 (<1)	X	X		X		3
3 (<1)	X		X	X		3
3 (<1)	X		X			2
2 (<1)						0
2 (<1)		X	X	X		3
2 (<1)		X				1
1 (<1)		X	X		X	3
1 (<1)	X			X	X	3
1 (<1)			X	X	X	3
1 (<1)		X		X	X	3
1 (<1)	X	X			X	3

Footnote: 50%, 20%, 15%, 10% and 5% of participants had data at 5, 4, 3, 2, 1 time points respectively

6.5.2 Loss to follow-up

Loss to follow-up of the CAS-HA population as a whole was considered in Chapter 3 where it was shown that the baseline characteristics of participants in the baseline cohort were largely similar to those remaining in the cohort at the 7.5-year follow-up time point (with the exception that those remaining in the cohort at 7.5-years were of marginally better health). In addition, in this chapter, the data in Table 6-6 show that participants responding at time point t on average have lower AUSCAN scores at time t-1 than those who did not respond at time point t, suggesting higher levels of pain and function in those with missing data or lost to follow-up at the next time point. Therefore, missing data cannot be assumed to be missing completely at random (MCAR) as its presence depends on observed variables in the dataset[75].

[75] Although the data are shown to depend on the AUSCAN this is still a test of whether the data are missing at random (MAR) as the data at the preceding time-point is an observed variable. It is difficult to test whether data are missing not at random (MNAR) as this relates to knowing the values of the precise outcome data that has not been collected as part of the study (Twisk et al. 2013)

Table 6-6: Assessing selective loss to follow-up with case selection based on AUSCAN pain and function (0-10)

	AUSCAN pain present at this time point?							
	18-months		3-years		5-years		7.5-years	
	Yes	No	Yes	No	Yes	No	Yes	No
	N =577	N = 46	N =523	N = 100	N = 417	N = 206	N=384	N =239
AUSCAN pain								
Baseline	2.5 (1.5, 4.5)	3.5 (1.5, 6.0)						
18-months			3.0 (0.5, 5.0)	4.0 (0.8, 5.0)				
3-years					3.0 (1.5, 5.0)	4.0 (2.0, 5.0)		
5-years							3.0 (1.5, 4.5)	4.5 (2.0, 5.0)
	AUSCAN function present at this time point?							
	18-months		3-years		5-years		7.5-years	
	Yes	No	Yes	No	Yes	No	Yes	No
	N =576	N = 47	N = 537	N = 86	N = 418	N = 205	N=384	N =239
AUSCAN function								
Baseline	2.2 (0.8, 4.4)	3.9 (1.1, 6.4)						
18-months			2.5 (0.3, 4.7)	3.1 (0.0, 4.7)				
3-years					2.5 (0.8, 4.7)	3.6 (1.7, 6.1)		
5-years							2.5 (0.8, 4.4)	3.9 (0.8, 6.1)

Values are medians (interquartile ranges)

6.6 Minimum important change (MIC)

Although MIC for the AUSCAN was considered in Chapter 4, one single value was not recommended for MIC as it was not possible to derive a reliable MIC value for the CAS-HA sample using the methods described therein. To respond to this, two further sources of evidence were considered, which, in themselves, are not methods to calculate MIC, but may support interpretation of the data and help to consider how “large” change in the AUSCAN should be before meaningful change has occurred. The two sources of evidence were population normative data derived for the AUSCAN (Bellamy et al. 2011) and treatment responder criteria as based on the AUSCAN and a global assessment of change measure (Pham et al. 2004).

6.6.1 Population normative data

Population data was initially considered to try and parallel the concept of normative data used in other disease areas: e.g. to test how many people starting with high blood pressure then resolve to being in the “normal range” for adults in their age and gender group. This concept could not be easily translated to the population data for the AUSCAN as these data have been based on a mixture of participants with and without hand pain. Given this, population data are only used in this thesis to test the expectation that participants in CAS-HA have higher AUSCAN scores than a general population sample (which is supported by data in Table 6-7 and Table 6-8), and also to test, in general, by how many AUSCAN points their problems are more severe (differences in median AUSCAN scores between the normative and CAS-HA data are typically around 2 to 3 points for AUSCAN pain and 1 to 2 points for AUSCAN function (Table 6-7 and Table 6-8)).

Table 6-7: Comparing AUSCAN pain measures in CAS-HA to percentile values derived from the normal population

	Normative data				CAS-HA																		
					Baseline				18-months				3-years				5-years				7.5-years		
Males	N$^{\alpha}$	50	75	90	N	50	75	90	N	50	75	90	N	50	75	90	N	50	75	90	N	50	75	90
50-54	296	0.0	1.1	3.2	17	2.5	4.0	6.5	11	2.0	4.0	4.5	7	3.0	6.0	7.5	-	-	-	-	-	-	-	-
55-59	385	0.0	1.4	3.8	50	2.5	4.5	7.3	42	2.0	4.0	5.0	24	2.5	4.8	6.0	8	0.5	2.8	4.5	5	2.5	3.5	4.5
60-64	333	0.0	1.0	3.6	60	2.8	5.0	6.0	57	3.0	4.5	6.9	51	3.5	5.0	6.0	36	2.5	4.5	5.5	22	2.8	4.5	6.5
65-69	237	0.0	1.6	4.8	37	2.5	4.4	5.0	47	3.5	5.0	6.0	46	3.3	4.5	5.5	48	3.0	4.5	5.5	43	2.5	3.5	5.0
70-74	298	0.1	1.6	3.6	24	2.5	3.0	6.0	25	2.5	4.0	5.5	27	2.5	4.0	5.0	25	2.5	5.0	5.5	34	2.5	4.5	6.0
75-79	259	0.0	1.6	3.8	21	3.0	5.0	6.0	22	3.0	5.0	6.5	25	2.5	5.0	6.5	19	2.5	4.0	5.0	18	2.3	4.0	5.0
80+	687	0.4	2.6	5.0	11	2.5	5.5	7.0	17	3.0	5.5	7.5	19	4.0	5.0	6.0	19	3.5	6.0	7.5	23	3.0	5.0	6.0
Females																								
50-54	308	0.0	1.4	3.8	62	2.5	4.5	6.5	34	2.5	4.0	4.5	10	3.3	4.5	6.3	-	-	-	-	-	-	-	-
55-59	296	0.4	2.4	5.0	88	2.5	4.3	6.0	84	4.0	5.0	6.5	72	3.5	5.0	7.0	40	2.5	4.5	5.5	16	2.5	4.3	5.5
60-64	367	0.6	2.8	5.8	67	4.0	5.5	6.5	70	3.5	5.0	6.5	72	3.3	5.0	7.0	75	3.5	5.0	6.5	58	3.3	5.0	6.5
65-69	277	0.6	2.6	5.0	57	2.5	5.0	6.0	62	3.5	4.5	5.5	56	3.5	5.0	6.0	45	4.0	5.0	7.0	60	3.3	5.0	6.5
70-74	350	0.6	2.8	6.2	55	3.0	4.5	5.5	58	4.0	5.5	7.0	59	3.0	5.0	6.5	51	3.5	5.0	6.0	44	3.0	5.0	7.0
75-79	269	0.8	3.0	5.4	28	3.3	5.0	6.0	33	4.5	5.0	6.0	41	4.0	5.0	5.5	34	4.0	5.0	6.5	36	4.0	5.0	7.0
80+	666	1.2	4.2	6.6	12	4.0	5.5	6.5	15	2.5	4.5	7.0	14	4.3	5.5	8.0	17	5.0	5.5	7.5	25	5.0	5.5	6.5

Normative data reproduced from (Bellamy et al. 2011). Dash indicates that no participants met the criteria to be in the group due to study design. Note that age has been recalculated at each time point therefore participants can change age-group over time. α = Stratified sampling was used to gain the same level of statistical precision for each age and gender strata, sample size therefore does not reflect the age distribution in the general population. N = number of participants, 50 = 50th percentile; 75 = 75th percentile, 90 = 90th percentile.

Table 6-8: Comparing AUSCAN function measures in CAS-HA to percentile values derived from the normal population

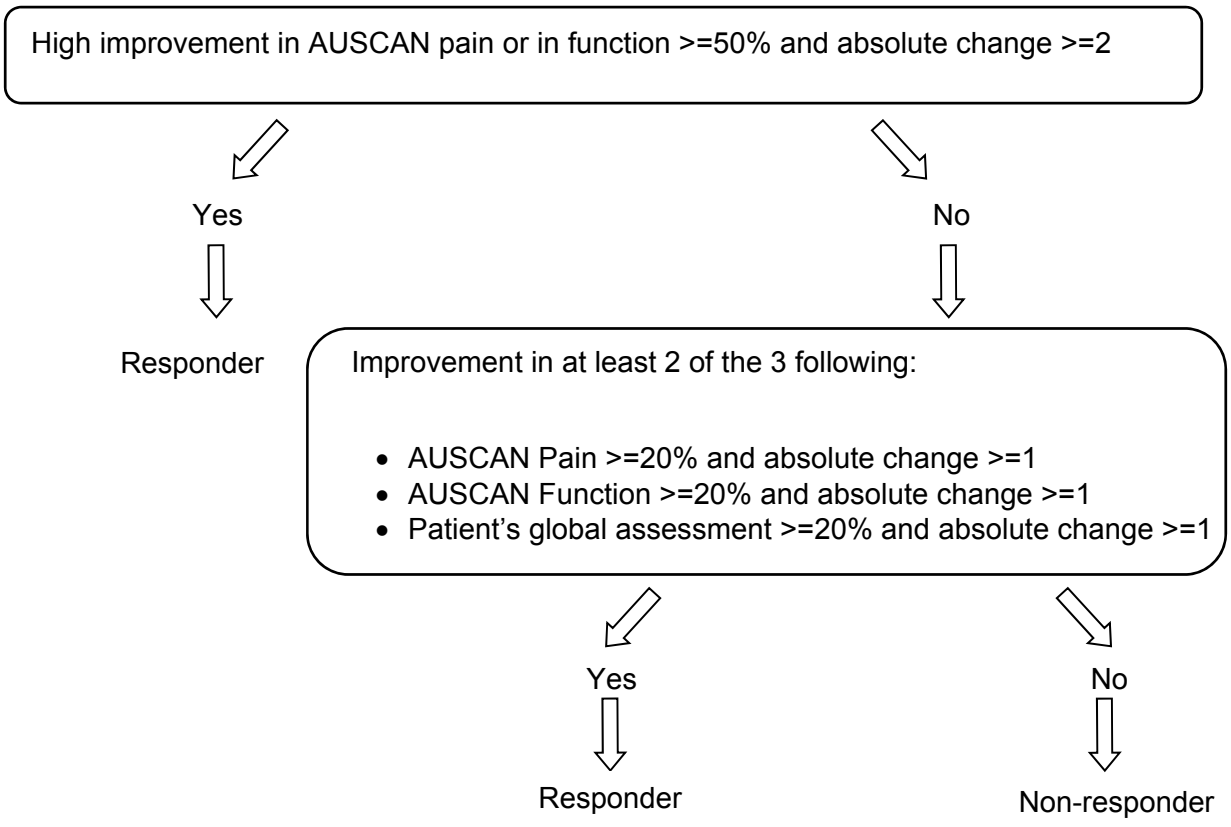
	Normative data				CAS-HA																			
					Baseline				18-months				3-years				5-years				7.5-years			
Males	N[α]	50	75	90	N	50	75	90	N	50	75	90	N	50	75	90	N	50	75	90	N	50	75	90
50-54	294	0.0	0.7	2.4	17	0.6	1.4	5.3	11	0.6	1.9	2.2	7	0.8	1.7	6.7	-	-	-	-	-	-	-	-
55-59	382	0.0	0.9	3.1	51	1.4	4.7	6.9	42	0.3	2.8	5.3	24	1.3	3.5	5.0	8	0.4	1.8	3.9	5	0.8	0.8	1.7
60-64	329	0.1	1.1	3.3	61	0.8	2.8	5.0	57	1.1	2.8	5.3	53	1.4	3.6	5.7	36	1.4	3.2	5.6	22	1.3	3.9	4.7
65-69	236	0.2	1.7	4.6	37	1.4	2.8	5.0	47	2.2	3.9	5.6	48	1.5	4.0	5.8	48	1.4	3.9	5.0	44	1.0	2.8	5.0
70-74	296	0.3	1.4	3.3	24	1.6	3.5	5.6	25	0.8	3.1	4.4	26	1.3	2.5	3.9	26	1.7	3.9	6.1	34	1.7	3.1	5.0
75-79	256	0.4	1.7	3.8	22	1.4	5.0	6.4	22	2.9	5.0	6.1	28	2.5	5.0	7.5	19	1.9	4.2	5.3	18	1.5	3.6	4.4
80+	669	1.0	3.4	6.0	11	3.6	4.7	5.0	16	1.8	4.9	5.0	19	3.6	5.0	6.3	20	3.5	5.8	7.4	23	2.5	4.7	6.1
Females																								
50-54	307	0.2	1.4	4.2	62	2.1	4.7	6.4	34	1.3	3.1	5.6	10	2.5	4.7	5.7	-	-	-	-	-	-	-	-
55-59	290	0.7	2.8	5.9	87	2.2	3.9	5.6	84	3.3	4.9	6.1	73	2.5	5.0	7.2	40	2.5	4.0	5.8	16	1.9	3.5	3.9
60-64	361	0.9	3.8	6.4	66	4.0	5.8	6.4	70	2.9	5.0	6.8	74	3.1	5.3	6.7	74	2.9	4.4	6.1	58	2.9	5.6	6.7
65-69	271	1.0	3.3	5.9	57	2.8	4.2	6.4	62	2.9	4.7	6.1	58	3.8	5.8	6.4	44	3.9	5.8	6.7	60	2.8	5.3	6.1
70-74	339	1.3	3.9	7.2	57	3.3	5.0	6.1	58	3.9	5.6	7.2	62	3.1	5.3	6.1	49	3.1	5.0	6.9	44	3.3	5.0	5.8
75-79	267	1.6	4.2	6.7	28	4.3	5.7	6.9	34	4.2	5.0	7.2	40	4.4	5.3	6.3	36	3.9	6.4	7.5	35	3.6	5.6	8.1
80+	650	2.7	6.1	8.3	13	2.8	4.7	7.5	14	2.4	5.6	6.1	15	5.6	7.8	8.1	18	5.6	6.4	8.6	25	5.0	5.6	7.5

Normative data reproduced from (Bellamy et al. 2011). Dash indicates that no participants met the criteria to be in the group due to study design. Note that age has been recalculated at each time point therefore participants can change age-group over time. α = Stratified sampling was used to gain the same level of statistical precision for each age and gender strata, sample size therefore does not reflect the age distribution in the general population. N = number of participants, 50 = 50th percentile; 75 = 75th percentile, 90 = 90th percentile.

6.6.2 Osteoarthritis Research Society International (OARSI) treatment responder criteria

The OARSI treatment responder criteria are an alternative source that can be used to determine if participants in CAS-HA show AUSCAN changes over time that are "large" enough to be considered a "response" to treatment in a clincial trial setting. Although intially developed for participants with knee and hip osteoarthritis (OA), the criteria have been used as the primary outcome measure in several trials of hand OA (Stukstette et al. 2013, Dziedzic et al. 2013) so are relevant to consider in this thesis. The OARSI responder criteria work by allocating participants into groups based on the algorithm shown in Figure 18 (reproduced from Pham et al. 2004 and adapted to relate to the AUSCAN measured on a 0-10 scale in this thesis[76]).

Figure 18: OARSI responder criteria for hand pain and function measured on a 0-10 scale



[76] The OARSI responder criteria presented in Pham et al. 2004 related to an outcome that was measured on a 0-100 scale. All absolute values were therefore divided by 10 to make them relevant to the 0-10 AUSCAN scale used in this thesis.

Although the OARSI responder criteria are designed to measure improvement in symptoms over time, both symptom 'improvement' and 'deterioration' are considered in this chapter as both are of interest, and could occur, in the CAS-HA study[77]. In particular, for hand pain, on average, 23% and 30%[78] of participants met the 20% cut-off component of the OARSI responder criteria for improvement or deterioration respectively (range 20 - 25% for improvement and 28 – 32% for deterioration) (Table 6-9). These figures are reduced to 8% and 15% when the 50% cut-off was applied, i.e. high improvement/deterioration. The corresponding figures for hand function are slightly lower than for hand pain. This suggests that although the overall trajectory of hand pain and function appears stable over time, there are still significant proportions of participants showing changes over time that are large enough to meet criteria for improvement or deterioration according to this algorithm, suggesting that there is variability in the sample to be further explored.

[77] Global assessment of change is not being considered in this chapter as the AUSCAN is the focus of analysis for this thesis

[78] Figure obtained by averaging data across the four sets of adjacent time points

Table 6-9: Participants in CAS-HA meeting the hand pain and function components of the OARSI responder criteria

	Time period				
Criteria	BL to 18-months N(%)	18-months to 3-years N(%)	3-years to 5-years N(%)	5-years to 7.5 years N(%)	BL to 7.5 years N(%)
Improvement					
Pain					
>=20% and absolute change >=1	138 (25)	101 (20)	88 (22)	83 (24)	79 (22)
>=50% and absolute change >=2	61 (11)	35 (7)	28 (7)	24 (7)	32 (9)
Function					
>=20% and absolute change >=1	95 (17)	77 (15)	60 (14)	60 (17)	51 (14)
>=50% and absolute change >=2	30 (5)	14 (3)	11 (3)	16 (5)	15 (4)
Pain and Function					
>=20% and absolute change >=1 for both measures	68 (13)	41 (8)	32 (8)	40 (12)	37 (10)
Deterioration					
Pain					
>=20% and absolute change >=1	177 (32)	148 (29)	124 (31)	98 (28)	133 (36)
>=50% and absolute change >=2	106 (19)	79 (16)	57 (14)	37 (11)	68 (19)
Function					
>=20% and absolute change >=1	129 (23)	124 (24)	93 (22)	68 (20)	116 (32)
>=50% and absolute change >=2	51 (9)	45 (9)	28 (7)	21 (6)	49 (13)
Pain and Function					
>=20% and absolute change >=1 for both measures	80 (15)	71 (14)	49 (12)	40 (12)	62 (17)

Percentage change calculated as (baseline – follow-up)/baseline; absolute change (baseline – follow-up). AUSCAN scaled from 1 to 11 to avoid dividing by 0 when calculating relative change over time. BL = baseline

6.7 Discussion

6.7.1 Summary of findings

In this chapter it has been shown that both AUSCAN pain and function have skewed outcome distributions at each time point and that the mean/median trajectory of hand pain and function is relatively stable over time. The growth and GEE models show a linear trend for hand pain and a quadratic trend for hand function, however when viewed against the range of the AUSCAN scale, and also against the suggested values for MIC in Chapter 4, the small annual increases in hand pain and function, although statistically significant, at the mean level, may not be “large” enough to be considered clinically meaningful to the participant.

By considering the data at the individual level, it has been shown that a greater proportion of participants show symptom deterioration over time than improvement, with the 20% cut-off used in the algorithm for the OARSI responder criteria being similar to the differences that were considered as potential values for MIC using the distribution method in Chapter 4. This data, along with the statistical models and descriptive plots, demonstrate a wide degree of variability both in baseline values of AUSCAN pain and function and also in the shapes of the trajectories over time, and support the objectives of the later thesis chapters to potentially explain such variability using (measured) predictors of interest or reflect different (unobserved) latent groups (as is later explored in Chapters 7 and 8, respectively).

6.7.2 Comparison with the literature

The AUSCAN has been used as an outcome in several community-based studies of hand pain and functional limitation, e.g. (Cole et al. 2011, Aslam et al. 2014), however few of these studies were comparable to CAS-HA by measuring change in hand pain and function over a longer-term follow-up time period. Two exceptions were the Genetics ARthrosis and Progression (GARP) study (Bijsterbosch et al. 2011) and the Oslo hand OA

cohort (Haugen et al. 2013) that measured changes in the AUSCAN from baseline to a 6 – and 7-year follow-up, respectively.

Although these studies recruited patients with a diagnosis of symptomatic hand OA from patient records (so are not completely comparable to CAS-HA) both studies reported similar findings to those shown in this chapter, i.e. that on average over a 6- to 7-year time period, the change in mean AUSCAN hand pain and functional difficulty is small (Bijsterbosch et al. 2011, Haugen et al. 2013) and that individual variation around these estimates is large (Bijsterbosch et al. 2011). Both studies also used pre-specified cut-offs for minimal important change and found that a greater percentage of participants showed worsening of hand pain and functional difficulty over time than improvement, which is consistent with the analysis of the OARSI responder criteria in this chapter (i.e. the analysis that looks at change between baseline and the 7.5 year follow-up time-point)[79].

The distribution of the AUSCAN presented in this chapter differs from other studies that have used differing methods to recruit study participants. For example the AUSCAN scores were normally distributed, rather than positively skewed, in the Genetics of Generalized Osteoarthritis (GOGO) study that recruited participants who had at least two siblings with bilateral hand OA defined from clinical assessment and x-ray scoring (Allen et al. 2006a). This is to be expected however, as the CAS-HA study recruited participants from a population-based sample reporting hand pain or problems in the last 12-months rather than from a clinical setting. It is therefore likely that several participants would have a score of zero on the AUSCAN as the time frame for this measure is “in the last week” rather than “in the last 12-months”. Also, as participants were not recruited from a clinical

[79] In the GARP study, 40% and 50% of participants showed worsening of hand pain and function over time respectively versus 26% showing improvement for each of hand pain and function. Worsening and improvement was defined as having a change score in excess of the positive (for worsening) and negative values (for improvement) of the minimum clinically important change values of 1.49 and 1.25 for hand pain and function respectively (i.e. 7 and 3% of the scale scores respectively when hand pain is measured on a scale of 0-20 and hand function on a scale of 0-36) (Bijsterbosch et al. 2011). The same cut-offs were used in the Oslo hand OA cohort and the corresponding figures for hand pain and function were 39% and 47% of patients for worsening and 21% and 32% for improvement (Haugen et al. 2013)

setting, they represent a broad spectrum of participants with a broad range of condition severity, so it may be that some participants with milder problems resolved to present with no pain at the time the questionnaire was completed by participants. This highlights the need for a cautious interpretation when generalizing findings from the CAS-HA study to other populations of interest.

The AUSCAN response rates in the CAS-HA study were reasonable when compared to other studies that have also used the AUSCAN as the primary measure of hand pain and function, e.g. the GARP and Oslo hand OA cohort (AUSCAN response rate: CAS-HA at 7.5 years = 62%, GARP at 6-years = 76% (Bijsterbosch et al. 2011), Oslo cohort at 7-years = 52% (Haugen et al. 2013)).

6.7.3 Limitations

A limitation of the data presented so far is the potential for selective loss-to-follow as, for both outcomes, there is a small increase in mean hand pain and functional difficulty between the 18-month and 3-year follow-up that is not replicated when the plot is repeated for only those participants with data at all time-points. This could imply that participants with more severe problems that have a chronic time course, are those that have remained in the cohort, however this is not in line with the finding that the AUSCAN score at the time-point prior to exiting the study is greater for those lost-to-follow-up at that time-point than those that remained in the study. Selective loss to follow-up is also an issue as only 62% of the original CAS-HA cohort completed the AUSCAN questionnaire at the 7.5-year follow-up. This has previously been explored in Chapter 3 on study design and recruitment.

6.7.4 Model extensions

In this chapter, the trajectory of hand pain and function over time has been described, and although it has been explored whether the baseline values of AUSCAN pain and function are related to rates of change over time, the models do not address any research

questions around which baseline factors are important predictors of outcome change over time. It is this topic that is explored in more detail in the next chapter (Chapter 7).

7 Predicting the course of hand pain and function over time

7.1 Introduction

In Chapter 6, GEE and growth models were used to model the trajectory of hand pain and function and, for both outcomes, their respective optimal growth models showed significant (unexplained) variation around the model fixed effects. The objective of this chapter is therefore to explore whether such variation can be explained by extending the models in Chapter 6 to include factors (or combinations of factors) measured at baseline that could potentially predict prognosis of hand pain and function over time.

As the GEE and growth models in Chapter 6 showed very similar results, the analysis in this chapter is only conducted using growth modelling. Growth models were chosen over GEE models as they give additional information on outcome variation between individuals and also because the missing data in CAS-HA is unlikely to be missing completely at random.

This chapter is structured into three key sections: the first to describe the modelling strategy used to determine which factors, out of those listed in Chapter 3, are strong predictors of hand pain and function over time (Section 7.2), the second to present the model results (Section 7.3) and the third to discuss the findings, strengths and weaknesses of the analysis and implications for later thesis chapters (Section 7.4).

7.2 Modelling strategy

7.2.1 Outcome and trajectory shape

In this chapter, hand pain and function were modelled as two separate outcomes. A linear trajectory (with a random intercept and slope) was initially used to model both outcomes to aid comparability of results between the two outcomes and to simplify the modelling process. The justification for, and impact of, this simplification is discussed further in Section 7.2.4.

7.2.2 Measurement of predictors

The same core set of potential predictors as proposed previously (Chapter 3) was used to model hand pain and function however “baseline hand pain” and “baseline hand function” were removed from the hand pain and hand function models respectively to avoid duplication of outcome and predictor information. The baseline for the outcome of interest was still included in the model however, but just expressed as the degree of variability around the model intercept term rather than as a predictor *per se*. The list of potential predictors was presented in detail in Table 3-1 in Chapter 3, so is summarised only briefly here in Figure 19.

Figure 19: Summary of potential predictors considered for inclusion in the hand pain and hand function models

Block 1 – Demographic
Age Gender Marital Status Occupation/Social Class Employment status Education Income
Block 2 – Lifestyle
Alcohol consumption Smoking status Social networks
Block 3 – Health
Self-rated health Number of comorbidities Pain in other body areas
Block 4 - Characteristics of hand condition
Hand pain severity Hand functional difficulty Side affected Time since onset of hand problem Sudden onset of hand problem Onset of hand condition following accident/injury to the hand Physical load on hands during work and leisure

Block 5 – Psychological factors
Anxiety Depression Illness perceptions Frustration with hand condition
Block 6 – Clinical Assessment
Body-mass index Hand grip-ability Muscle strength Hand osteoarthritis Carpal tunnel syndrome Dupuytren's contracture De Quervain's tenosynovitis Trigger finger
Block 7 – X-ray
Severity of radiographic hand osteoarthritis

All predictors in Figure 19 were initially modelled using their highest level of measurement, i.e. with no categorisation for continuous predictors, and using the maximum number of response options available (so as measured on the questionnaire or at clinical assessment) for categorical predictors. These decisions were taken to maximise the amount of predictive information contained within each variable (Royston et al. 2009) as the objective of this particular analysis was to predict the trajectory of hand pain and function as precisely and accurately as possible. A limitation of this approach for categorical variables, however, was the potential for some response categories to contain only a small number of participants and produce unstable parameter estimates. To guard against this, throughout the modelling process, parameter estimates were checked to ensure they were plausible (e.g. all variance estimates were positive) and if not, a minimal amount of sensible category merging was considered to ensure plausible estimates were achieved.

7.2.3 Selection of model predictors

Two alternative modelling strategies were used in this chapter to explore whether the choice of modelling strategy impacted on the selection of variables retained. The first modelling strategy aimed to fit the most basic model to the data and then subsequently added model complexity (Bliese et al. 2002), whereas the second modelling strategy aimed to start with a maximal model to be reduced to achieve model parsimony (Cheng et al. 2010). Both modelling approaches were supported by the principles reported in Singer et al. 2003 who state that: “a sound statistical model includes all necessary predictors and no unnecessary ones” and by Preacher et al. 2008 who recommend that models should be tested in an order that is supported by theory and driven by the research question of interest, hence the decision to use the “block structure” of the predictors within the modelling strategy below, rather than relying on a fully automated forward or backwards selection approach. As the AUSCAN outcome for hand pain and function has a skewed distribution, comparisons of the model log-likelihoods were conducted using the Satorra-

Bentler Scaled Chi-square (SBSC) test. For all analyses, a predictor was considered to add value to a model if the SBSC test was statistically significant with a p-value <0.05[80].

Modelling strategy 1

Modelling strategy 1 was conducted in three stages, the first to identify an initial pool of potentially important predictors of either the model intercept or slope, the second to check whether all initial predictors were still statistically significant when all predictors in the initial pool were included in a multivariable model, and the third to explore whether the predictors did indeed predict both the model intercept and slope, or just the intercept alone.

- Stage 1

As shown in Figure 19, the potential predictors in this chapter have been grouped into seven blocks based on their method of measurement (i.e. questionnaire, clinical assessment, or x-ray) and it is this block structure that was used as a basis to select the important predictors at Stage 1. This is because it was highlighted in Chapter 2 that it would be clinically useful to group the measures to determine if a good prediction model could be developed using data that was simple to collect, i.e. from a questionnaire, or whether data needing more complex collection methods, such as a clinical assessment, were then needed to improve model prediction.

More specifically, forward selection procedures were undertaken to select an optimum model firstly from the predictors in the demographic block (block 1). This was achieved by comparing the "overall value" of including the predictor in the model, i.e. by comparing the fit of the model without the predictor in it to one where the additional predictor was included as a potential predictor of both the model intercept and slope. As forward selection procedures were used to select the model predictors in the demographic block,

[80] A value of 0.05 was chosen (rather than a more inclusive threshold of 0.10) because the number of potential model predictors was large so could be statistically significant by chance alone.

predictors remained in the model even if they became non-significant later when other predictors in the same block were added (a constraint that was later relaxed in Stage 2 of the modelling strategy, see below). The optimum[81] model from the demographic predictors was then carried forward as the null model to which predictors in the lifestyle block (block 2) were then tested against (using the same forward selection procedure as defined for block 1) to see if they improved model fit. A new optimum model (that could potentially include predictors from both the demographic and lifestyle blocks) was then carried forward as the new null model to which predictors from the health block (block 3) were tested. This procedure continued iteratively until all predictors in all of the blocks were tested.

- Stage 2

In Stage 1 above, forward selection was chosen to select predictors of interest over stepwise selection procedures. This approach was used to minimise the number of models that needed to be fitted to the data as the SBSCT test was not available as an automated process in Mplus so all models had to be fitted "by hand". This method also avoided potentially removing predictors prematurely in the modelling process (i.e. ones that later became significant when other predictors were added to the model)[82]. A consequence of this decision was the potential for not all predictors included in Stage 1 to improve model fit as predictors were not continually tested for their removal from the model when other predictors were added. In Stage 2, backwards deletion was therefore used to remove any predictors from the model that did not significantly improve model fit. As in Stage 1, the predictors were tested to see whether they were statistically significant predictors of the model intercept and/or slope, i.e. by comparing the fit of the model

[81] Optimum refers to "optimum at Stage 1 in the modelling". It does not mean optimal overall as the model at this stage could potentially include within it predictors that are non-significant

[82] This decision was taken for practical reasons as the number of models to potentially fit to the data is excessively large even for forward selection. Forward selection reduces the number of models that need to be fitted to the data, compared to stepwise methods, as the subsequent removal of model predictors is not tested after predictors have been initially included in the model

without the predictor in it to one where the additional factor was included as a potential predictor of both the model intercept and slope.

- Stage 3

After Stages 1 and 2, all predictor variables had been tested for significant contribution to the model, either as a predictor of the model intercept or slope. Given this strategy, there was the potential that a predictor could be included in the model that was a significant predictor of the model intercept, but not a significant predictor of the model slope. As a final stage, to reduce the number of parameters estimated to ensure parsimony, the model was further "pruned" to remove any non-significant predictors of the model slope. The model slopes, which represent the interaction between time and the predictor, were considered for removal prior to the model intercepts, i.e. the main effects, as the latter are needed in the model for the interaction to be interpreted[83]. This was achieved by using backwards deletion and comparing a model where the predictor is only predictive of the model intercept to that where it is potentially predictive of both the intercept and slope. After the slope terms were tested, any predictors of the intercept terms that were not identified as a predictor of the slope in the model were also tested for their removal - it may be that the estimates for some of the intercepts change as predictors of the slope are removed so this was tested in the final modelling stage.

Modelling strategy 2

Modelling strategy 2 mirrored modelling strategy 1, except that model predictors were selected using backwards deletion (rather than forward selection) at Stage 1. At Stage 1, for modelling strategy 2, all predictors in the first block were included in the model and were iteratively removed (if necessary) from the model using backwards deletion. The optimum model from the first block was then taken forward and all predictors from block 2

[83] This also fits with the objective of the model to find predictors that were good predictors of change over time, rather than those that were just good predictors of participants' initial baseline starting point

added. The block 2 predictors were then subsequently removed from the model (if necessary) to form a new null model. The new null model was then taken forward and used as a null model for all block 3 predictors to be added and then tested for their removal. This procedure continued until all of the predictors in all of the blocks had been tested. Stages 2 and 3 as described above were then applied.

Contrasting modelling strategies

In general, modelling strategy 1 was preferred over modelling strategy 2 as it fitted more closely with the overarching objective of the model (i.e. to start with a simple model that was built up to become more complex as more predictors were added). It also minimised the potential for data drop-out[84]. If the two strategies, however, provided different sets of optimal predictors and both are of similar goodness-of-fit, both sets of predictors were considered as two potentially competing optimal models. This was to fit in with the philosophy that several "optimal models" may exist, each with differing sets of predictors in them (Singer et al. 2003).

7.2.4 Linear trajectory assumption for hand function

In the modelling process described above a linear trajectory was assumed for both hand pain and function, despite there being some evidence in Chapter 6 that a quadratic term for hand function was statistically significant. The decision to focus the analysis on a linear model was taken to aid comparability between the models for hand pain and function, to encourage model convergence (i.e. by estimating the trajectory shape with as few parameters as possible) and to simplify the modelling process (i.e. to not also consider whether the factors of importance were predictors of the quadratic term in the model as well). This was (in part) justifiable as the quadratic term, although statistically significant in the hand function model, was small in magnitude and so was unlikely to have a great impact on the overall trajectory shape.

However, to test the impact of ignoring the potential for a quadratic component to contribute to the model, two sensitivity analyses were conducted for the "final" hand function model(s). The first involved adding a quadratic term for time as a fixed effect to the linear model previously defined as the "final model"; the second, increasing further the model complexity, involved adding a quadratic term for time as both a fixed and random

[84] Modelling strategy 1 uses forward selection so will (in general) be testing models with fewer predictors in them. As participants can only be included in the model if they have data present for all predictors in the model, a model with fewer predictors will maximise the sample size analysed

effect. These sensitivity analyses were used to check to what extent parameter estimates in the final model changed when a quadratic term was added to the hand function model and if shown to be similar, in part, support the use of a linear trajectory when selecting predictors for the model.

7.2.5 Model assumptions

After the optimal model, or a small set of optimal models, had been determined from the data, the strategies described in Chapter 5 were used to explore whether the model assumptions were satisfied for the models fitted in this chapter. For presentation purposes, to ensure that visual inspection of the plots was manageable, a random sample of 50 participants was drawn independently for hand pain and hand function (to avoid assessing goodness of fit on the same participants for both outcomes) and the observed and predicted trajectories plotted. The random sample was selected from participants with complete AUSCAN data at all time-points and with baseline data for all predictors in the model (to make assessment of model fit easier to derive from the plots presented). Model assumptions were only tested for the "final" models derived at the end of modelling stage 3.

7.2.6 Scaling of the predictor variables

At the end of the modelling process, the interpretability of the model estimates derived was considered by identifying whether there were any continuous predictors in the model that could not be (by definition), or were unlikely to be, zero in the dataset (e.g. age, body mass index or length of time with a hand condition) and whether the sign for the fixed model intercept was negative (a negative sign lacks interpretability for AUSCAN pain and function as zero is the minimum value for these measures), and if beneficial, the model was re-run after relevant predictors had been centered using the method previously described in Chapter 5.

7.3 Modelling results

The modelling results in this chapter are presented for hand pain and function using similar headings as defined in the methodology section above, i.e. selection of model predictors (Section 7.3.1), linear trajectory assumption for hand function (Section 7.3.2), model fit (Section 7.3.3), checking model assumptions (Section 7.3.4) and scaling of predictors (Section 7.3.5). Only the results from modelling strategy 1 are presented as, for both outcomes, modelling strategy 2 selected the same core set of model predictors.

7.3.1 Selection of model predictors

Predictors included at each modelling stage

Models A, B and C in Table 7-1 (for hand pain) and Table 7-2 (for hand function) show parameter estimates and 95% confidence intervals for predictors remaining in the model at the end of each of the three modelling stages. At the end of Stage 1, fifteen and eighteen predictors remained in the model for hand pain and function, respectively[85]. For both outcomes there was evidence that not all of the predictors in the model at the end of Stage 1 were needed, e.g. for both hand pain and function the 95% confidence interval for "Age when left school" contains zero for both the model intercept and slope, justifying the need for modelling Stage 2.

At modelling Stage 2, three predictors were excluded for hand pain (age when left school, income and emotional representation) and six for hand function (employment status, age when left school, income, pain in other body areas, consequences and treatment control). Although the estimate of the model intercept changed considerably between the models at Stages 1 and 2, only small changes in all other estimates were observed.

[85] During modelling Stage 1 it was found that "onset following accident or injury" could not be modelled alongside "sudden onset" due to multicollinearity (spearman's rho between these two predictors was 0.74) so to resolve this, "sudden onset" was chosen as the single predictor for inclusion in the model as it gave a slightly higher log-likelihood than that for "onset following accident or injury" (Hand pain log-likelihood: -3631.158 vs -3631.760; Hand function log-likelihood: -3196.537 vs -3197.759) enabling model estimates to be derived

At Stage 3, a relatively large number of model slope predictors were removed from the model (seven for each of hand pain and function, though only three overlapped for the two outcomes: gender, general health, and hand osteoarthritis from clinical assessment). This simplified the model greatly. In addition the intercepts for employment status and general health were also removed from the hand pain model. Although this simplification degraded some of the goodness-of-fit statistics slightly, overall, the simplified models were a good fit to the data.

Table 7-1: Predictors in the hand pain model at each modelling stage (predictors in Figure 19 not included in this table are those that were not statistically significant after Stage 1 had been applied)

	Model A "Predictors after Stage 1" N = 432	Model B "Predictors after Stage 2" N = 436	Model C "Predictors after Stage 3" N = 456
Fixed part			
Initial status			
Intercept (μ_α)	-0.73 (-2.69, 1.23)	0.32 (-0.99, 1.63)	0.44 (-0.50, 1.37)
Gender			
Female	Ref	Ref	
Male	0.58 (0.30, 0.87)	0.56 (0.28, 0.85)	0.49 (0.25, 0.73)
Employment status			
Employed	Ref	Ref	N/A
Not working due to ill-health/unemployment	0.47 (0.03, 0.91)	0.46 (0.01, 0.90)	N/A
Retired	-0.23 (-0.49, 0.03)	-0.22 (-0.48, 0.04)	N/A
Housewife	-0.26 (-0.65, 0.13)	-0.23 (-0.64, 0.19)	N/A
Other	0.45 (-0.53, 1.42)	0.45 (-0.60, 1.51)	N/A
Age when left school	0.03 (-0.06, 0.11)	N/A	N/A
Income			
Find it a strain to get by from week to week	Ref	N/A	N/A
Have to be careful with money	0.44 (-0.09, 0.96)	N/A	N/A
Able to manage without much difficulty	0.52 (-0.01, 1.06)	N/A	N/A
Quite comfortably off	0.60 (0.05, 1.15)	N/A	N/A
General health			
Excellent	Ref	Ref	N/A
Very good	0.04 (-0.49, 0.56)	0.10 (-0.42, 0.61)	N/A
Good	0.05 (-0.49, 0.60)	0.08 (-0.45, 0.61)	N/A
Fair	0.20 (-0.46, 0.85)	0.22 (-0.41, 0.86)	N/A

Poor	-0.60 (-1.49, 0.29)	-0.64 (-1.51, 0.23)	N/A
Physical component score (SF-12)	0.00 (-0.01, 0.02)	0.00 (-0.01, 0.02)	0.00 (-0.01, 0.01)
Pain in other body areas			
No other pain	Ref	Ref	Ref
Regional pain	0.36 (-0.01, 0.73)	0.32 (-0.05, 0.70)	0.34 (0.03, 0.66)
Widespread pain	0.76 (0.33, 1.19)	0.75 (0.32, 1.19)	0.76 (0.39, 1.13)
Number of days in the last 12-months with hand pain			
Less than 7-days	Ref	Ref	Ref
1-4 weeks	0.06 (-0.36, 0.49)	0.00 (-0.43, 0.43)	-0.05 (-0.50, 0.41)
>1-month but <3-months	0.11 (-0.35, 0.58)	0.11 (-0.35, 0.57)	0.10 (-0.37, 0.57)
3-months or more	0.55 (0.17, 0.94)	0.54 (0.16, 0.93)	0.57 (0.17, 0.98)
Baseline AUSCAN function	0.60 (0.52, 0.68)	0.61 (0.53, 0.69)	0.61 (0.53, 0.68)
Sudden onset of hand condition			
Bilateral problem: both hands sudden onset	Ref	Ref	Ref
Bilateral problem: one hand sudden onset	-0.16 (-0.68, 0.36)	-0.14 (-0.66, 0.38)	-0.01 (-0.48, 0.46)
Bilateral problem: neither hand of sudden onset	0.26 (-0.04, 0.57)	0.25 (-0.06, 0.57)	0.36 (0.09, 0.63)
Unilateral problem: of sudden onset	-0.22 (-0.83, 0.39)	-0.26 (-0.88, 0.35)	-0.35 (-0.86, 0.16)
Unilateral problem: not of sudden onset	0.22 (-0.22, 0.67)	0.23 (-0.22, 0.68)	0.24 (-0.15, 0.64)
IPQR - treatment control	-0.04 (-0.07, -0.00)	-0.03 (-0.07, 0.00)	-0.04 (-0.07, -0.01)
IPQR- emotional representation	0.03 (-0.00, 0.06)	N/A	N/A
Average pinch strength	0.01 (-0.03, 0.04)	0.01 (-0.02, 0.04)	0.02 (-0.01, 0.05)
Meets ACR criteria for hand osteoarthritis (OA)			
No	Ref	Ref	Ref
Yes	0.42 (0.15, 0.69)	0.47 (0.19, 0.74)	0.38 (0.14, 0.61)
Number of joints with radiographic hand OA	0.01 (-0.02, 0.03)	0.01 (-0.01, 0.03)	-0.00 (-0.02, 0.02)
Rate of change			
Time (μ_β)	0.55 (0.05, 1.06)	0.42 (0.05, 0.78)	0.52 (0.32, 0.72)
Gender			
Female	Ref	Ref	N/A
Male	-0.06 (-0.14, 0.02)	-0.06 (-0.13, 0.01)	N/A

Employment status			
Employed	Ref	Ref	N/A
Not working due to ill-health/unemployment	-0.08 (-0.19, 0.03)	-0.06 (-0.16, 0.05)	N/A
Retired	0.00 (-0.06, 0.07)	0.01 (-0.06, 0.07)	N/A
Housewife	0.02 (-0.08, 0.12)	0.02 (-0.08, 0.13)	N/A
Other	-0.20 (-0.34, -0.05)	-0.19 (-0.35, -0.03)	N/A
Age when left school	0.00 (-0.02, 0.02)	N/A	N/A
Income			
Find it a strain to get by from week to week	Ref	N/A	N/A
Have to be careful with money	-0.10 (-0.20, 0.00)	N/A	N/A
Able to manage without much difficulty	-0.14 (-0.25, -0.03)	N/A	N/A
Quite comfortably off	-0.14 (-0.25, -0.03)	N/A	N/A
General health			
Excellent	Ref	Ref	N/A
Very good	0.08 (-0.08, 0.24)	0.06 (-0.09, 0.22)	N/A
Good	0.04 (-0.13, 0.20)	0.02 (-0.14, 0.18)	N/A
Fair	0.06 (-0.13, 0.26)	0.06 (-0.13, 0.25)	N/A
Poor	0.19 (-0.06, 0.45)	0.21 (-0.03, 0.46)	N/A
Physical component score (SF-12)	-0.01$^{\alpha}$ (-0.01, -0.00)	-0.01$^{\alpha}$ (-0.01, -0.00)	-0.01$^{\alpha}$ (-0.01, -0.00)
Pain in other body areas			
No other pain	Ref	Ref	N/A
Regional pain	-0.02 (-0.10, 0.07)	-0.01 (-0.10, 0.07)	N/A
Widespread pain	-0.01 (-0.11, 0.10)	-0.00 (-0.11, 0.10)	N/A
Number of days in the last 12-months with hand pain			
Less than 7-days	Ref	Ref	Ref
1-4 weeks	0.05 (-0.06, 0.17)	0.06 (-0.06, 0.18)	0.07 (-0.06, 0.19)
>1-month but <3-months	0.06 (-0.05, 0.18)	0.07 (-0.05, 0.18)	0.05 (-0.07, 0.18)
3-months or more	-0.02 (-0.12, 0.09)	-0.01 (-0.12, 0.09)	-0.03 (-0.14, 0.08)
Baseline AUSCAN function	-0.04 (-0.06, -0.02)	-0.04 (-0.06, -0.02)	-0.04 (-0.06, -0.02)
Sudden onset of hand condition			
Bilateral problem: both hands sudden onset	Ref	Ref	N/A
Bilateral problem: one hand sudden onset	0.02 (-0.12, 0.15)	0.01 (-0.13, 0.15)	N/A
Bilateral problem: neither hand of sudden onset	0.06 (-0.03, 0.14)	0.05 (-0.03, 0.14)	N/A

Unilateral problem: of sudden onset	-0.06 (-0.18, 0.06)	-0.06 (-0.18, 0.06)	N/A
Unilateral problem: not of sudden onset	0.03 (-0.09, 0.15)	0.02 (-0.10, 0.14)	N/A
IPQR - treatment control	-0.00 (-0.01, 0.01)	-0.00 (-0.01, 0.01)	N/A
IPQR- emotional representation	-0.01[α] (-0.01, 0.00)	N/A	N/A
Average pinch strength	-0.01 (-0.02, -0.00)	-0.01[α] (-0.02, -0.00)	-0.02[α] (-0.02, -0.01)
Meets ACR criteria for hand osteoarthritis (OA)			
No	Ref	Ref	N/A
Yes	-0.04 (-0.10, 0.02)	-0.05 (-0.11, 0.01)	N/A
Number of joints with radiographic hand OA	0.01[α] (0.00, 0.01)	0.01[α] (0.00, 0.01)	0.01[α] (0.00, 0.01)
Random part			
Variance			
Intercept (σ_{α}^{2})	0.39 (0.22, 0.70)	0.41 (0.24, 0.71)	0.43 (0.25, 0.72)
Slope (σ_{β}^{2})	0.02 (0.01, 0.04)	0.02 (0.01, 0.04)	0.02 (0.01, 0.04)
Covariance			
Intercept and slope ($\sigma_{\alpha\beta}^{2}$)	0.02 (-0.03, 0.06)	0.02 (-0.03, 0.06)	0.01 (-0.03, 0.05)
Residual – variance ($\sigma_{\varepsilon t}^{2}$)			
Baseline	1.06 (0.83, 1.37)	1.11 (0.87, 1.43)	1.17 (0.91, 1.49)
18-months	2.76 (2.27, 3.34)	2.72 (2.25, 3.30)	2.80 (2.33, 3.38)
3-years	1.99 (1.65, 2.41)	1.99 (1.65, 2.40)	2.02 (1.68, 2.43)
5-years	1.43 (1.15, 1.77)	1.44 (1.16, 1.78)	1.48 (1.22, 1.81)
7.5-years	1.39 (1.01, 1.93)	1.41 (1.03, 1.95)	1.38 (1.00, 1.89)
Information criteria model fit			
TLI	0.97	0.97	0.96
CFI	0.98	0.98	0.97
SRMR	0.01	0.01	0.02
RMSEA[β]	0.02	0.03	0.04

Akaike (AIC)	6545	6598	6879
Bayesian (BIC)	6822	6834	7015
Sample-size adjusted (BIC)	6606	6650	6910

Unless otherwise stated, figures are parameter estimates and 95% confidence intervals (based on robust standard errors) in brackets.
Employment categories of “Not working due to ill health” (N = 56) and “Unemployed” (N=3) and comorbidity categories of “Six” (N=7) and “Seven” (N=3) were merged during the analysis as the model could not be estimated due to small N. α = The estimate is the same as the upper or lower confidence interval limit due to rounding of model estimates. β = a 95% confidence interval around the RMSEA could only be calculated for certain analysis options in Mplus e.g. confirmatory factor analysis and was not available for ANALYSIS TYPE = RANDOM in Mplus. It therefore could not be reported in this table. TLI = Tucker-Lewis index, CFI = Comparative fit index, SRMR = Standardised root mean-square residual, RMSEA = Root mean square error of approximation, p = p-value, N/A = not applicable, Ref = reference category.

Table 7-2: Predictors in the hand function model at each modelling stage (predictors in Figure 19 not included in this table are those that were not statistically significant after Stage 1 had been applied)

	Model A "Predictors after Stage 1" N = 457	Model B "Predictors after Stage 2" N = 502	Model C "Predictors after Stage 3" N = 502
Fixed part			
Initial status			
Intercept (μ_α)	0.37 (-2.10, 2.83)	2.07 (0.60, 3.53)	2.03 (0.60, 3.46)
Age	0.01 (-0.01, 0.03)	0.01 (-0.00, 0.03)	0.01 (-0.00, 0.03)
Gender			
Female	Ref	Ref	Ref
Male	-0.48 (-0.82, -0.15)	-0.40 (-0.73, -0.08)	-0.41 (-0.73, -0.10)
Employment status			
Employed	Ref	N/A	N/A
Not working due to ill-health/unemployment	0.28 (-0.12, 0.68)	N/A	N/A
Retired	0.12 (-0.16, 0.40)	N/A	N/A
Housewife	0.59 (0.24, 0.94)	N/A	N/A
Other	-0.13 (-0.99, 0.73)	N/A	N/A
Age when left school	0.03 (-0.05, 0.12)	N/A	N/A
Income			
Find it a strain to get by from week to week	Ref	N/A	N/A
Have to be careful with money	0.10 (-0.58, 0.78)	N/A	N/A
Able to manage without much difficulty	0.26 (-0.43, 0.94)	N/A	N/A
Quite comfortably off	-0.06 (-0.76, 0.64)	N/A	N/A
Alcohol consumption			
Daily or most days	Ref	Ref	Ref
Once or twice a week	0.03 (-0.24, 0.29)	0.07 (-0.19, 0.33)	0.04 (-0.21, 0.29)
Once or twice a month	0.06 (-0.24, 0.35)	0.06 (-0.23, 0.35)	0.04 (-0.24, 0.32)
Once or twice a year	-0.00 (-0.35, 0.34)	0.01 (-0.33, 0.34)	-0.00 (-0.33, 0.32)

Never	0.75 (0.33, 1.17)	0.75 (0.37, 1.14)	0.71 (0.31, 1.10)
General health			
Excellent	Ref	Ref	Ref
Very good	-0.06 (-0.46, 0.35)	-0.05 (-0.45, 0.34)	0.01 (-0.36, 0.38)
Good	-0.32 (-0.75, 0.12)	-0.20 (-0.61, 0.21)	-0.12 (-0.50, 0.25)
Fair	-0.08 (-0.62, 0.46)	0.04 (-0.48, 0.55)	0.14 (-0.34, 0.62)
Poor	1.09 (0.25, 1.92)	1.27 (0.48, 2.06)	1.33 (0.61, 2.06)
Physical component score (SF-12)	-0.01 (-0.03, 0.00)	-0.02 (-0.03, -0.00)	-0.02 (-0.03, -0.00)
Number of co-morbidities			
0	Ref	Ref	Ref
1	0.01 (-0.21, 0.23)	0.02 (-0.19, 0.23)	0.02 (-0.20, 0.23)
2	0.10 (-0.20, 0.41)	0.09 (-0.22, 0.40)	0.09 (-0.22, 0.40)
3	-0.05 (-0.46, 0.37)	-0.04 (-0.46, 0.39)	-0.05 (-0.47, 0.38)
4	-0.68 (-1.25, -0.10)	-0.56 (-0.99, -0.13)	-0.55 (-0.98, -0.13)
5 or 6	-1.20 (-2.08, -0.31)	-1.16 (-1.78, -0.54)	-1.15 (-1.76, -0.54)
Pain in other body areas			
No other pain	Ref	N/A	N/A
Regional pain	-0.02 (-0.29, 0.25)	N/A	N/A
Widespread pain	0.02 (-0.34, 0.38)	N/A	N/A
Baseline AUSCAN pain	0.43 (0.35, 0.51)	0.43 (0.35, 0.50)	0.43 (0.35, 0.50)
Length of time with a hand problem	0.02 (0.00, 0.03)	0.01 (0.00, 0.03)	0.01 (0.00, 0.03)
IPQR - consequences	0.02 (-0.01, 0.04)	N/A	N/A
IPQR - treatment control	0.02 (-0.01, 0.05)	N/A	N/A
Frustration with hand condition			
All days	Ref	Ref	Ref
Most days	-0.12 (-0.58, 0.34)	-0.24 (-0.70, 0.23)	-0.24 (-0.66, 0.19)
Some days	-0.79 (-1.23, -0.35)	-0.87 (-1.32, -0.42)	-0.86 (-1.27, -0.45)
Few days	-0.97 (-1.41, -0.52)	-1.03 (-1.48, -0.58)	-1.01 (-1.42, -0.60)
No days	-1.37 (-1.85, -0.89)	-1.43 (-1.91, -0.94)	-1.39 (-1.84, -0.95)
Grip-ability test	0.02 (0.01, 0.03)	0.02 (0.01, 0.03)	0.02 (0.01, 0.03)
Average grip strength	-0.01 (-0.02, -0.00)	-0.01$^{\alpha}$ (-0.02, -0.01)	-0.01$^{\alpha}$ (-0.02, -0.01)
Meets ACR criteria for hand osteoarthritis (OA)			
No	Ref	Ref	Ref

Yes	0.36 (0.13, 0.60)	0.34 (0.12, 0.57)	0.35 (0.13, 0.57)
Rate of change			
Time (μ_β)	0.15 (-0.44, 0.74)	0.09 (-0.24, 0.43)	0.13 (-0.08, 0.35)
Age	0.00 (-0.00, 0.01)	0.00$^\alpha$ (0.00, 0.01)	0.00$^\alpha$ (0.00, 0.01)
Gender			
Female	Ref	Ref	N/A
Male	0.01 (-0.07, 0.09)	-0.01 (-0.09, 0.06)	N/A
Employment status			
Employed	Ref	N/A	N/A
Not working due to ill-health/unemployment	-0.09 (-0.20, 0.02)	N/A	N/A
Retired	-0.03 (-0.10, 0.05)	N/A	N/A
Housewife	-0.04 (-0.13, 0.05)	N/A	N/A
Other	-0.14 (-0.24, -0.03)	N/A	N/A
Age when left school	0.01 (-0.01, 0.03)	N/A	N/A
Income			
Find it a strain to get by from week to week	Ref	N/A	N/A
Have to be careful with money	-0.04 (-0.16, 0.07)	N/A	N/A
Able to manage without much difficulty	-0.09 (-0.21, 0.04)	N/A	N/A
Quite comfortably off	-0.07 (-0.19, 0.06)	N/A	N/A
Alcohol consumption			
Daily or most days	Ref	Ref	N/A
Once or twice a week	-0.04 (-0.09, 0.01)	-0.03 (-0.09, 0.02)	N/A
Once or twice a month	-0.06 (-0.12, 0.00)	-0.02 (-0.08, 0.04)	N/A
Once or twice a year	-0.02 (-0.10, 0.06)	-0.01 (-0.09, 0.06)	N/A
Never	-0.07 (-0.15, 0.02)	-0.05 (-0.14, 0.03)	N/A
General health			
Excellent	Ref	Ref	N/A
Very good	0.08 (-0.03, 0.19)	0.06 (-0.05, 0.17)	N/A
Good	0.09 (-0.02, 0.20)	0.07 (-0.04, 0.18)	N/A
Fair	0.11 (-0.03, 0.26)	0.09 (-0.05, 0.23)	N/A

Poor	0.10 (-0.09, 0.28)	0.06 (-0.12, 0.24)	N/A
Physical component score (SF-12)	-0.00[α] (-0.01, -0.00)	-0.00[α] (-0.01, -0.00)	-0.00[α] (-0.01, -0.00)
Number of co-morbidities			
0	Ref	Ref	Ref
1	0.04 (-0.01, 0.10)	0.04 (-0.01, 0.10)	0.05 (-0.00, 0.10)
2	0.01 (-0.05, 0.07)	0.02 (-0.03, 0.08)	0.02 (-0.03, 0.08)
3	0.09 (-0.02, 0.20)	0.09 (-0.01, 0.19)	0.10 (0.01, 0.20)
4	-0.09 (-0.22, 0.04)	-0.02 (-0.13, 0.10)	-0.02 (-0.13, 0.09)
5 or 6	0.38 (0.23, 0.54)	0.45 (0.33, 0.57)	0.45 (0.35, 0.56)
Pain in other body areas			
No other pain	Ref	N/A	N/A
Regional pain	-0.04 (-0.11, 0.03)	N/A	N/A
Widespread pain	-0.02 (-0.11, 0.06)	N/A	N/A
Baseline AUSCAN pain	-0.03 (-0.04, -0.01)	-0.02 (-0.04, -0.01)	-0.03 (-0.04, -0.01)
Length of time with a hand problem	0.00[α] (-0.00, 0.00)	0.00[α] (-0.00, 0.00)	N/A
IPQR - consequences	0.00 (-0.01, 0.01)	N/A	N/A
IPQR - treatment control	-0.01 (-0.02, -0.00)	N/A	N/A
Frustration with hand condition			
All days	Ref	Ref	N/A
Most days	0.04 (-0.08, 0.16)	-0.00 (-0.13, 0.12)	N/A
Some days	0.07 (-0.04, 0.18)	0.00 (-0.12, 0.12)	N/A
Few days	0.09 (-0.02, 0.21)	0.02 (-0.11, 0.15)	N/A
No days	0.11 (-0.01, 0.23)	0.03 (-0.10, 0.16)	N/A
Grip-ability test	-0.00 (-0.01, -0.00)	-0.00[α] (-0.01, -0.00)	-0.00[α] (-0.01, -0.00)
Average grip strength	-0.00[α] (-0.00, 0.00)	-0.00[α] (-0.00, 0.00)	N/A
Meets ACR criteria for hand osteoarthritis (OA)			
No	Ref	Ref	N/A
Yes	0.00 (-0.05, 0.06)	0.00 (-0.05, 0.05)	N/A

Random part

Variance

Intercept (σ_{α}^{2})	0.50 (0.35, 0.70)	0.57 (0.42, 0.78)	0.58 (0.43, 0.78)
Slope (σ_{β}^{2})	0.01 (0.00, 0.03)	0.01 (0.01, 0.03)	0.01 (0.01, 0.03)
Covariance			
Intercept and slope ($\sigma_{\alpha\beta}^{2}$)	0.05 (0.01, 0.09)	0.04 (0.01, 0.08)	0.04 (0.00, 0.08)
Residual – variance ($\sigma_{\varepsilon t}^{2}$)			
Baseline	0.65 (0.47, 0.90)	0.66 (0.48, 0.91)	0.66 (0.49, 0.90)
18-months	2.20 (1.75, 2.75)	2.21 (1.79, 2.73)	2.21 (1.79, 2.73)
3-years	1.18 (0.91, 1.53)	1.22 (0.96, 1.54)	1.22 (0.96, 1.54)
5-years	0.92 (0.68, 1.24)	0.93 (0.71, 1.23)	0.94 (0.71, 1.23)
7.5-years	1.10 (0.77, 1.59)	0.97 (0.68, 1.40)	0.96 (0.67, 1.37)
Information criteria model fit			
TLI	1.00	0.96	0.97
CFI	1.00	0.97	0.98
SRMR	0.01	0.01	0.01
RMSEA$^{\beta}$	0.00	0.04	0.03
Akaike (AIC)	6453	7079	7055
Bayesian (BIC)	6800	7332	7241
Sample-size adjusted (BIC)	6533	7142	7101

Unless otherwise stated, figures are parameter estimates and 95% confidence intervals (based on robust standard errors) in brackets.
Employment categories of “Not working due to ill health” (N = 56) and “Unemployed” (N=3) and comorbidity categories of “Six” (N=7) and “Seven” (N=3) were merged during the analysis as the model could not be estimated due to small N. α = The estimate is the same as the upper or lower confidence interval limit due to rounding of model estimates. β = a 95% confidence interval around the RMSEA could only be calculated for certain analysis options in Mplus e.g. confirmatory factor analysis and was not available for ANALYSIS TYPE = RANDOM in Mplus. It therefore could not be reported in this table. TLI = Tucker-Lewis index, CFI = Comparative fit index, SRMR = Standardised root mean-square residual, RMSEA = Root mean square error of approximation, p = p-value, N/A = not applicable, Ref = reference category.

Comparison of predictors in the hand pain and function models

Only three factors (gender, the physical component score (PCS) of the SF-12, and hand osteoarthritis as measured by the ACR criteria) were significant predictors in both the hand pain and function models, either as a predictor of the model intercept and/or model slope (see Table 7-3 and Table 7-4 for a simplified version of model C containing only those predictors in the "final" model). At least one predictor from each block was included in all models, except for the lifestyle block for hand pain and the x-ray block in the hand function model – no predictors from these blocks were included in the models, respectively. For gender, the direction of the association differed between the outcomes of hand pain and function, with men having worse hand pain scores on average at baseline than women, but better hand function. The direction and magnitude of the parameter estimates associated with meeting the ACR criteria for hand osteoarthritis on the model intercept, and increasing PCS scores on the model intercept and slope were similar for both outcomes.

Although baseline hand pain and baseline hand function were not directly included in the model as predictors of hand pain and function respectively, the estimate of the correlation between the random intercept and slope was non-significant in each model (hand pain correlation 0.12 (95% confidence interval: -0.44, 0.61), hand function correlation 0.45 (95% confidence interval: -0.23, 0.83)). This suggested that participants' rate of change in hand pain or hand function does not depend on the respective baseline starting value for each of the measures, however it is noted that the 95% confidence intervals around these estimates are wide.

Table 7-3: "Final" hand pain prediction model

	Model C "Predictors after Stage 3" N = 456
Fixed part	
Initial status	
Intercept (μ_α)	0.44 (-0.50, 1.37)
Gender	
Female	
Male	0.49 (0.25, 0.73)
Physical component score (SF-12)	0.00 (-0.01, 0.01)
Pain in other body areas	
No other pain	Ref
Regional pain	0.34 (0.03, 0.66)
Widespread pain	0.76 (0.39, 1.13)
Number of days in the last 12-months with hand pain	
Less than 7-days	Ref
1-4 weeks	-0.05 (-0.50, 0.41)
>1-month but <3-months	0.10 (-0.37, 0.57)
3-months or more	0.57 (0.17, 0.98)
Baseline AUSCAN function	0.61 (0.53, 0.68)
Sudden onset of hand condition	
Bilateral problem: both hands sudden onset	Ref
Bilateral problem: one hand sudden onset	-0.01 (-0.48, 0.46)
Bilateral problem: neither hand of sudden onset	0.36 (0.09, 0.63)
Unilateral problem: of sudden onset	-0.35 (-0.86, 0.16)
Unilateral problem: not of sudden onset	0.24 (-0.15, 0.64)
IPQR - treatment control	-0.04 (-0.07, -0.01)

Average pinch strength	0.02 (-0.01, 0.05)
Meets ACR criteria for hand osteoarthritis (OA)	
No	Ref
Yes	0.38 (0.14, 0.61)
Number of joints with radiographic hand OA	-0.00 (-0.02, 0.02)
Rate of change	
Time (μ_β)	0.52 (0.32, 0.72)
Physical component score (SF-12)	-0.01$^{\alpha}$ (-0.01, -0.00)
Number of days in the last 12-months with hand pain$^{\beta}$	
Less than 7-days	Ref
1-4 weeks	0.07 (-0.06, 0.19)
>1-month but <3-months	0.05 (-0.07, 0.18)
3-months or more	-0.03 (-0.14, 0.08)
Baseline AUSCAN function	-0.04 (-0.06, -0.02)
Average pinch strength	-0.02$^{\alpha}$ (-0.02, -0.01)
Number of joints with radiographic hand OA	0.01$^{\alpha}$ (0.00, 0.01)
Random part	
Variance	
Intercept (σ_α^2)	0.43 (0.25, 0.72)
Slope (σ_β^2)	0.02 (0.01, 0.04)
Covariance	
Intercept and slope ($\sigma_{\alpha\beta}^2$)	0.01 (-0.03, 0.05)
Residual – variance ($\sigma_{\varepsilon t}^2$)	
Baseline	1.17 (0.91, 1.49)
18-months	2.80 (2.33, 3.38)
3-years	2.02 (1.68, 2.43)

5-years	1.48 (1.22, 1.81)
7.5-years	1.38 (1.00, 1.89)
Information criteria model fit	
TLI	0.96
CFI	0.97
SRMR	0.02
RMSEA$^{\gamma}$	0.04
Akaike (AIC)	6879
Bayesian (BIC)	7015
Sample-size adjusted (BIC)	6910

Unless otherwise stated, figures are parameter estimates and 95% confidence intervals (based on robust standard errors) in brackets.

α = The estimate is the same as the upper or lower confidence interval limit due to rounding of model estimates. β = Coefficients for this variable are not significantly different from zero (i.e. the reference category) even though the term is entered into the prediction model. This is because the reference category defined was not the category with the smallest regression coefficient when coefficients are ordered by magnitude γ = a 95% confidence interval around the RMSEA could only be calculated for certain analysis options in Mplus e.g. confirmatory factor analysis and was not available for ANALYSIS TYPE = RANDOM in Mplus. It therefore could not be reported in this table. p = p-value, N/A = not applicable, Ref = reference category, TLI = Tucker-Lewis index, CFI = Comparative fit index, SRMR = Standardised root mean-square residual, RMSEA = Root mean square error of approximation

Table 7-4: "Final" hand function prediction model and relevant sensitivity analyses

	Model C "Predictors after Stage 3" N = 502	Model D "Adding a fixed effect quadratic term" N= 502	Model E "Adding a fixed and random effect quadratic term" N= 502
Fixed part			
Initial status			
Intercept (μ_α)	2.03 (0.60, 3.46)	1.98 (0.54, 3.41)	1.88 (0.44, 3.33)
Age	0.01 (-0.00, 0.03)	0.01 (-0.00, 0.03)	0.01 (-0.00, 0.03)
Gender			
Female	Ref	Ref	Ref
Male	-0.41 (-0.73, -0.10)	-0.42 (-0.73, -0.10)	-0.48 (-0.79, -0.16)
Alcohol consumption			
Daily or most days	Ref	Ref	Ref
Once or twice a week	0.04 (-0.21, 0.29)	0.04 (-0.21, 0.29)	0.01 (-0.24, 0.25)
Once or twice a month	0.04 (-0.24, 0.32)	0.04 (-0.25, 0.32)	0.03 (-0.25, 0.31)
Once or twice a year	-0.00 (-0.33, 0.32)	-0.00 (-0.33, 0.32)	-0.01 (-0.33, 0.32)
Never	0.71 (0.31, 1.10)	0.70 (0.31, 1.10)	0.69 (0.33, 1.06)
General health			
Excellent	Ref	Ref	Ref
Very good	0.01 (-0.36, 0.38)	0.02 (-0.35, 0.39)	0.07 (-0.31, 0.45)
Good	-0.12 (-0.50, 0.25)	-0.12 (-0.50, 0.26)	-0.08 (-0.47, 0.30)
Fair	0.14 (-0.34, 0.62)	0.14 (-0.34, 0.62)	0.16 (-0.31, 0.63)
Poor	1.33 (0.61, 2.06)	1.34 (0.61, 2.07)	1.41 (0.68, 2.14)
Physical component score (SF-12)	-0.02 (-0.03, -0.00)	-0.02 (-0.03, -0.00)	-0.01 (-0.03, -0.00)
Number of co-morbidities			
0	Ref	Ref	Ref

1	0.02 (-0.20, 0.23)	0.02 (-0.20, 0.23)	0.03 (-0.18, 0.24)
2	0.09 (-0.22, 0.40)	0.09 (-0.22, 0.40)	0.13 (-0.17, 0.44)
3	-0.05 (-0.47, 0.38)	-0.04 (-0.47, 0.38)	-0.02 (-0.43, 0.40)
4	-0.55 (-0.98, -0.13)	-0.55 (-0.97, -0.12)	-0.49 (-0.89, -0.08)
5 or 6	-1.15 (-1.76, -0.54)	-1.14 (-1.74, -0.54)	-1.18 (-1.76, -0.60)
Baseline AUSCAN pain	0.43 (0.35, 0.50)	0.43 (0.35, 0.50)	0.44 (0.36, 0.52)
Length of time with a hand problem	0.01 (0.00, 0.03)	0.01 (0.00, 0.03)	0.01 (0.00, 0.02)
Frustration with hand condition			
All days	Ref	Ref	Ref
Most days	-0.24 (-0.66, 0.19)	-0.24 (-0.67, 0.18)	-0.33 (-0.76, 0.10)
Some days	-0.86 (-1.27, -0.45)	-0.86 (-1.27, -0.45)	-1.01 (-1.42, -0.61)
Few days	-1.01 (-1.42, -0.60)	-1.01 (-1.42, -0.60)	-1.12 (-1.53, -0.70)
No days	-1.39 (-1.84, -0.95)	-1.39 (-1.84, -0.95)	-1.54 (-1.99, -1.10)
Grip-ability test	0.02 (0.01, 0.03)	0.02 (0.01, 0.03)	0.02 (0.01, 0.03)
Average grip strength	-0.01$^{\alpha}$ (-0.02, -0.01)	-0.01$^{\alpha}$ (-0.02, -0.01)	-0.01$^{\alpha}$ (-0.02, -0.01)
Meets ACR criteria for hand osteoarthritis (OA)			
No	Ref	Ref	Ref
Yes	0.35 (0.13, 0.57)	0.35 (0.13, 0.57)	0.34 (0.12, 0.55)
Rate of change			
Time ($\mu_{\beta 1}$)	0.13 (-0.08, 0.35)	0.21 (-0.00, 0.43)	0.21 (-0.01, 0.42)
Time squared ($\mu_{\beta 2}$)	N/A	-0.01 (-0.02, -0.00)	-0.01 (-0.02, -0.00)
Age	0.00$^{\alpha}$ (0.00, 0.01)	0.00$^{\alpha}$ (0.00, 0.01)	0.00$^{\alpha}$ (0.00, 0.01)
Physical component score (SF-12)	-0.00$^{\alpha}$ (-0.01, -0.00)	-0.00$^{\alpha}$ (-0.01, -0.00)	-0.00$^{\alpha}$ (-0.01, -0.00)
Number of co-morbidities			
0	Ref	Ref	Ref
1	0.05 (-0.00, 0.10)	0.05 (-0.00, 0.10)	0.05 (-0.00, 0.10)
2	0.02 (-0.03, 0.08)	0.02 (-0.03, 0.08)	0.03 (-0.02, 0.08)
3	0.10 (0.01, 0.20)	0.10 (0.01, 0.20)	0.11 (0.01, 0.20)
4	-0.02 (-0.13, 0.09)	-0.03 (-0.13, 0.08)	-0.02 (-0.12, 0.08)
5 or 6	0.45 (0.35, 0.56)	0.45 (0.33, 0.56)	0.46 (0.36, 0.56)
Baseline AUSCAN pain	-0.03 (-0.04, -0.01)	-0.03 (-0.04, -0.01)	-0.02 (-0.04, -0.01)

Grip-ability test	-0.00[α] (-0.01, -0.00)	-0.00[α] (-0.01, -0.00)	-0.00[α] (-0.01, -0.00)
Random part			
Variance			
Intercept (σ_{α}^{2})	0.58 (0.43, 0.78)	0.58 (0.43, 0.78)	0.61 (0.32, 1.17)
Slope ($\sigma_{\beta 1}^{2}$)	0.01 (0.01, 0.03)	0.01 (0.01, 0.03)	0.17 (0.08, 0.38)
Quadratic ($\sigma_{\beta 2}^{2}$)	N/A	N/A	0.00 (0.00[α], 0.01)
Covariance			
Intercept and slope ($\sigma_{\alpha\beta 1}^{2}$)	0.04 (0.00, 0.08)	0.04 (0.00, 0.08)	0.05 (-0.16, 0.25)
Intercept and quadratic $\sigma_{\alpha\beta 2}^{2}$)	N/A	N/A	-0.01 (-0.03, 0.02)
Slope and quadratic $\sigma_{\beta 1\beta 2}^{2}$)	N/A	N/A	-0.02 (-0.03, -0.00)
Residual – variance ($\sigma_{\varepsilon t}^{2}$)			
Baseline	0.66 (0.49, 0.90)	0.66 (0.48, 0.90)	0.53 (0.25, 1.10)
18-months	2.21 (1.79, 2.73)	2.21 (1.79, 2.74)	2.18 (1.75, 2.70)
3-years	1.22 (0.96, 1.54)	1.20 (0.96, 1.51)	0.91 (0.68, 1.22)
5-years	0.94 (0.71, 1.23)	0.94 (0.71, 1.23)	0.76 (0.56, 1.04)
7.5-years	0.96 (0.67, 1.37)	0.95 (0.66, 1.36)	1.21 (0.70, 2.09)
Information criteria model fit			
TLI	0.97	0.97	0.99
CFI	0.98	0.98	0.99
SRMR	0.01	0.01	0.02
RMSEA[β]	0.03	0.03	0.01
Akaike (AIC)	7055	7051	7018
Bayesian (BIC)	7241	7241	7221
Sample-size adjusted (BIC)	7101	7098	7068

Unless otherwise stated, figures are parameter estimates and 95% confidence intervals (based on robust standard errors) in brackets.

Comorbidity categories of “Six” (N=7) and “Seven” (N=3) were merged during the analysis as the model could not be estimated due to small N. α = The estimate is the same as the upper or lower confidence interval limit due to rounding of model estimates, β = a 95% confidence interval around the RMSEA could only be calculated for certain analysis options in Mplus e.g. confirmatory factor analysis and was not available for ANALYSIS TYPE = RANDOM in Mplus. It therefore could not be reported in this table. p = p-value, N/A = not applicable, Ref = reference category, TLI = Tucker-Lewis index, CFI = Comparative fit index, SRMR = Standardised root mean-square residual, RMSEA = Root mean square error of approximation

7.3.2 Linear trajectory assumption for hand function

Incorporating a fixed quadratic term for time in the hand function model made minimal difference to the magnitude of the parameter estimates for the predictors of interest (see model D in Table 7-4). Differences in the parameter estimates from the linear model were slightly greater when both a fixed and random effect for the quadratic term were added to the model, however differences were still small (model E in Table 7-4). This limits the probability that for hand function a different set of predictors would have been selected from those in the linear model and that predicted values would vary greatly from those obtained from the linear model presented.

7.3.3 Model fit

For both hand pain and function, the linear model gave goodness-of-fit statistics that were above the thresholds defined in Chapter 5 for the model to be considered a good fit to the data, i.e. the Tucker-Lewis index (TLI) was >0.95, the Comparative fit index (CFI) >0.95, the standardised root mean-square residual (SRMR) <0.05 and the root mean square error of approximation (RMSEA) < 0.06. For the hand function model, all three information criterion, i.e. Akaike, Bayesian and Sample-size adjusted Bayesian, were slightly lower for the quadratic models than for the linear model (Table 7-4). This indicated that the quadratic models were a better fit to the data, but this difference was not large, so the extra complexity of the quadratic term did not greatly improve model fit over and above what could be achieved from the linear model alone.

7.3.4 Checking model assumptions

Observed versus predicted trajectories

Figure 20 and Figure 21 show that when participants' observed values for hand pain and function are roughly linear over time, the model fits the data well, with intercept and slope predictions reflecting the true trajectory over time (e.g. participants 30751, 33066, 34381 for hand pain and participants 32735, 33794, 35991 for hand function). However, for those

with non-linear changes over time, where fluctuations around the fitted (straight) line are more than just measurement error alone, as expected, the constraint of the linear model is too restrictive, providing poor model fit for some participants (e.g. participants 31594, 32735, and 30354 for hand pain and 32245, 34627, and 35758 for hand function) (plots for all participants examined, data not shown).

Although observed, this assessment of model fit is considered given the results from the hand function model showing that the magnitude of the parameter estimates did not differ greatly between the linear and quadratic model. This suggests that although a linear model may not fit the data well for some participants, the overall impact of such lack of fit may be minimal in the CAS-HA data when the data are used to explore the strength of the relationship between predictor variables and the outcome of interest.

Figure 20: Observed and predicted values for AUSCAN pain for a random sample of 50 participants with complete data at all time-points

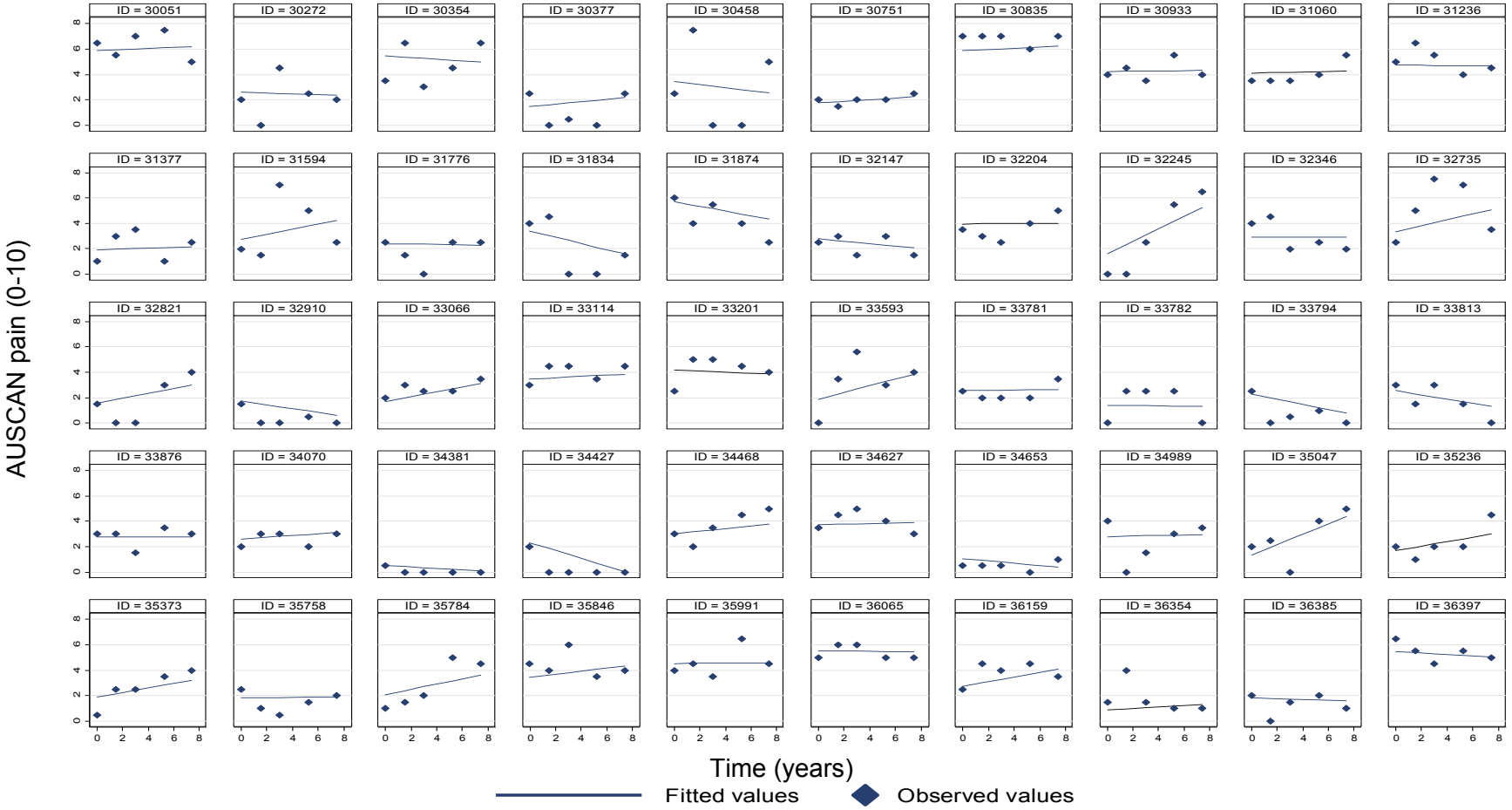
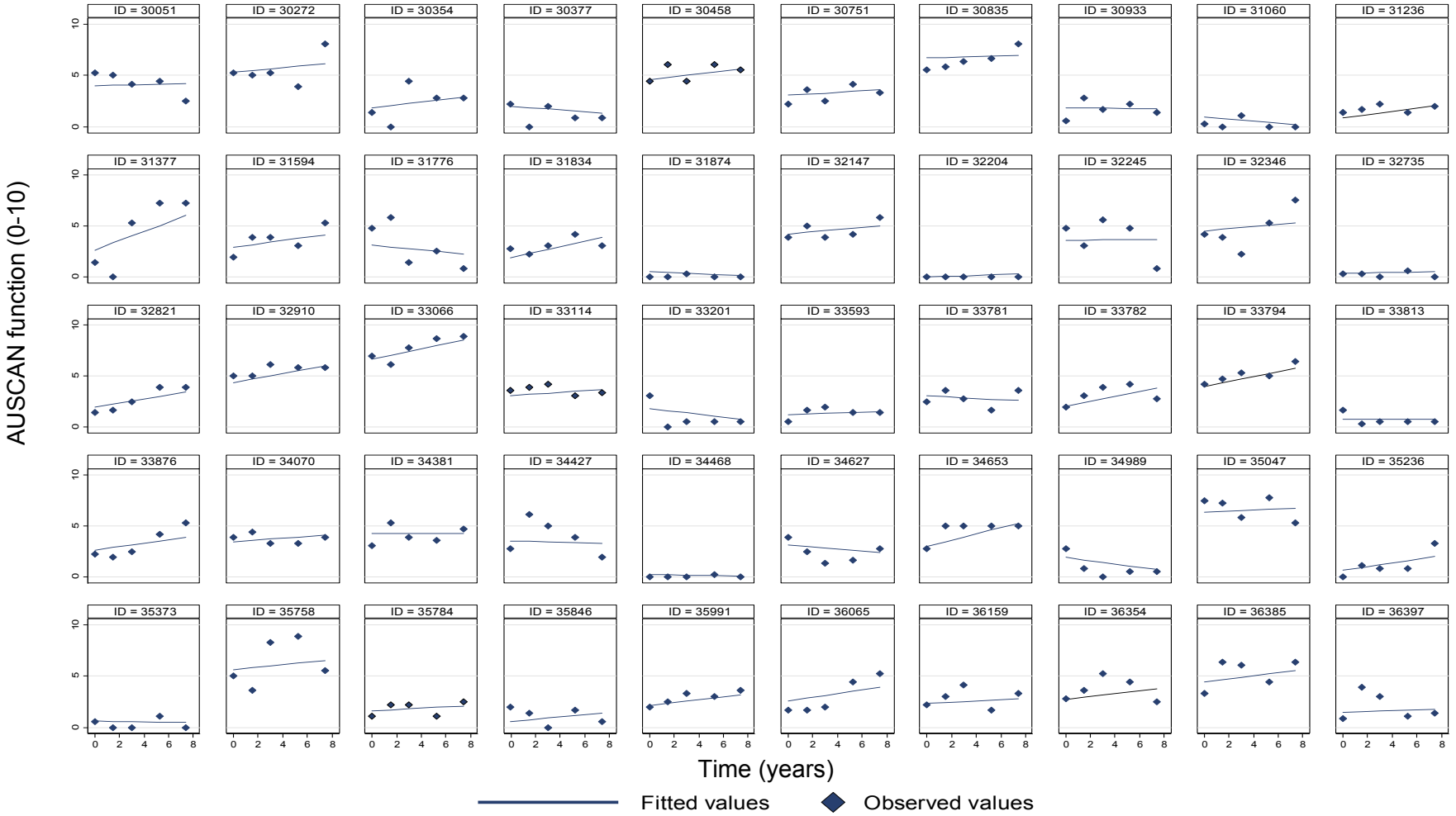


Figure 21: Observed and predicted values for AUSCAN function for a random sample of 50 participants with complete data at all time-points

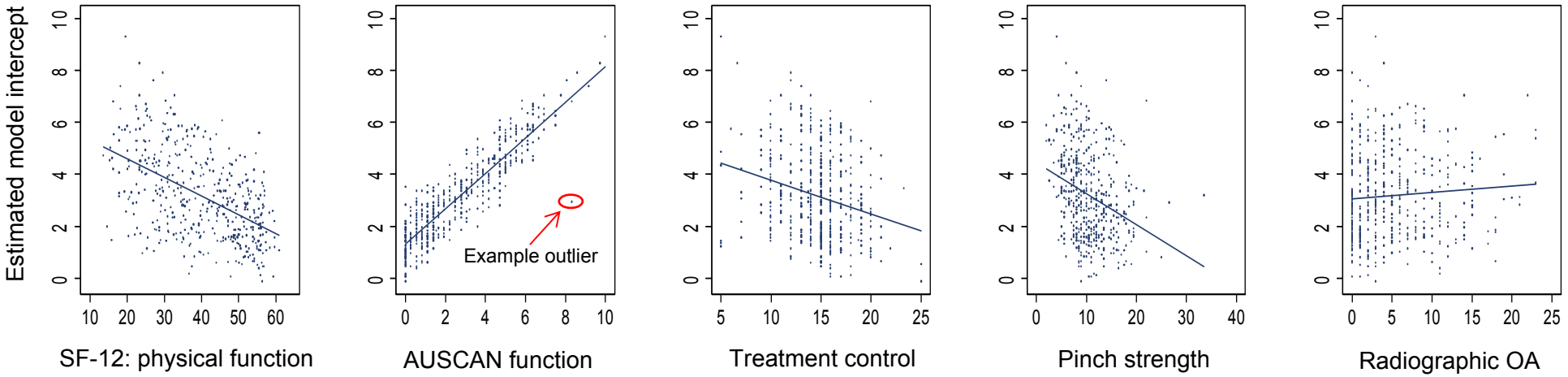


Relationship of model predictors to the growth parameters (i.e. the estimate of the intercept and slope for each participant)

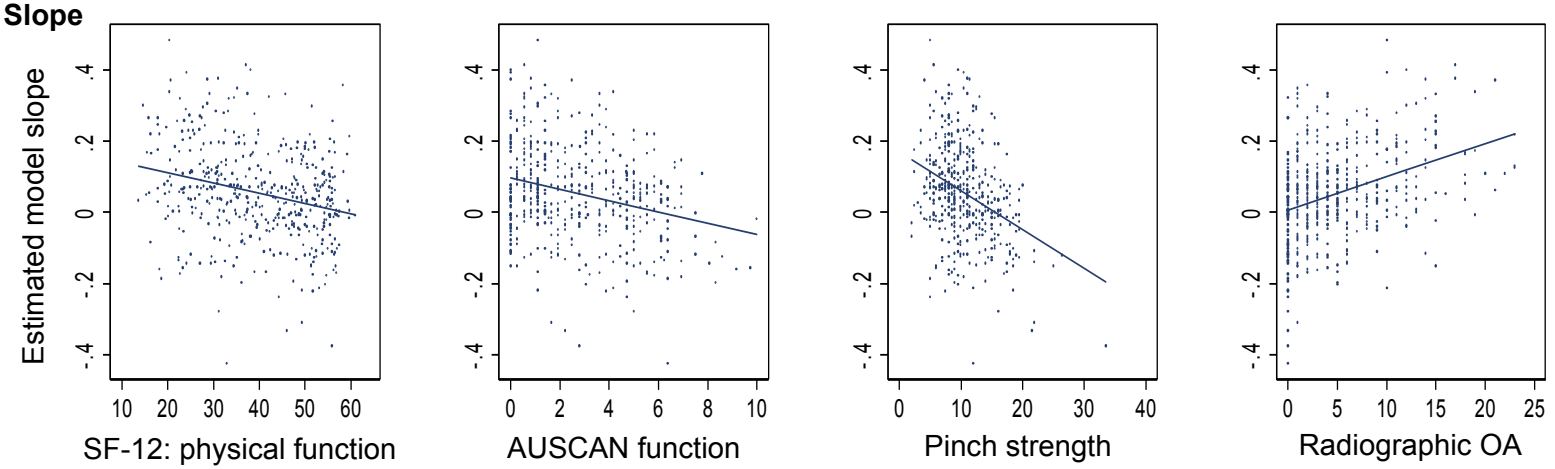
A linear relationship exists between estimates of the model intercept and slope for all continuous predictors in the hand pain and function models, however, the strength of the linear relationship varied depending on the predictor (Figure 22 and Figure 23). Indirectly, these plots also highlight potential outliers in the data, i.e. those whose predicted intercepts are not in keeping with their baseline values for one particular variable of interest: e.g. one participant has an intercept of 3 in the hand pain model, but a raw baseline AUSCAN function score of 8 (indicated by a red circle on Figure 22), another participants has an intercept of 12 in the hand function model, yet a raw baseline GAT score of 280 (indicated by a green circle on Figure 23). After checking that the potential outliers had not occurred due to data entry errors, it was considered whether to re-run the analysis removing such participants from the dataset. This approach however was not taken as the objective of the analysis was to identify factors that were predictive of the outcome given questionnaire and clinical assessment responses that were likely to arise in practice when completed by participants (i.e. a pragmatic analysis). Also, the number of data points that could be clearly identified as outliers was small, lessening their impact on the analysis as presented.

Figure 22: Plots of estimated model growth parameters by continuous model predictors for AUSCAN pain

Intercept

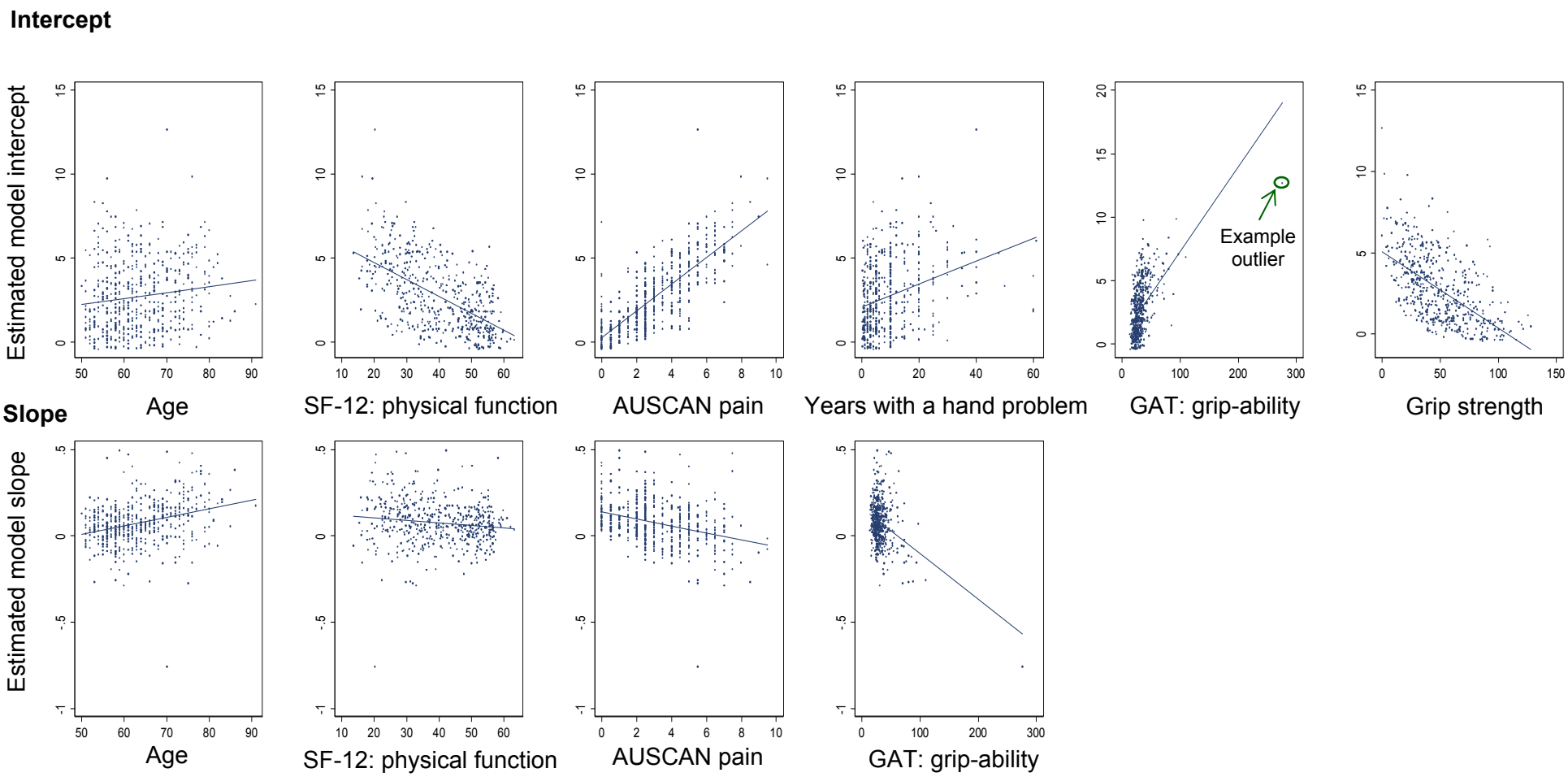


Slope



Blue line = least squares line of best fit

Figure 23: Plots of estimated model growth parameters by continuous model predictors for AUSCAN function



Blue line = least squares line of best fit

Residual plots

For both hand pain and function, the overall residual and the estimate of the random intercept and slope were normally distributed and their magnitudes did not depend on the participants' survey identification number, i.e. the order that participants were recruited to the study (Figure 24 and Figure 25). The random intercept and slope were not highly dependent on the individual values of any model predictor; however, the overall residual was more varied at the 18-month follow-up time point so was estimated separately for each time point in the model (Figure 26 and Figure 27).

The correlations between residuals at each time point were small (typically around -0.2) however the correlation between the last two time-points for pain and function was larger (-0.41 for hand pain and -0.49 for hand function). Correlations between the residuals and each growth parameter in the model (α_i and β_i) were relatively small for both hand pain and function with all correlations less than 0.34. Overall this means that model assumptions related to the independence of the residuals in the model were met.

Figure 24: Residual plots for the hand pain model

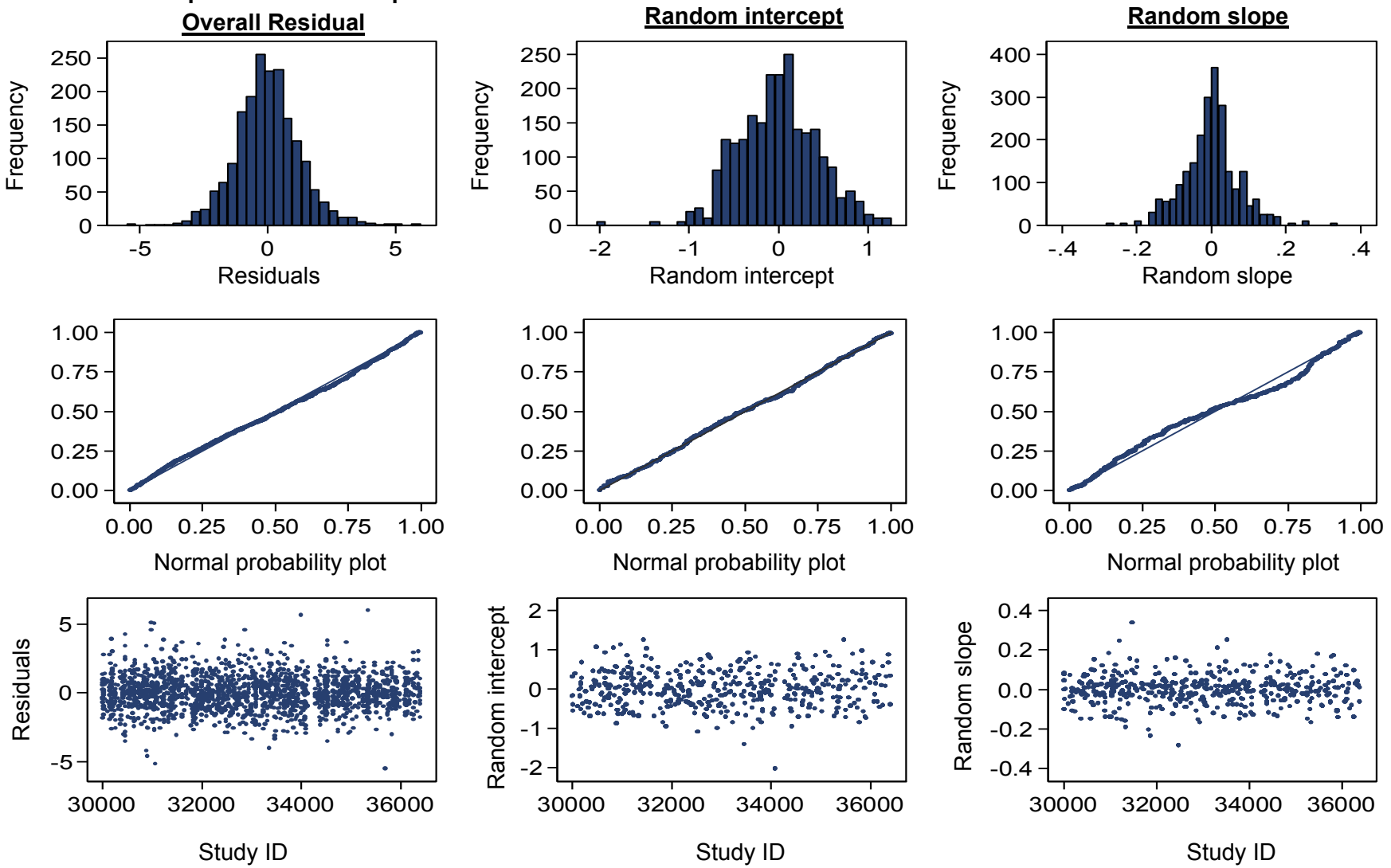


Figure 25: Residual plots for the hand function model

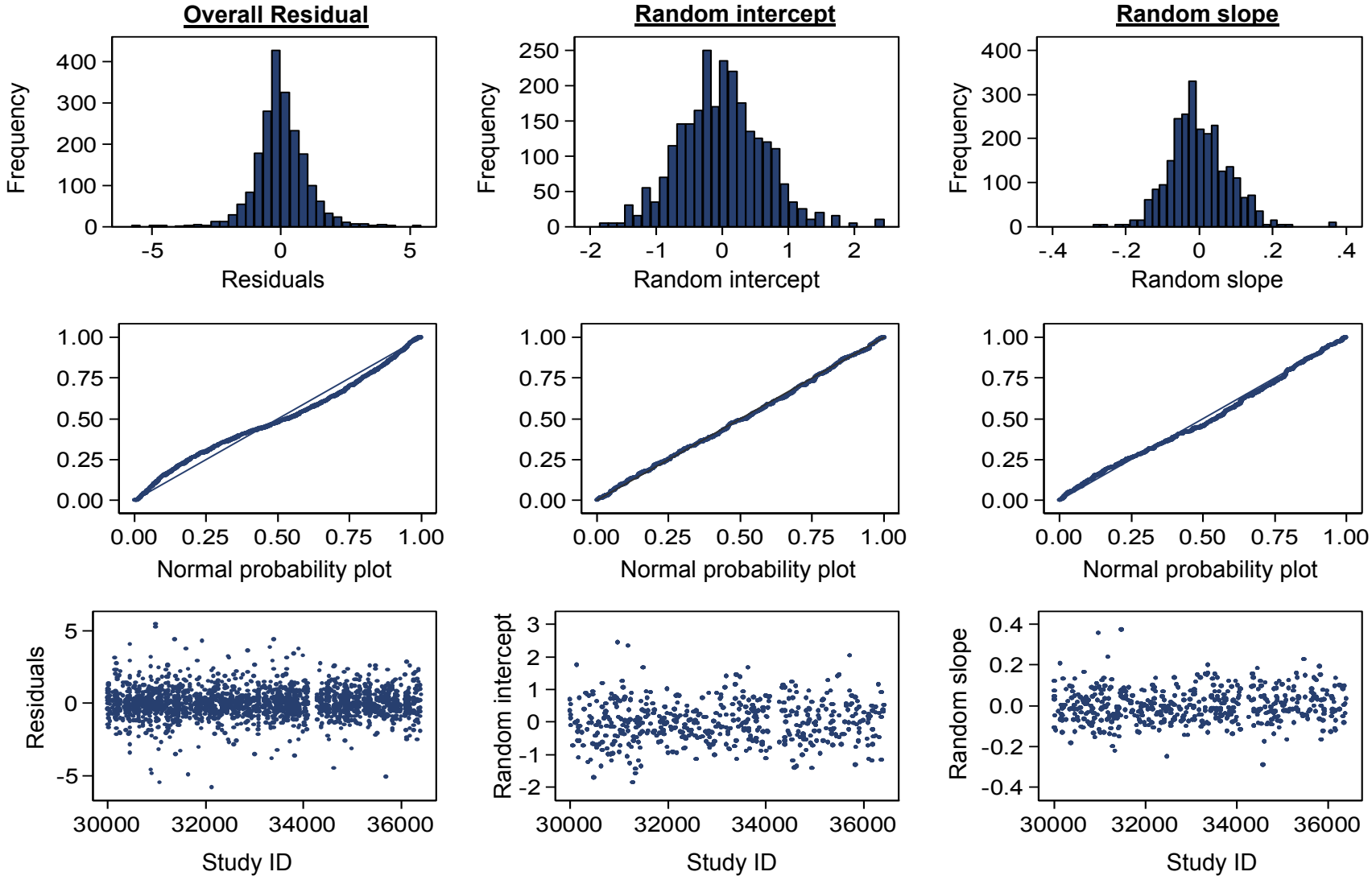


Figure 26: Plots of the random intercept and slope terms against each predictor variable in the hand pain model and for the overall residual by time

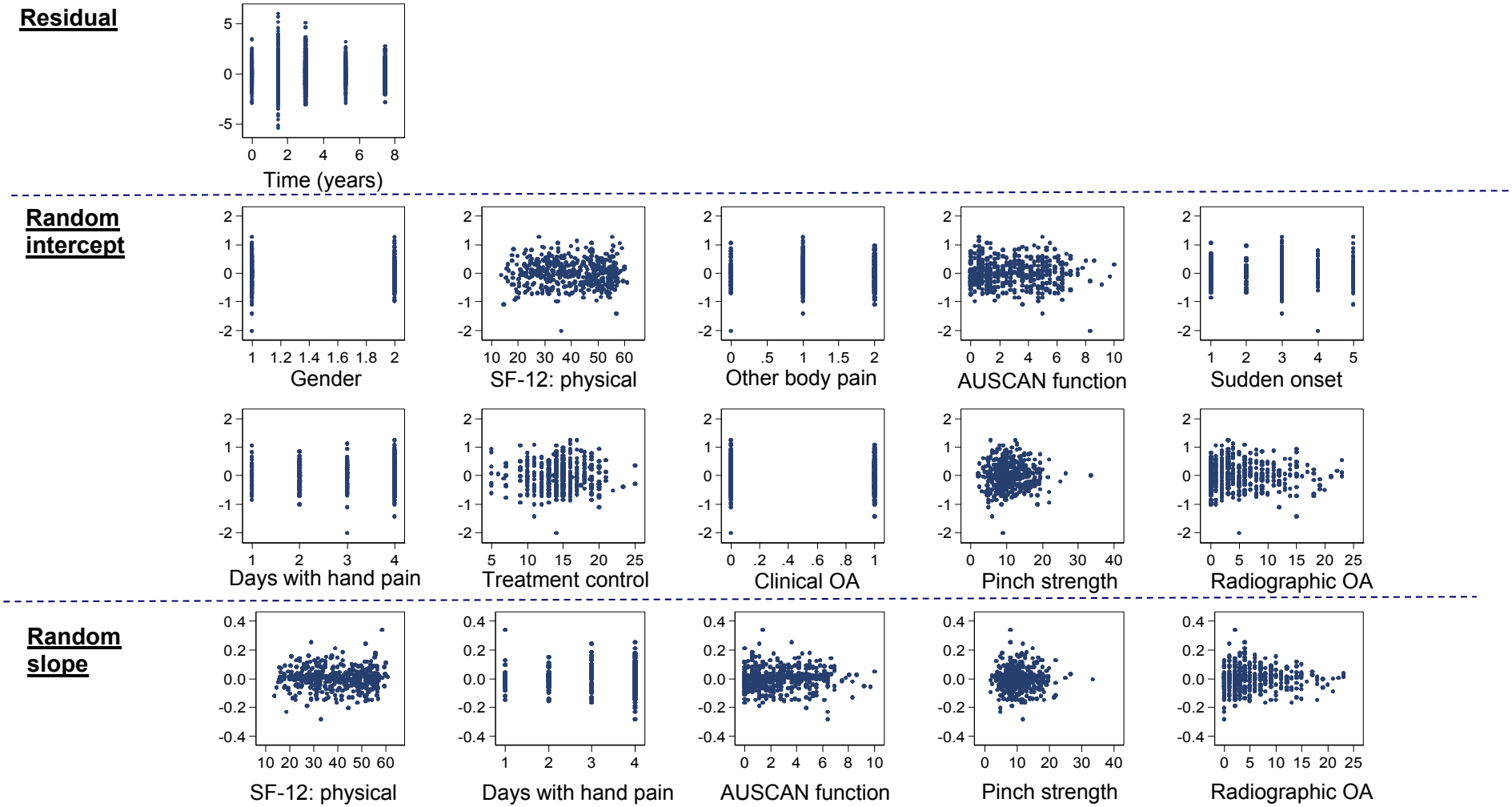
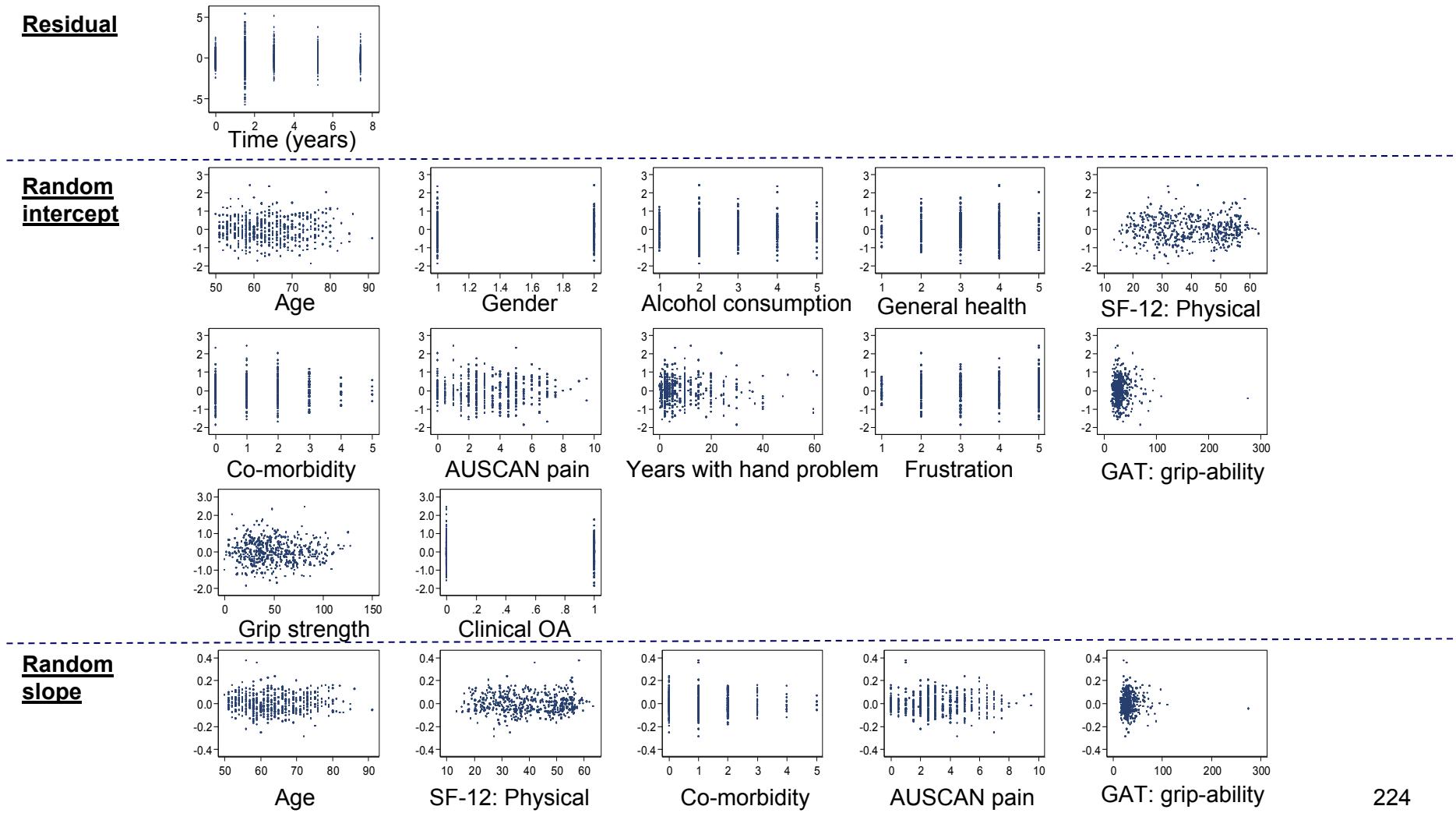


Figure 27: Plots of the random intercept and slope terms against each predictor variable in the hand function model and for the overall residual by time



7.3.5 Scaling of the predictor variables in the model

The estimate of the model intercept was positive in both the hand pain and function model. Only the hand function model contained predictors that (theoretically) could not be zero for CAS-HA participants, i.e. age and length of time with a hand condition (Table 7-4). By centering these two predictors, and re-running the hand function model, a new estimate of the model intercept was derived of 2.86 (95% confidence interval (1.84, 3.87)) and is interpreted as the mean hand function score for a participant aged 64 years, who had a hand problem for 9 years and who has all remaining model predictors at their lowest possible level[86]. As the majority of predictors in the model did not need centering, and to ease comparison with an un-centered hand pain model, for simplicity, the un-centred model has been reported in this thesis.

7.4 Discussion

7.4.1 Summary of the key findings

The prediction models in this chapter have been used to show that a wide range of baseline factors, when considered in combination, are significantly associated with the model random intercept (baseline status) for hand pain, with factors including: gender, physical function, body pain, number of days with hand pain in the last 12-months, hand function, gradual symptom onset, treatment control, pinch strength and evidence of clinical and radiographic hand OA. Only three of these baseline predictors were also shown to be associated with baseline status in the hand function model: gender, the physical component score (PCS) of the SF-12, and clinical evidence of hand OA. However this was estimated after adjustment for nine other variables that were also significantly associated with the random intercept in the hand function model, i.e. age, alcohol consumption, general health, number of co-morbidities, hand pain, length of time

[86] As expected, all other parameter estimates in the model remained unchanged, except for the model intercept

with a hand condition, frustration with the hand condition and two objective measures of hand function (the grip-ability test and grip strength).

A smaller number of predictors were identified that predicted the model slope (i.e. the rate of change of hand pain or function over time) and were predictors only when viewed in combination with other predictors in the model. These included physical function, number of days with hand pain in the last 12-months, hand function, pinch strength, and the number of joints with radiographic hand OA for hand pain. Age, physical function, number of co-morbidities, hand pain and the grip-ability test were important predictors of the rate of change of hand function. Evidence has been presented that the models for hand pain and function fit the data well.

7.4.2 Comparison with the literature

As highlighted in Chapter 6, few longitudinal studies exist in community dwelling older adults and no studies have been identified that have used similar analysis techniques to those in this chapter. A direct comparison to the published literature is therefore limited, not only through lack of publication, but also by differences in the selection criteria used to recruit participants to studies, the potential pool of predictors from which predictors are drawn and those that are included in the models, and the measurement tools used to collect the primary outcome of interest. However, despite all these limitations, data collected in cross-sectional studies still provide some support for the factors identified as predictors of the model intercept in this chapter, e.g. age and gender have each been shown to be associated with levels of functional difficulty for patients with hand OA when recruited from rheumatology clinics at a single time point (Jones et al. 2001) and there is mixed support for an association between hand pain and functional difficulty and severity

of radiographic OA although this was highly dependent on the definition used to define radiographic OA[87] (Haugen et al. 2013, Dahaghin et al. 2005a).

As referred to in Chapter 6, two particular studies have been identified that include a longitudinal element to their data collection procedures over a longer-term follow-up period: the Genetics Arthrosis and Progression study (GARP) (Bijsterbosch et al. 2011) and the Oslo hand OA cohort (Haugen et al. 2013) with longitudinal follow-ups of 6- and 7-years respectively. Although recruiting participants from those with a diagnosis of symptomatic hand OA, the GARP study (Bijsterbosch et al. 2011) showed similar findings to those presented in this chapter; that is that a poor outcome for hand pain at 6-years was associated with greater baseline functional difficulty and that a poor outcome for hand function at 6-years was associated with greater hand pain at baseline. Both studies also showed mixed findings on the role of radiographic OA in predicting change in hand pain and function over time[88]. In addition, a Dutch study found age, gender and having complaints for longer than 3-months at baseline to be predictive of poor outcome for a combined score across pain and function (Spies-Dorgelo et al. 2007), which, with the exception of gender is consistent with the findings in this study, despite the study follow-up period being shorter than CAS-HA, 12-months versus 7.5 years, and adults being recruited only if they had consulted with a hand problem.

Taken together, these studies, although small in number, with varying study designs and outcome measure collection, still offer some support that the factors tested in this thesis

[87] E.g. whether the definition was based on a simple count of the number of joints with radiographic OA or more complex analyses of radiographic subgroups e.g. erosive versus non-erosive OA

[88] The studies explored whether radiographic OA at baseline was predictive of poor outcome for hand pain and function. The GARP study showed that structural abnormalities and specific subsets of hand OA (e.g. erosive vs non-erosive OA) were not predictive of poor outcome for either hand pain or hand function (Bijsterbosch et al. 2011). However, the Oslo hand OA cohort showed that increasing radiographic sum scores at baseline, as measured by summing the Kellgren and Lawrence scores for each joint, was associated with increasing hand functional difficulty over a 7-year time period (Haugen et al. 2013). This latter definition of radiographic OA is consistent with that used in this thesis

are plausible candidates as predictors of change in hand pain and function over time and that the results in this thesis are broadly in line with the literature as currently published.

7.4.3 Strengths and limitations

The number of predictors modelled

A major limitation of the analysis in this chapter is the large number of predictors that have been modelled, with the potential for model over-fitting, unstable results to be produced, and for some predictors to be significant by chance alone due to multiple testing. Although this is a limitation of the analysis, it was somewhat unavoidable as it was not possible to define a smaller set of potential predictors given the lack of evidence available in the literature to define a smaller set. These predictors, although large in number, were included in the model after consultation with clinical experts who judged it was an important clinical question to model predictors split into blocks depending on how they were measured in the data.

Although multiple testing was an issue for this analysis (as it was the significance level that was used to decide on the predictors to include in the model), this was somewhat guarded against as the threshold for including items in the forward selection process was reduced to 0.05, rather than 0.1 that is a threshold commonly used to avoid removing predictors from the model prematurely. It was also minimised by defining the decision rules for inclusion/exclusion of predictors from the model *a priori*, so decisions were not influenced by successively looking at the data, and is of somewhat less concern as the focus of the analysis was on exploring how well factors predict the outcome of interest in combination rather than on the precise statistical significance of the individual predictors *per se*. It is acknowledged however, that some authors recommend that separate exploratory and confirmatory analyses should be conducted, either in separate data samples or by splitting a single dataset into two parts, to test whether predictors in the model remain significant when applied to a separate dataset not used in the development

of the model (Cheng et al. 2010). Although using an external dataset was not possible due to lack of available data, it would be possible to use the split-half method (Cheng et al. 2010), or other similar methodologies such as cross-validation or bootstrapping (Steyerberg et al 2001), to explore the reliability of the predictors included in the model and assess the overall internal validity of the model by estimating any shrinkage in the R-square value that has occurred when using a single dataset to develop the model. This remains an option for further analysis.

Method used to select the predictors of interest

The modelling strategies used in this chapter were chosen as those that were practical to address the research question of interest and that could be applied to a large number of potential predictors. Two modelling strategies were also used, rather than one, to limit the possibility that the selection of predictors was dependent on the strategy used. It is acknowledged, however, that other modelling strategies could have been used to select the predictors of interest (e.g. those that did not consider the block structure imposed on the predictors of interest, see Appendix 23 for further details) so a different set of predictors could have been included in the models if a different model strategy had been used.

It is also acknowledged that other techniques could have been used to select the predictors of interest, especially as forward and backwards selection techniques are reported to have limitations, e.g. that they are more likely to include predictors in the model with relatively large, rather than relatively small, regression coefficients (Steyerberg et al. 2000), they overestimate the magnitude of the parameter estimates and underestimate their associated standard errors, and, as previously mentioned, have problems due to multiple testing (Flom et al. 2007). Other possible techniques include Least Absolute Shrinkage and Selection Operator (LASSO) regression, which aims to combine variable selection and shrinkage of the regression parameters in the same model

so that it more accurately reflects data from an external data set (Tibshirani 1996). This method could be explored further to see if estimates for the predictors of interest were similar using this new method, but this is a focus for further work as it is less well used in the literature (Walter et al. 2009) and it would need to be explored how to use this method for a longitudinal random effects model with a non-normal outcome.

In the modelling strategy as presented, the questionnaire items have been split into four blocks (demographic, lifestyle, health, characteristics of the hand condition and psychological factors) and it is possible that if these blocks were analysed in a different order then a different set of predictors may have emerged from the data. The blocks were generally ordered from generic to more specific questions on health and their hand condition, but this is not the only order that could be considered. However, as the analysis presented in this chapter required over a 1000 models to be fitted to the data by hand, exploring the impact of changing the order that the blocks were analysed was not feasible.

In this chapter the analysis has focussed on exploring which factors predict the intercept and slope and in the model, however the model could be extended to include further interactions between the predictors of interest, i.e. additional interactions between the model predictors not just the interaction of the predictor with time. In keeping with the philosophy of the model this would imply that all interactions (i.e. 2-way, 3-way, and higher order interactions) between all predictors should be included and tested to see if they improve model fit (without clinical justification for their presence), however, practically this could be very complex and add many more terms to the model that may not be reliably estimated given the sample size that is available for analysis. Further interactions between predictors were therefore not explored, however there is potential that such interactions could improve model fit. Additionally, when the sensitivity analysis for hand function was conducted for the quadratic model, it was not explored whether any of the predictors in the model were significant predictors of the quadratic term in the model (along with the intercept and slope) as, similarly to adding in interactions between

predictors, this could make the model more complex to fit, but it is acknowledged that this also has the potential to improve model fit.

Interpretation of model predictors

A key objective of the models in this chapter was to explore which combinations of factors best predict each outcome trajectory over time, rather than to explain why the factors were predictive *per se*. Interest is therefore in how well the model predicts the data rather than on the interpretation of the coefficients for individual factors in the model. All model coefficients are therefore presented "after adjustment" for all other factors so are highly dependent on the other predictors that are included in the model (see Appendix 24 for an illustration related to gender as a predictor of the model intercept and the SF-12 physical function scale as predictor of the model slope). This is a particular limitation of comparing the predictors that are common to both the hand pain and function models as the full list of model predictors differs between the models compared.

7.4.4 Model replication and application

As the work in this chapter has been developed on an exploratory basis, the next step for the analysis would be to explore if the selection of model predictors replicated in a separate dataset of older adults with hand pain and problems and if other predictors were needed to improve model fit (external validation and updating). In addition, it could also be tested in an independent data set to assess how close the observed values of hand pain and function were to those predicted by the models in this chapter. Alternatively, bootstrapping could be used to sample (with replacement) a large number of datasets to explore the number of times each predictor was significant across the bootstrapped datasets of interest (internal validation). This latter analysis was considered for this thesis however was not developed further, as it would be computationally difficult to do without an automated process for model selection in each of the bootstrapped datasets.

7.4.5 Clinical Applicability

A key challenge of the models presented in this chapter is to consider how the results can be translated into a prediction tool that can be used in clinical practice to identify those patients whose condition is likely to worsen over time and hence those who would benefit from early intervention and onward referral. Although it is a strength that the outcomes in this chapter have not been simplified into categorical variables representing “changed over time”/“not changed over time”, to avoid loss of information in the outcome predicted, this has a trade-off as such models are more challenging to apply in a clinical setting.

One possibility may be to generate a predicted value based on patients’ responses to the questions that predict the model slope. Inspection of the distribution of the predicted values could therefore be used to produce a threshold whereby predicted values less than this threshold represent those benefitting from further onward treatment. This is limited however as it does not take into account the baseline value for the outcome of interest which is likely to be another important determinant when considering which patients are likely to benefit from onward referral.

Models in this Chapter (Chapter 7) are a first step to exploring and understanding which factors may be important to predict change over time in the outcomes. The role of the baseline value of the outcome is explored further in the analysis presented in chapter 8, which aims to provide an alternative approach to identifying important predictors of change over time. This chapter explores if distinct subgroups of participants can be identified that have similar within-group trajectory shapes over time.

8 Trajectory subgroups for hand pain and function in CAS-HA

8.1 Introduction

The previous chapter explored whether key baseline predictors of the trajectory of hand pain and function could be identified when variation around a single mean trajectory curve was modelled. The tenability of this assumption is explored in this chapter by testing whether a small number of subgroups exist in the data that have differing trajectory shapes of hand pain and function over time for each outcome modelled separately in the data (Objective 1). If such subgroups exist, their validity is then explored by examining whether the baseline characteristics of participants differ between the groups identified (Objective 2). It is also tested whether any combinations of baseline characteristics can be identified that predict trajectory group membership well as measured by goodness-of-fit and predictive validity of the model derived (Objective 3).

The chapter is structured by describing the methods used to address Objectives 1-3 in Section 8.2. The results are then presented in Section 8.3 and take into account any revisions to the methods that were needed in light of the results that were found at earlier stages in the modelling process. The chapter then concludes with a discussion of the results and the strengths and limitations of the analysis as presented (Section 8.4).

8.2 Methods

8.2.1 Identifying the optimum number of trajectory groups and polynomial form

Modelling strategy

LCGM (as described in Chapter 5) was used to identify the optimum number of subgroups and polynomial form to model in the CAS-HA data[89], with a search for the optimum model needed, as prior to this current analysis, no published studies in older adults with hand pain/problems were identified to suggest what the optimum model might be for this

[89] GMM was only used in a later stage of the modelling process (see Section 8.2.2 for details)

population of interest (published studies on trajectories of pain or function identified have focused on knee pain, back pain or other musculoskeletal conditions rather than people with hand problems).

The search for the optimum model was conducted by initially fitting a one-group linear model to the data[90] and then subsequently increasing the number of groups (by one) until model fit no longer improved, with this approach to model fitting being recommended by Curran et al. 2003. Model fit was assessed using the full range of fit indices described in Chapter 5 along with a judgement as to whether the model was clinically useful, which, in the context of this chapter, was defined as a model where at least one or more groups had a rate of change over time that differed from all other groups in the model. This definition was used so that groups from the optimum model could be used as an outcome to describe the characteristics of participants with varying rates of change over time and aligns with the philosophy that the selection of the optimum model should be based both on clinical usefulness and statistical goodness-of-fit (Jung et al. 2008, Curran et al. 2003)[91]. Clinically useful models were only considered optimal if they also showed statistical fit indices that were not greatly inferior to the other alternative models considered. This was done to avoid labelling models as optimum if they were over-fitted and would hence be unlikely to be reproduced in other external data samples.

When the optimal linear models for hand pain and function had been identified it was explored whether the addition of quadratic and cubic terms modified trajectory shapes and improved model fit to the data (Curran et al. 2003). For simplicity, if a quadratic term was added to the model, it was added to all groups in the model, irrespective of whether the quadratic term was non-significant in some of the trajectory groups. Also, when a cubic term was fitted to the model, the cubic term was added to a model that already contained

[90] A linear model was chosen (initially) for simplicity, to encourage model convergence by reducing the number of parameters estimated, and also because the mean trajectory of both outcomes in the CAS-HA sample was roughly linear

[91] If the optimum model included within-group mean trajectories that were stable over time and differentiated only by their intercept, predicting group membership as an outcome would offer little additional insight in to the data than was gained from the growth/GEE models shown in Chapter 6

a linear and quadratic component within it (irrespective as to whether these terms were significant in the model or not). Therefore, in the results section, “the quadratic model” refers to a model with both a linear and quadratic term in it for all trajectory groups, and “the cubic model” refers to a model with a linear, quadratic and cubic term in it for all trajectory groups.

Sensitivity analysis

The strategy described above was applied to all participants irrespective of the amount of AUSCAN data they had present and was considered the primary analysis as the amount of information to estimate each model parameter was maximised. However, before a model could be concluded as optimal at any stage of analysis, a sensitivity analysis was carried out using only those participants that had AUSCAN data at all time-points, to test if a model derived from participants with complete data gave similar trajectory shapes, and if participants with complete data were allocated to the same trajectory groups. In addition, it was checked if the whole decision making process (i.e. the search for the optimum number of groups and polynomial form) would have been the same had only participants with complete-case data been analysed.

8.2.2 Growth mixture modelling

GMM was used to test how clearly group membership was defined within the optimal LCGM derived in Section 8.2.1. This was explored by testing whether model fit improved if the optimal LCGM was re-defined as a GMM with the same number of groups and polynomial form as the optimal LCGM. If model fit was shown to improve using the GMM, the GMM was then considered as the optimal model for the remaining analysis in this chapter.

8.2.3 Describing and predicting trajectory subgroup membership

After the optimum number of latent groups and polynomial form were identified using the strategy described in Sections 8.2.1 to 8.2.2, descriptive statistics were used to describe

the baseline characteristics of participants in each subgroup, i.e. using count and percentage data (for categorical outcomes), means and standard deviations (for continuous, normally distributed data) and median and interquartile range (for continuous, skewed data). The baseline characteristics used for this analysis were those listed in Table 3-1 of Chapter 3 and the analysis was conducted separately for hand pain and function.

Multinomial logistic regression, as described in Appendix 25, was then used to explore whether there were any specific combinations of baseline characteristics that predicted trajectory group membership well. This analysis technique was used as the outcome of interest (trajectory group membership) was measured on a nominal scale, however if only two trajectory groups were identified from the previous analysis, logistic regression would be used as a simpler alternative. The potential predictors of interest were selected using forward-stepwise selection modelling techniques only after taking into account the number of events observed per regression coefficient estimated (see Section 8.3.9 for further details). Likelihood ratio tests were used to compare the log likelihoods of the models with and without the predictor in them to identify variables that were statistically significant predictors of the outcome of interest (Long et al. 2006) (cut-offs of $p < 0.05$ were used for entry of the predictor into the model and $p > 0.10$ for removal)[92].

[92] It is possible to simultaneously model predictors of trajectory group membership within an LCGM using a one-step or three-step maximum likelihood approach (Vermunt 2010, Asparouhov 2014a). The one-step approach was not used in this thesis as trajectory group membership can change depending on which specific predictors are included in the model, hence making the definition of a single optimum LCGM difficult. The three-step approach was preferred to the one-step approach, as uncertainty in class membership allocation can be included in the multinomial model and estimated without changes to trajectory group membership, but it was not possible to apply this method to this analysis problem as a model would not run whereby a key predictor of interest i.e. the baseline in the outcome of interest, was also included as a key variable to define the trajectory groups. It therefore remains a limitation of the multinomial logistic regression presented that perfect class allocation is assumed.

8.2.4 Assessing model goodness-of-fit and predictive validity for multinomial/logistic regression

Goodness of fit

Model goodness of fit[93] was assessed using both Nagelkerke's pseudo R-squared value (range 0 to 1 (Hu et al. 2006)) and the Brier score (range 0 to 2 for a multinomial logistic regression (Biesheuvel et al. 2008)) (see Appendix 26 for details). The Nagelkerke's[94] pseudo R-square was chosen from a range of potential pseudo R-square statistics available in the literature (e.g. Cox and Snell, McFadden's, Effron's pseudo R-square (Institute for Digital Research and Education 2011)) as this measure is commonly used (Hu et al. 2006) and, as it is measured on a scale from 0 to 1, is intuitively appealing as an approximation to an R-square statistic derived from linear regression[95] (Institute for Digital Research and Education 2011). A higher value of the Nagelkerke's pseudo R-square, and a lower value on the Brier score, indicated better model fit.

Along with goodness-of-fit, the following concepts were also considered to assess the predictive validity of the model: model discrimination (i.e. can the model discriminate between participants belonging to, or not belonging to, each trajectory group (Steyerberg et al. 2010)), model calibration (i.e. is there agreement between observed outcomes and the predictions made from a model (Steyerberg et al. 2010)), and model accuracy (i.e. is there agreement between trajectory group membership derived from the LCGM and that based on the predicted probabilities from the multinomial/logistic model (i.e. the

[93] Goodness of fit is defined as the distance between the predicted outcome and the actual outcome of interest (Steyerberg et al. 2010)

[94] This is the same as the Cragg and Uhler pseudo statistic that is calculated in STATA version 13.0 (Institute for Digital Research and Education 2011)

[95] The percentage of variance explained by the model (R-square) cannot be directly calculated for a multinomial logistic regression model as estimates are derived using maximum likelihood (an iterative technique) rather than ordinary least squares as used for linear regression (Institute for Digital Research and Education 2011). Several pseudo R-square statistics have therefore been proposed to approximate to a true R-square value, but they themselves are approximations, and not exact measures of the percentage of variance explained by the model (University of Strathclyde Humanities and Social Sciences)

"predicted" group membership). The methods used to test each concept are explained further below.

Model discrimination

Model discrimination was assessed by generating box-plots of the predicted probabilities obtained from the multinomial/logistic model stratified by observed trajectory group membership (i.e. the trajectory group that the participant was assigned to from the LCGM). If the model discriminated well between trajectory groups, the predicted probabilities would be higher for those in the trajectory group that directly related to the predicted probabilities that were plotted. In addition, a set of C-statistics were also calculated to explore how well each trajectory group could discriminate from a reference trajectory group that was defined after the number and form of the trajectory groups was determined[96]. The C-statistic used in binary logistic regression (defined as the area under the ROC curve[97]) was extended to the multinomial model using the 'conditional risk method', as recommended by Van Calster et al. (2012a) and was used to overcome the difficulty that predicted probabilities do not sum to one if only two categories are considered from the multinomial logistic model when calculating the C-statistic. An example of how the 'conditional risk method' was applied is shown in Appendix 27.

Calibration

Model calibration was assessed separately for each predicted probability generated from the multinomial/logistic model[98] and was calculated by grouping the predicted values into

[96] It was considered whether to test (using C-statistics) how well the model discriminated each trajectory group from all others, but this approach was not used as it would blur information about which particular trajectory groups were well discriminated in the model and which were not

[97] The ROC curve is generated using the methodology described in Chapter 4 and is calculated by comparing the true outcome (measured as 0 and 1) to multiple cut-offs on the predicted probability scale. The area under the ROC curve is interpreted as the probability that the model gives a higher probability if the participant has the event than a non-event (Van Calster et al. 2012a)

[98] For model calibration, the predicted probabilities were calculated directly from the multinomial model rather than those generated from the conditional risk method to explore model discrimination

deciles[99] and counting (for each decile) the number of people in the trajectory group with the same label as given for the predicted probability, e.g. if the predicted probability of being in trajectory group one was categorised, it would be the number of people in trajectory group one from the LCGM that would be counted. This data was plotted on a graph and the data points inspected to see how closely they lay to a straight 45^{o} line from the origin (a 45^{o} line would indicate a perfectly calibrated model)[100]. To quantify the calibration of the model, a linear regression model was also fitted to the data points to see how far the intercept and slope estimates were from the perfectly calibrated values of 0 for the intercept and 1 for the slope.

Accuracy

Model accuracy was assessed by calculating the percentage of agreement between trajectory group membership derived from the LCGM and that based on the predicted probabilities from the model[101]. Although the accuracy of the model was reported, it was not considered in great detail, as model accuracy has the conflict that it can be improved even when the usefulness of the model as a prediction tool is degraded (see Appendix 28 for details on the accuracy paradox).

8.3 Results

8.3.1 Identifying the optimum number of trajectory groups and polynomial form

For both hand pain and function the optimum number of latent groups was not consistent across (a) the different goodness of fit indices considered, (b) the polynomial form

[99] Ten equally spaced groups along the probability continuum i.e. 0 to 0.1, 0.1 to 0.2, 0.2 to 0.3 etc

[100] The Hosmer-Lemeshow goodness-of-fit test was not used to statistically compare the observed and expected proportions from the calibration plot (with expected numbers calculated by summing the probabilities for all participants in the decile) as it produces a p-value sensitive to sample size i.e. any small differences between observed and expected frequencies will be statistically significant if the sample size is large (University of Strathclyde Humanities and Social Sciences). It also provides limited information as to where (i.e. for what range of predicted probabilities) the model does, or does not, fit the data well (Afifi et al. 1996) and has the potential to depend on the number of groups used to stratify the data prior to calculating the observed and expected counts (Allison 2013)

[101] The "predicted" group membership was derived by assigning participants to the trajectory group that they had highest predicted probability for across their individual probabilities derived from the multinomial model

modelled and (c) between an unrestricted and a complete case analysis. Consequently, the considerations described in Table 8-1 were used to determine an optimal model for hand pain, and function, respectively.

Table 8-1: Decision process used to select the optimal LCGM for hand pain and function

Hand pain	Hand function
Stage 1: Optimal <u>linear</u> model **5-, 6- and 7-group models** 7-group model potentially optimal from the AIC and sample size adjusted BIC (Table 8-2). Reject as optimal as nearly half of all groups have a small N (<5%) and one posterior probability is less than 0.7. 5-group model preferred over the 6-group model from three fit indices: the BIC, the VLMR LRT and the LMR LRT, although the BIC value is tied for the 5- and 6-group models. Reject as optimal as trajectory plot gave virtually parallel lines so lacked clinical usefulness (Figure 28). 6-group model considered potentially optimal at the end of stage 1 (Figure 28)	**Stage 1: Optimal <u>linear</u> model** **5- 6- and 7-group models** 7-group model potentially optimal from the AIC, BIC and sample size adjusted BIC (Table 8-3). Reject as optimal as one group with small N (<5%). 5-group model preferred over the 6-group model as the VLMR LRT and adjusted LMR LRT are only significant for the 5-group model. 5-group model rejected as optimal though as trajectory plot gave virtually parallel lines so lacked clinical usefulness (Figure 29). 6-group model considered potentially optimal at the end of stage 1 (Figure 29)
Stage 2: Addition of quadratic and cubic terms 6-group model: quadratic term significant ($p < 0.05$) for four out of the six groups. Cubic term significant in one group when cubic term added to the quadratic model (Figure 28). 6-group cubic model considered potentially optimal at the end of stage 2 as cubic term significant in one group	**Stage 2: Addition of quadratic and cubic terms** 6-group model: quadratic term significant ($p < 0.05$) for one out of the six groups (Figure 29). Cubic term significant in one group when cubic term added to the quadratic model (Figure 29). 6-group cubic model considered potentially optimal at the end of stage 2 as cubic term significant in one group
Stage 3: Sensitivity analysis using the complete$^{\alpha}$ data set 6-group cubic model was not robust in the complete data (Figure 28). 6-group cubic model rejected as optimal.	**Stage 3: Sensitivity analysis using the complete$^{\alpha}$ data set** 6-group cubic model was not robust in the complete case data (Figure 29).

6-group quadratic model considered 6-group quadratic model rejected as optimal as not robust in the complete data (Figure 28) 6-group linear model rejected as optimal as not robust in the complete data (Figure 28) 5-group model re-considered	6-group cubic model rejected as optimal. 6-group quadratic model considered 6-group quadratic model was robust in the complete case data (Figure 29) and gave reasonable agreement in group membership for those with complete-case data. Group 1 = Moderate (agreement = 98%) Group 2 = Severe (agreement = 83%) Group 3 = Improving (agreement = 94%) Group 4 = Progressively deteriorating (agreement = 92%) Group 5 = Mild/moderate (agreement = 99%) Group 6 = Mild (agreement = 79%) 6-group quadratic model considered optimal at the end of stage 3
Stage 4: Reconsidering the 5-group model Adding a quadratic term to the 5-group model gave trajectory groups with virtually parallel lines (Figure 28) – reject as lacking clinical usefulness 5-group cubic model gave clinically useful groups and the cubic term was significant in one of the five groups (p=0.001). Model robust in the complete case data and gave reasonable agreement in group membership for those with complete-case data (Figure 28). Group 1 = Severe (agreement = 100%) Group 2 = Mild deterioration (agreement = 81%) Group 3 = Moderate (agreement = 88%) Group 4 = Mild (agreement = 91%) Group 5 = Episodic (agreement = 92%)	**Stage 4: Reconsidering the 5-group model** Adding quadratic and cubic terms to the 5-group model gave trajectory groups with virtually parallel lines (Figure 29) – reject as lacking clinical usefulness

Stage 5: Fitting a 5-group cubic GMM Unreliable model estimates - implausible negative estimates of variance around each of the fixed polynomial estimates. Global model solution not achieved. Robust model solution also not obtained in the complete case data	**Stage 5: Fitting a 6-group quadratic GMM** Unreliable model estimates - implausible correlation estimates between the latent intercept, slope and quadratic terms that were greater than one. Robust model solution also not obtained in the complete case data
Optimal model: LCGM 5-group cubic	**Optimal model: LCGM 6-group quadratic**

$^{\alpha}$ The complete dataset was defined as the group of participants with AUSCAN pain/function data at all time-points. The decision on the optimal model would not have differed had the complete case data been initially used and the test for replication applied to all study participants. For a full list of fit indices on the complete dataset see Appendix 29 and Appendix 30.

Table 8-2: Goodness-of-fit Statistics for LCGM fitted to AUSCAN Pain (N=621[α])

Model type	AIC	BIC	ABIC	Entropy	VLMR LRT	Adjusted LMR LRT	PB LRT	Group N	Average posterior probability
Linear									
1	11125	11156	11134	N/A	N/A	N/A	N/A	621	1.0
2	10231	10276	10244	0.80	p<0.001	p<0.001	p<0.001	343, 278	0.95, 0.94
3	9958	10016	9974	0.78	p<0.001	p<0.001	p<0.001	286, 120, 215	0.88, 0.90, 0.92
4	9885	9956	9905	0.72	p=0.001	p=0.001	p<0.001	165, 72, 187, 197	0.87, 0.88, 0.81, 0.81
5	9866	9950	9890	0.74	p=0.009	p=0.012	p<0.001	189, 85, 7, 155, 185	0.78, 0.83, 0.91, 0.87, 0.80
6	9852	9950	9880	0.73	p=0.315	p=0.330	p<0.001	7, 186, 177, 152, 14, 85	0.92, 0.80, 0.74, 0.85, 0.75, 0.84
7	9846	9956	9877	0.75	p=0.607	p=0.616	p=0.020	173, 88, 10, 182, 7, 154, 7	0.76, 0.83, 0.81, 0.81, 0.92, 0.85, 0.69
Quadratic									
5	9863	9969	9893	0.74	p=0.030	p=0.034	p<0.001	86, 185, 189, 8, 153	0.84, 0.80, 0.79, 0.89, 0.87
6	9842	9966	9877	0.73	p=0.060	p=0.068	p<0.001	8, 158, 167, 27, 87, 174	0.89, 0.86, 0.73, 0.74, 0.84, 0.81
Cubic									
5	9859	9987	9895	0.71	p=0.137	p=0.145	p<0.001	81, 159, 198, 144, 39	0.88, 0.73, 0.83, 0.87, 0.72
6	9835	9985	9877	0.73	p=0.106	p=0.113	p<0.001	140, 90, 8, 49, 198, 136	0.86, 0.85, 0.94, 0.72, 0.80, 0.73

For the linear model, highlighting indicates models with the lowest AIC, BIC, ABIC values, models with one group less than the model with a non-significant LRT p-value, group frequencies less than 5% of the sample (i.e. N <30) and posterior probabilities <0.7. All models achieved a global solution as in each model the largest log-likelihood was replicated for more than two random starting values. For a full list of fit indices for groups 1 to 4 and 7 for the quadratic and cubic models for hand pain see Appendix 31. α = two participants were excluded from the analysis as they had no data at all time-points. AIC = Akaike Information Criteria, BIC = Bayesian Information Criteria, ABIC = Sample-size adjusted BIC, VLMR LRT = Vuong-Lo-Mendell-Rubin likelihood ratio test, LMR LRT = Lo-Mendell-Rubin likelihood ratio test, PBLRT = parametric bootstrapped likelihood ratio test, N/A = not applicable, p = p-value.

Figure 28: Trajectory plots explored whilst searching for an optimal hand pain model

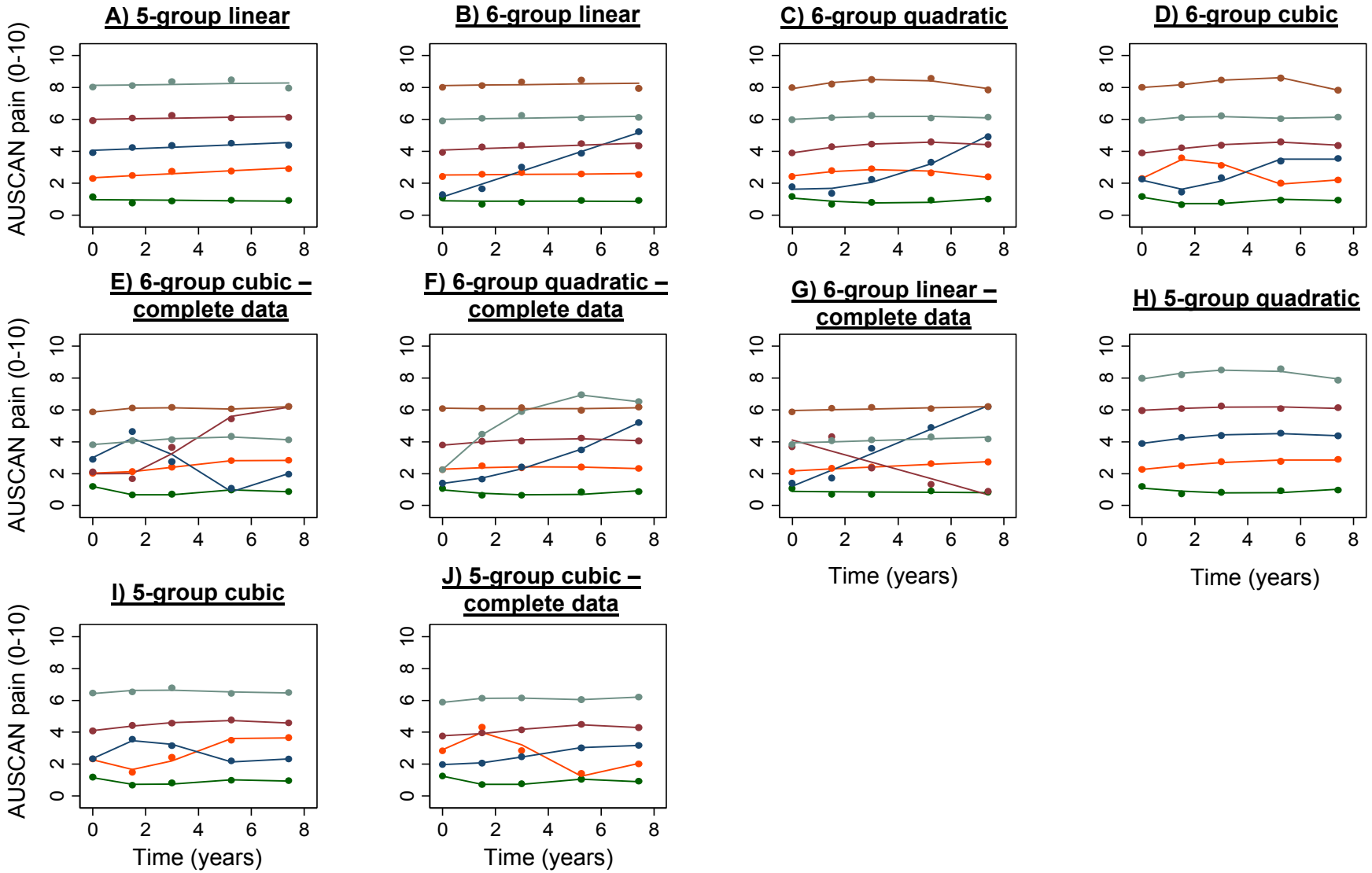
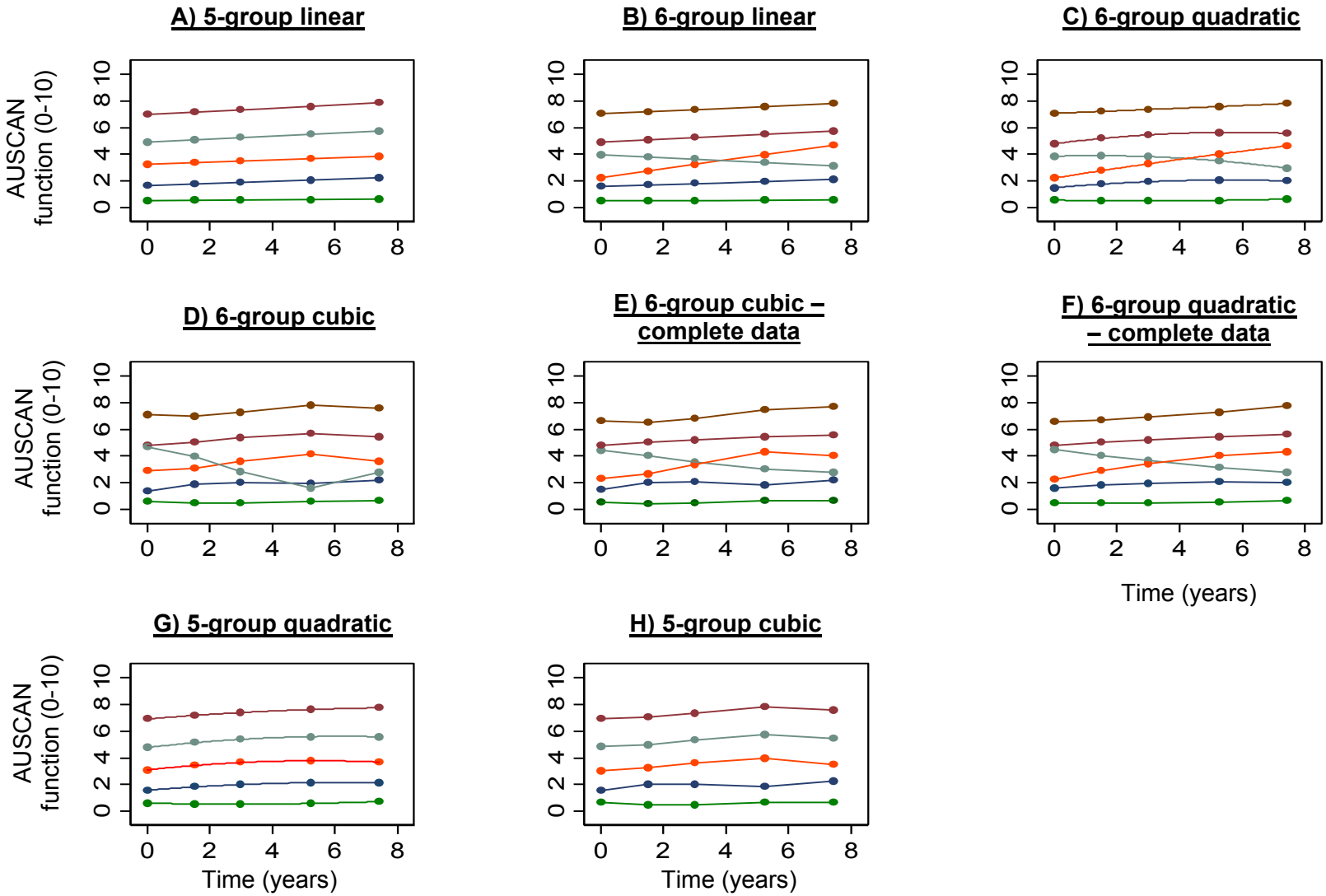


Table 8-3: Goodness-of-fit Statistics for LCGM fitted to AUSCAN Function (N=621[α])

Model type	AIC	BIC	ABIC	Entropy	VLMR LRT	Adjusted LMR LRT	PB LRT	Group N	Average posterior probability
Linear									
1	11519	11550	11528	N/A	N/A	N/A	N/A	621	1.0
2	10059	10104	10072	0.89	p<0.001	p<0.001	p<0.001	247, 374	0.96, 0.97
3	9588	9646	9604	0.86	p=0.077	p=0.084	p<0.001	123, 201, 297	0.94, 0.90, 0.96
4	9415	9486	9435	0.84	p<0.001	P=0.001	p<0.001	258, 171, 128, 64	0.95, 0.87, 0.90, 0.91
5	9346	9430	9370	0.78	p=0.008	p=0.010	p<0.001	129, 60, 173, 117, 142	0.84, 0.91, 0.88, 0.88, 0.78
6	9307	9404	9335	0.77	p=0.243	p=0.258	p<0.001	139, 119, 65, 72, 59, 167	0.78, 0.87, 0.74, 0.73, 0.91, 0.87
7	9278	9388	9309	0.78	p=0.026	p=0.030	p<0.001	70, 67, 114, 6, 59, 138, 167	0.74, 0.74, 0.86, 0.92, 0.88, 0.79, 0.87
Quadratic									
5	9339	9446	9370	0.79	p=0.101	p=0.109	p<0.001	118, 61, 137, 177, 128	0.88, 0.90, 0.79, 0.88, 0.84
6	9302	9427	9338	0.77	p=0.589	p=0.598	p<0.001	121, 58, 70, 64, 138, 170	0.86, 0.91, 0.73, 0.74, 0.78, 0.87
Cubic									
5	9335	9464	9372	0.79	p=0.017	p=0.019	p<0.001	128, 137, 177, 61, 118	0.84, 0.79, 0.88, 0.91, 0.88
6	9299	9450	9342	0.81	p=0.440	p=0.450	p<0.001	131, 118, 60, 175, 124, 13	0.79, 0.84, 0.92, 0.87, 0.87, 0.86

For the linear model, highlighting indicates models with the lowest AIC, BIC, ABIC values, models with one group less than the model with a non-significant LRT p-value, group frequencies less than 5% of the sample (i.e. N <30) and posterior probabilities <0.7. All models achieved a global solution as in each model the largest log-likelihood was replicated for more than two random starting values. For a full list of fit indices for groups 1 to 4 and 7 for the quadratic and cubic models for hand pain see Appendix 32.
α = two participants were excluded from the analysis as they had no data at all time-points. AIC = Akaike Information Criteria, BIC = Bayesian Information Criteria, ABIC = Sample-size adjusted BIC, VLMR LRT = Vuong-Lo-Mendell-Rubin likelihood ratio test, LMR LRT = Lo-Mendell-Rubin likelihood ratio test, PBLRT = parametric bootstrapped likelihood ratio test, N/A = not applicable, p = p-value.

Figure 29: Trajectory plots explored whilst searching for an optimal hand function model



8.3.2 Strategies explored to investigate why GMM could not be fitted to the data

As described in Table 8-1, GMM did not produce reliable estimates when fitted to either the hand pain or function data. By inspecting the model solutions, it was considered whether model convergence could be achieved if it were assumed that any negative estimates of variance were zero, and, any correlations greater than one were one (i.e. by fixing the variance and correlation estimates in the model to be at these pre-defined values). This suggestion was rejected however, as the variance estimates were not close to zero and the correlation estimates were considerably higher than one. Also, as some of the negative estimates of variance related to the intercept term, it seemed unrealistic, given the plots in Figure 30 and Figure 31, that the estimate of variance for the intercept would be precisely zero.

It was also considered whether the GMM could be fitted to the data if the models were simplified to only include a random effect term for the model intercept and slope, rather than for all the higher order polynomial terms included in the model. This was not an option however as negative estimates of variance continued to be produced for the hand pain model and estimates in the hand function model were implausible for one group derived (i.e. average hand function was estimated to be -4 and -38 at the latter two time points so included values out of the range of the 0-10 hand function scale).

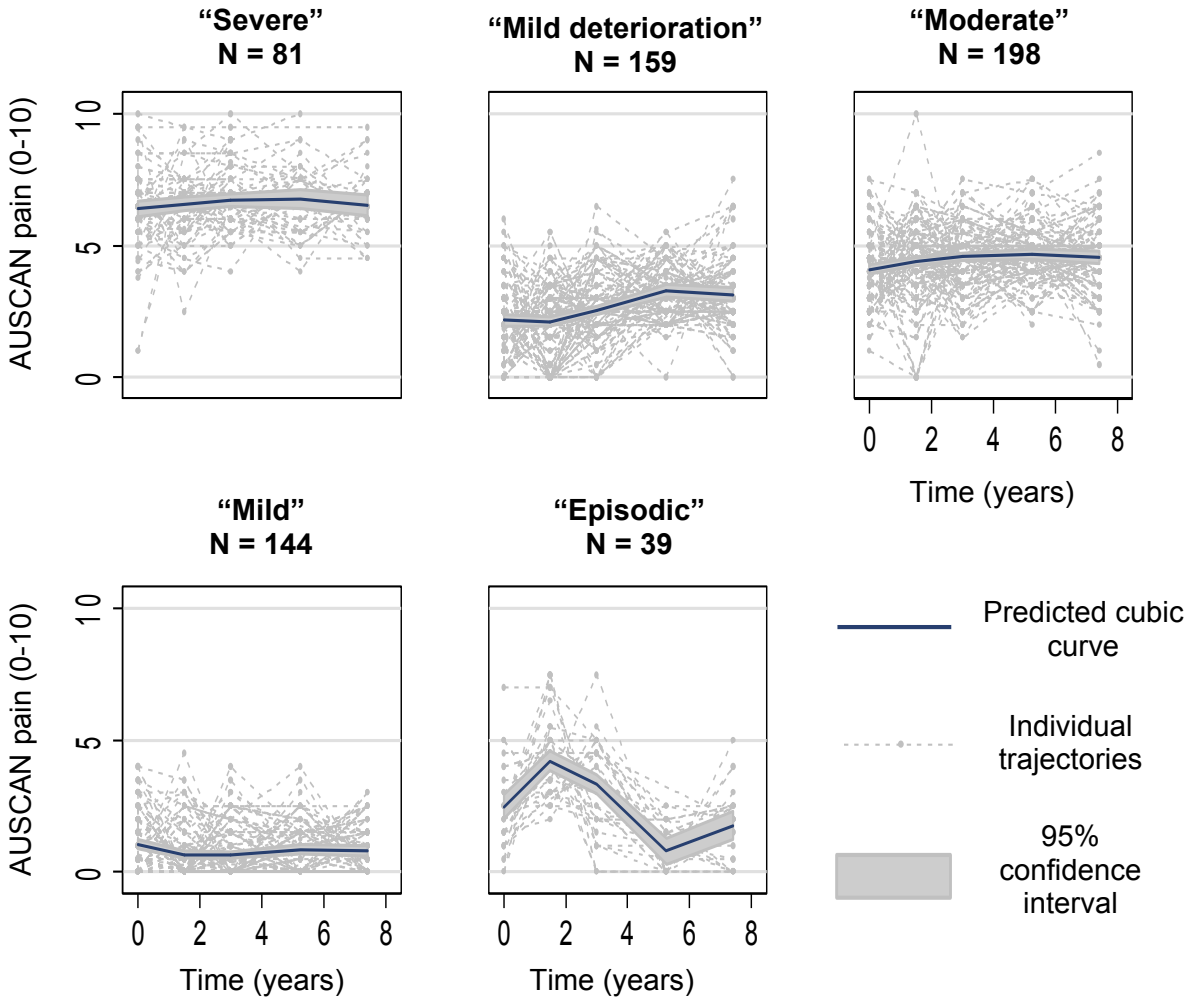
Conversely, increasing the complexity of the model, by relaxing the assumption that the variance estimates around each polynomial term were equal in each group, also did not produce plausible model results, possibly because the model was too complex to be fitted to the data resulting in a local implausible solution being obtained. The GMM approach was therefore not used further in this thesis.

8.3.3 The optimal models for hand pain and function

The optimal LCGM for hand pain was a five-group cubic model with groups labelled as shown in Figure 30. The corresponding graph for the complete-case analysis is in

Appendix 33[102]. The group with the highest prevalence was the "Moderate" group (N = 198; 32%), but this was only slightly higher than the "Mild deterioration" and "Mild groups" (26% and 23% respectively). Participants in the "Severe" group and the "Episodic" groups were less prevalent, making up 13 and 6% of the CAS-HA sample, respectively. In all groups however, there is considerable variation around the mean trajectory curves plotted.

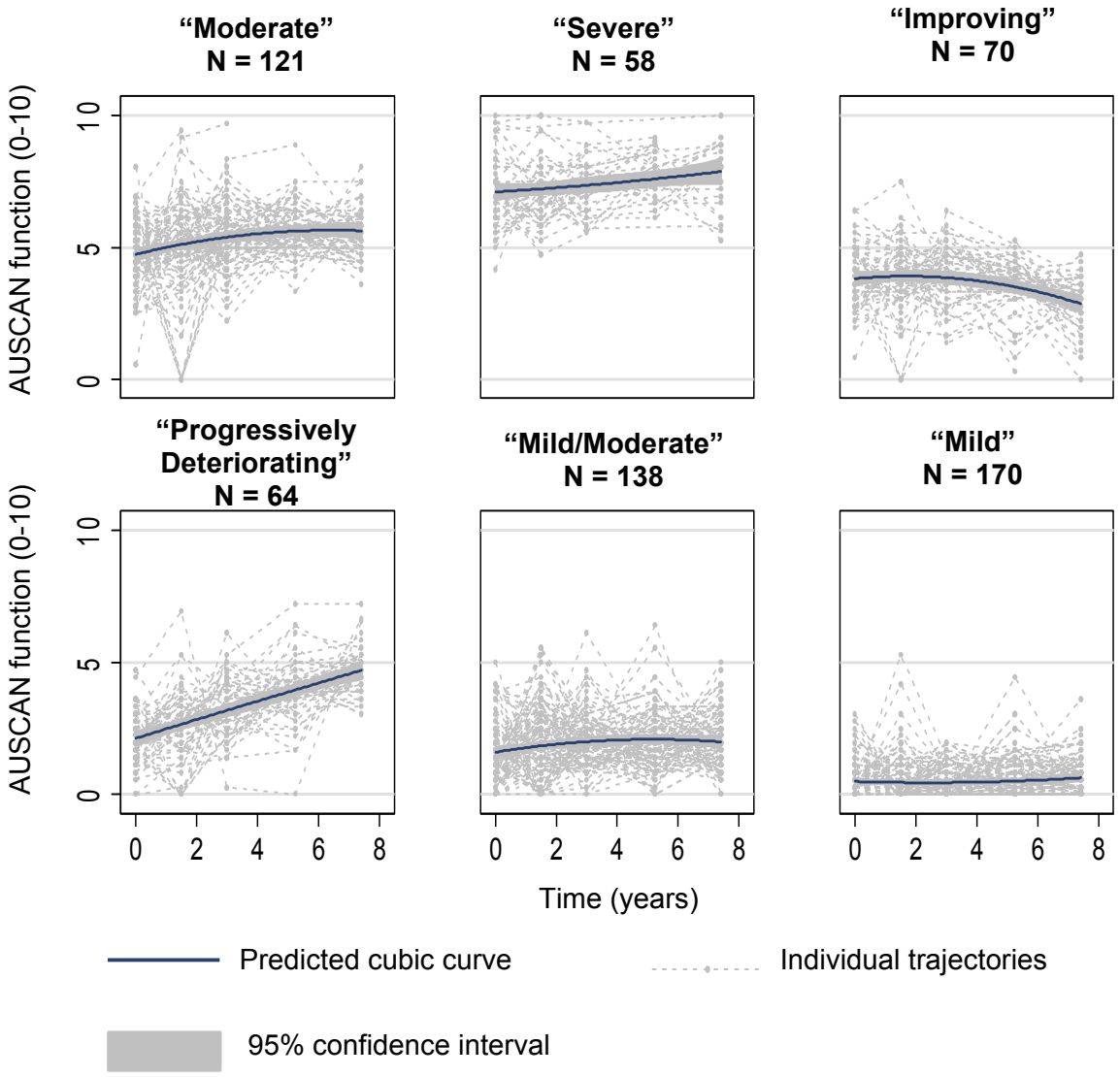
Figure 30: Trajectory plots for hand pain from a 5-group cubic LCGM



[102] The labels were chosen as group descriptors, but it is acknowledged that the labels refer to group mean trajectories rather than to all possible individual trajectories in that group.

The optimal LCGM for hand function was a 6-group quadratic model as shown in Figure 31. The corresponding graph for the complete-case analysis is in Appendix 34[103]. The "Mild" and "Mild/moderate" groups were the groups with the highest prevalence constituting 27% and 22% of the sample, respectively. Only 10% of the sample was in the "Progressively deteriorating" group in the hand function model. Although all groups showed variation in the individual trajectories around the mean curve, this variation was smaller in the "Mild" group than for the other groups presented.

Figure 31: Trajectory plots for hand function from a 6-group quadratic LCGM



[103] The labels were chosen as group descriptors, but it is acknowledged that the labels refer to group mean trajectories rather than to all possible individual trajectories in that group.

A key similarity between the models for hand pain and function is that both contain a group of participants with relatively low levels of hand pain severity/functional difficulty at study entry, which is maintained at this relatively low level throughout the 7.5 year follow-up period (constituting 23% and 27% of the hand pain and function samples, respectively) and a contrasting group of participants where the majority enter the study with an AUSCAN score >= 5 that does not resolve to a level lower than 5 at any point during the study follow-up (constituting 13% and 9% of the samples, respectively).

A key difference between the two outcomes however, is that a single "progressively deteriorating" group was only evident in the hand function model (i.e. a group showing, on average, a continual increase in functional difficulty over time). This group however was relatively small, making up only 10% of the total sample size. Although no equivalent group was found in the hand pain data, two groups showing potential deterioration did emerge: the first showing deterioration between the baseline and the 18-month follow-up followed by improvement in the remaining follow-up periods (Episodic N = 39, 6%), and the second showing mild deterioration in the second part of the follow-up period (i.e. 3-years to 7.5-years) (N = 159, 26%)[104]. Only the hand function model contained a small group of participants whose hand function improved (on average) over the 7.5 year follow-up (N = 70, 11%).

8.3.4 Baseline descriptive characteristics of the trajectory groups

The potential predictors listed in Chapter 3 were used to describe the characteristics of participants in each trajectory group. As the focus of the analysis was to describe the groups, no statistical tests were undertaken to test whether such differences were statistically significant. However, to facilitate a description of the groups especially as the number of predictors and groups were large, two key comparisons were made in the data,

[104] Evidence for deterioration was less strong in the "Mild deterioration" group than the "Episodic" group as the average rate of deterioration in the "Mild deterioration" group is small over time

with only the largest between-group differences reported[105]. The first was to contrast those in the "Mild" and "Severe" trajectory groups so that the magnitude of any differences between the two most extreme trajectory groups could be quantified. The second was to contrast those with a similar range of baseline scores but with differing trajectories over time to potentially identify important predictors of future outcome, i.e. to compare the "Mild deterioration", "Mild" and "Episodic" trajectory groups for hand pain and the "Progressively deteriorating" and "Mild/Moderate" groups for hand function. The results from these comparisons are described in Sections 8.3.5 to 8.3.7 below and relate to Table 8-4 and Table 8-5. In addition, the rates of missing data at the 6-year follow-up time-point were also used to explore if there was any differential loss to follow-up between the trajectory groups derived (Section 8.3.8).

8.3.5 Comparing participants in the "Severe" and "Mild" trajectory groups for hand pain and function

Participants in the "Severe" group for hand pain (when compared to those in the "Mild" group) had worse baseline hand-related characteristics as evidenced by having a longer duration of symptoms, worse hand function (measured by the AUSCAN, GAT, grip and pinch strength), worse hand pain (measured by the AUSCAN), both hands affected, and a greater likelihood of having either OA (measured by the ACR criteria on clinical examination and increased number of joints with OA on x-ray), carpal tunnel syndrome, De Quervain's tenosynovitis or trigger finger. Looking at psychological measures of health, participants in the "Severe" group experience greater levels of anxiety and depression (as measured by the HADS, the emotional representation subscale of the IPQR and mental component score of the SF-12), more frustration with their hand condition and experienced a condition with greater impact on their life. They also have

[105] For categorical characteristics this was defined as difference of greater than 15% and for continuous outcomes as those differences greater than 15% on the range of the scale if no clinical information was available to judge if differences were clinically significant

poorer general health (measured by self-reported and the physical component score of the SF-12), more likely to report pain in other areas of the body, and are less likely to drink on a regular basis. Demographically, participants in this group are more likely to be female, not in current employment and have to be careful with money to manage on the income they have.

The characteristics described above were also the main characteristics that differentiated the “Severe” and “Mild” trajectory groups in the hand function model.

8.3.6 Exploring potential predictors from the hand pain model

Compared to participants in the “Mild” group, participants in the “Mild deterioration” and “Episodic” groups were more likely to a) report worse general health (poorer self-rated health, lower SF-12 physical component score), b) have higher rates of pain elsewhere (regional or widespread), and c) have hand problems that appear to be of longer duration, have a gradual rather than acute onset, and report more days where they were frustrated by their hand problem. A small number of characteristics differentiated between the “Mild deterioration” and “Episodic” groups: those in the “Mild deterioration” group were more likely to have both hands affected and more joints with radiographic hand OA, but less likely to be married or cohabiting.

8.3.7 Exploring potential predictors from the hand function model

For hand function, differences in baseline characteristics between the “Progressively deteriorating” and “Mild/Moderate” trajectory groups were generally smaller than those for hand pain. The largest differences observed were that participants in the “Progressively deteriorating” group were more likely to have widespread pain and carpal tunnel syndrome than those in the “Mild/Moderate” group.

8.3.8 Differential missing data rates

The percentage of participants with missing data at the 6-year follow-up differed depending on trajectory group membership, with participants in the “Severe” and

“Moderate” trajectory groups being more likely to have data missing at the 6-year follow-up than those in trajectory groups with less extreme symptom severity.

Table 8-4: Key baseline characteristics by trajectory group membership for hand pain

	Hand pain				
Baseline characteristics	Severe	Mild deterioration	Moderate	Mild	Episodic
	N=81	N=159	N=198	N=144	N=39
Percentage of missing data for AUSCAN pain at the 6-year follow-up	49%	33%	43%	34%	26%
Age[β]	64 (58, 71)	64 (58, 70)	64 (58, 71)	61 (57, 71)	60 (55, 66)
Female gender	57 (70)	99 (62)	131 (66)	76 (53)	22 (56)
Married/cohabiting[γ]	62 (78)	120 (76)	154 (78)	115 (80)	36 (92)
Manual occupation[δ]	40 (58)	73 (49)	93 (50)	77 (56)	21 (54)
Currently employed	11 (14)	44 (29)	41 (22)	53 (38)	16 (42)
Age when left school (years)[β]	15 (15, 16)	15 (15, 16)	15 (15, 16)	15 (15, 16)	15 (15, 15)
Go from school to full time education	11 (14)	22 (14)	30 (15)	31 (22)	3 (8)
Income					
Find it a strain to get by from week to week	5 (6)	6 (4)	11 (6)	1 (1)	0 (0)
Have to be careful with money	35 (44)	54 (35)	85 (43)	39 (27)	14 (37)
Able to manage without much difficulty	30 (38)	66 (42)	71 (36)	68 (48)	16 (42)
Quite comfortably off	9 (11)	30 (19)	29 (15)	35 (24)	8 (21)
Alcohol consumption					
Daily or most days	12 (15)	40 (25)	36 (18)	39 (27)	9 (23)
Once or twice a week,	23 (29)	56 (35)	76 (39)	43 (30)	19 (49)
Once or twice a month	12 (15)	29 (18)	35 (18)	32 (22)	7 (18)
Once or twice a year	19 (24)	26 (16)	29 (15)	21 (15)	2 (5)
Never	14 (18)	7 (4)	21 (11)	8 (6)	2 (5)
Smoking status					
Never	34 (42)	78 (49)	102 (52)	80 (57)	20 (51)
Previously smoked	36 (44)	68 (43)	81 (41)	57 (40)	15 (38)
Currently smoke	11 (14)	12 (8)	14 (7)	4 (3)	4 (10)
Lives alone	17 (22)	30 (20)	35 (18)	23 (17)	3 (8)
General Health					
Excellent	0	11 (7)	2 (1)	11 (8)	5 (13)

Very good	8 (10)	38 (24)	24 (12)	53 (37)	9 (23)
Good	15 (19)	76 (48)	85 (44)	66 (46)	14 (36)
Fair	38 (47)	31 (20)	67 (35)	13 (9)	8 (21)
Poor	20 (25)	3 (2)	16 (8)	1 (1)	3 (8)
Physical component score of the SF-12[β]	26.9 (23.0, 33.8)	44.3 (31.9, 51.4)	34.6 (27.3, 45.0)	51.4 (41.7, 55.3)	46.0 (33.0, 51.7)
Number of comorbidities[β]	1.0 (0.0, 2.0)	1.0 (0.0, 2.0)	1.0 (0.0, 2.0)	0.5 (0.0, 1.0)	0.0 (0.0, 1.0)
Pain in other body areas					
No other pain	3 (4)	21 (13)	26 (13)	45 (31)	6 (15)
Regional pain	35 (43)	110 (69)	98 (49)	88 (61)	25 (64)
Widespread pain	43 (53)	28 (18)	74 (37)	11 (8)	8 (21)
AUSCAN pain[β]	6.5 (6.0, 7.0)	2.0 (1.1, 3.0)	4.0 (3.0, 5.0)	0.5 (0.0, 2.0)	2.5 (1.5, 3.0)
Number of days in the last 12-months with hand pain					
less than 7-days	2 (3)	8 (6)	13 (7)	15 (15)	2 (5)
1-4 weeks	0 (0)	24 (17)	13 (7)	23 (22)	6 (16)
>1-mth but <3-mths	1 (1)	29 (20)	24 (13)	27 (26)	4 (11)
3-mths or more	77 (96)	82 (57)	140 (74)	38 (37)	26 (68)
AUSCAN function[β]	6.4 (5.0, 7.5)	1.4 (0.6, 2.5)	3.9 (2.5, 5.0)	0.6 (0.0, 1.1)	1.1 (0.6, 1.9)
Side affected					
Dominant hand only	2 (2)	13 (8)	17 (9)	34 (24)	9 (23)
Non-dominant hand only	1 (1)	8 (5)	5 (3)	11 (8)	3 (8)
One hand affected but participant ambidextrous	0 (0)	2(1)	2 (1)	3 (2)	0 (0)
Both hands affected	78 (96)	136 (86)	174 (88)	96 (67)	27 (69)
Time since hand problem onset (years)[β]	11 (5, 20)	5 (3, 10)	6 (3, 15)	3 (1, 8)	6 (2, 10)
At least one hand of sudden onset	19 (23)	31 (20)	38 (19)	54 (38)	8 (21)
At least one hand onset following accident or injury	9 (11)	20 (13)	19 (10)	14 (10)	3 (8)
Past or present job, hobbies or pastimes involved excessive hand use	70 (89)	121 (80)	163 (84)	94 (72)	32 (84)
HADS - Anxiety[β]	8.0 (6.0, 10.0)	5.0 (3.0, 7.0)	7.0 (4.0, 10.0)	5.0 (2.0, 8.0)	6.0 (4.0, 9.0)
HADS - Depression[β]	6.0 (3.0, 8.0)	3.0 (1.0, 5.0)	5.0 (3.0, 7.0)	2.0 (1.0, 4.0)	3.0 (2.0, 6.0)
Mental component score of the SF-12[β]	42.6 (34.7, 57.8)	55.2 (49.4, 59.5)	52.4 (41.2, 58.2)	55.4 (48.1, 57.9)	54.0 (46.9, 58.7)
Illness perceptions					

Long disease time course[β]	25.0 (24.0, 29.0)	24.0 (21.0, 26.0)	24.0 (22.0, 27.0)	22.0 (17.0, 24.0)	23.0 (21.0, 25.0)
Consequences[β]	20.0 (17.0, 23.0)	12.0 (9.0, 15.0)	15.0 (13.0, 18.0)	12.0 (9.0, 13.2)	12.0 (10.0, 15.0)
Personal Control[α]	16.6 (4.2)	18.4 (4.2)	17.9 (3.9)	18.2 (4.6)	18.0 (4.5)
Treatment Control[α]	13.0 (3.4)	14.9 (3.1)	14.4 (3.2)	15.4 (3.3)	14.3 (3.3)
Illness coherence[β]	12.0 (10.0, 16.0)	11.0 (10.0, 15.0)	11.0 (10.0, 14.0)	12.0 (10.0, 15.0)	11.0 (10.0, 15.0)
Cyclical time course[β]	11.0 (8.0, 14.0)	12.0 (8.5, 14.0)	12.0 (9.0, 15.0)	12.0 (8.0, 14.0)	10.7 (8.0, 15.0)
Emotional representation[α]	17.4 (5.1)	12.1 (3.7)	14.5 (4.3)	11.8 (3.9)	12.9 (3.8)
Frustration with hand condition in the last month					
All days	23 (29)	2 (1)	13 (7)	2 (2)	0 (0)
Most days	29 (37)	12 (8)	27 (14)	5 (4)	2 (5)
Some days	16 (20)	23 (15)	57 (30)	4 (3)	7 (18)
Few days	6 (8)	33 (22)	48 (26)	19 (15)	11 (29)
No days	5 (6)	82 (54)	43 (23)	94 (76)	18 (47)
Body-mass index[α]	28.6 (5.0)	28.7 (5.0)	28.7 (5.3)	27.2 (3.6)	27.2 (3.7)
Hand grip-ability (GAT)[β]	37.3 (31.3, 49.8)	27.1 (23.7, 33.6)	31.2 (25.0, 38.1)	24.3 (21.8, 28.6)	26.4 (22.8, 30.5)
Grip strength[α]	30.8 (20.9)	53.3 (24.4)	42.4 (22.3)	62.5 (26.1)	54.5 (25.6)
Pinch strength[α]	7.4 (3.7)	10.9 (3.8)	9.3 (3.7)	12.8 (4.5)	11.4 (4.9)
Meets the ACR criteria for hand OA	53 (65)	34 (21)	77 (39)	18 (13)	9 (23)
Has carpal tunnel syndrome	56 (75)	58 (37)	106 (57)	41 (29)	12 (32)
Has Dupuytren's contracture	22 (27)	37 (23)	55 (28)	38 (26)	13 (33)
Has De Quervain's tenosynovitis	30 (46)	30 (19)	51 (28)	19 (14)	7 (18)
Has trigger finger	30 (37)	28 (18)	42 (21)	15 (10)	8 (21)
Number of joints with Kellgren-Lawrence x-ray grade >=2[β]	5.0 (2.0, 9.0)	4.0 (1.0, 8.5)	4.0 (2.0, 8.0)	2.0 (1.0, 5.0)	2.0 (0.0, 5.0)

Figures are numbers and percentages unless otherwise stated. α = Mean (standard deviation), β = Median (inter-quartile range), γ = Dichotomised as Married/Cohabiting versus Separated/Divorced/Widowed/Single due to small N, δ = Manual occupation (i.e. Lower supervisory/technical, Semi-routine occupations or Routine occupation categories of the SOC 2000 coding (Office for National Statistics (ONS) 2002)). SF-12 = Short-form 12, AUSCAN = Australian/Canadian Hand Osteoarthritis Index, HADS = Hospital Anxiety and Depression Scale, ACR = American College of Rheumatology, OA = Osteoarthritis

Table 8-5: Key baseline characteristics by trajectory group membership for hand function

	Hand function					
Baseline characteristics	Moderate N=121	Severe N=58	Improving N=70	Progressively Deteriorating N=64	Mild/Moderate N=138	Mild N=170
Percentage of missing data for AUSCAN function at the 6-year follow-up	48%	62%	34%	28%	32%	34%
Age[β]	66 (59, 73)	64 (58, 72)	62 (57, 68)	66 (58, 73)	62 (58, 70)	61 (56, 67)
Female gender	92 (76)	44 (76)	49 (70)	41 (64)	92 (67)	67 (39)
Married/cohabiting[γ]	87 (73)	43 (74)	57 (81)	51 (80)	114 (84)	135 (80)
Manual occupation[δ]	56 (51)	27 (55)	30 (45)	30 (52)	73 (55)	88 (54)
Currently employed	14 (12)	2 (4)	23 (34)	15 (25)	45 (34)	66 (40)
Age when left school (years)[β]	15 (14, 15)	15 (14, 16)	15 (15, 16)	15 (15, 16)	15 (15, 16)	15 (15, 16)
Go from school to full time education	14 (12)	8 (15)	16 (23)	8 (13)	21 (16)	30 (18)
Income						
Find it a strain to get by from week to week	5 (4)	5 (9)	2 (3)	4 (6)	6 (4)	1 (1)
Have to be careful with money	51 (43)	27 (49)	26 (38)	19 (30)	54 (40)	50 (30)
Able to manage without much difficulty	48 (40)	19 (35)	31 (45)	29 (45)	46 (34)	78 (46)
Quite comfortably off	16 (13)	4 (7)	10 (14)	12 (19)	30 (22)	39 (23)
Alcohol consumption						
Daily or most days	17 (14)	7 (12)	13 (19)	13 (21)	37 (27)	49 (29)
Once or twice a week,	36 (30)	14 (25)	33 (47)	26 (41)	47 (34)	61 (36)
Once or twice a month	31 (26)	9 (16)	9 (13)	10 (16)	21 (15)	35 (21)
Once or twice a year	23 (19)	15 (26)	8 (11)	9 (14)	23 (17)	19 (11)

Never	14 (12)	12 (21)	7 (10)	5 (8)	9 (7)	5 (3)
Smoking status						
Never	58 (48)	29 (50)	42 (60)	34 (53)	74 (54)	77 (46)
Previously smoked	51 (42)	23 (40)	22 (31)	26 (41)	54 (40)	81 (49)
Currently smoke	12 (10)	6 (10)	6 (9)	4 (6)	8 (6)	9 (5)
Lives alone	26 (22)	14 (26)	11 (16)	11 (18)	17 (13)	29 (18)
General Health						
Excellent	0 (0)	0 (0)	4 (6)	2 (3)	7 (5)	16 (9)
Very good	15 (13)	1 (2)	10 (14)	11 (17)	36 (26)	59 (35)
Good	37 (31)	12 (21)	37 (54)	30 (48)	63 (46)	77 (45)
Fair	54 (45)	25 (43)	14 (20)	19 (30)	28 (20)	17 (10)
Poor	14 (12)	20 (34)	4 (6)	1 (2)	3 (2)	1 (1)
Physical component score of the SF-12[β]	30.0 (23.5, 37.1)	25.2 (21.7, 31.9)	41.4 (31.9, 52.1)	40.4 (31.3, 49.2)	43.9 (34.5, 50.9)	51.0 (40.8, 55.5)
Number of comorbidities[β]	1.0 (0.0, 2.0)	2.0 (1.0, 3.0)	1.0 (0.0, 2.0)	1.0 (0.0, 2.0)	1.0 (0.0, 2.0)	1.0 (0.0, 1.0)
Pain in other body areas						
No other pain	13 (11)	4 (7)	8 (11)	7 (11)	23 (17)	46 (27)
Regional pain	61 (50)	19 (33)	44 (63)	33 (52)	88 (64)	111 (65)
Widespread pain	47 (39)	35 (60)	18 (26)	24 (38)	27 (20)	13 (8)
AUSCAN pain[β]	4.5 (3.5, 5.5)	6.5 (5.5, 7.5)	3.5 (3.0, 4.5)	3.0 (2.5, 4.5)	2.0 (1.0, 3.0)	1.0 (0.0, 2.5)
Number of days in the last 12-months with hand pain						
less than 7-days	5 (4)	3 (5)	4 (6)	6 (10)	7 (6)	15 (11)
1-4 weeks	7 (6)	0 (0)	6 (9)	5 (8)	20 (17)	28 (21)
>1-mth but <3-mths	9 (8)	1 (2)	11 (16)	10 (16)	26 (22)	28 (21)
3-mths or more	92 (81)	53 (93)	46 (69)	41 (66)	66 (55)	65 (48)
AUSCAN function[β]	5.0 (4.2, 5.6)	6.9 (6.1, 8.1)	3.6 (3.1, 4.7)	1.9 (1.4, 3.1)	1.4 (0.8, 2.2)	0.3 (0.0, 0.8)
Side affected						
Dominant hand only	7 (6)	3 (5)	7 (10)	3 (5)	20 (14)	35 (21)
Non-dominant hand only	3 (2)	0 (0)	2 (3)	4 (6)	9 (7)	10 (6)
One hand affected but participant ambidextrous	0 (0)	0 (0)	1 (1)	1 (2)	3 (2)	2 (1)
Both hands affected	111 (92)	55 (95)	60 (86)	56 (88)	106 (77)	123 (72)

Time since hand problem onset (years)[β]	9.0 (3.0, 18.0)	12.0 (6.0, 20.0)	6.0 (4.0, 15.0)	5.0 (2.0, 10.0)	5.0 (2.0, 10.0)	4.0 (2.0, 10.0)
At least one hand of sudden onset	27 (22)	17 (29)	12 (17)	10 (16)	34 (25)	50 (29)
At least one hand onset following accident or injury	12 (10)	9 (16)	5 (7)	8 (13)	14 (10)	17 (10)
Past or present job, hobbies or pastimes involved excessive hand use	96 (82)	49 (88)	57 (83)	50 (79)	106 (83)	122 (77)
HADS - Anxiety[β]	7.0 (5.0, 10.0)	8.0 (6.0, 12.0)	7.0 (4.0, 9.0)	7.0 (4.0, 9.0)	6.0 (4.0, 8.0)	5.0 (2.0, 7.0)
HADS - Depression[β]	5.0 (3.0, 8.0)	6.5 (4.0, 9.0)	3.5 (1.0, 6.0)	3.0 (2.0, 7.0)	3.0 (2.0, 5.0)	2.0 (1.0, 4.0)
Mental component score of the SF-12[β]	52.3 (40.5, 59.1)	42.4 (33.9, 56.6)	55.5 (48.0, 58.8)	52.6 (40.6, 57.6)	53.5 (47.4,58.7)	56.0 (49.0, 58.8)
Illness perceptions						
Long disease time course[β]	24.0 (22.0, 28.0)	25.0 (24.0, 29.0)	24.0 (22.0, 26.0)	24.0 (22.0, 27.0)	24.0 (20.0, 25.0)	23.0 (18.0, 24.0)
Consequences[β]	16.0 (14.0, 20.0)	21.0 (17.0, 24.0)	15.0 (12.0, 18.0)	13.0 (11.5, 15.0)	12.0 (9.0, 15.0)	12.0 (9.0, 14.0)
Personal Control[α]	17.9 (4.0)	16.1 (4.0)	18.3 (3.8)	18.1 (4.3)	17.8 (4.6)	18.5 (4.2)
Treatment Control[α]	14.1 (3.4)	13.0 (3.2)	14.5 (2.9)	14.2 (3.3)	14.8 (3.3)	15.3 (3.2)
Illness coherence[β]	12.0 (10.0, 14.0)	12.5 (10.0, 17.0)	10.0 (10.0, 15.0)	12.0 (10.0, 15.0)	11.0 (10.0, 15.0)	11.0 (10.0, 15.0)
Cyclical time course[β]	12.0 (8.0, 14.0)	11.5 (8.0, 14.0)	12.0 (9.0, 14.0)	12.0 (8.0, 15.0)	12.0 (8.0, 15.0)	12.0 (8.0, 14.0)
Emotional representation[α]	15.3 (4.3)	18.3 (4.7)	13.8 (3.8)	13.2 (4.0)	12.5 (3.8)	11.6 (4.2)
Frustration with hand condition in the last month						
All days	16 (14)	21 (38)	0 (0)	1 (2)	1 (1)	1 (1)
Most days	31 (27)	20 (36)	6 (9)	2 (3)	10 (8)	6 (4)
Some days	39 (34)	6 (11)	19 (29)	15 (24)	18 (14)	10 (6)
Few days	18 (16)	5 (9)	22 (33)	18 (29)	30 (24)	24 (15)
No days	11 (10)	4 (7)	19 (29)	27 (43)	66 (53)	115 (74)
Body-mass index[α]	29.4 (5.6)	28.7 (5.5)	28.1 (4.0)	27.8 (4.7)	28.3 (4.9)	27.4 (3.9)
Hand grip-ability (GAT)[β]	35.3 (28.8, 45.1)	44.4 (33.5, 62.0)	30.3 (25.1, 35.9)	27.0 (22.8, 32.6)	26.3 (23.0, 31.1)	24.5 (21.5, 29.6)
Grip strength[α]	34.4 (19.9)	25.7 (15.6)	43.1 (20.8)	48.7 (22.0)	50.5 (20.5)	69.0 (25.2)
Pinch strength[α]	8.1 (3.3)	6.8 (3.4)	8.9 (3.7)	10.1 (3.2)	10.9 (4.1)	13.5 (4.0)
Meets the ACR criteria for hand	56 (46)	36 (62)	27 (39)	16 (25)	39 (28)	17 (10)

OA						
Has carpal tunnel syndrome	56 (50)	46 (88)	44 (64)	33 (52)	47 (36)	47 (28)
Has Dupuytren's contracture	38 (31)	12 (21)	21 (30)	14 (22)	34 (25)	46 (27)
Has De Quervain's tenosynovitis	37 (34)	17 (43)	17 (26)	16 (26)	25 (18)	25 (15)
Has trigger finger	32 (26)	20 (34)	14 (20)	16 (25)	24 (17)	17 (10)
Number of joints with kellgren-Lawrence x-ray grade >=2[β]	4.5 (2.0, 9.0)	5.0 (2.0, 9.0)	4.0 (1.0, 7.0)	4.0 (2.0, 7.0)	3.0 (1.0, 8.0)	3.0 (1.0, 6.0)

Figures are numbers and percentages unless otherwise stated. α = Mean (standard deviation), β = Median (inter-quartile range), γ = Dichotomised as Married/Cohabiting versus Separated/Divorced/Widowed/Single due to small N, δ = Manual occupation (i.e. Lower supervisory/technical, Semi-routine occupations or Routine occupation categories of the SOC 2000 coding (Office for National Statistics (ONS) 2002)). SF-12 = Short-form 12, AUSCAN = Australian/Canadian Hand Osteoarthritis Index, HADS = Hospital Anxiety and Depression Scale, ACR = American College of Rheumatology, OA = Osteoarthritis

8.3.9 Predicting trajectory subgroup membership

The number of groups in the optimal models for hand pain and function influenced how feasible it was to apply the modelling strategy to predict trajectory group membership. In particular, as the number of groups was relatively large in each model this resulted in some groups containing only a small number of participants (the smallest group size for hand pain was 39, and for hand function, 58). As it is recommended for multinomial logistic regression that at least 10 events are needed per regression coefficient estimated (Biesheuvel et al. 2008), this implies that a minimum of 40 or 50 people are needed in each outcome category per regression coefficient estimated for hand pain and function respectively to avoid over-fitting the data[106]. This suggests that with the group sizes given only a single predictor with one degree of freedom could be reliably estimated as a predictor in the model.

Although not ideal, to go some way to address this, the following three steps were taken: (1) to restrict the number of predictors that could potentially enter the model, (2) to simplify the number of response options that the potential predictors are measured on and (3) to bootstrap the model estimates derived. These steps are described below and the limitations of this discussed further in Section 8.4.3.

Restricting the number of predictors included in the model

It was initially stated in the introduction to this chapter that the objective of this analysis was to find the "best" combination of predictors to predict trajectory group membership. This objective was modified by restricting the number of predictors included in any single model to no more than three predictors over and above the baseline value for the outcome of interest. The baseline value was automatically included in all models considered as it is clearly an important predictor of trajectory group membership given the

[106] Minimum values of 40 and 50 were derived as 4 or 5 parameters are estimated in a 5- and 6-group multinomial logistic regression model respectively for hand pain and function when a predictor with one degree of freedom is fitted to the data. This is because the reference category for the parameter estimate is fixed at zero.

trajectory plots in Figure 30 and Figure 31. The limit of three additional predictors was chosen for practical reasons to give scope for additional information to be added to the model over and above baseline, but not so much that the model is greatly over-fitted and unlikely to fit well in another data set; an approach needed due to a lack of evidence in the literature and clinical *a priori* knowledge to define a small set of predictors to include in the analysis.

Simplification of the response categories

The pool of potential predictors listed in Table 3-1 in Chapter 3 were reconsidered to see if any of the categorical predictor variables with more than two response options could be simplified by dichotomising them or trichotomising them if dichotomisation was not an option. As far as possible the cut-off points were chosen to be clinically meaningful, though in some instances the amount of clinical knowledge to choose the cut-point was limited. The response frequency in each category was also considered; if this was small in one particular category then this guided the decision that this category could be merged with an adjacent response category. The categorisation of the potential predictors that was applied to the data prior to analysis is shown in Table 8-6.

Table 8-6: Categorisation of potential baseline predictors if they were measured on a categorical scale

Concept	Full response categories	Simplified response categories
Marital status	Married, Separated, Divorced, Widowed, Cohabiting, Single	1= Married/Cohabiting; 2 = All other categories
Occupation/Social Class	Higher managerial, Higher professional, Lower managerial/professional, Intermediate occupations, Self-employed, Lower supervisory/technical, Semi-routine occupations, Routine occupations	1= Non-manual (all categories listed before self-employed); 2=Self-employed; 3=Manual (all categories listed after self-employed)
Employment status	Employed, Not working due to ill health, Retired, Unemployed, Housewife, Other	1= Employed; 2 = All other categories
Income	Find it a strain to get by from week to week, Have to be careful with money, Able to manage without much difficulty, Quite comfortably off	1= Find it a strain to get by from week to week/have to be careful with money; 2=All other categories
Alcohol consumption	Daily or most days, Once or twice a week, Once or twice a month, Once or twice a year, Never	1=Daily or most days; 2= All other categories
Smoking status	Never, Previously smoked, Currently smoke	1=Never; 2=Previously or currently smoke
Self-rated health	General health: Excellent, Very good, Good, Fair, Poor	1= Excellent, Very good, Good; 2=Fair, Poor

Number of comorbidities	0-7	1 = No comorbidities, 2 = One comorbidity, 3 = 2 or more comorbidities
Pain in other body areas	Manchester definition of regional pain (Macfarlane et al. 1996): No other pain, Regional pain, Widespread pain	1=No other pain; 2 =Regional or widespread pain
Hand pain severity	Number of days in the last 12-months with hand pain: less than 7-days, 1-4 weeks, >1-month but <3-months, 3-months or more	1= < 3-mths of pain in last 12-mths; 2= >=3-mths of pain in the last 12-mths
Side affected	Dominant hand only, Non-dominant hand only, One hand affected but participant ambidextrous, both hands affected	1=One hand affected; 2=Both hands affected
Sudden onset of hand problem	Both hands sudden onset, One hand sudden onset$^{\beta}$, Neither hand sudden onset	1=One or both hands of sudden onset; 2 = Neither hand of sudden onset
Onset of hand condition following accident or injury to the hand	Both hands onset following accident or injury, One hand onset following accident or injury$^{\infty}$, Neither hand onset following accident or injury	1=One or both hands onset following accident or injury; 2=Neither hand of onset following accident or injury
Frustration with hand condition	All days, most days, some days, few days, no days	1 = All days or most days; 2 = Some, few or no days

Bootstrapping of model regression estimates

Bootstrapping was used to explore the stability of the relative risk ratios (RRRs) that were generated from the "final" multinomial model (see Appendix 25 for details of the multinomial logistic model) and to explore the impact that a small sample size may have on the reliability of RRRs obtained. The aim of the bootstrapping procedure was to explore the stability of the model estimates rather than to adjust for any over-estimation in model fit due to the lack of an external data set used to test the performance of the model (see Appendix 36 for an example of how bootstrapping could be used to adjust the C-statistic for over-estimation). The bootstrapping procedure was undertaken by sampling, with replacement, 5000 random data sets that were the same size as the dataset of interest (i.e. after participants with missing data for any of the predictor variables of interest had been excluded). In each of the bootstrapped samples (N=5000), the multinomial logistic regression models for hand pain and function were applied and the relative risk ratios recorded.

The distribution of each RRR was then examined, and from this, 95% normal-based bootstrapped confidence intervals were calculated, defined as the RRR in the original sample[107] +/- 1.96 x standard deviation of the RRR across all bootstrapped samples (with the standard deviation used to estimate the standard error of the RRR distribution). If the lower and upper limit of the bootstrapped confidence interval were both above one, or both below one, this added to the evidence that the predictor was an important predictor of the outcome of interest. It was also checked that the conclusions from the analysis would not change if percentile based, or bias corrected confidence intervals were used to calculate the bootstrap confidence intervals[108].

[107] The RRR in the original sample is used as the point estimate for calculating the confidence interval (rather than the mean of the RRR across the bootstrap samples), as any bias in the original estimate will only be exaggerated when the data are bootstrapped (StataCorp 2013e)

[108] 95% percentile based confidence intervals were calculated as the 2.5th and 97.5th percentiles of the RRR bootstrapped distribution. Bias-corrected confidence intervals adjust these percentile values to account for any difference between the RRR estimate in the original sample and the

8.3.10 Selection of model predictors

The “Mild” group was chosen as the reference category for all multinomial logistic regression models. A reference category was needed so that the model could be estimated, but its choice was somewhat arbitrary, as it only influenced which between group comparisons could be inferred from the model, rather than the p-values derived from the likelihood ratio tests to assess predictors to include in the model.

An automated forward stepwise procedure was applied for multinomial logistic regression using STATA version 13.0, but this was problematic. Although the aim of the model was to search for only three additional predictors, so any model itself would only include up to four predictors including baseline level of the outcome, a full list of potential predictors needed to be specified before the model could be run. The model ran when 35 predictors were defined, but not when the 36 predictors considered here were listed suggesting that a maximum number of variables (35 variables) had been reached in STATA’s stepwise command[109].

A forward stepwise procedure was therefore conducted “by hand” and was used to identify the three strongest predictors of trajectory group membership for each outcome, over and above the baseline measure. As the objective of this model was to identify a small set of strong predictors, the p-value cut-off used to remove predictors from the model was reduced from 0.10 as originally specified, to 0.05 to select only strong predictors rather than develop an all-inclusive model containing predictors of borderline significance. In addition, the item domain structure used to group the predictors in Chapter 3 was not used to restrict the order that predictors were added to the model.

---

mean of the RRR estimate across the bootstrapped samples (i.e. adjust for bias) (StataCorp 2013e). Accelerated bias-corrected 95% confidence intervals (i.e. that adjust the percentile values for both bias and skewness in the bootstrapped distribution (Frangos 1990)) could not be calculated as the multinomial logistic regression model could not be fitted for a small number of bootstrapped samples (<0.4%) due to no participants being selected in that bootstrap sample who were in the trajectory group with the smallest group size

[109] This also occurred when forward selection, rather than forward stepwise, was used as the method for model selection

This approach, although practical, was not ideal as it restricted the range of bootstrapping options available to test the stability of the parameter estimates initially derived (i.e. only the model with a fixed set of predictors in it could be bootstrapped, rather than re-selecting the three most important predictors separately for each bootstrap sample)[110]. It was preferred, however, over restricting the analysis to only those predictors that showed clear differences between trajectory groups in the descriptive data as this could mask potential predictors that may become significant when assessed alongside other predictors in the model.

8.3.11 Multinomial logistic regression – model results

The three strongest predictors of the hand pain trajectory group, over and above baseline hand pain, were: AUSCAN hand function, pinch strength and sudden symptom onset (Table 8-7). As an illustration of the RRR's in Table 8-7, the RRR of 6.6 for the predictor of sudden onset in the "Severe" trajectory group shows that the relative risk of being in the "Severe" group compared to the "Mild group was 6.6 times higher for those who did not have a sudden onset of symptoms (i.e. had a gradual onset) compared to those reporting a sudden onset. The RRRs were statistically significant for all predictors when compared to the "Mild" group, with the exception of AUSCAN function. For this predictor the relative risk was not significantly inflated for a one-unit increase in AUSCAN functional difficulties for the "Mild deterioration" and "Episodic" categories when compared to the "Mild" group. The magnitude of the RRRs were comparable across trajectory groups for pinch strength, however for AUSCAN pain, function and sudden onset of symptoms the RRRs were larger for the "Severe" and "Moderate" groups compared to the remaining within-predictor trajectory groups of interest.

[110] It was explored whether the stepwise command could be used in the statistical software SPSS to achieve a model that fitted the data well. This was a possibility, but as a Centre, we did not have a subscription for the SPSS bootstrap add-in module, so the bootstrapping aspect of the analysis could not be applied in SPSS

Table 8-7: Multinomial logistic regression results for hand pain with the "Mild" group as the reference category (N=577)

Predictor	Relative risk ratio (RRR) (95% CI)	Relative risk ratio (RRR) (95% normal-based bootstrapped CI)
Severe		
AUSCAN pain	13.8 (8.5, 22.4)	13.8 (7.7, 24.6)
AUSCAN function	2.0 (1.4, 2.9)	2.0 (1.4, 3.1)
Average pinch strength (lbs)	0.8 (0.7, 0.9)	0.8 (0.7. 0.9)
Sudden onset of hand condition		
One or both hands of sudden onset	Ref	Ref
Neither hand of sudden onset	6.6 (2.1, 20.6)	6.6 (2.0, 22.2)
Mild deterioration		
AUSCAN pain	2.1 (1.6, 2.7)	2.1 (1.6, 2.7)
AUSCAN function	1.0 (0.8, 1.2)	1.0 (0.8, 1.3)
Average pinch strength (lbs)	0.9 (0.8, 0.9)	0.9 (0.8, 0.9)
Sudden onset of hand condition		
One or both hands of sudden onset	Ref	Ref
Neither hand of sudden onset	3.4 (1.9, 6.4)	3.4 (1.9, 6.4)
Moderate		
AUSCAN pain	4.6 (3.4, 6.2)	4.6 (3.4, 6.1)
AUSCAN function	1.5 (1.2, 1.9)	1.5 (1.1, 2.0)
Average pinch strength (lbs)	0.9 (0.8, 0.9)	0.9 (0.8, 0.9)
Sudden onset of hand condition		
One or both hands of sudden onset	Ref	Ref
Neither hand of sudden onset	6.6 (3.0, 14.6)	6.6 (2.8, 15.3)
Mild (reference group)	1	1
Episodic		
AUSCAN pain	2.5 (1.8, 3.6)	2.5 (1.8, 3.5)
AUSCAN function	0.9 (0.6, 1.2)	0.9 (0.6, 1.3)
Average pinch strength (lbs)	0.9 (0.8, 1.0)	0.9 (0.8, 1.0)
Sudden onset of hand condition		
One or both hands of sudden onset	Ref	Ref
Neither hand of sudden onset	3.2 (1.3, 8.0)	3.2 (1.2, 8.7)

When hand function was considered as an outcome, the key predictors were: the physical component score of the SF-12, frustration with the hand condition and grip strength, along with baseline hand function (Table 8-8). The predictor, frustration with your hand condition, although included in the model overall, could only discriminate between subgroups developing “Mild” and “Improving” trajectories (when the “Mild” group was used as the reference category in the model). Removal of this predictor from the model did not greatly change estimates of the RRR for other predictors in the model (Table 8-8). The RRRs for the physical component score of the SF-12 and grip strength were similar for all within-predictor trajectory groups. The RRRs for AUSCAN function however were higher for the “Moderate” and “Severe” trajectory groups compared to the other RRRs for that predictor. In particular, the RRR of belonging to the “Severe” trajectory group compared to the “Mild” was excessively large for a one-unit change in baseline hand function, but this could be plausible given that there is no overlap in the baseline levels of hand function between these two groups (Figure 31).

Bootstrapping the 95% confidence intervals did not change the confidence limits greatly from those generated directly for the hand pain model, however for hand function, there were several instances where the range of the 95% confidence intervals was wider when bootstrapping was applied, e.g. for the coefficients associated with the variable “frustration with their hand condition”.

Table 8-8: Multinomial logistic regression results for hand function with the "Mild" group as the reference category (N=526 for the model with "Frustration" in it; N = 543 for the model with "Frustration" omitted)

	Hand function		
Predictor	Relative risk ratio (RRR) (95% CI)	Relative risk ratio (RRR) (95% normal-based bootstrapped CI)	Relative risk ratio (RRR) omitting frustration with hand condition (95% normal-based bootstrapped CI)
Moderate			
AUSCAN function	40.3 (20.8, 78.0)	40.3 (17.6, 92.0)	39.9 (18.4, 86.5)
Physical component score of the SF-12 (0-100)	0.9 (0.8, 0.9)	0.9 (0.8, 0.9)	0.9 (0.8, 0.9)
Average grip strength (lbs)	0.9 (0.9, 1.0)	0.9 (0.9, 1.0)	0.9 (0.9, 1.0)
Frustration with hand condition			
All days or most days	Ref	Ref	N/A
Some, few or no days	2.1 (0.3, 13.1)	2.1 (0.2, 22.4)	N/A
Severe			
AUSCAN function	212.9 (86.5, 523.9)	212.9 (70.9, 638.9)	206.9 (74.0, 578.2)
Physical component score of the SF-12 (0-100)	0.8 (0.8, 0.9)	0.8 (0.8, 0.9)	0.8 (0.8, 0.9)
Average grip strength (lbs)	0.9 (0.9, 1.0)	0.9 (0.9, 1.0)	0.9 (0.9, 1.0)
Frustration with hand condition			
All days or most days	Ref	Ref	N/A
Some, few or no days	2.6 (0.3, 21.9)	2.6 (0.2, 38.4)	N/A
Improving			
AUSCAN function	25.0 (13.5, 46.3)	25.0 (12.5, 50.2)	20.5 (10.9, 38.6)
Physical component score of the SF-12 (0-100)	1.0 (0.9, 1.0)	1.0 (0.9, 1.0)	1.0 (0.9, 1.0)
Average grip strength (lbs)	1.0 (1.0, 1.0)	1.0 (0.9, 1.0)	1.0 (0.9, 1.0)
Frustration with hand condition			
All days or most days	Ref	Ref	N/A
Some, few or no days	11.8 (1.6, 85.6)	11.8 (0.4, 385.3)	N/A
Progressively deteriorating			
AUSCAN function	6.3 (4.0, 10.2)	6.3 (3.9, 10.5)	6.0 (3.7, 9.6)
Physical component score of the SF-12 (0-100)	0.9 (0.9, 1.0)	0.9 (0.9, 1.0)	0.9 (0.9, 1.0)

Average grip strength (lbs)	1.0 (1.0, 1.0)	1.0 (1.0, 1.0)	1.0 (1.0, 1.0)
Frustration with hand condition			
All days or most days	Ref	Ref	N/A
Some, few or no days	3.6 (0.6, 22.4)	3.6 (0.0, 2519.5)	N/A
Mild/Moderate			
AUSCAN function	3.0 (2.0, 4.5)	3.0 (2.0, 4.6)	3.2 (2.1, 4.9)
Physical component score of the SF-12 (0-100)	1.0 (0.9, 1.0)	1.0 (0.9, 1.0)	1.0 (0.9, 1.0)
Average grip strength (lbs)	1.0 (1.0, 1.0)	1.0 (1.0, 1.0)	1.0 (1.0, 1.0)
Frustration with hand condition			
All days or most days	Ref	Ref	N/A
Some, few or no days	0.7 (0.2, 2.7)	0.7 (0.2, 3.4)	N/A
Mild (reference group)	1	1	1

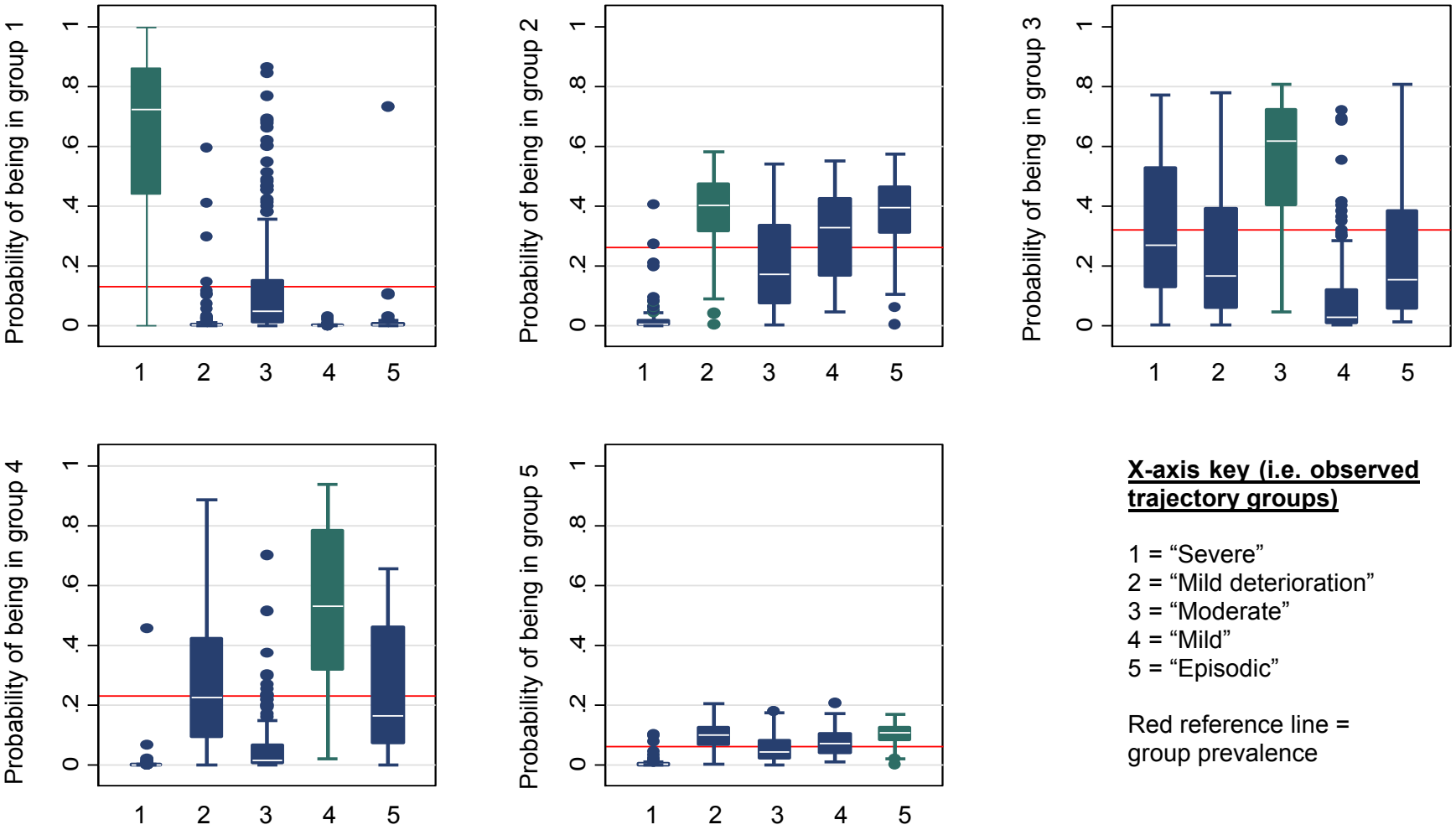
8.3.12 Model goodness-of-fit and model performance

The Nagelkerke's pseudo R-square and the Brier scores for the hand pain and function models suggest that the models are a reasonable fit to the data (Table 8-9) and discriminate well between the respective "Mild" reference categories and each trajectory group, as determined by high C-statistics (Table 8-9) and median predicted probabilities that are consistently higher in the observed group related to the predicted probability (i.e. the box shown in green on Figure 32 and Figure 33). As expected, the model is better able to discriminate between "extreme" groups, i.e. between the "Severe" and "Mild" groups than those that are more in the middle range, for example, ability of the model to discriminate those in the "Episodic" group for the hand pain model and the "Progressively deteriorating" group for the hand function model, is less apparent on the plots presented (Figure 32, panel 5; Figure 33, panel 4). High C-statistics also reflect this as they are each calculated relative to the "Mild" reference group so discrimination is likely to be high for those trajectory groups with baseline scores (especially) that differ greatly from the "Mild" reference group.

Table 8-9: Overall model goodness of fit and C-statistics comparing each trajectory group to the "Mild" reference category

	Nagelkerke's pseudo R-square	Brier score	C-statistic (95% confidence interval)
Hand pain			
Mild (reference group)			
Severe			0.99 (0.98, 1.00)
Mild deterioration	0.70	0.51	0.78 (0.73, 0.84)
Moderate			0.97 (0.95, 0.98)
Episodic			0.79 (0.72, 0.87)
Hand function			
Mild (reference group)			
Moderate			1.00 (0.99, 1.00)
Severe	0.83	0.49	1.00 (1.00, 1.00)
Improving			0.99 (0.98, 1.00)
Progressively deteriorating			0.93 (0.89, 0.96)
Mild/Moderate			0.83 (0.78, 0.88)

Figure 32: Discrimination box plots for the trajectory groups derived from the 5-group LCGM for hand pain

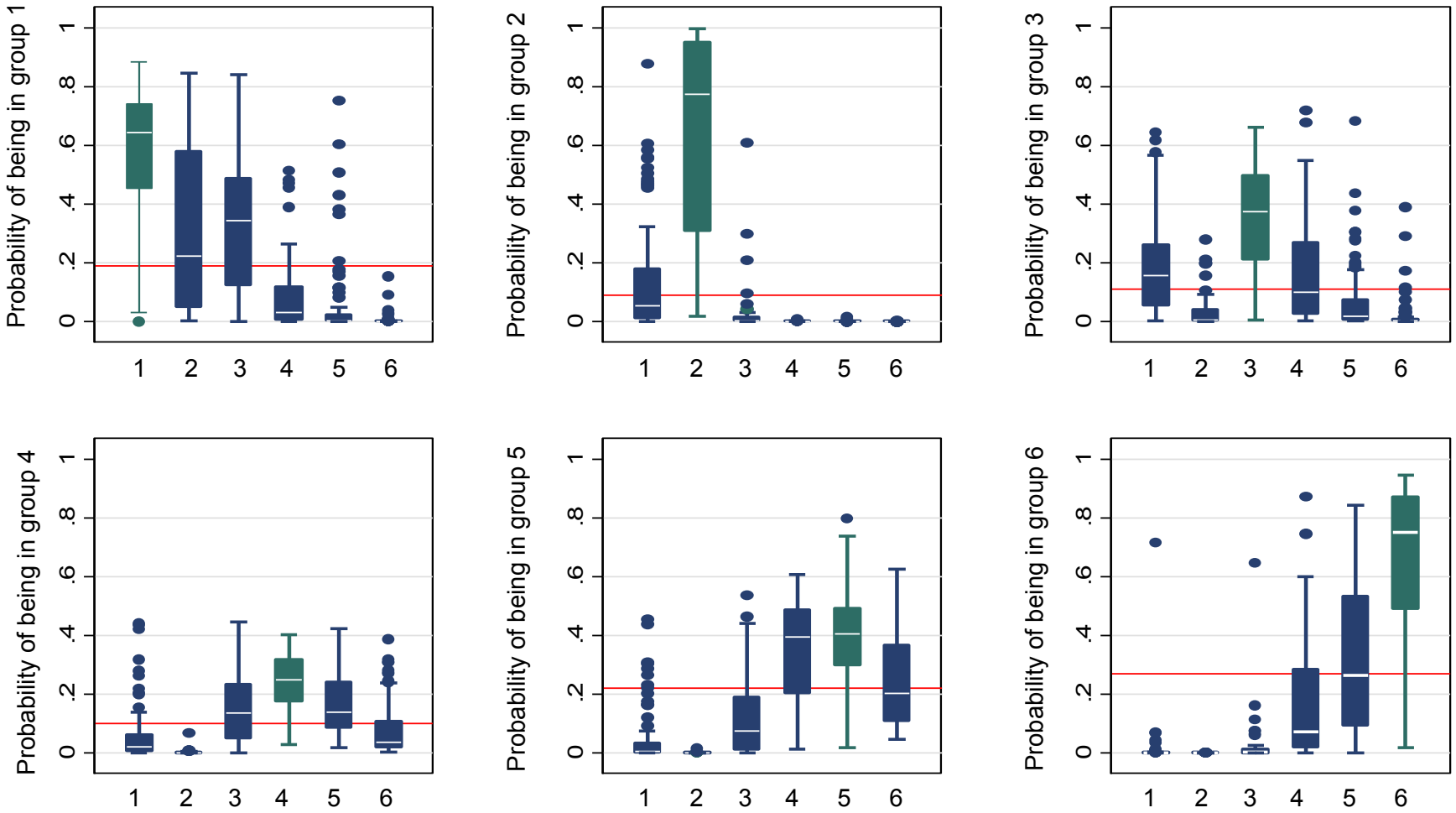


X-axis key (i.e. observed trajectory groups)

1 = "Severe"
2 = "Mild deterioration"
3 = "Moderate"
4 = "Mild"
5 = "Episodic"

Red reference line = group prevalence

Figure 33: Discrimination box plots for the trajectory groups derived from the 6-group LCGM for hand function



X-axis key (i.e. observed trajectory groups): 1 = “Moderate”, 2 = “Severe”, 3 = “Improving”, 4 = “Progressively deteriorating”, 5 = “Mild/moderate”, 6 = “Mild”. Red reference line = group prevalence

Reasonable model calibration was also achieved for both hand pain and function with data points lying close to the 45 degree line (Figure 34 and Figure 35). Exceptions were the “Episodic” group for hand pain and the “Improving” and “Progressively deteriorating” groups for hand function (where some points deviated from the 45 degree line), but this could be due to a small sample size in the group for which the observed probability was calculated. Model accuracy was also reasonable, although only 61% of participants belonged to the same trajectory group as given in the LCGM when a maximum probability rule was used to assign participants to groups based on the predicted probabilites from the multinomial model i.e. (354/577 for hand pain and 320/526 for hand function) (Table 8-10 and Table 8-11). It was also observed that in the hand pain model, no participants were assigned to the “Episodic” group suggesting that when a “maximum probability” rule is used to assign participants to groups, the “Episodic” group is not clearly identified.

Figure 34: Calibration plots for each trajectory group derived from the 5-group LCGM for hand pain

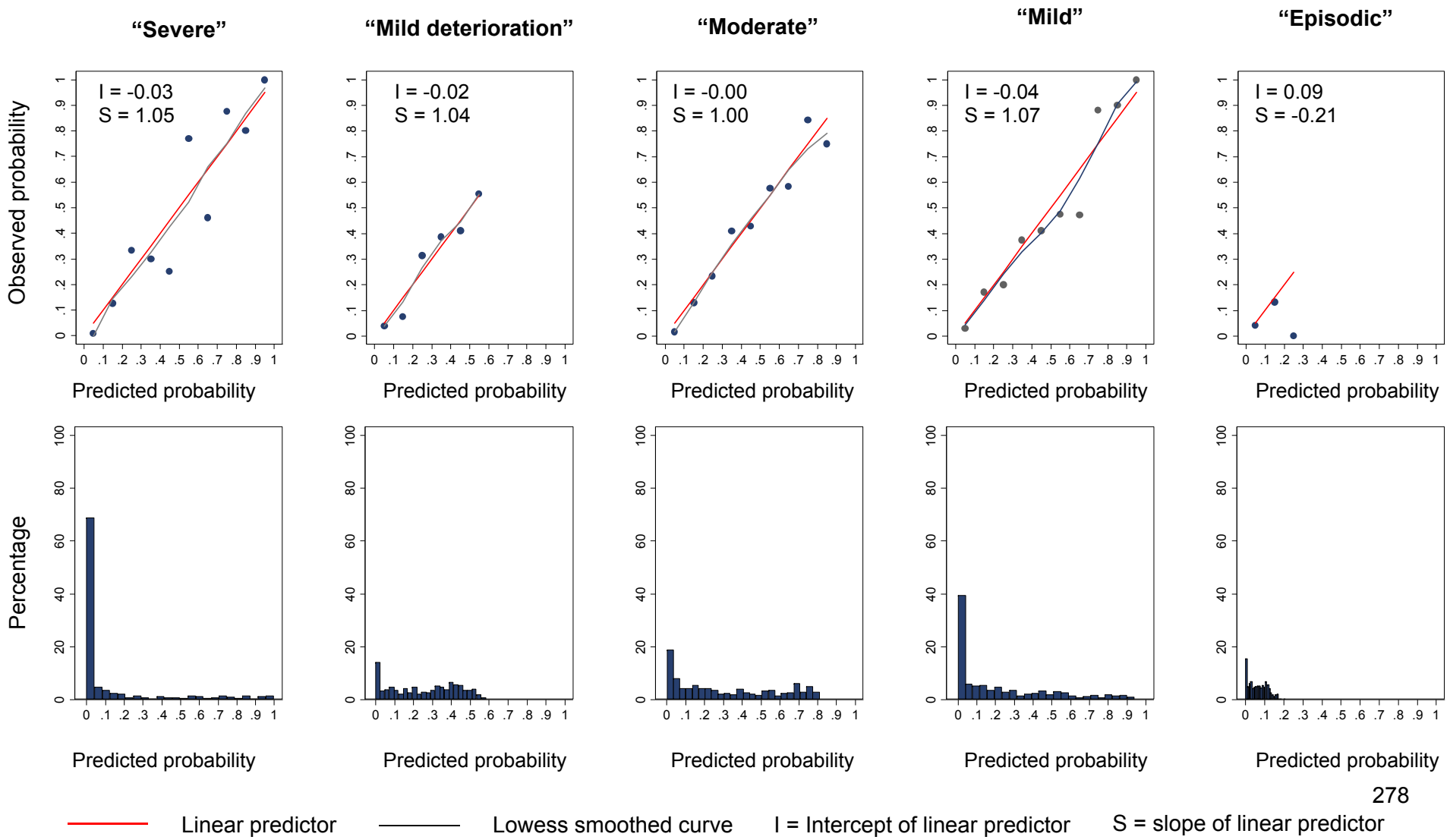


Figure 35: Calibration plots for each trajectory group derived from the 6-group LCGM for hand function

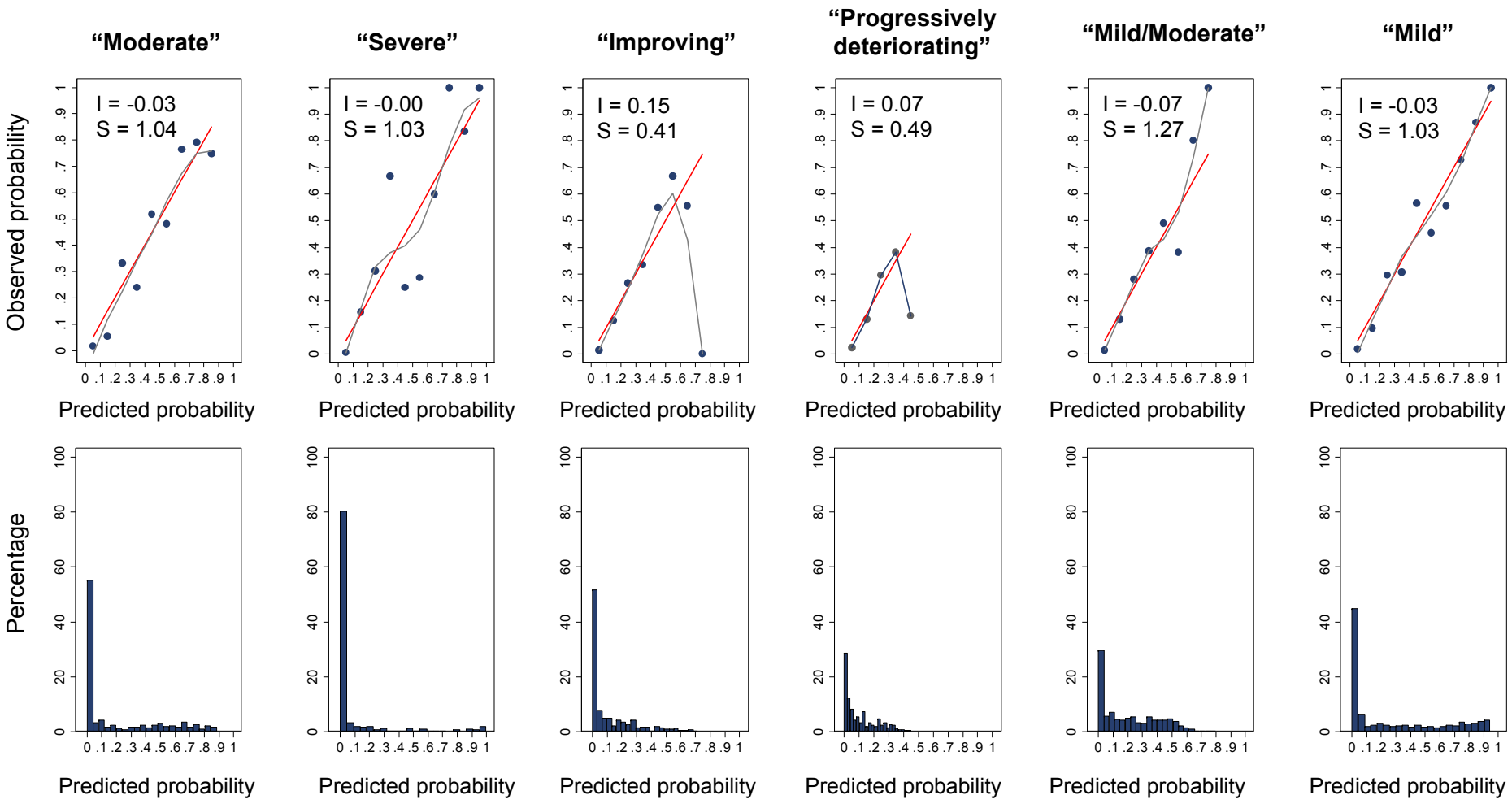


Table 8-10: Model Accuracy from the hand pain model

Trajectory group from the LCGM	Trajectory group membership predicted from the multinomial logistic regression					Total
	Severe	Mild deterioration	Moderate	Mild	Episodic	
Severe	54	0	19	1	0	74
Mild deterioration	1	77	36	37	0	151
Moderate	12	30	139	2	0	183
Mild	0	40	6	84	0	130
Episodic	1	15	9	14	0	39
Total	68	162	209	138	0	577

Table 8-11: Model Accuracy from the hand function model

Trajectory group from the LCGM	Trajectory group membership predicted from the multinomial logistic regression						Total
	Moderate	Severe	Improving	Progressively Deteriorating	Mild/ Moderate	Mild	
Moderate	80	9	9	5	2	1	106
Severe	15	28	0	0	0	0	43
Improving	21	1	26	4	8	1	61
Progressively Deteriorating	5	0	8	7	32	6	58
Mild/Moderate	6	0	3	3	62	38	112
Mild	0	0	1	1	27	117	146
Total	127	38	47	20	131	163	526

A key finding from the results of the LCGM is that for both hand pain and function, the baseline measure in the outcome of interest is a strong predictor of trajectory group membership, so the question arises: by how much does model goodness of fit and predictive validity improve by including the extra predictors in the model? This was tested by repeating the assessment of model goodness of fit and predictive validity for the model where only the baseline value for the outcome of interest was included as a predictor. This analysis showed that the fit and performance of the model was not greatly degraded by removing the additional predictors from the model (data given in Appendix 35).

8.4 Discussion

8.4.1 Summary of the key findings

In this chapter, LCGM have been used to identify groups of participants that, between groups, have differing trajectory shapes over time for hand pain and function, with a 5-group cubic model found to be optimal for hand pain and a 6-group quadratic model for hand function. For both hand pain and function, groups of participants were identified with "Mild" and "Severe" problems that were maintained at a similar level throughout the course of the study, with the "Severe" group characterised at baseline as being more likely to be female, to have worse hand-related characteristics, more psychological difficulties, poorer general health, to not be in current employment, to be less likely to drink on a regular basis and have to be careful with money to manage on the income they have than those in the "Mild" group.

Progressively deteriorating groups were shown for each outcome however the pattern of deterioration depended on the outcome considered. Progressively deteriorating groups were characterised by having poorer general health, higher rates of pain elsewhere in the body, longer duration of symptoms, gradual rather than acute onset, more frustration with their hand problem, having a bilateral problem, having more joints with radiographic OA, not being married or co-habiting, and having carpal tunnel syndrome. A group of participants showing improvement over time was identified in the hand function model only.

When the predictors were assessed in combination to explore those that predicted trajectory group membership, the key predictors emerging for hand pain were: AUSCAN hand function, pinch strength and sudden onset of symptoms (over and above baseline hand pain) and for hand function: the physical component score of the SF-12, frustration with the hand condition and grip strength (over and above baseline hand function). Although the models are limited in their derivation (as discussed earlier), they each show

a reasonable fit to the data, however, this goodness of fit is not greatly improved by having the additional predictors in the model over and above the baseline value for the outcome of interest.

8.4.2 Comparison with the literature

As discussed in the previous chapter, there is a lack of evidence around the likely course of hand pain and function in older adults so consequently, no studies have been found that use LCGM to explore whether subgroups of trajectories of pain and function exist in a similar population to CAS-HA. LCGM, however, have been used to explore symptom course in other musculoskeletal populations and although based in other body sites, there are some parallels to the findings for hand pain and function presented in this chapter. For example, both Collins at al. (2014) and Holla et al. (2014) have shown groups of participants with (on average) stable trajectories over a 6- and 5-year time frame for knee pain and activity limitations respectively, whereas, along with groups of participants that remained stable over time, Verkleij et al. (2012) also showed two groups of participants with regularly and highly progressively symptom deterioration in participants with hip OA over a period of 2-years. A further study explored replication of trajectory groups in two cohorts of patients with knee pain that were sampled to each have matching baseline characteristics. They showed three groups of participants with stable trajectories over time ("Mild, non-progressive", "Moderate" and "Severe, non-improving"), and two further groups showing "Improving" and "Progressing" symptoms, that although identified, were more difficult to replicate in an external (matched) data sample (Nicholls et al. 2014).

When the baseline characteristics of the trajectory groups were explored the findings were plausible and in line with several cross-sectional studies of pain and function, albeit in differing populations to CAS-HA[111]. Examples include cross-sectional evidence for the relationship between pain severity and depression (Denkinger et al. 2014), anxiety (Swain

[111] Cross-sectional studies are relevant here as a key factor that distinguishes several trajectory groups is their baseline starting value

et al. 2014), and grip and pinch strength for hand pain specifically (Dominick et al. 2005). Differences in baseline characteristics therefore support the rationale for grouping participants based on their trajectories over time as, although not possible to show reliably in this data, it can then be explored whether combinations of these variables can then be used to give helpful information to participants on the course of their problem over time.

8.4.3 Strengths and limitations

Choice on the number of groups to include in the LCGM

A multi-stage and multi-criteria approach was used to select the optimum number of groups for the LCGM in this chapter, which highlights the exploratory nature of this type of analysis. Using a different emphasis to balance the multiple model selection criteria could have produced differing optimal models than those presented (e.g. if statistical fit indices were used in isolation then this may have indicated that a different model be regarded as optimal). Both statistical fit and clinical interpretability were considered in this chapter to choose the optimal model, to try and gain further understanding of the data over and above what had been gained in the analysis of the data using growth modelling in Chapter 6[112].

The plots of the trajectories within each latent group show that although each trajectory group is characterised by a mean trajectory over time, there is still a lot of variation of the trajectories around each mean trajectory curve. Ideally this variation would have been modelled using GMM, however, these models failed to produce interpretable results when they were applied. This therefore highlights the difficulties and limitation of considering the groups as distinct entities as it needs to be recognised that not all participants in the group exactly follow the average trajectory line as given for the group, however, they have been allocated to the group that most likely represents their trajectory, thus providing a useful

[112] It was recognised that if optimal models were chosen whose groups were only differentiated by their intercept then this would offer little more understanding of the data than could be gained by including a random intercept term in a growth model.

descriptive technique to view the data and explore the range of trajectory shapes contained within it.

In this chapter clinical interpretability was evident if there was at least one group in the model that showed a trajectory shape that changed over time. Alternative definitions of clinical interpretability could have been used, such as defining groups that differ by the treatment options available or potentially suitable to them. In this population-based sample, where not all participants have sought healthcare for their hand problems, a prognostic definition was considered useful in order to define and characterise a group of participants whose hand pain and function are likely to remain at a low level over a long time course, to offer them information about their prognosis, in contrast to others with more severe problems where onward referral for help and support may be needed.

Sample size may also influence decisions regarding the optimal model. For example, with a sample size of around 600 participants there will only be a finite number of times that the sample can be split before one group is extracted that has a small sample size. It is therefore likely that given the sample size there is a maximum number of groups that can be extracted from that data before within-group sample sizes become too small to be reliable. This is a plausible problem, however as multiple indices are used to select the optimal model (not just the sample sizes in the groups of interest) it could still be that a model with 2-4 groups could be optimal. In addition, the criteria on sample size is regarding the percentage of participants in each group not falling below 1-5%, rather than the actual number of participants in each group *per se* so it is less likely to depend on the actual size of the sample analysed.

Defining the polynomial form in the data

In the analysis presented, when higher order polynomials were included in the model they were included in all groups irrespective as to whether they were non-significant in some groups (even though this was counter to the approach taken in Chapter 7 where all terms

were removed from the model if they are non-significant). This approach was taken as programming difficulties were encountered in Mplus when it was attempted to fit a model with a different form of polynomial in each group[113]. It was also highlighted on the discussion board associated with the Mplus software that it was often simpler to leave non-significant polynomial terms in the model unless there was a strong theoretical reason why a lower order model needs to be assumed for a particular trajectory group of interest (Muthen 2008b) and hence this approach was followed.

The use of multinomial logistic regression to predict the outcome of interest

Multinomial logistic regression was used in this chapter to model trajectory group membership and has the advantage over the models fitted in Chapter 7 that model performance can more easily be tested when participants are divided into groups of interest. The results of this model also have greater clinical applicability as it is plausible that model results could be used to predict how likely it is that a particular patient would have an unfavourable clinical course given their key characteristics that are in the model at baseline; a model that could either be applied in a consultation setting or possibly as a web-based tool for patients to complete online to provide guidance regarding self-management or consulting healthcare. Although this is the ultimate goal for analysis, this was limited by difficulties that arose when applying the multinomial model to the data.

Firstly, as previously mentioned, a small sample size in some of the trajectory groups is a severe limitation to the number of predictors that can be explored in the data. Although bootstrapping was used as a potential way to address this issue, the only true way to

[113] It was difficult to program different polynomial forms for each group in Mplus as when the polynomial term was set to 0, the variance term associated with it was still modelled in the data. It was not intuitive how to remove it from the model, and also it was hard to specify code such as "if group = 1 then x^2 = 0 because group 1 isn't always group 1 when a new model is run with a different random start i.e. the order that groups are extracted from the data can change depending on what random start is used. The proctraj program (available in the SAS computer program and with conversion code to STATA) was explored as an alternative analysis program as it was straightforward to program different polynomials for each group using this program. This wasn't an option however as the Satorra-Bentler Scaled Chi-square test was not available using this program and this was needed to analyse the skewed outcome

address this problem is to run the model on a larger data set with a larger number of participants than those in the CAS-HA study. An alternative solution would be to try and reduce the number of trajectory groups that are modelled, e.g. by pooling some of the groups together, however, it would be difficult to justify which groups are similar to each other to be pooled when they had previously be shown to be distinct using LCGM.

Using stepwise-regression is also a limitation of the analysis for reasons similar to those described in Chapter 7. One option for the multinomial analysis could have been to use the descriptive data in Table 8-4 and Table 8-5 to pre-select important predictors of interest to then take forward as a subset of variables to the multinomial logistic regression stage. This approach however was not pursued as some predictors may become significant/non-significant when assessed in the presence of other predictors of interest.

In addition, as previously alluded to, a further difficulty with this analysis is in the variation that occurs within the trajectory groupings and also the potential for overlap between participants at “the edges” of each trajectory group. In the multinomial model participants are assumed to either be “in the group” or “not in the group”, but the data show that even when participants are “in the group” there is still variability of trajectory shapes that occur even though they have all been given the same label of being “in the group”. This issue will always occur however, when a measure that is essentially continuous, is made into a categorical variable for the purpose of being a useful variable to aid clinical decision making.

A further limitation could be around the reliability of some of the estimates that have been produced. It is clear, even without using a multinomial model, that the baseline level of the outcome of interest is a clear predictor of trajectory group membership. As the baseline in the outcome of interest is such a strong predictor it may be overly dominating the model so that other predictors were not reliably estimated after such a strong predictor is

considered[114]. The baseline was included in the model however, as it clearly is a predictor of trajectory group membership and, if it were not included, then the predictors selected to be in the model were more likely to be those that were markers of the baseline starting value, i.e. variables that are highly correlated with the baseline in the outcome of interest, rather than those that were predictive of the course of the outcome over time.

A further limitation to the analysis presented is that the performance of the model has been tested on the same data that was used to develop it, which is a key limitation of the findings. The performance of the model is likely to be over-optimistic than if it were tested on an external dataset. It was considered whether to use bootstrapping to adjust for over-optimism as per the procedure outlined in Appendix 36, but this approach was not used as it would be computationally difficult to write a program to apply this method to a multinomial model[115].

Despite the comprehensive discussion of the limitations of this analysis, the results have been included in this thesis as the process of developing and testing a prognostic model has provided useful training and professional development during the course of writing this thesis. It also demonstrates research skills that can be applied in future research projects. A fuller description of the predictive ability of the model is therefore needed in an external data set with a larger sample size to test for replication of the trajectory groups and where a fuller exploration of potential predictors, with reliable parameter estimates can be conducted. Consequently the assessment of goodness of fit and predictive validity

[114] An example of this is that the estimate for baseline hand function for the "Severe" group, compared to the "Mild" group in the hand function model is 212.9 (95% confidence interval: 86.5, 523.9). This is interpreted as the factor by which the relative risk increases for a one point increase in the AUSCAN. This is very large, but somewhat understood, as it is observed that virtually everyone in the "Severe" group has a baseline AUSCAN function score > 5 and everyone in the "Mild" group a score <=5. Baseline hand function is therefore acting as a near-perfect predictor of trajectory group membership for the "Severe" group compared to the "Mild" group, which may be leading to unreliable parameter estimates especially when the confidence intervals around the RRRs are also very large

[115] It would be computationally difficult to adjust model fit for overestimation as the conditional risk method would need to be applied to generate multiple C-indexes, and, in the context of this model, a forward stepwise process could only be applied "by hand". This would make automation of the process difficult to apply to a large set of bootstrapped samples

of the model has been kept brief, so other, newer, or more complex measures of model fit have not been explored, e.g. using the polychotomous C-index which aims to combine the information across all of the pairwise C-indexes presented in this chapter into an overall index of model fit (Van Calster et al. 2012b).

A final potential limitation of the work presented in this chapter (and also that presented in Chapters 6 and 7) is that it assumes that hand pain and hand function are two separate independent outcomes, whereas in reality these outcomes are likely to be related within any individual. It is this concept that is therefore explored further in the next chapter (Chapter 9).

9 Joint trajectory modelling

9.1 Introduction

In the previous results (Chapters 6 - 8), hand pain and function were modelled as two separate (independent) outcomes. However, given knowledge of the constructs, and several cross-sectional studies that have shown that these two outcome measures are related at a single time-point (Baron et al. 1987, Dahaghin et al. 2005b), it is plausible that the trajectories over time of these outcomes are associated. To explore this further, two objectives are addressed in this chapter:

Objective 1: To quantify the strength of the relationship between the trajectory curves for hand pain and function and to specifically explore whether the cross-sectional associations previously reported extend to an association between the longitudinal trajectories for these outcomes over time.

Objective 2: To explore whether groups of participants can be identified that have differing trajectories of hand pain and function over time, and, if so, whether characteristics of such participants can be identified that explain why the trajectories for pain and function differ within each participant identified.

9.2 Methods

9.2.1 Objective 1 – association between the trajectory curves for hand pain and function

To address Objective 1, parallel process growth models (as described in Chapter 5) were fitted to the data to quantify the magnitude of the association between the random coefficients included in the individual growth models for hand pain and function. Initially a linear model was fitted to the data (i.e. to explore the associations between the random intercepts and random slope terms for hand pain and function respectively). The linear model was then extended to include two quadratic terms (one for hand pain and one for

hand function) and was used to explore whether the findings from the linear model were replicated when a quadratic (rather than linear) trajectory was assumed for each outcome of interest. In addition, it was also explored whether there was a relationship between the quadratic term for hand pain and the quadratic term for hand function.

9.2.2 Objective 2 – can distinct groups of participants be identified with similar trajectories of hand pain and function over time?

Parallel process latent class growth models (PPLCGM) were used to identify an optimum number of latent groups to describe the joint trajectory of hand pain and function over time[116]. Initially a linear model was assumed for both hand pain and function and a search for the optimum number of groups undertaken using the strategies described for LCGM in Chapter 5. It was then tested whether the addition of a quadratic or cubic term improved model fit and also whether the choice of the optimum number of groups was the same irrespective of the polynomial form used to model the data.

9.3 Results

9.3.1 Objective 1 - association between the trajectory curves for hand pain and function

When a linear parallel process growth model was fitted to the data it did not produce plausible estimates as the correlation between the random slope for hand pain and the random slope for hand function was greater than 1 (correlation = 1.231; 95% confidence interval: 1.03, 1.43). A personal communication with one of the developers of the Mplus software (Linda Muthen) suggested that this type of problem often happens if it is incorrectly assumed that the correlation between two outcomes at a single time-point is zero (i.e. the correlation between hand pain and function at each time point is zero).

As it is plausible that hand pain and function would be correlated at each time point these additional correlations were added to the model. The results of this additional analysis are

[116] PPLCGM was used rather than a parallel process growth mixture model (PPGMM) as the PPGMM produced models with correlations greater than one that were implausible

shown in column 2 of Table 9-1 and, where relevant, are presented as variances and co-variances as this is the primary output given in the Mplus software. Only the key coefficients of interest from the model are converted into correlation coefficients for ease of interpretation, giving an estimated correlation between the random intercept for hand pain and function of 0.88 (95% confidence interval: 0.84, 0.93) and a correlation between the random slopes as 0.78 (0.52, 1.04).

The results of this model were stable when a quadratic term was included in the model (Table 9-1, column 3), with correlations of: 0.88 (95% confidence interval: 0.82, 0.94) between the random intercept terms, 0.88 (0.58, 1.19) between the random slopes and 0.76 (0.36, 1.16) between the random quadratic terms. However, the confidence intervals associated with slope and quadratic terms were wide and their upper bound exceeded one (see Appendix 37 for details on the process used to define the quadratic model used in the analysis).

Table 9-1: Parallel process model of hand pain and functional difficulty trajectories over time

	Column 1	Column 2	Column 3
	Linear model	Linear model with correlation between concurrent time points	Quadratic model with correlation between concurrent time points
Relationship between hand pain and function			
Prediction			
Hand pain intercept → hand function slope	-0.01 (-0.03, -0.00)	0.02 (-0.00, 0.03)	0.01 (-0.06, 0.09)
Hand function intercept → hand pain slope	-0.02 (-0.03, -0.00)	0.01 (-0.01, 0.02)	0.02 (-0.03, 0.08)
Hand pain intercept → hand function quadratic	N/A	N/A	-0.00 (-0.01, 0.01)
Hand function intercept → hand pain quadratic	N/A	N/A	-0.00 (-0.01, 0.00)
Covariance[α] between polynomial model terms			
Hand pain intercept and hand function intercept	3.76 (3.29, 4.22)	3.26 (2.75, 3.77)	3.23 (2.56, 3.90)
Hand pain slope and hand function slope	0.04 (0.03, 0.05)	0.01 (-0.00, 0.02)	0.09 (-0.03, 0.20)
Hand pain quadratic and hand function quadratic	N/A	N/A	0.00 (-0.00, 0.00)
Hand pain slope and hand function quadratic	N/A	N/A	-0.01 (-0.02, 0.00)
Hand function slope and hand pain quadratic	N/A	N/A	-0.01 (-0.02, 0.00)
Hand pain slope and hand pain quadratic	N/A	N/A	-0.01 (-0.03, 0.00)
Hand function slope and hand function quadratic	N/A	N/A	-0.01 (-0.02, 0.01)
Covariance[α] between concurrent time-points			
Baseline hand pain and baseline hand function	N/A	0.84 (0.55, 1.13)	0.72 (0.27, 1.17)
18-month hand pain and 18-month hand function	N/A	1.62 (1.25, 2.00)	1.63 (1.25, 2.02)
3-year hand pain and 3-year hand function	N/A	0.67 (0.47, 0.87)	0.55 (0.33, 0.76)

5-year hand pain and 5-year hand function	N/A	0.74 (0.53, 0.95)	0.71 (0.49, 0.93)
7.5 year hand pain and 7.5-year hand function	N/A	0.73 (0.43, 1.03)	0.61 (-0.02, 1.23)
Hand pain model			
Fixed estimates			
Intercept (α_Z)	3.12 (2.96, 3.29)	3.14 (2.98, 3.31)	3.10 (2.93, 3.27)
Time (β_{Z1})	0.10 (0.06, 0.14)	0.03 (-0.02, 0.08)	0.05 (-0.11, 0.21)
Time squared (β_{Z2})	N/A	N/A	-0.00 (-0.02, 0.02)
Variance			
Intercept ($\sigma^2_{\alpha z}$)	3.52 (3.07, 3.97)	3.13 (2.67, 3.59)	3.05 (2.49, 3.60)
Slope ($\sigma^2_{\beta z1}$)	0.03 (0.02, 0.04)	0.01 (0.00, 0.03)	0.10 (-0.01, 0.21)
Quadratic ($\sigma^2_{\beta z2}$)	N/A	N/A	0.00 (0.00, 0.00)
Residual – variance ($\sigma^2_{\varepsilon zt}$)			
Baseline	1.43 (1.15, 1.70)	1.80 (1.48, 2.12)	1.72 (1.31, 2.14)
18-months	2.27 (1.85, 2.70)	2.42 (1.99, 2.84)	2.38 (1.93, 2.82)
3-years	1.61 (1.34, 1.88)	1.63 (1.36, 1.90)	1.47 (1.17, 1.77)
5-years	1.33 (1.09, 1.57)	1.39 (1.13, 1.64)	1.40 (1.12, 1.68)
7.5-years	1.30 (0.98, 1.62)	1.49 (1.11, 1.87)	0.96 (0.17, 1.75)
Hand function model			
Fixed estimates			
Intercept (α_Y)	2.79 (2.61, 2.98)	2.80 (2.61, 2.98)	2.73 (2.55, 2.92)
Time (β_{Y1})	0.12 (0.08, 0.16)	0.03 (-0.03, 0.08)	0.11 (-0.10, 0.33)
Time squared (β_{Y2})	N/A	N/A	-0.01 (-0.03, 0.02)

Variance			
Intercept ($\sigma^2_{\alpha Y}$)	4.79 (4.26, 5.31)	4.36 (3.83, 4.90)	4.42 (3.70, 5.13)
Slope ($\sigma^2_{\beta Y1}$)	0.03 (0.02, 0.04)	0.01 (0.00, 0.03)	0.09 (-0.03, 0.21)
Quadratic ($\sigma^2_{\beta Y2}$)	N/A	N/A	0.00 (-0.00, 0.00)
Residual – variance ($\sigma^2_{\varepsilon t}$)			
Baseline	0.91 (0.67, 1.14)	1.17 (0.85, 1.48)	1.02 (0.53, 1.50)
18-months	1.72 (1.35, 2.09)	1.85 (1.46, 2.23)	1.87 (1.47, 2.26)
3-years	1.00 (0.79, 1.20)	0.96 (0.75, 1.18)	0.86 (0.62, 1.09)
5-years	0.96 (0.74, 1.18)	0.99 (0.74, 1.24)	0.94 (0.70, 1.19)
7.5-years	0.91 (0.62, 1.19)	1.06 (0.72, 1.40)	0.97 (0.29, 1.65)
SB scaled chi-square test			
Model 1 vs Model 2	χ^2 = 566.06 (d.f. = 5); p < 0.001		
Model 2 vs Model 3			χ^2 = 20.15 (d.f. = 11); p = 0.043
Model fit			
Akaike (AIC)	18341	17520	17519
Bayesian (BIC)	18439	17639	17688
Sample-size adjusted Bayesian (BIC)	18369	17554	17567

Unless otherwise stated, figures are parameter estimates and 95% confidence intervals (based on robust standard errors) in brackets. [α] Note that co-variances are given in the table rather than correlations, so it's not appropriate to test whether the correlation is significant by assessing whether the 95% confidence interval for the covariance spans zero. Z = hand pain process, Y = Hand function process, p = p-value, N/A = not applicable, SB = Satorra-Bentler, → = predicts.

9.3.2 Objective 2 – can groups of participants be identified with similar trajectories of hand pain and function over time?

The optimum number of latent groups was not clear from the goodness-of-fit statistics presented in Table 9-2. Both the unadjusted and adjusted VLMR LRTs suggested that a 2-group model was optimal whereas the AIC, BIC and ABIC did not reach a minimum value when models with up to seven groups were considered (models with greater than seven groups were not considered as this resulted in several groups containing <5% of the total sample size). All entropy values and posterior probabilities were greater than 0.8 and 0.7, respectively, so this did not give clear guidance as to the optimal model.

However, inspection of the trajectory plots from all models revealed a consistent pattern as within all trajectory groups, in all models, the trajectory shape for hand pain and function was similar, e.g. no trajectory groups were identified that contained participants whose hand pain increased over time but where their hand function had remained stable, nor those with increasing hand pain followed by an increase in functional difficulty (see Figure 36 and Figure 37 for a graphical display of the linear model). Although some groups were defined by having slightly higher hand pain than functional difficulty and vice versa, this is a tentative comparison as although the two outcomes are measured on a scale of 0-10, the content of the questions that make up each scale still differs between the outcomes. It was therefore not possible to select an optimal PPLCGM in the data that could be used to identify groups of participants with differing trajectories of hand pain and function over time; a result that did not change when a quadratic or cubic curve was assumed in the data (data not shown).

Table 9-2: Goodness-of-fit Statistics for PPLCGM fitted to AUSCAN Pain and function (N=621[α])

Model type	AIC	BIC	ABIC	Entropy	VLMR LRT	Adjusted LMR LRT	PB LRT	Group N	Average posterior probability
Linear									
1	22644	22706	22662	N/A	N/A	N/A	N/A	621	1.0
2	19896	19980	19920	0.93	p<0.001	p<0.001	p<0.001	359, 262	0.98, 0.97
3	19070	19177	19100	0.90	p=0.079	p=0.083	p<0.001	122, 284, 215	0.94, 0.98, 0.93
4	18730	18859	18767	0.86	p=0.021	p=0.022	p<0.001	176, 178, 183, 84	0.92, 0.92, 0.89, 0.96
5	18577	18728	18620	0.84	p=0.212	p=0.218	p<0.001	142, 58, 113, 185, 123	0.88, 0.93, 0.89, 0.88, 0.91
6	18507	18680	18556	0.83	p=0.109	p=0.113	p<0.001	125, 103, 178, 25, 67, 123	0.85, 0.83, 0.88, 0.88, 0.87, 0.90
7	18435	18630	18490	0.82	p=0.278	P=0.287	p<0.001	122, 93, 109, 21, 173,76, 27	0.89, 0.82, 0.81, 0.89, 0.87, 0.84, 0.89
Quadratic									
1	22645	22716	22665	N/A	N/A	N/A	N/A	621	1.0
2	19892	19994	19921	0.93	p<0.001	p<0.001	p<0.001	356, 265	0.98, 0.97
3	19065	19198	19102	0.90	p=0.097	p=0.100	p<0.001	286, 121, 214	0.97, 0.94, 0.93
4	18720	18884	18767	0.86	p=0.022	p=0.023	p<0.001	179, 86, 182, 174	0.92, 0.96, 0.89, 0.92
5	18569	18764	18624	0.84	p=0.437	p=0.441	p<0.001	58, 113, 126, 138, 186	0.92, 0.89, 0.91, 0.89, 0.87
6	18499	18725	18563	0.83	p=0.268	p=0.271	p<0.001	174, 117, 64, 119, 24, 123	0.84, 0.90, 0.93, 0.87, 0.90, 0.90
7	18421	18678	18494	0.82	p=0.131	p=0.134	p<0.001	95, 123, 25, 107, 172, 71, 28	0.83, 0.90, 0.91, 0.83, 0.84, 0.86, 0.89
Cubic									
1	22648	22728	22671	N/A	N/A	N/A	N/A	621	1.0
2	19896	20015	19930	0.93	p<0.001	p<0.001	p<0.001	356, 265	0.98, 0.97
3	19074	19234	19119	0.90	p=0.110	p=0.113	p<0.001	120, 215, 286	0.95, 0.93, 0.97
4	18720	18920	18777	0.86	p=0.036	p=0.037	p<0.001	182, 87, 175, 177	0.91, 0.95, 0.92, 0.90
5	18575	18814	18642	0.84	p=0.664	p=0.665	p<0.001	178, 137, 145, 53, 108	0.89, 0.90, 0.89, 0.91, 0.89
6	18488	18767	18567	0.84	p=0.331	p=0.332	p<0.001	152, 40, 114, 63, 100, 152	0.91, 0.91, 0.84, 0.82, 0.89, 0.91
7	18403	18722	18493	0.84	p=0.215	p=0.216	p<0.001	43, 122, 125, 146, 38, 52, 95	0.91, 0.86, 0.86, 0.91, 0.90, 0.81, 0.89

Highlighting indicates models with the lowest AIC, BIC, ABIC values, models with one group less than the model with a non-significant LRT p-value, group frequencies less than 5% of the sample (i.e. N <30) and posterior probabilities <0.7. All models achieved a global solution as in each model the largest log-likelihood was replicated for more than two random starting values [α] Two participants were excluded from the analysis as they had no data at all time-points. AIC = Akaike Information Criteria, BIC = Bayesian Information Criteria, ABIC = Sample-size adjusted BIC, VLMR LRT = Vuong-Lo-Mendell-Rubin likelihood ratio test, LMR LRT = Lo-Mendell-Rubin likelihood ratio test, PBLRT = parametric bootstrapped likelihood ratio test, N/A = not applicable, p = p-value

Figure 36: Linear PPLCGM for AUSCAN Pain (blue) and function (green) stratified by the number of trajectory groups in the model

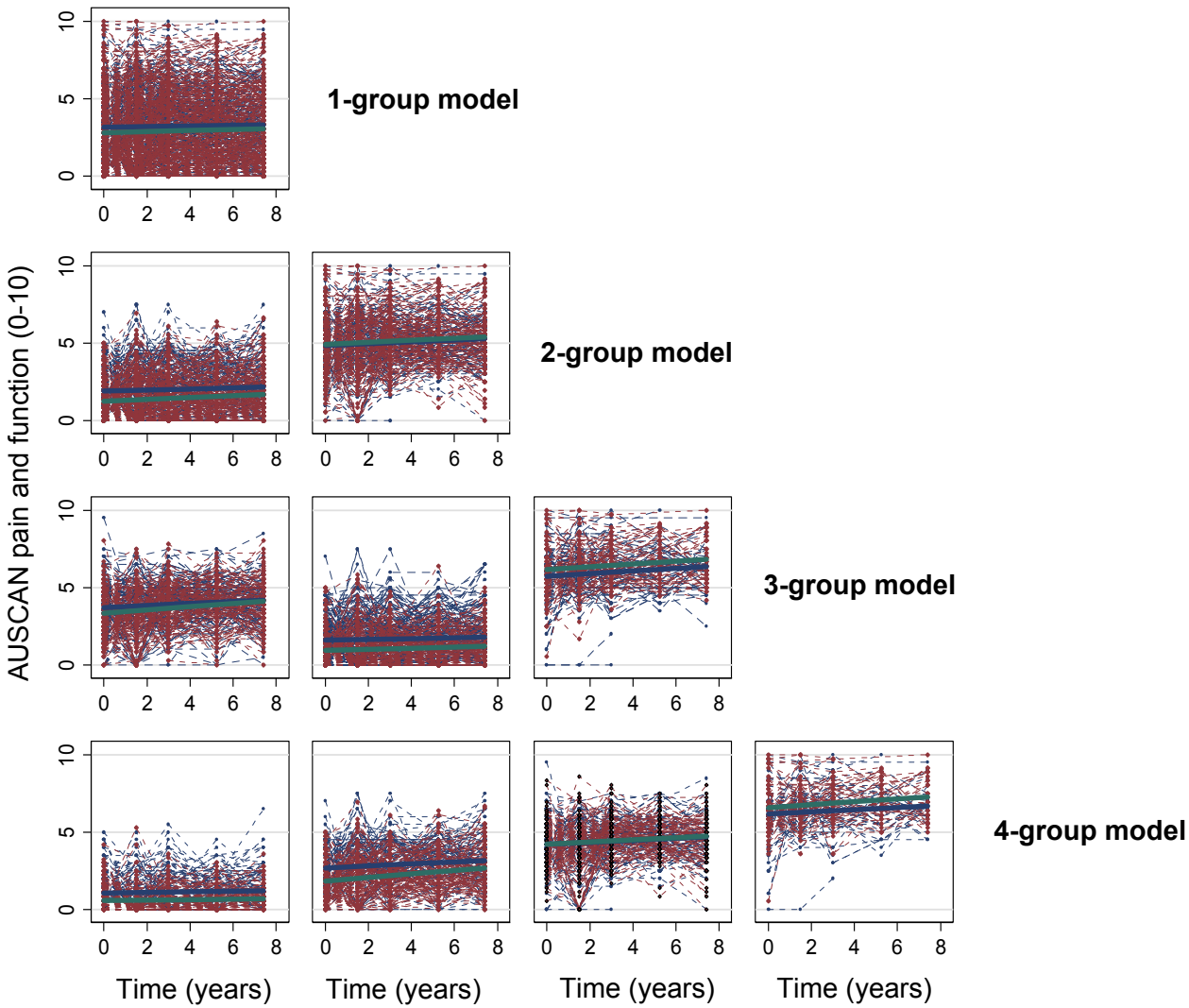
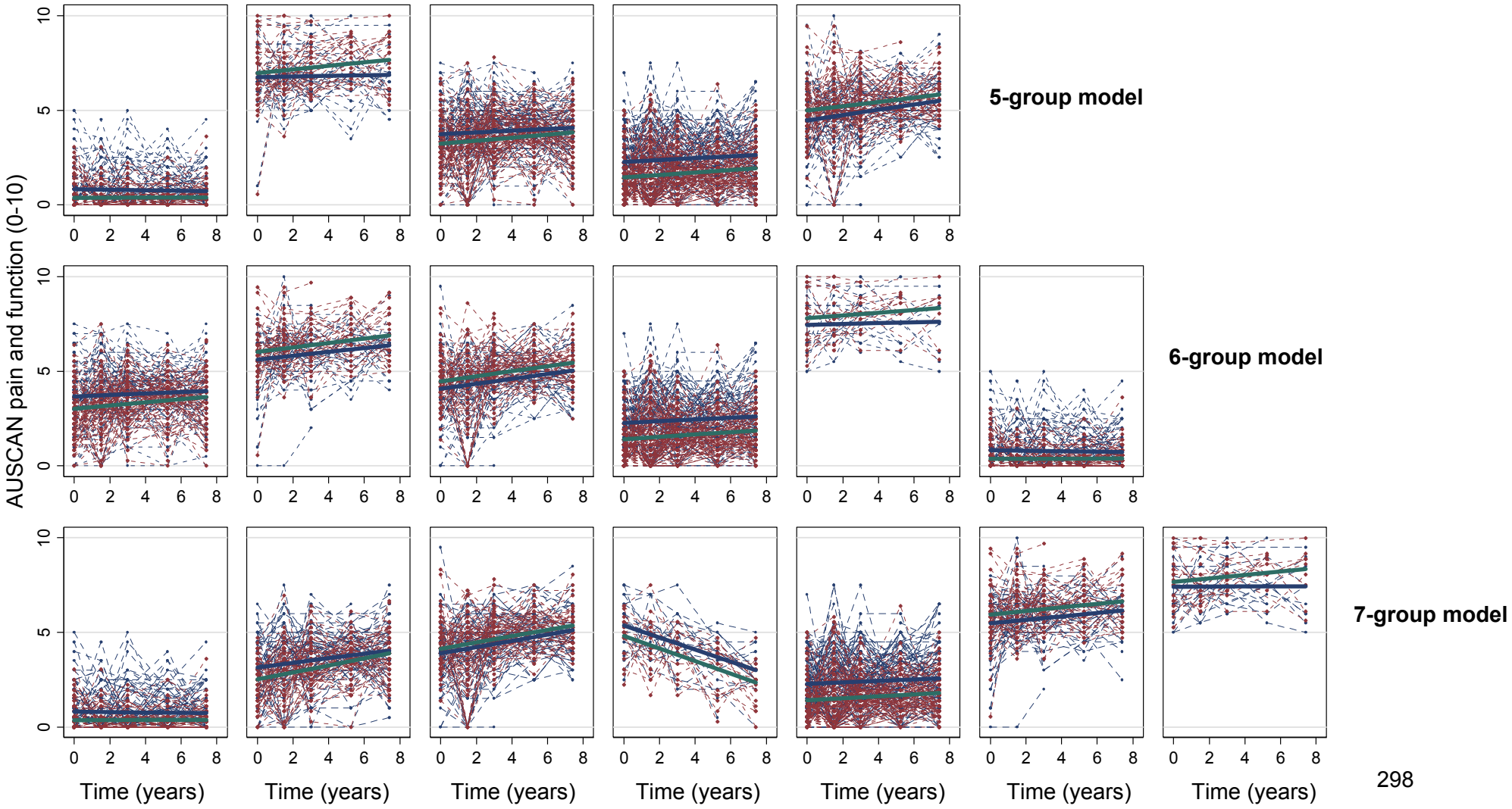


Figure 37: Linear PPLCGM for AUSCAN Pain (blue) and function (green) stratified by the number of trajectory groups in the model (continued)



9.4 Discussion

9.4.1 Summary of the key findings

It has been shown in this chapter that there is a close relationship between the course of hand pain and function over time; however, when expressed in a longitudinal model, there is more certainty around the strength of the relationship between the model's intercepts (severity of pain & limitations at baseline) than their slope or quadratic terms (changes over time), with the 95% confidence intervals for the latter associations being wide and with an upper limit slightly above one. The association between the model intercepts is plausible given that previous cross-sectional studies have shown that hand pain and function are associated at a single time point (Baron et al. 1987, Dahaghin et al. 2005b). As no groups of participants were identified in the PPLCGM that had differing trajectories of hand pain and function over time, this gives further support that hand pain and function are closely related at baseline and over time.

9.4.2 Comparison with the literature

No other studies were found that have used PPGM or PPLCGM to model the relationship between hand pain and function over time, however PPGM have been used in a study of knee OA to explore the relationship between statin use and knee pain, knee function and OA structural progression (Riddle et al. 2013). This is therefore a relatively new statistical technique applied in this field so comparisons with other studies in the literature are limited.

A comparison can however be made to other data in this thesis by questioning why it is that different baseline predictors are selected in Chapter 7 as predictors of the trajectory of hand pain and function over time when it has been shown in this chapter that the two trajectories are highly correlated. This could occur as it is plausible that there could be several competing models, each with different predictors in them that could be equally good at predicting the outcome. However, another alternative explanation could be that

even though the point estimate for the correlations are high, the 95% confidence intervals around these correlation estimates are particularly wide, especially for the model slope and quadratic terms, so within these models there is still uncertainty in the data as to the true estimate of the correlations between the trajectory parameters of interest for hand pain and function. This lack of certainty in the correlation could therefore express itself by selection of differing sets of predictors for hand pain and function.

9.4.3 Strengths and weaknesses

A key strength of the analysis in this chapter is that data from all time-points have been included in the model to explore the relationship between hand pain and functional difficulty over time. This is preferred over looking at the correlation (for example) between change scores for hand pain and function (when measured between two arbitrary points in time) as in this model more information is used to estimate the "true" symptom course over time. In addition, PPGM are preferred over choosing one measure to be the outcome of interest and then fitting the other outcome as a time-varying covariate as it avoids having to choose which outcome is the "primary" measure and also allows both outcomes to be modelled as having a trajectory of interest over time, which is more in line with how the outcomes are measured and the research question that is being addressed. This method also allows parameter estimates to be modelled simultaneously, limiting the problem of multiple testing in a single dataset and the number of models that need to be fitted to the data to address specific research questions around potential predictors of each trajectory over time.

A weakness of the analysis presented is that the estimates from the model have not been adjusted for other covariates that may influence the trajectory of hand pain and function over time (e.g. age and gender). Although it is plausible to add such predictors to the model, this may produce unstable model estimates if the number of model predictors is large given the sample size available for analysis. A further weakness of the data from the PPLCGM is that it does not provide a useful tool with which to explore the characteristics

of participants with clearly differing trajectories of hand pain and function over time as, given the analysis technique used, such groups were not clearly identified in the data.

9.4.4 Clinical Applicability

The models in this chapter provide data to test whether the hypothesised associations between levels of, and change in, hand pain and function are observed, with a view to identifying a group of patients, who, despite an increasing level of pain, continue to maintain a good level of hand function. Identification of such patients could be useful to explore whether there is a particular coping strategy that such patients are using to maintain good hand function, and if so to explore whether this could be included in a treatment package for future patients. As it was not possible for a group of such patients to be identified in the analysis, the clinical applicability of the models presented is limited. They do, however, provide evidence for an association between hand pain and function that is supported by data (rather than simply hypothesised) and have the potential to be used in further work that efficiently explores predictors of hand pain and function simultaneously in one model rather than considering hand pain and function as completely separate outcomes.

10 Discussion and conclusions

The aim of this chapter is to provide a summary of the main findings presented in this thesis (Section 10.1) and to discuss further the strengths and weakness of the methodologies used that are applicable to all results presented in this thesis (Section 10.2). Further sections will also be presented to discuss further work and alternative analysis approaches (Sections 10.3 - 10.5).

10.1 Summary of the main findings

It has been shown that there is a lack of longitudinal studies exploring the course of hand pain and function over time in community-dwelling older adults, which is important to address, as both pain and functional difficulties are key consequences of having a hand condition, like OA, and are potential drivers of healthcare utilisation. Longitudinal data has therefore been collected and analysed in the Clinical Assessment Study of the Hand (the CAS-HA cohort study), and showed that, on average, hand pain and functional difficulty did not greatly change when assessed over a 7.5 year follow-up period in a sample of participants reporting hand pain in the last 12-months. This finding is also reflected by showing that the majority of participants had relatively stable trajectories of hand pain and functional difficulty over the 7.5 year follow-up period, with 68% and 78% of participants having trajectories that fluctuated around mean levels described as “Severe”, “Mild” and “Moderate” for hand pain and “Severe”, “Mild” and “Moderate” and “Mild/Moderate” for hand function respectively, hence clinical deterioration of symptoms is not always inevitable for all older people reporting hand pain and functional difficulty.

When individual trajectories were considered, key baseline factors to predict those individuals at greater risk of symptom deterioration were: hand pain; physical function, number of days with hand pain in the last 12-months, hand function, pinch strength, and the number of joints with radiographic hand OA, and for hand function; age, physical function, number of co-morbidities, hand pain and the grip-ability test. It was recognised

however, that such models, although suitable for predicting individual risk of deterioration, were limited in their ability to predict overall outcome course over a 7.5 year follow-up time-period. When predictors of the overall outcome course were explored using multinomial logistic regression to predict trajectory group membership it was found that the strongest predictor of overall trajectory course was the baseline measure in the outcome of interest, hence this may be the single strongest predictor of future symptom course.

10.2 Strengths and weakness

As described in Chapter 3, participants were recruited to the CAS-HA study from two general practices in North Staffordshire and were recruited from those reporting hand pain or problems in the last 12-months on a postal survey questionnaire. The design of the study raises several issues to be discussed in relation to the results presented, as below:

Recruitment of a population-based sample

A key strength of the study design is that participants in CAS-HA were recruited irrespective of their consultation patterns with a general practitioner for their hand problem as the study research questions, e.g. the population prevalence of a range of hand conditions, the prognosis of hand pain and function over time, the causes and consequences of this, could be addressed in the widest population possible and not depend on consultation behaviour[117]. As general practice (GP) records cover a large proportion of the UK population, they were therefore used as the sampling frame.

A key advantage of this recruitment method is that it allows the course of symptoms to be described for the population as a whole so it may be more reflective of "true" symptom course in the population, rather than in only a subset of participants that have consulted. As participants were recruited based on whether they reported hand pain in the last 12-

[117] Recruitment based on previous consultation with a general practitioner is practically possible (as illustrated in previous trials conducted in our research centre (e.g the APEX trial (Hay et al. 2004)), however this also relies on accurate recording of the consultation for a hand problem in the GP medical records

months on a survey questionnaire, they were only required to have hand pain rather than to be diagnosed with a specific hand condition. This approach was used as pain and functional difficulty are often the key reason for health care consultation rather than other clinical aspects of disease severity (e.g. x-ray evidence of hand OA).

Symptom onset

As participants have been recruited from a population survey they are not necessarily recruited at the point of symptom onset so little is known about the course of participants' symptoms prior to study enrolment other than a self-report question on the length of time the participant had a hand condition prior to baseline (a question that in itself may be subject to recall bias as participants with recent symptom onset may be more able to accurately report their time of onset than those whose symptoms first occurred a long time ago). It is therefore assumed in the modelling in Chapters 6, 8 and 9 that the rate of symptom change from baseline is the same for someone who has had their condition for a long time as for someone who has only had it for a short time, which may not be plausible.

This assumption is considered using data from Chapter 8 that shows that participants in the "Severe" trajectory groups for hand pain and function report, on average, having their hand problem for longer than for participants in all other trajectory groups. Participants may therefore be in this group because they have had more time to develop more severe symptoms rather than necessarily implying that the participants in this group developed severe symptoms more quickly. Evidence in Chapter 7 also supports this, as the predictor "length of time with a hand condition" did not predict the model slope for hand pain or function (i.e. the rate of change over time), albeit this is interpreted in light of the other variables in the model.

This is a limitation of the study design: that symptom severity data is not collected at regular intervals from the point of symptom onset. It would however be methodologically difficult to construct a "symptom onset", i.e. inception, cohort, as it would be difficult to

identify participants at the point of symptom onset. Another option would be to define the inception cohort based on the point of first consultation for a hand problem. This would be limited, however, as the decision to consult is not always motivated by symptom duration or severity. Alternatively, using repeated surveys to regularly monitor all adults at age 50-years onwards to define their point of symptom onset is unlikely to be feasible.

The analysis in this thesis therefore leaves unanswered questions around whether the trajectory groups in Chapter 8 simply represent groups with different levels of symptom severity or whether they reflect different stages along the long term progression of hand pain and function over time. For example, it remains unclear whether it would be possible for participants in the "Mild" trajectory group to remain in a "Mild" group for the rest of their lives, or whether progression beyond 7.5 years is inevitable for participants in this group. This could only be explored by extending the length of follow-up for each participant to be beyond 7.5 years.

Regression to the mean

Regression to the mean[118] potentially could have occurred in the CAS-HA study if participants were more motivated to attend the clinical assessment when their symptoms were severe, relative to their usual symptom level. There is slight evidence of this in Chapter 3 as participants attending the clinical assessment have slightly raised levels of AUSCAN pain and functional difficulty compared to those that did not attend. This would impact on the study findings if, for example, participants' symptoms are fluctuating between an upper and lower bound prior to, and between, the baseline and 18-month follow-up time points, and if the baseline measure is recorded at the upper bound of the fluctuating symptoms. It is then likely that the measure taken at 18-months is lower than at baseline, indicating that improvement over time has occurred, when in reality this reflects fluctuation of symptoms around a mean value that is stable over time. Regression to the

[118] This is a bias that occurs when participants are recruited at a time when their symptoms are most severe so are more likely to have readings closer to the mean when measured on a second occasion (Everitt et al. 2005)

mean is less likely in the CAS-HA cohort however as participants are not recruited at the point of consultation with a general practitioner when it is more likely that symptoms of pain or functional difficulty are more severe than usual.

Follow-up time periods

The timing of measurements in the CAS-HA study was planned to be at regular 18-month intervals over a 6-year follow-up time period. A 6-year follow-up period was chosen so it could be explored how symptoms change over a relatively long time-period as, at the time of applying for study funding, there was a lack of studies with longer term follow-up (>3-years) for conditions such as osteoarthritis that were likely to be chronic and last for several years (Jordan et al. 2009). Although the relatively long follow-up time period is a key strength of this study, it is also acknowledged that results can only apply to this follow-up time window so it cannot be inferred from the data how participants' symptoms changed beyond the last follow-up time point. A longer term follow-up beyond 6-years would be of interest in this cohort however, to examine whether the rates of symptom change over time increase if participants are observed for longer. This is especially relevant as, overall, the rate of symptom change over time in this study is small (Chapter 6).

Also, when considering the time frame for the study, regular 18-month follow-up time-periods were chosen to balance the cost of data collection over a long follow-up time-period, however, it is acknowledged that with a relatively long time period between two follow-up time points, the ability to identify participants whose symptoms are rapidly fluctuating on a regular basis, between weeks or between months, is limited. There is also the potential for changes over time to remain undetected in the data if they occur in periods of time that are not covered by the time-frames used in the questions on the questionnaires.

Related to the spacing of the time points, an “average” time was calculated to define the time-point for analysis. This was needed especially for the latter two time points as there was a delay in the mailing so an appropriate follow-up time needed to be estimated. Although an “average” time is commonly used for chronic diseases, a variable for the “exact” time since baseline can be included as a predictor in the growth models (i.e. by calculating a value for time for each participant that could potentially differ between participants at each time point rather than coding time as 0, 1.5, 3, 5.25 and 7.417 years). This approach was not used throughout this thesis due to the complexity of using the “exact” time variable in LCGM and PPGMM, however, when the “final” prediction models in Chapter 7 were re-run using the “exact” time approach, the results did not differ greatly from those that have been presented (data not shown).

Selection bias

The issue of selection bias was explored in Chapter 3 and it was shown that despite only 29% of participants attending the clinical assessment, the key characteristics of those in the CAS-HA cohort differed only slightly from those in the wider sample of interest. This is a key strength of the study design, and hence, the analysis presented in this thesis, especially that the AUSCAN (as the outcome of interest), is representative of the wider sample of participants with hand pain who completed the Regional Pain questionnaire.

In addition, the follow-up rates in CAS-HA were high at the first two time follow-up time points (>90%), and, despite this dropping to 71 and 66% at the latter two time-points, there is still evidence that participants’ baseline characteristics at the 7.5-year follow-up are only slightly different from those recruited at baseline. Although slight, it is acknowledged, however that those remaining in the cohort at the 7.5-year follow-up had less severe hand problems at baseline compared to all those who attended the baseline clinical assessment. This is a limitation of the study, and could have occurred if older members of the cohort were those who had more severe hand problems and were also

more likely to drop out of the cohort due to frailty, other serious health conditions, or death.

Suitability of the outcome measure

In Chapter 4, it was shown that the AUSCAN is a reliable and valid measure of hand pain and functional difficulty. However, it is acknowledged that the AUSCAN is a self-reported measure of hand pain and functional difficultly, and evaluates hand pain and functional difficulty in light of any adaptations that participants are using to manage their hand condition, e.g. taking medication, avoiding activities or using gadgets to help with tasks of daily living (Bellamy et al. 2002a). Throughout this thesis, it therefore cannot simply be assumed that when no change occurs it is because there has been no deterioration in symptoms. It could be that people have adapted to their pain problem and perceive symptoms differently (response shift (Streiner 2003)) or are using alternative strategies that help them to adapt and manage their hand pain or functional difficulty more effectively (Johnson et al. 2007, Nicholls et al. 2014).

One alternative approach to measuring functional difficulty could be to ask participants to identify their worst functional problem at baseline (e.g. picking up pans, doing up buttons) and for participants to rate the severity of this problem at each time point (Dziedzic et al. 2011, Beurskens et al. 1995). By holding the functional task "constant" in the measure, this would avoid people reporting that they had no functional difficulty with the task when the reason for this was because they had stopped doing the task specified in the closed form question. This alternative measure of hand function was not used in the CAS-HA study however as it is difficult to compare outcomes between groups of participants when the form of the outcome measure also differs between those participants being compared.

In addition, objective measures of grip and pinch strength and the grip-ability test were only included at baseline and the 6-year follow-up time-point as participants only came for a clinical assessment at these time-points, so it was not possible to test whether the

findings in this thesis are replicated using objective measures of hand function. The outcome measures used in this thesis focussed on hand pain and functional difficulty rather than using direct measures of the clinical features of specific hand conditions (e.g. x-ray evidence of hand OA, nerve conduction tests for carpal tunnel) as it is predominantly hand pain and functional difficulty that motivates consultation (Bijsterbosch et al. 2011, Peat, G., personal communication) and, for OA in particular, the course of symptoms often differs from that of structural progression of the disease (Haugen et al. 2013).

Accuracy of the outcome measure

In Chapter 6 it was highlighted that there exists an excess of participants with no hand pain or functional difficulty at the 18-month follow-up relative to the other time-points. This could have occurred due to an inconsistency in the instruction on the 18-month questionnaire compared to other time points. On the 18-month questionnaire participants were guided to complete the AUSCAN only if they had hand problems, whereas at all other time points participants completed the AUSCAN irrespective of whether they had a hand problem or not. To correct for this in the analysis, where participants had indicated that they had no hand problems their AUSCAN score was replaced from a missing value to be zero. Considering the data alongside the other time-points, this may have been too conservative as it may be that even when participants report no hand problems, their average AUSCAN score is greater than zero. The effect of this inconsistency on the questionnaire is a potential limitation of the findings of this study.

Measurement of the predictors

In Chapter 7, several potentially important predictors were excluded from the analysis because they had low prevalence and would not provide reliable estimates when analysed (e.g. ethnicity, previous stroke, Parkinson's disease, or rheumatoid arthritis). Although it is a strength of the CAS-HA study that participants have been recruited irrespective of their

clinical diagnosis and non-hand-specific characteristics[119] (to gain more accurate rates of prevalence), it limits the ability to explore the course of hand pain and functional difficulty in particular clinical subgroups of the population where the condition of interest is uncommon in the population.

For this, condition-specific cohorts would be needed to identify participants over a wider geographical area and from secondary care. That said though, the presence of five common clinical conditions were determined from the clinical assessment data (OA, carpal tunnel syndrome, Dupuytren's contracture, De Quervain's tenosynovitis and trigger finger) and were used in Chapters 7 and 8 to test whether the course of hand pain and functional difficulty differed depending on clinical cause. These results were interpreted, however, with the complexity that several participants have multiple potential causes for their hand problems (see Appendix 38). It is acknowledged however, that although multiple causes of hand pain have been included in this cohort, the majority of participants have hand OA (82% of participants have at least one hand joint with evidence of radiographic hand OA defined as Kellgren and Lawrence grade >=2) and that the proportion of people with the other conditions is similar for those with and without radiographic hand OA (Appendix 39).

Generalisability of study findings

The sampling frame for the CAS-HA study was selected from general practice records which is appropriate in the UK as 98% of the UK population are registered with a general practitioner (Bowling 2014). However as both practices are located in the North Staffordshire area of the UK, this may mean that results do not generalise to other areas of the UK or other countries. Although this is a limitation of the study, this was

[119] Recruiting participants based on diagnosis (either self-reported by the participant or based on medical records) could be problematic as it assumes that a previous consultation has occurred for a diagnosis to be obtained and that all medical records are accurate and reliable. Knowing a diagnosis may also be less useful if this will not change how the participant is managed in clinical practice and largely restricts people to be in a dichotomy of having the condition or not (Hemingway et al. 2013). This may not be appropriate for diseases that include a range of clinical presentations and severities e.g. OA

unavoidable, due to the practicalities of running the research clinics to collect the study data. However, to try and minimise this bias, when the general practices were chosen, they were purposely chosen so that one was urban and one a semi-rural practice. The findings in this study also only apply to the health care system in the UK as it has been provided over the time course of the study so do not generalise to other health care settings where service provisions may not reflect that given in the UK.

10.3 Further work

The analyses presented in this thesis are largely exploratory so represent only a first step in exploring trajectories of hand pain and function in this population. Several areas of further work therefore emerge, that could potentially be explored, both in terms of further research questions, but also in terms of differing methodological approaches that could be applied to the data.

Extending the pool of potential predictors

In Chapter 2, the process that was undertaken to select the pool of predictors of interest was described and it was highlighted that predictors could only be explored if they had been measured in the CAS-HA study with a low percentage of missing data. The pool of variables tested was therefore not exhaustive. For example, a measure of somatization[120] was not included in the analysis, although measured in the study, as the identity subscale of the IPQ-R had a high percentage of missing data in the CAS-HA baseline cohort, but this could be a potentially important predictor, as found in another study (Spies-Dorgelo et al. 2008), and other variables, perhaps less relevant to primary care, have not be explored (e.g. genetic variables relating to the underlying cause of the disease (Hamalainen et al. 2014)). It is acknowledged however, that when selecting a pool of predictors for testing in a prediction model that a small set of clinically plausible variables may be preferable,

[120] Somatization is defined as the unconscious process by which psychological distress is expressed as physical symptoms (MedicineNet 2012)

rather than being over inclusive to include all variables that could possibly be measured in a dataset.

In addition, when the role of radiographic hand OA was explored, only a simple definition of radiographic severity was used (number of hand joints with Kellgren-Lawrence grade >=2). Further work could be conducted to explore whether more complex definitions of radiographic hand OA, such as those based on joint location e.g. nodal OA, and type of OA, e.g. erosive OA, are more predictive of the outcomes of interest. This analysis has been completed in the CAS-HA cohort exploring the predictive ability of hand OA subsets from baseline to the 3-year follow-up (Marshall et al. 2013), but to date, this has not been completed using all of the time points in the study and explored alongside other potential model predictors.

Mechanisms for explaining changes in hand pain and function

The research questions in this thesis have been focussed on describing and predicting the course of hand pain and functional difficulty over time from baseline predictors without regard to the mechanism through which the variables might explain outcome. Further work could therefore be undertaken to explore more hypothesis-driven research questions around why, and how, factors explain the outcomes of interest. For example, by introducing time-varying covariates and time-lagged predictors into the prediction models in Chapter 7, it could be explored whether current and previous levels of anxiety and depression lead to an increase in hand pain and functional limitations over time, or models based on change scores could be used to test whether increasing participants' control over their condition improves future symptom course (e.g. to inform the content of a future intervention package). In addition, considering alternative outcomes in a parallel process growth model (e.g. body mass index, number of affected joints with hand OA) offers an alternative way to explore the relationship of each time-varying outcome with hand pain or hand function over time.

Adding time-varying or time-lagged covariates into the models in Chapter 7 could also be used to explore if their addition improves model fit. This would be useful as currently the models in Chapter 7 assume that the course of hand pain and function is solely governed by baseline variables (Curran et al. 2003). This latter assumption was made in Chapter 7 as the aim of the models was to identify baseline factors to predict the course of the outcome over time, with a view to providing information that could potentially allow the identification of people at increased risk of more severe or progressive trajectories. This assumption may not be appropriate, however, if the aim is to develop a model that provides an optimal explanation of the course of the outcome overall.

Although it would be useful to address the research questions such as those above, one possible limitation around testing for time-lag in the CAS-HA data would be the long time period between follow-up data collection points (as several other factors could also be simultaneously changing between the time points). Other sources of data, such as that arising from clinical trials or other epidemiological studies with a shorter time-period between follow-ups, may therefore be more appropriate to test more complex questions around predictor time-lag and the outcome of interest.

Trajectory groups as predictor variables

LCGM and GMM can be extended to model, simultaneously, not only the trajectory of the outcome of interest, but also how that trajectory predicts an external outcome of interest, when the external outcome is measured either at the last, or at a later time-point to the outcome variable (Wang et al. 2007). This model extension could be useful to explore further research questions such as how the long term trajectories of hand pain and function predict long-term progression of x-ray evidence of hand OA over time. This could extend work previously published by Bijsterbosch et al. (2011) who explored whether mean change in the AUSCAN from baseline to a 6-year follow-up differed between those having, or not having, radiographic progression over the same time period. This extension

of the analysis would allow data from all follow-up time points to be used in the analysis rather than defining a single time point to anchor how change in the AUSCAN is calculated over time.

Outcome of interest

In Chapter 9 it was shown that there is a relationship between hand pain and hand function, so it may be possible that further analysis could repeat the prediction models in Chapter 7, but using a total score for hand pain and function as the outcome of interest, rather than using separate scores for hand pain and function. This would enable the analysis to focus on an overall summary score of the participants' problem and to identify the predictors that predict an overall summary score rather than having different predictors for hand pain and function. This may be useful if the aim is to identify a small list of predictors that could be put forward as potentially useful for clinical practice where the problems of the whole person are taken into account.

Regression to the mean

It has been discussed earlier in this chapter that regression to the mean is less of a concern in the CAS-HA study than for studies measuring outcomes using very short time windows, e.g. on a single day. However, one way to explore the possible influence of regression to the mean could be to repeat the analysis excluding data at baseline (e.g. to explore whether the predictors measured at the 18-month follow-up predict the outcome across the subsequent time points). This may be a useful sensitivity analysis to conduct, as participants in CAS-HA were recruited at an arbitrary time point along their symptom trajectory, so using a different time point as the starting point for the analysis could be a partial test of model replication. It could also be useful to extend this model into a time-lag model to explore the predictive value of a previous measure of hand pain and functional difficulty, as in the models as presented, the only way to test whether baseline levels of

hand pain and function predicted the rate of progression over time was to examine the correlation between the model intercept and model slope.

Health care use

A key aim of this thesis was to model the course of hand pain and function irrespective of consultation patterns and the treatments that participants are currently using to treat their hand pain and functional difficulty. No descriptive information has therefore been given on the type and frequency of treatments that people were using or received over the course of the study. This information could be obtained, however, using a combination of data from the follow-up questionnaires in the CAS-HA study and also information from a review of the general practice medical records of participants that gave consent, and could be used to explore if the types of treatments that participants are currently using, or have previously used, differ between the trajectory groups.

Although this may be an interesting issue to explore, it is complicated by the wide range of treatments that participants could be trying for their hand condition and also by whether the option to use each treatment was equally available or offered to all participants. This could be influenced by whether a healthcare practitioner had been consulted, previous treatment experience, treatment preferences, access to services, or other treatments that are being used for other health conditions. Also, as this is not a randomised trial, analysis of questions exploring the effectiveness of a particular treatment would need to address the high risk of selection bias and confounding.

Predicting treatment response

A further area of work could also consider whether the factors identified in this thesis are also predictive of effective response to treatment; it cannot be concluded that factors that predict prognosis are necessarily those that predict response to a specific type of treatment (Hingorani et al. 2013, Riley et al. 2013). Moreover, an examination of which factors are potentially modifiable and could be targeted by treatment may be undertaken

however this would require analysis of data from a large clinical trial with a potentially effective treatment.

10.4 Alternative analysis approaches

A key feature of the outcomes modelled in this thesis is their skewed distribution which has been accounted for by estimating the model parameters using maximum likelihood with robust standard errors. Alternative approaches to deal with the analysis of a skewed outcome are possible, e.g. generalised growth models could be fitted to the data using a gamma link function to relate the linear predictor to the outcome of interest (Azuero et al. 2010). Although this approach is possible for growth models, it would be complex to extend this to LCGM and PPGMM and the software to do this is not currently available in Mplus, so this approach was not used throughout this thesis.

An alternative approach to the analysis of skewed outcomes in the SEM framework has been proposed by Asparouhov et al. (2014b) and is currently under development in Mplus version 7.2. This approach aims to model the data using a skew t-distribution to estimate not only the mean and covariance structure in the data, but also the higher order moments of skew and kurtosis. This approach is promising as an alternative analysis technique as all of the SEM models described in this thesis could potentially be fitted in this framework. It was not used in this thesis however as the approach is still under development and at the time of analysis, only version 7.11 of the Mplus software was available.

Along with a skewed distribution, a further feature of the outcomes modelled is that they are derived by summing together scores for the individual items on the questionnaire. This therefore makes the assumption that the items are measured on an interval-level scale and are additive (e.g. that the difference between a score of 1 and 2 is the same as the difference between a score of 3 and 4) and that the factor structure of the AUSCAN items remains constant over time, which may not be appropriate. An alternative analysis approach could be to use Rasch analysis (Rasch 1960, Fischer 1995) to develop a truly

interval-level score from the AUSCAN items and use these revised scales as the outcome of interest as has been carried out for other self-completed measures, e.g. Muller et al. (2009).

Such scale development has been conducted by Haugen et al (2011b) for the AUSCAN measure using data from the Oslo Hand OA cohort where it was found that the AUSCAN may be improved by dropping the item “pain at rest” and splitting the function scale into two: one for high precision and one for grip strength tasks. A revised version of the AUSCAN scoring could be developed using Rasch modelling techniques in the CAS-HA data and then explored whether the results presented in this thesis are dependent on how the AUSCAN is scored. This, however, has the limitation that between-study comparisons are more complex when several different scoring methods have been used to scale the AUSCAN items into a total score.

In addition, an alternative analysis could be to include the AUSCAN in the SEM framework as a continuous latent variable that is marked by the items that make up each sub-scale to form a multiple indicator latent growth curve model (Curran et al. 2003). In this model, a separate estimate would be obtained at each time point to express the relationship between each AUSCAN item and its underlying latent variable, which could be compared at each time point to explore whether the factor structure of the AUSCAN was constant over time. However, as the number of time points and the number of items in the AUSCAN would be large for this analysis to be applied in this thesis, this approach was not used to reduce the number of estimates required of the data.

The outcome used in this thesis aimed to incorporate all of the data at all of the follow-up time points to fully characterise participants’ patterns of symptom change over time and to make best use of the longitudinal data collected. It was therefore difficult to implement some of the standard approaches to prediction modelling that involve defining a clear start- and end-point in the data and then using data collected at the start point to predict

data at the future endpoint (Steyerberg et al. 2013). An alternative analysis could be to define key endpoints in the CAS-HA data (possibly those that represent a short, medium, or long-term follow-up), and then use standard multivariable regression models to predict the outcome at these key points in time (with the outcome analysed either as a continuous measure or categorised to define a group of participants meeting a pre-defined threshold for minimum clinically important symptom deterioration). It could therefore be explored whether the same set of predictor variables are identified using this approach to those in this thesis, as this could offer a simpler approach to communicate information on prognosis to participants and clinicians than when prediction scores are derived by simultaneously estimating likely prognosis using all of the longitudinal data collected within a study.

10.5 Model replication and future use

The models in this thesis have been developed to provide descriptive information on how participants' hand symptoms change over time and to provide preliminary evidence on factors that may be important to consider when predicting how participants' hand pain and functional difficulty change over time. The evidence in this thesis is preliminary, so before predictors can be used for the identification of high risk groups a separate test of replication in an external cohort would be required, either in participants who have been recruited using methods similar to those in CAS-HA or in participants where the data is collected at the point of GP consultation[121]. It is therefore acknowledged that the analysis that has been presented in this thesis represents only the first stage of the full process needed to develop a full prediction model, and has focussed largely on model development rather than testing fully the internal and external validity of the models and

[121] It was considered whether to re-run the models in this thesis on a sub-group of CAS-HA participants that were defined if they had a recent consultation with their GP to test for model replication. This analysis was not conducted as the baseline variables would not have been collected directly at the point of consultation and the sub-group would not be an independent sample from the participants in CAS-HA so this was not an ideal sample with which to test for model replication

evaluating the impact that such models can have on clinical practice (Steyerberg et al. 2013).

In addition, before any prediction models could be applied, the feasibility of the measurement of the predictors would need to be considered. For example, the physical component score of the SF-12 was included in the models for both hand pain and functional difficulty but is measured using 12 questions, which may not be practical to apply, in practice. It could first be tested whether model fit is greatly inferior if only one or two questions from this measure were included in the model. This could then lead to the development and testing of a model that could be more practical to apply to a “real life” clinical setting.

References

Adebajo, A. (2010). *ABC of rheumatology* (4th ed.). Chichester: Wiley-Blackwell BMJ/Books.

Afifi, A., & Clark, V. (1996). *Computer-aided multivariate analysis* (Third ed.). London: Chapman and Hall.

Akaike, H. (1974). A new look at the statistical model identification. *IEE Transactions on Automatic Control, 19*(6), 716.

Allen, K. D., Jordan, J. M., Renner, J. B., & Kraus, V. B. (2006a). Relationship of global assessment of change to AUSCAN and pinch and grip strength among individuals with hand osteoarthritis. *Osteoarthritis and Cartilage, 14*(12), 1281-1287.

Allen, K. D., Jordan, J. M., Renner, J. B., & Kraus, V. B. (2006b). Validity, factor structure, and clinical relevance of the AUSCAN osteoarthritis hand index. *Arthritis and Rheumatism, 54*(2), 551-556.

Allen, K. D., DeVellis, R. F., Renner, J. B., Kraus, V. B., & Jordan, J. M. (2007). Validity and factor structure of the AUSCAN osteoarthritis hand index in a community-based sample. *Osteoarthritis and Cartilage, 15*(7), 830-836.

Allison, P. (2013). Why I don't trust the hosmer-lemeshow goodness of fit test for logistic regression. Retrieved from http://www.statisticalhorizons.com/hosmer-lemeshow. Last accessed: 15/4/15

Altman D. (1999). *Practical statistics for medical research* (First CRC Press reprint ed.). London: Chapman & Hall/CRC press.

Altman, R., Alarcon, G., Appelrouth, D., Bloch, D., Borenstein, D., Brandt, K. et al. (1990). The American College of Rheumatology criteria for the classification and reporting of osteoarthritis of the hand. *Arthritis and Rheumatism, 33*(11), 1601-1610.

American Accreditation HealthCare Commission. (2014a). Carpal tunnel syndrome. Retrieved from http://www.nlm.nih.gov/medlineplus/ency/article/000433.htm. Last accessed: 15/4/15

American Accreditation HealthCare Commission. (2014b). De quervain's tendinitis. Retrieved from http://www.nlm.nih.gov/medlineplus/ency/patientinstructions/000537.htm. Last accessed: 15/4/15

Andruff, H., Carraro, N., Thompson, A., & Gaudreau, P. (2009). Latent class growth modelling: A tutorial. *Tutorials in Quantitative Methods for Psychology, 5*(1), 11-24.

Arbuckle, J. L. (2006). Amos (Version 20.0). Chicago: SPSS.

Arreguin Reyes, R., Lopez Lopez, C. O., Alvarez Hernandez, E., Medrano Ramirez, G., Montes Castillo Mde, L., & Vazquez-Mellado, J. (2012). Evaluation of hand function in rheumatic disease. Validation and usefulness of the Spanish version AUSCAN, m-SACRAH and Cochin questionnaires. *Reumatologia Clinica, 8*(5), 250-254.

Arthritis Research UK. (2013). What is osteoarthritis? Retrieved from http://www.arthritisresearchuk.org/arthritis-information/conditions/osteoarthritis/what-is-osteoarthritis.aspx. Last accessed: 15/4/15

Arthritis Research UK. (2014). Key facts about arthritis. Retrieved from http://www.arthritisresearchuk.org/arthritis-information/data-and-statistics.aspx. Last accessed: 15/4/15

Aslam, I., Perjar, I., Shi, X. A., Renner, J. B., Kraus, V. B., Golightly, Y. M. et al. (2014). Associations between biomarkers of joint metabolism, hand osteoarthritis, and hand pain and function: The Johnston County Osteoarthritis Project. *The Journal of Rheumatology, 41*(5), 938-944.

Asparouhov, T., & Muthén, B. (2014a). Auxiliary variables in mixture modeling: Three-step approaches using Mplus. *Structural Equation Modeling: A Multidisciplinary Journal, 21*(3), 329-341.

Asparouhov, T., & Muthen, B. (2014b). Structural equation models and mixture models with continuous non-normal skewed distributions. Retrieved from http://www.statmodel.com/download/Skew.pdf. Last accessed: 15/4/15

Azuero, A., Pisu, M., McNees, P., Burkhardt, J., Benz, R., & Meneses, K. (2010). An application of longitudinal analysis with skewed outcomes. *Nursing Research, 59*(4), 301-307.

Ballinger, G. A. (2004). Using generalized estimating equations for longitudinal data analysis. *Organizational Research Methods, 7*(2), 127-150.

Baron, M., Dutil, E., Berkson, L., Lander, P., & Becker, R. (1987). Hand function in the elderly: relation to osteoarthritis. *The Journal of Rheumatology, 14*(4), 815-819.

Bellamy, N., Campbell, J., Haraoui, B., Buchbinder, R., Hobby, K., Roth, J. H., & MacDermid, J. C. (2002a). Dimensionality and clinical importance of pain and disability in hand osteoarthritis: development of the Australian/Canadian (AUSCAN) osteoarthritis hand index. *Osteoarthritis and Cartilage, 10*(11), 855-862.

Bellamy, N., Campbell, J., Haraoui, B., Gerecz-Simon, E., Buchbinder, R., Hobby, K., & MacDermid, J. C. (2002b). Clinimetric properties of the AUSCAN osteoarthritis hand

index: an evaluation of reliability, validity and responsiveness. *Osteoarthritis and Cartilage, 10*(11), 863-869.

Bellamy N. (2007). International estimation of minimally clinically important improvement (MCII75): The reflect study [Abstract]. *Intern Med J, 37* A36.

Bellamy, N., Wilson, C., & Hendrikz, J. (2011). Population-based normative values for the Australian/Canadian (AUSCAN) hand osteoarthritis index: Part 2. *Seminars in Arthritis and Rheumatism, 41*(2), 149-156.

Bentler, P. (1995). *EQS: Structural equations program manual*. Encino, CA: Multivariate software.

Bentler, P. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*(2), 238-246.

Bergner, M., Bobbitt, R. A., Carter, W. B., & Gilson, B. S. (1981). The sickness impact profile: Development and final revision of a health status measure. *Medical Care, 19*(8), 787-805.

Berkman, L. F., & Syme, S. L. (1979). Social networks, host resistance, and mortality: A nine-year follow-up study of Alameda county residents. *American Journal of Epidemiology, 109*(2), 186-204.

Berrington, A., Smith, P. W. F., & Sturgis, P. (2006). *An overview of methods for the analysis of panel data.* National Centre for Research Methods, School of Social Sciences, University of Southampton.

Beurskens, A. J., de Vet, H. C., Koke, A. J., Lindeman, E., Regtop, W., van der Heijden, G. J., & Knipschild, P. G. (1995). Efficacy of traction for non-specific low back pain: A randomised clinical trial. *Lancet, 346*(8990), 1596-1600.

Biesheuvel, C. J., Vergouwe, Y., Steyerberg, E. W., Grobbee, D. E., & Moons, K. G. (2008). Polytomous logistic regression analysis could be applied more often in diagnostic research. *Journal of Clinical Epidemiology, 61*(2), 125-134.

Bijsterbosch, J., Watt, I., Meulenbelt, I., Rosendaal, F. R., Huizinga, T. W., & Kloppenburg, M. (2010). Clinical burden of erosive hand osteoarthritis and its relationship to nodes. *Annals of the Rheumatic Diseases, 69*(10), 1784-1788.

Bijsterbosch, J., Watt, I., Meulenbelt, I., Rosendaal, F. R., Huizinga, T. W., & Kloppenburg, M. (2011). Clinical and radiographic disease course of hand osteoarthritis and determinants of outcome after 6 years. *Annals of the Rheumatic Diseases, 70*(1), 68-73.

Birrell, F., Croft, P., Cooper, C., Hosie, G., Macfarlane, G. J., & Silman, A. (2000). Radiographic change is common in new presenters in primary care with hip pain. PCR hip study group. *Rheumatology (Oxford, England), 39*(7), 772-775.

Bland, J. M., & Altman, D. G. (1999). Measuring agreement in method comparison studies. *Statistical Methods in Medical Research, 8*(2), 135-160.

Bland, M. D., Beebe, J. A., Hardwick, D. D., & Lang, C. E. (2008). Restricted active range of motion at the elbow, forearm, wrist, or fingers decreases hand function. *Journal of Hand Therapy: Official Journal of the American Society of Hand Therapists, 21*(3), 268-74.

Bliese, P., & Ployhart, R. (2002). Growth modelling using random coefficient models: Model building, testing and illustrations. In W. Paul Vogt (Ed.), *SAGE quantitative research methods.* (Vol. 5, pp. 363-91). Thousand Oaks, CA: SAGE

Bollen, K., & Curran, P. (2006). *Latent curve models: A structural equation perspective*. Hoboken, New Jersey: John Wiley & Sons, Inc.

Bone and joint decade's musculoskeletal portal. Retrieved from http://www.boneandjointdecade.org/. Last accessed 15/4/15

Botha-Scheepers, S., Riyazi, N., Watt, I., Rosendaal, F. R., Slagboom, E., Bellamy, N. et al. (2009). Progression of hand osteoarthritis over 2 years: A clinical and radiological follow-up study. *Annals of the Rheumatic Diseases, 68*(8), 1260-1264.

Bowling, A. (2014). *Research methods in health* (Fourth ed.). Maidenhead: Open University Press.

Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthly Weather Review, 78*(1).

Byrne B. (2001). *Structural equation modeling with AMOS. Basic concepts, applications, and programming*. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Byrne, B. M., & Crombie, G. (2003). *Modeling and testing change: An introduction to the latent growth curve model.* Understanding Statistics 2(3).

Cano-de-la-Cuerda, R., Perez-de-Heredia, M., Miangolarra-Page, J. C., Munoz-Hellin, E., & Fernandez-de-Las-Penas, C. (2010). Is there muscular weakness in parkinson's disease? *American Journal of Physical Medicine & Rehabilitation/Association of Academic Physiatrists, 89*(1), 70-76.

Centre for Reviews and Dissemination. (2009). Centre for reviews and dissemination UoY systematic reviews: CRD's guidance for undertaking reviews in healthcare. Retrieved from http://www.york.ac.uk/inst/crd/index_guidance.htm. . Last accessed 15/4/15

Cheng J, Edwards LJ, Maldonado-Molina MM, et al. Real longitudinal data analysis for real people: building a good enough mixed model. Stat Med. 2010;29(4):504-20.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (Second edition ed.). Hillsdale, NJ: Lawrence Erlbaum.

Cole, M. (2008). *Investigating the cognitative processes that mediate protection motivation theory: A parallel process latent growth modeling analysis. PhD thesis. Wayne State University. USA*

Cole, A., Gill, T. K., Taylor, A. W., & Hill, C. L. (2011). Prevalence and associations of hand pain in the community: Results from a population-based study. *Scandinavian Journal of Rheumatology, 40*(2), 145-149.

Collett, D. (1999). *Modelling binary data* (Second ed.). United States of America: Chapman and Hall/CRC.

Collins, J. E., Katz, J. N., Dervan, E. E., & Losina, E. (2014). Trajectories and risk profiles of pain in persons with radiographic, symptomatic knee osteoarthritis: Data from the Osteoarthritis Initiative. *Osteoarthritis and Cartilage, 22*(5), 622-630.

Conaghan, P. G., Kloppenburg, M., Schett, G., Bijlsma, J. W., & EULAR osteoarthritis ad hoc committee. (2014). Osteoarthritis research priorities: A report from a EULAR ad hoc expert committee. *Annals of the Rheumatic Diseases, 73*(8), 1442-1445.

Coolican, H. (2014). *Research methods and statistics in psychology* (6th ed.). London: Psychology Press, Taylor & Francis Group.

Croft, P. R., Lewis, M., Papageorgiou, A. C., Thomas, E., Jayson, M. I. V., Macfarlane, G. J., & Silman, A. J. (2001). Risk factors for neck pain: A longitudinal study in the general population. *Pain, 93*(3), 317-325.

Crosby, R. D., Kolotkin, R. L., & Williams, G. R. (2003). Defining clinically meaningful change in health-related quality of life. *Journal of Clinical Epidemiology, 56*(5), 395-407.

Cui, J. (2007). QIC program and model selection in GEE analyses. *The Stata Journal, 7*(2), 209.

Curran, P. J., & Hussong, A. M. (2003). The use of latent trajectory models in psychopathology research. *Journal of Abnormal Psychology, 112*(4), 526-544.

Curran, P. J., Obeidat, K., & Losardo, D. (2010). Twelve frequently asked questions about growth curve modeling. *Journal of Cognition and Development: Official Journal of the Cognitive Development Society, 11*(2), 121-136.

Dahaghin, S., Bierma-Zeinstra, S. M., Ginai, A. Z., Pols, H. A., Hazes, J. M., & Koes, B. W. (2005a). Prevalence and pattern of radiographic hand osteoarthritis and association with pain and disability (the Rotterdam study). *Annals of the Rheumatic Diseases, 64*(5), 682-687.

Dahaghin, S., Bierma-Zeinstra, S. M., Reijman, M., Pols, H. A., Hazes, J. M., & Koes, B. W. (2005b). Prevalence and determinants of one month hand pain and hand related disability in the elderly (Rotterdam study). *Annals of the Rheumatic Diseases, 64*(1), 99-104.

de Vet, H. C., Terwee, C. B., Knol, D. L., & Bouter, L. M. (2006a). When to use agreement versus reliability measures. *Journal of Clinical Epidemiology, 59*(10), 1033-1039.

de Vet, H., Terwee, C. B., Ostelo, R. W., Beckerman, H., Knol, D. L., & Bouter, L. M. (2006b). Minimal changes in health status questionnaires: Distinction between minimally detectable change and minimally important change. *Health Qual Life Outcomes, 4*, 54.

de Vet, H. C., Ostelo, R. W., Terwee, C. B., van der Roer, N., Knol, D. L., Beckerman, H. et al. (2007). Minimally important change determined by a visual method integrating an anchor-based and a distribution-based approach. *Quality of Life Research: An*

International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation, 16(1), 131-142.

Dellhag, B., & Bjelle, A. (1995). A grip ability test for use in rheumatology practice. *The Journal of Rheumatology, 22*(8), 1559-1565.

Denkinger, M. D., Lukas, A., Nikolaus, T., Peter, R., Franke, S., & ActiFE study group. (2014). Multisite pain, pain frequency and pain severity are associated with depression in older adults: Results from the ActiFE ulm study. *Age and Ageing, 43*(4), 510-514.

Diggle, P., Heagerty, P., Liang, K., & Zeger, S. (2002). *Analysis of longitudinal data* (Second ed.). United States: Oxford University Press.

Dominick, K. L., Jordan, J. M., Renner, J. B., & Kraus, V. B. (2005). Relationship of radiographic and clinical variables to pinch and grip strength among individuals with osteoarthritis. *Arthritis and Rheumatism, 52*(5), 1424-1430.

Dziedzic, K. S., Thomas, E., & Hay, E. M. (2005). A systematic search and critical review of measures of disability for use in a population survey of hand osteoarthritis (OA). *Osteoarthritis and Cartilage, 13*(1), 1-12.

Dziedzic, K. S., Thomas, E., Myers, H., Hill, S., & Hay, E. M. (2007a). The Australian/Canadian osteoarthritis hand index in a community-dwelling population of older adults: Reliability and validity. *Arthritis and Rheumatism, 57*(3), 423-428.

Dziedzic, K., Thomas, E., Hill, S., Wilkie, R., Peat, G., & Croft, P. R. (2007b). The impact of musculoskeletal hand problems in older adults: Findings from the North Staffordshire osteoarthritis project (NorStOP). *Rheumatology (Oxford, England), 46*(6), 963-967.

Dziedzic, K. S., Hill, S., Nicholls, E., Hammond, A., Myers, H., Whitehurst, T. et al. (2011). Self management, joint protection and exercises in hand osteoarthritis: A randomised controlled trial with cost effectiveness analyses. *BMC Musculoskeletal Disorders, 12*, 156.

Dziedzic, K., Nicholls, E., Hill, S., Hammond, A., Handy, J., Thomas, E., & Hay, E. (2013). Self-management approaches for osteoarthritis in the hand: A 2x2 factorial randomised trial. *Annals of the Rheumatic Diseases,* 74(1), 108-118.

Everitt, B., & Palmer, C. (2005). *Encyclopaedic companion to medical statistics* (First ed.). London: Hodder Education.

Ferguson, E., & Cox, T. (1993). Exploratory factor analysis: A users' guide. *International Journal of Selection and Assessment, 1*(2), 84-94.

Fischer G. H. (1995). Derivations of the Rasch model. In: G.H. Fischer, I. W. Molenaar, editor. Rasch models: foundations, recent developments, and applications. New York: Springer-Verlag; 15-38.

Flaherty, B. P., & Kiff, C. J. (2012). Latent class and latent profile models. In: H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, D. Rindskopf & K. J. Sher (Eds.). Washington, DC, US: American Psychological Association; 391 - 404

Flom, P., & Cassell, D., L. (2007). Stopping stepwise: Why stepwise and similar selection methods are bad, and what you should use. Retrieved from http://www.nesug.org/proceedings/nesug07/sa/sa07.pdf. Last accessed 15/4/15

Foster, N. E., Bishop, A., Thomas, E., Main, C., Horne, R., Weinman, J., & Hay, E. (2008). Illness perceptions of low back pain patients in primary care: What are they, do they change and are they associated with outcome? *Pain, 136*(1-2), 177-187.

Frangos, C. C., & Schucany, W. R. (1990). Jackknife estimation of the bootstrap acceleration constant. *Computational Statistics & Data Analysis, 9*(3), 271-281.

Medical dictionary by Farlex. Retrieved from http://medical-dictionary.thefreedictionary.com/prognosis. Last accessed 15/4/15

Frey B., B. (2015). *100 questions and answers about tests and measurement* (1st ed.). United States of America: SAGE Publications, Inc.

Fries, J. F., Spitz, P. W., & Young, D. Y. (1982). The dimensions of health outcomes: the Health Assessment Questionnaire, disability and pain scales. *The Journal of Rheumatology, 9*(5), 789-793.

Gagliese, L. (2009). Pain and aging: The emergence of a new subfield of pain research. *The Journal of Pain: Official Journal of the American Pain Society, 10*(4), 343-353.

Ghisletta, P., & Spini, D. (2004). An introduction to generalized estimating equations and an application to assess selectivity effects in a longitudinal study on very old individuals. *Journal of Educational and Behavioral Statistics, 29*(4), 421-437.

Goldstein, H., Browne, W., & Rasbash, J. (2002). Multilevel modelling of medical data. *Statistics in Medicine, 21*(21), 3291-3315.

Goldstein, H., Browne, W., & Rasbash, J. (2004). Multilevel modelling of medical data. In R. D'Agostino (Ed.), *Tutorials in biostatistics* (Volume 2 ed., pp. 69) John Wiley and Sons Ltd.

Greenhalgh, T. (1997). How to read a paper. getting your bearings (deciding what the paper is about). *BMJ (Clinical Research Ed.), 315*(7102), 243-246.

Grimes, D. A., & Schulz, K. F. (2002). Bias and causal associations in observational research. *Lancet, 359*(9302), 248-252.

Grobbee, D. E., & Hoes, A. W. (1997). Confounding and indication for treatment in evaluation of drug treatment for hypertension. *BMJ (Clinical Research Ed.), 315*(7116), 1151-1154.

Gutierrez, R. G. (2005). Use of the term 'relative risk ratio'. Retrieved from: http://www.stata.com/statalist/archive/2005-04/msg00678.html. Last accessed 15/4/15

Guyatt, G. H., Bombardier, C., & Tugwell, P. X. (1986). Measuring disease-specific quality of life in clinical trials. *CMAJ: Canadian Medical Association Journal, 134*(8), 889-895.

Hamalainen, S., Solovieva, S., Vehmas, T., Luoma, K., Leino-Arjas, P., & Hirvonen, A. (2014). Genetic influences on hand osteoarthritis in Finnish women - a replication study of candidate genes. *PloS One, 9*(5), e97417.

Hart, D. J., & Spector, T. D. (2000). Definition and epidemiology of osteoarthritis of the hand: A review. *Osteoarthritis and Cartilage, 8 Suppl A*, S2-7.

Haugen, I. K., Slatkowsky-Christensen, B., Lessem, J., & Kvien, T. K. (2009). The responsiveness of joint counts, patient-reported measures and proposed composite scores in hand osteoarthritis: Analyses from a placebo-controlled trial. *Annals of the Rheumatic Diseases, 69(8),* 1436-40.

Haugen, I. K., Englund, M., Aliabadi, P., Niu, J., Clancy, M., Kvien, T. K., & Felson, D. T. (2011a). Prevalence, incidence and progression of hand osteoarthritis in the general population: The Framingham osteoarthritis study. *Annals of the Rheumatic Diseases,* 70(9), 1581-6.

Haugen, I. K., Moe, R. H., Slatkowsky-Christensen, B., Kvien, T. K., van der Heijde, D., & Garratt, A. (2011b). The AUSCAN subscales, AIMS-2 hand/finger subscale, and FIOHA were not unidimensional scales. *Journal of Clinical Epidemiology, 64*(9), 1039-1046.

Haugen, I. K., Slatkowsky-Christensen, B., Boyesen, P., van der Heijde, D., & Kvien, T. K. (2013). Cross-sectional and longitudinal associations between radiographic features and measures of pain and physical function in hand osteoarthritis. *Osteoarthritis and Cartilage, 21*(9), 1191-1198.

Haugen, I. K., Slatkowsky-Christensen, B., Boyesen, P., Sesseng, S., van der Heijde, D., & Kvien, T. K. (2014). MRI findings predict radiographic progression and development of erosions in hand osteoarthritis. *Annals of the Rheumatic Diseases.* doi: 10.1136/annrheumdis-2014-205949. [Epub ahead of print]

Hay, E., Barlas, P., Foster, N., Hill, J., Thomas, E., & Young, J. (2004). Is acupuncture a useful adjunct to physiotherapy for older adults with knee pain?: The "acupuncture, physiotherapy and exercise" (APEX) study. *BMC Musculoskeletal Disorders, 5*, 31.

Hayden J.A., Cote P., & Bombardier C. (2006). Evaluation of the quality of prognosis studies in systematic reviews. *Annals of Internal Medicine, 144*(6), 427-437.

Hayden, J. A., Chou, R., Hogg-Johnson, S., & Bombardier, C. (2009). Systematic reviews of low back pain prognosis had variable methods and results: Guidance for future prognosis reviews. *Journal of Clinical Epidemiology, 62*(8), 781-796.

Hayden, J. A., Dunn, K. M., van der Windt, D. A., & Shaw, W. S. (2010). What is the prognosis of back pain? *Best Practice & Research. Clinical Rheumatology, 24*(2), 167-179.

Hayden, J. A., van, d. W., Cartwright, J. L., Côté, P., & Bombardier, C. (2013). Assessing bias in studies of prognostic factors. *Annals of Internal Medicine, 158*(4), 280-286.

Hemingway, H., Croft, P., Perel, P., Hayden, J. A., Abrams, K., Timmis, A. et al. (2013). Prognosis research strategy (PROGRESS) 1: A framework for researching clinical outcomes. *BMJ, 346* doi:10.1136/bmj.e5595

Heyde, C. (1997). *Quasi-likelihood and its application: A general approach to optimal parameter estimation* (First ed.). New York: Springer.

Higgins, J. P., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ (Clinical Research Ed.), 327*(7414), 557-560.

Hill, S., Dziedzic, K., Thomas, E., Baker, S. R., & Croft, P. (2007). The illness perceptions associated with health and behavioural outcomes in people with musculoskeletal hand problems: Findings from the North Staffordshire osteoarthritis project (NorStOP). *Rheumatology (Oxford, England), 46*(6), 944-951.

Hill, S., Dziedzic, K. S., & Ong, B. N. (2010). The functional and psychological impact of hand osteoarthritis. *Chronic Illness, 6*(2), 101-110.

Hingorani, A. D., Windt, D. A. v. d., Riley, R. D., Abrams, K., Moons, K. G. M., Steyerberg, E. W., et al. (2013). Prognosis research strategy (PROGRESS) 4: Stratified medicine research. *BMJ, 346* doi:10.1136/bmj.e5793

Holla, J. F., van der Leeden, M., Heymans, M. W., Roorda, L. D., Bierma-Zeinstra, S. M., Boers, M., et al. (2014). Three trajectories of activity limitations in early symptomatic knee osteoarthritis: A 5-year follow-up study. *Annals of the Rheumatic Diseases, 73*(7), 1369-1375.

Hooper, D., Coughlan, J., & Mullen, M. (2008). Structural equation modelling: Guidelines for determining model fit. *The Electronic Journal of Business Research Methods, 6*(1), 53.

Hox, J., & Stoel, R. (2005). *Multi-level and SEM approaches to growth curve modeling*. Chichester: John Wiley & Sons, Ltd.

Hu, B., Shao, J., & Palta, M. (2006). Pseudo-R^2 in logistic regression model. *Statistica Sinica, 16*(3), 847.

Hu, F. B., Goldberg, J., Hedeker, D., Flay, B. R., & Pentz, M. A. (1998a). Comparison of population-averaged and subject-specific approaches for analyzing repeated binary outcomes. *American Journal of Epidemiology, 147*(7), 694-703.

Hu, L., & Bentler, P. M. (1998b). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods, 3*(4), 424-453.

Hunter, S., & Crome, P. (2002). Hand function and stroke. *Reviews in Clinical Gerontology, 12*(01), 68.

IBM Corp (2011). In Arbuckle JL (Ed.), *IBM SPSS Amos 20 user's guide*

IBM Corp (2013). Released 2013. IBM SPSS Statistics for Windows, Version 22.0. Armonk, NY: IBM Corp.

Institute for Digital Research and Education. (2011). FAQ: What are pseudo R-squares? Retrieved from http://www.ats.ucla.edu/stat/mult_pkg/faq/general/Psuedo_RSquareds.htm. Last accessed 15/4/15

Jaeschke, R., Singer, J., & Guyatt, G. H. (1989). Measurement of health status. Ascertaining the minimal clinically important difference. *Controlled Clinical Trials, 10*(4), 407-415.

Jiang, X. (2004). *Nonparametric quasi-likelihood in longitudinal data analysis. PhD thesis. University of Maryland, USA.*

Johnson, S. R., Archibald, A., Davis, A. M., Badley, E., Wright, J. G., & Hawker, G. A. (2007). Is self-reported improvement in osteoarthritis pain and disability reflected in objective measures? *The Journal of Rheumatology, 34*(1), 159-164.

Jones, G., Cooley, H. M., & Bellamy, N. (2001). A cross-sectional study of the association between heberden's nodes, radiographic osteoarthritis of the hands, grip strength, disability and pain. *Osteoarthritis and Cartilage, 9*(7), 606-611.

Jordan, K. P., Wilkie, R., Muller, S., Myers, H., Nicholls, E., & Arthritis Research Campaign National Primary Care Centre. (2009). Measurement of change in function and disability in osteoarthritis: Current approaches and future challenges. *Current Opinion in Rheumatology, 21*(5), 525-530.

Jung, T., & Wickrama, K. A. S. (2008). An introduction to latent class growth analysis and growth mixture modeling. *Social and Personality Psychology Compass, 2*(1), 302-317.

Kahn, J. H. (2011). Multilevel modeling: Overview and applications to research in counseling psychology. *Journal of Counseling Psychology, 58*(2), 257-271.

Kamper, S. J., Ostelo, R. W., Knol, D. L., Maher, C. G., de Vet, H. C., & Hancock, M. J. (2010). Global perceived effect scales provided reliable assessments of health transition in people with musculoskeletal disorders, but ratings are strongly influenced by current status. *Journal of Clinical Epidemiology, 63*(7), 760-766.e1.

Kellgren, J. H., & Lawrence, J. S. (1957). Radiological assessment of osteo-arthrosis. *Annals of the Rheumatic Diseases, 16*(4), 494-502.

Kenny D. (2014). Measuring model fit. Retrieved from http://davidakenny.net/cm/fit.htm. Last accessed 15/4/15.

Kirkwood BR, & Sterne AC. (2003). Systematic reviews and meta analysis. *Essential medical statistics* (Second ed., pp. 371). United Kingdom: Blackwell Publishing.

Kuchibhatla, M., & Fillenbaum, G. G. (2003). Comparison of methods for analyzing longitudinal binary outcomes: Cognitive status as an example. *Aging & Mental Health, 7*(6), 462-468.

Kuijpers, T., van der Windt, D. A., van der Heijden, G. J., & Bouter, L. M. (2004). Systematic review of prognostic cohort studies on shoulder disorders. *Pain, 109*(3), 420-431.

Kwok, W. Y., Plevier, J. W., Rosendaal, F. R., Huizinga, T. W., & Kloppenburg, M. (2013). Risk factors for progression in hand osteoarthritis: A systematic review. *Arthritis Care & Research, 65*(4), 552-562.

Lawrence, J., S. (1977). Osteo-arthrosis. *Rheumatism in populations* (pp. 98). London: William Heinemann Medical Books.

Liang, K., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika, 73*(1), 13-22.

Lister, G. (1978). *The hand: Diagnosis and indications* (2nd ed.). Edinburgh: Churchill Livingstone.

Little, R. J. A., & Rubin, D. B. (1987). *Statistical analysis with missing data*. New York: Wiley.

Long, J. S., & Freese, J. (2006). *Regression models for categorical dependent variables using STATA* (Second ed.). United States of America: Stata press.

MacDermid, J. C., Wessel, J., Humphrey, R., Ross, D., & Roth, J. H. (2007). Validity of self-report measures of pain and disability for persons who have undergone arthroplasty for osteoarthritis of the carpometacarpal joint of the hand. *Osteoarthritis and Cartilage, 15*(5), 524-530.

Macfarlane, G. J., Croft, P. R., Schollum, J., & Silman, A. J. (1996). Widespread pain: Is an improved classification possible? *The Journal of Rheumatology, 23*(9), 1628-1632.

Macfarlane, G. J. (2008). Changing patient perceptions of their illness: Can they contribute to an improved outcome for episodes of musculoskeletal pain? *Pain, 136*(1-2), 1-2.

Mallen, C. D., Peat, G., Thomas, E., Dunn, K. M., & Croft, P. R. (2007a). Prognostic factors for musculoskeletal pain in primary care: A systematic review. *The British Journal of General Practice: The Journal of the Royal College of General Practitioners, 57*(541), 655-661.

Mallen, C. D., Peat, G., Thomas, E., Lacey, R., & Croft, P. (2007b). Predicting poor functional outcome in community-dwelling older adults with knee pain: Prognostic value of generic indicators. *Annals of the Rheumatic Diseases, 66*(11), 1456-1461.

Mallen, C. D., & Peat, G. (2009). Discussing prognosis with older people with musculoskeletal pain: A cross-sectional study in general practice. *BMC Family Practice, 10*, 50.

Marshall, M., van der Windt, D., Nicholls, E., Myers, H., Hay, E., & Dziedzic, K. (2009). Radiographic hand osteoarthritis: Patterns and associations with hand pain and function in a community-dwelling sample. *Osteoarthritis and Cartilage, 17*(11), 1440-1447.

Marshall, M., Peat, G., Nicholls, E., van der Windt, D., Myers, H., & Dziedzic, K. (2013). Subsets of symptomatic hand osteoarthritis in community-dwelling older adults in the United Kingdom: Prevalence, inter-relationships, risk factor profiles and clinical characteristics at baseline and 3-years. *Osteoarthritis and Cartilage, 21*(11), 1674-1684.

Massy-Westropp, N., Krishnan, J., & Ahern, M. (2004). Comparing the AUSCAN osteoarthritis hand index, Michigan hand outcomes questionnaire, and sequential

occupational dexterity assessment for patients with rheumatoid arthritis. *The Journal of Rheumatology, 31*(10), 1996-2001.

Mathiowetz, V., Weber, K., Volland, G., & Kashman, N. (1984). Reliability and validity of grip and pinch strength evaluations. *The Journal of Hand Surgery, 9*(2), 222-226.

Matthews,J., N.,S. (2005). Summary measures analysis of longitudinal data. *Encyclopedia of biostatistics.* John Wiley & Sons, Ltd.

McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods, 1*(1), 30-46.

MedicineNet. (2012). Definition of somatization. Retrieved from http://www.medterms.com/script/main/art.asp?articlekey=9463. Last accessed 15/4/15

Medline Plus (2015). Finger injuries and disorders. Retrieved from http://vsearch.nlm.nih.gov/vivisimo/cgi-bin/query-meta?v%3Aproject=medlineplus&query=dupuytren+contracture. Last accessed 15/4/15

Meenan, R. F., Mason, J. H., Anderson, J. J., Guccione, A. A., & Kazis, L. E. (1992). AIMS2. The content and properties of a revised and expanded arthritis impact measurement scales health status questionnaire. *Arthritis and Rheumatism, 35*(1), 1-10.

Microsoft Office. Official website for microsoft ACCESS. Retrieved from http://office.microsoft.com/en-gb/access/. Last accessed 15/4/15

Moe, R. H., Garratt, A., Slatkowsky-Christensen, B., Maheu, E., Mowinckel, P., Kvien, T. K. et al. (2010). Concurrent evaluation of data quality, reliability and validity of the Australian/Canadian osteoarthritis hand index and the functional index for hand osteoarthritis. *Rheumatology (Oxford, England), 49*(12), 2327-2336.

Moon, K. W., Lee, S. S., Kim, J. H., Song, R., Lee, E. Y., Song, Y. W. et al. (2012). Cross-cultural adaptation, validation, and responsiveness of the Korean version of the AUSCAN osteoarthritis index. *Rheumatology International, 32*(11), 3551-3557.

Mora, P. A., Bennett, I. M., Elo, I. T., Mathew, L., Coyne, J. C., & Culhane, J. F. (2009). Distinct trajectories of perinatal depressive symptomatology: Evidence from growth mixture modeling. *American Journal of Epidemiology, 169*(1), 24-32.

Moss-Morris, R., Weinman, J., Petrie, K. J., Horne, R., Cameron, L. D., & Buick, D. (2002). The revised illness perception questionnaire (IPQ-R). *Psychology & Health, 17*(1), 1-16.

Muller, S., Thomas, E., & Peat, G. (2009). Derivation and testing of an interval-level score for measuring locomotor disability in epidemiological studies of middle and old age. *Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation, 18*(10), 1341-1355.

Muthen, L. (2008a). How to get more fit index. Retrieved from http://www.statmodel.com/discussion/messages/9/3008.html?1236817441. Last accessed 15/4/15

Muthen, B. (2008b). Latent variable mixture modeling: different growth curve models in GMM. Retrieved from http://www.statmodel.com/discussion/messages/13/3406.html?1224686032. Last accessed 15/4/15

Muthen, B., & Asparouhov, T. (2009). In Fitzmaurice, G Davidian, M Verbeke, G Molenberghs,G. (Ed.), *Growth mixture modeling: Analysis with non-gaussian random effects*

Muthen, L., & Muthen, B. (2010). *Mplus user's guide* (6th ed.). Los Angeles.

Myers, H., Nicholls, E., Handy, J., Peat, G., Thomas, E., Duncan, R. et al. (2007). The clinical assessment study of the hand (CAS-HA): A prospective study of musculoskeletal hand problems in the general population. *BMC Musculoskeletal Disorders, 8*, 85.

Myers, H., Nicholls, E., & Dziedzic, K. (2008). The effect of adaptation behaviour on hand function in older adults with self-reported hand problems. *Rheumatology, 47*, II23-II24.

Myers, H., Thomas, E., Hay, E. M., & Dziedzic, K. S. (2011). Hand assessment in older adults with musculoskeletal hand problems: A reliability study. *BMC Musculoskeletal Disorders, 12:3*.

Nagin, D. S., & Odgers, C. L. (2010). Group-based trajectory modeling in clinical research. *Annual Review of Clinical Psychology, 6*, 109-138.

National Collaborating Centre for Chronic Conditions. (2008). *Osteoarthritis: National clinical guideline for care and management in adults.* London: Royal College of Physicians.

Nicholls, E., van der Windt, D. A., Jordan, J. L., Dziedzic, K. S., & Thomas, E. (2012). Factors associated with the severity and progression of self-reported hand pain and functional difficulty in community-dwelling older adults: A systematic review. *Musculoskeletal Care, 10*(1), 51-62.

Nicholls, E., Thomas, E., van der Windt, D. A., Croft, P. R., & Peat, G. (2014). Pain trajectory groups in persons with, or at high risk of, knee osteoarthritis: Findings from the Knee Clinical Assessment Study and the Osteoarthritis Initiative. *Osteoarthritis and Cartilage, 22(12), 2041-50.*

Niu, J., Zhang, Y., LaValley, M., Chaisson, C. E., Aliabadi, P., & Felson, D. T. (2003). Symmetry and clustering of symptomatic hand osteoarthritis in elderly men and women: The Framingham study. *Rheumatology (Oxford, England), 42*(2), 343-348.

Nylund, K. L., Asparouhov, T., & Muthen, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A monte carlo simulation study. *Structural Equation Modeling: A Multidisciplinary Journal, 14*(4), 535-569.

Office for National Statistics (ONS). (2002). *The national statistics socioeconomic classification user manual.* (No. Version 1). London: Office for National Statistics.

Office for National Statistics (ONS). (2009). *General lifestyle survey.* Retrieved from: http://www.ons.gov.uk/ons/rel/ghs/general-lifestyle-survey/index.html. Last accessed 15/4/15

Office for National Statistics (ONS). (2012) Ageing - fastest increase in the 'oldest old'. Retrieved from: http://www.ons.gov.uk/ons/dcp171776_258607.pdf. Last accessed 15/4/15

Ottenbacher, K. J., Johnson, M. B., & Hojem, M. (1988). The significance of clinical change and clinical change of significance: Issues and methods. *American Journal of Occupational Therapy, 42*(3), 156-163.

Palmer, K., Walker-Bone, K., Linaker, C., Reading, I., Kellingray, S., Coggon, D., & Cooper, C. (2000). The Southampton examination schedule for the diagnosis of musculoskeletal disorders of the upper limb. *Annals of the Rheumatic Diseases, 59*(1), 5-11.

Palmer, K., (2003). Regional musculoskeletal conditions: Pain in the forearm, wrist and hand. *Best Practice & Research. Clinical Rheumatology, 17*(1), 113-135.

Papageorgiou, A. C., Croft, P. R., Ferry, S., Jayson, M. I., & Silman, A. J. (1995). Estimating the prevalence of low back pain in the general population. Evidence from the south Manchester back pain survey. *Spine, 20*(17), 1889-1894.

Parsons, S., Ingram, M., Clarke-Cornwell, A., & Symmons, D. (2011). *A heavy burden. The occurrence and impact of musculoskeletal conditions in the United Kingdom today.* Retrieved from: https://www.escholar.manchester.ac.uk/uk-ac-man-scw:123774. Last accessed 15/4/15

Payne, B. R., Gross, A. L., Parisi, J. M., Sisco, S. M., Stine-Morrow, E., Marsiske, M., & Rebok, G. W. (2014). Modelling longitudinal changes in older adults' memory for spoken discourse: Findings from the ACTIVE cohort. *Memory* 22(8): 990-1001.

Peat, G., Thomas, E., Handy, J., Wood, L., Dziedzic, K., Myers, H., et al. (2004). The Knee Clinical Assessment Study - CAS(K). A prospective study of knee pain and knee osteoarthritis in the general population. *BMC Musculoskeletal Disorders, 5*, 4.

Pham, T., van der Heijde, D., Altman, R. D., Anderson, J. J., Bellamy, N., Hochberg, M. et al. (2004). OMERACT-OARSI initiative: Osteoarthritis Research Society International set of responder criteria for osteoarthritis clinical trials revisited. *Osteoarthritis and Cartilage, 12*(5), 389-399.

Preacher, K. J., Wichman, A. L., MacCallum, R. C., & Briggs, N. E. (2008). Latent growth curve modeling. *Thousand Oaks*, CA: Sage Publications.

Rabe-Hesketh, S., & Skrondal, A. (2008). *Multilevel and longitudinal modeling using STATA* (Second ed.). USA: Stata Press.

Raftery A., E. (1995). Bayesian model selection in social research. In P. Marsden (Ed.), *Sociological methodology* (pp. 111). Cambridge, MA: Blackwell Sci.

Rasbash, J., Steele, F., Browne, W. J., & Goldstein, H. (2012). *A user's guide to MLwiN, v2.26. Centre for multilevel modelling, University of Bristol*

Rasch G. Probabilistic model for some intelligence and attainment tests. Chigaco: The University of Chicago Press; 1960.

Riddle, D. L., Moxley, G., & Dumenci, L. (2013). Associations between statin use and changes in pain, function and structural progression: A longitudinal study of persons with knee osteoarthritis. *Annals of the Rheumatic Diseases, 72*(2), 196-203.

Rigdon, E. (1997). Non positive definite matrices - causes and cures. Retrieved from http://www2.gsu.edu/~mkteer/npdmatri.html. Last accessed 15/4/15

Riley, R. D., Hayden, J. A., Steyerberg, E. W., Moons, K. G. M., Abrams, K., Kyzas, P. A. et al. (2013). Prognosis research strategy (PROGRESS) 2: Prognostic factor research. *PLoS Med, 10*(2), e1001380.

Royston, P., Moons, K. G., Altman, D. G., & Vergouwe, Y. (2009). Prognosis and prognostic research: Developing a prognostic model. *BMJ (Clinical Research Ed.), 338*, b604. doi:10.1136/bmj.b604

Schwarz,G. (1978). Estimating the dimension of a model. *The Annals of Statistics,* 6(2), 461-464

Sclove, s. (1987). Application of model-selection criteria to some problems in multivariate-analysis. *Psychometrika, 52*(3), 333-343.

Scottish Intercollegiate Guidelines Network. Retrieved from http://www.sign.ac.uk/methodology/filters.html. Last accessed 15/4/15

Silman, A. (1995). *Epidemiological studies: A practical guide* (1st ed.). Great Britain: Cambridge University Press.

Sim, J., Jordan, K., Lewis, M., Hill, J., Hay, E. M., & Dziedzic, K. (2006). Sensitivity to change and internal consistency of the Northwick Park neck pain questionnaire and derivation of a minimal clinically important difference. *The Clinical Journal of Pain, 22*(9), 820-826.

Simpson, C. (2002). *Hand assessment: A clinical guide for therapists* (1st ed.). Wiltshire: APS Publishing.

Singer, J., & Willet, J. (2003). *Applied longitudinal data analysis* (First ed.). USA: Oxford University Press.

Slatkowsky-Christensen, B., Kvien, T. K., & Bellamy, N. (2005). Performance of the Norwegian version of AUSCAN - a disease-specific measure of hand osteoarthritis. *Osteoarthritis and Cartilage, 13*(7), 561-567.

Smith, H. B. (1973). Smith hand function evaluation. *The American Journal of Occupational Therapy: Official Publication of the American Occupational Therapy Association, 27*(5), 244-251.

Spies-Dorgelo, M. N., van der Windt, D. A., van der Horst, H. E., Prins, A. P., & Stalman, W. A. (2007). Hand and wrist problems in general practice - patient characteristics and factors related to symptom severity. *Rheumatology (Oxford, England), 46*(11), 1723-1728.

Spies-Dorgelo, M. N., van der Windt, D. A., Prins, A. P., Dziedzic, K. S., & van der Horst, H. E. (2008). Clinical course and prognosis of hand and wrist problems in primary care. *Arthritis and Rheumatism, 59*(9), 1349-1357.

StataCorp. (2013a). *Stata 13 Longitudinal data/panel data reference manual*. College Station, TX: Stata Press.

StataCorp. (2013b). *Stata Statistical Software: Release 13*. College Station, TX: StataCorp LP.

StataCorp. (2013c). *Stata 13 Multiple imputation reference manual*. College Station, TX: Stata Press.

StataCorp. (2013d). *Stata 13 Base reference manual*. College Station, TX: Stata Press. http://www.stata.com/manuals13/rbrier.pdf. Last accessed 15/4/15

StataCorp. (2013e). *Stata 13 Base reference manual*. College Station, TX: Stata Press. Retrieved from http://www.stata.com/manuals13/rbootstrap.pdf#rbootstrap. Last accessed 15/4/15

Steiger, J. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research, 25*(2), 173.

Steyerberg, E. W., Eijkemans, M. J., Harrell, F. E.,Jr, & Habbema, J. D. (2000). Prognostic modelling with logistic regression analysis: A comparison of selection and estimation methods in small data sets. *Statistics in Medicine, 19*(8), 1059-1079.

Steyerberg, E. W., Harrell, F. E.,Jr, Borsboom, G. J., Eijkemans, M. J., Vergouwe, Y., & Habbema, J. D. (2001). Internal validation of predictive models: Efficiency of some procedures for logistic regression analysis. *Journal of Clinical Epidemiology, 54*(8), 774-781.

Steyerberg, E. W., Vickers, A. J., Cook, N. R., Gerds, T., Gonen, M., Obuchowski, N. et al. (2010). Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology (Cambridge, Mass.), 21*(1), 128-138.

Steyerberg, E. W., Moons, K. G. M., van, d. W., Hayden, J. A., Perel, P., Schroter, S. et al. (2013). Prognosis research strategy (PROGRESS) 3: Prognostic model research. *PLoS Med, 10*(2), e1001381.

Stratford, P. W., & Riddle, D. L. (2005). Assessing sensitivity to change: Choosing the appropriate change coefficient. *Health and Quality of Life Outcomes, 3*, 23-23.

Streiner D, N. G. (2003). *Health measurement scales: A practical guide to their development and use*. United States: Oxford University Press.

Stukstette, M. J., Dekker, J., den Broeder, A. A., Westeneng, J. M., Bijlsma, J. W., & van den Ende, C. H. (2013). No evidence for the effectiveness of a multidisciplinary group based treatment program in patients with osteoarthritis of hands on the short term; results of a randomized controlled trial. *Osteoarthritis and Cartilage, 21*(7), 901-910.

Swain, N., & Johnson, M. (2014). Chronic pain in New Zealand: A community sample. *The New Zealand Medical Journal, 127*(1388), 21-30.

Tas, U., Verhagen, A. P., Bierma-Zeinstra, S. M., Odding, E., & Koes, B. W. (2007). Prognostic factors of disability in older people: A systematic review. *The British Journal of General Practice: The Journal of the Royal College of General Practitioners, 57*(537), 319-323.

Terwee, C. B., Bot, S. D., de Boer, M. R., van der Windt, D. A., Knol, D. L., Dekker, J. et al. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology, 60*(1), 34-42.

Terwee, C. B., Roorda, L. D., Knol, D. L., De Boer, M., & de Vet, H. (2009). Linking measurement error to minimal important change of patient-reported outcomes. *J Clin Epidemiol, 62*, 1062-1067.

Terwee, C. B., Roorda, L. D., Dekker, J., Bierma-Zeinstra, S. M., Peat, G., Jordan, K. P. et al. (2010). Mind the MIC: Large variation among populations and methods. *Journal of Clinical Epidemiology, 63*(5), 524-534.

Thomas, E., Peat, G., Harris, L., Wilkie, R., & Croft, P. R. (2004a). The prevalence of pain and pain interference in a general population of older adults: Cross-sectional findings from the North Staffordshire osteoarthritis project (NorStOP). *Pain, 110*(1-2), 361-368.

Thomas, E., Wilkie, R., Peat, G., Hill, S., Dziedzic, K., & Croft, P. (2004b). The North Staffordshire osteoarthritis project - NorStOP: Prospective, 3-year study of the epidemiology and management of clinical osteoarthritis in a general population of older adults. *BMC Musculoskeletal Disorders, 5*, 2.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological), 58*(1), 267-288.

Tofighi, D., & Enders, C. (2008). Identifying the correct number of classes in growth mixture models. *Advances in latent variable mixture models* (pp. 317) Information Age Publishing.

Tucker, L., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika, 38*(1), 1-10.

Twisk, J. (2003). *Applied longitudinal data analysis for epidemiology*. United States of America: Cambridge University Press.

Twisk, J., de Boer, M., de Vente, W., & Heymans, M. (2013). Multiple imputation of missing values was not necessary before performing a longitudinal mixed-model analysis. *Journal of Clinical Epidemiology, 66*(9), 1022-1028.

University of Strathclyde Humanities and Social Sciences. Goodness of fit measures. Retrieved from http://www.strath.ac.uk/aer/materials/5furtherquantitativeresearchdesignandanalysis/unit6/goodnessoffitmeasures/. Last accessed 15/4/15

Urwin, M., Symmons, D., Allison, T., Brammah, T., Busby, H., Roxby, M. et al. (1998). Estimating the burden of musculoskeletal disorders in the community: The comparative prevalence of symptoms at different anatomical sites, and the relation to social deprivation. *Annals of the Rheumatic Diseases, 57*(11), 649-655.

Van Calster, B., Vergouwe, Y., Looman, C. W., Van Belle, V., Timmerman, D., & Steyerberg, E. W. (2012a). Assessing the discriminative ability of risk models for more than two outcome categories. *European Journal of Epidemiology, 27*(10), 761-770.

Van Calster, B., Van Belle, V., Vergouwe, Y., Timmerman, D., Van Huffel, S., & Steyerberg, E. W. (2012b). Extending the c-statistic to nominal polytomous outcomes: The polytomous discrimination index. *Statistics in Medicine, 31*(23), 2610-2626.

van der Windt, D. A., van der Heijden, G. J., de Winter, A. F., Koes, B. W., Deville, W., & Bouter, L. M. (1998). The responsiveness of the shoulder disability questionnaire. *Annals of the Rheumatic Diseases, 57*(2), 82-87.

van Dijk, G. M., Dekker, J., Veenhof, C., van den Ende, C. H., & Carpa Study Group. (2006). Course of functional status and pain in osteoarthritis of the hip or knee: A systematic review of the literature. *Arthritis and Rheumatism, 55*(5), 779-785.

van Kampen, D., Willems, W., van Beers, L., Castelein, R., Scholtes, V., & Terwee, C. (2013). Determination and comparison of the smallest detectable change (SDC) and the minimal important change (MIC) of four-shoulder patient-reported outcome measures (PROMs). *Journal of Orthopaedic Surgery and Research, 8*(1), 40.

Verkleij, S. P., Hoekstra, T., Rozendaal, R. M., Waarsing, J. H., Koes, B. W., Luijsterburg, P. A., & Bierma-Zeinstra, S. M. (2012). Defining discriminative pain trajectories in hip osteoarthritis over a 2-year time period. *Annals of the Rheumatic Diseases, 71*(9), 1517-1523.

Vermunt JK. Latent Class Modeling with Covariates: Two Improved Three-Step Approaches. Political Analysis. 2010;18(4):450-69.

Vermunt J., K., Magidson J. Latent Gold 4.0 user's guide. Belmont, MA: Statistical Innovations Inc, 2013.

Vos, T., Flaxman, A., Naghavi, M., Lozano, R. et al (2012). Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990-2010: a systematic analysis for the global burden of disease study 2010. *The Lancet, 380*(9859), 2163-2196.

Walker-Bone, K. E., Palmer, K. T., Reading, I., & Cooper, C. (2003). Soft-tissue rheumatic disorders of the neck and upper limb: Prevalence and risk factors. *Seminars in Arthritis and Rheumatism, 33*(3), 185-203.

Walsh, W. B., & Betz, N. E. (1990). *Tests and assessment* (Second edition ed.). United States of America: Prentice-Hall Inc.

Walter, S., & Tiemeier, H. (2009). Variable selection: Current practice in epidemiological studies. *European Journal of Epidemiology, 24*(12), 733-736.

Wang, M., & Bodner, T. E. (2007). Growth mixture modeling. *Organizational Research Methods, 10*(4), 635-656.

Ware, J., Kosinski, M., & Keller, S. D. (1996). A 12-item short-form health survey: Construction of scales and preliminary tests of reliability and validity. *Medical Care, 34*(3), 220-233.

Wilkie, R., Peat, G., Thomas, E., & Croft, P. (2007). Factors associated with restricted mobility outside the home in community-dwelling adults aged fifty years and older with knee pain: An example of use of the international classification of functioning to investigate participation restriction. *Arthritis and Rheumatism, 57*(8), 1381-1389.

Wittoek, R., Cruyssen, B. V., Maheu, E., & Verbruggen, G. (2009). Cross-cultural adaptation of the Dutch version of the functional index for hand osteoarthritis (FIHOA) and a study on its construct validity. *Osteoarthritis and Cartilage, 17*(5), 607-612.

Woodward M. (1999). Fundamental issues. *Epidemiology: Study design and data analysis* (First ed., pp. 18). USA: Chapman and Hall.

Woolf, A. D., & Pfleger, B. (2003). *Burden of major musculoskeletal conditions.* (No. 81 (9)). Bulletin of the World Health Organization.

Wright, A., Hannon, J., Hegedus, E. J., & Kavchak, A. E. (2012). Clinimetrics corner: A closer look at the minimal clinically important difference (MCID). *The Journal of Manual & Manipulative Therapy, 20*(3), 160-166.

Wu, J., Witkiewitz, K., McMahon, R. J., Dodge, K. A., & Conduct Problems Prevention Research Group. (2010). A parallel process growth mixture model of conduct problems and substance use with risky sexual behavior. *Drug and Alcohol Dependence, 111*(3), 207-214.

Xie, H., McHugo, G. J., He, X., & Drake, R. E. (2010). Using the group-based dual trajectory model to analyze two related longitudinal outcomes. *Journal of Drug Issues, 40*(1), 45-61.

Zigmond, A. S., & Snaith, R. P. (1983). The hospital anxiety and depression scale. *Acta Psychiatrica Scandinavica, 67*(6), 361-370.

Zyluk, A., & Puchalski, P. (2013). A comparison of the results of carpal tunnel release in patients in different age groups. *Neurologia i Neurochirurgia Polska, 47*(3), 241-246.

Appendix 1: Ethical Approval letters

CAS-HA baseline

Shropshire and Staffordshire NHS
Health Authority

Local Research Ethics Committee
Heron House
120 Grove Road
Fenton
Stoke-on-Trent
ST4 4LX

Tel.: 01782 298013/298004

Fax: 01782 298298

NE/BAC/JEC

22 January 2004

Private and Confidential

Dr. G. Peat
Research Fellow
Primary Care Sciences Research Centre
Keele University
Keele
Staffordshire
ST5 5BG

Dear Dr. Peat

Project 1430
The clinical epidemiology of signs and symptoms of joint pain and osteoarthritis: a population based, cross sectional study of knee and hand pain

Thank you for your letter asking for a protocol amendment changing the documentation as this study moves from Phase I (relating to knee pain) to Phase II (relating to hand pain) to reflect this change.

This was approved at the recent meeting of the North Staffordshire Local Research Ethics Committee on 14 January 2004.

Yours sincerely

J E Clarke

Dr. N. Edelstyn pp
Vice Chair
Research Ethics Committee

CAS-HA 18-month follow-up

NHS

North Staffordshire Local Research Ethics Committee
Heron House
120 Grove Road
Fenton
Stoke-on-Trent
Staffordshire
ST4 4LX

Telephone: 01782 298013
Facsimile: 01782 298298

13 October 2005

Dr Krysia Dziedzic
arc Senior Lecturer in Physiotherapy
Primary Care Sciences Research Centre
Keele University
Keele
Staffordshire
ST5 5BG

Dear Dr Dziedzic

Full title of study: **The course of hand pain and hand problems in community-dwelling adults aged 50 years and over in North Staffordshire; 18-month follow-up of the Clinical Assessment Study of the Hand (CASHA).**

REC reference number: **05/Q2604/89**

Thank you for your letter of 27 September 2005, responding to the Committee's request for further information on the above research and submitting revised documentation.

The further information was considered at the meeting of the Sub-Committee of the REC held on 12 October 2005. A list of the members who were present at the meeting is attached.

Confirmation of ethical opinion

On behalf of the Committee, I am pleased to confirm a favourable ethical opinion for the above research on the basis described in the application form, protocol and supporting documentation as revised.

Ethical review of research sites

The favourable opinion applies to the research sites listed on the attached form.

Conditions of approval

The favourable opinion is given provided that you comply with the conditions set out in the attached document. You are advised to study the conditions carefully.

Approved documents

The final list of documents reviewed and approved by the Committee is as follows:

Document	*Version*	*Date*

Application		17 August 2005
Investigator CV		15 August 2005
Protocol	1	15 August 2005
Letter from Sponsor		12 August 2005
Peer Review		14 June 2005
Questionnaire		
Questionnaire Phone		
Questionnaire Postal		
Letter of invitation to participant Phase 4	1	15 August 2005
Letter of invitation to participant	1	15 August 2005
GP/Consultant Information Sheets	1	15 August 2005
Participant Information Sheet	3	19 September 2005
Participant Consent Form Baseline study (First Stage)	1	15 August 2005
Participant Consent Form Baseline questionnaire (Second Stage)	1	15 August 2005
Participant Consent Form Baseline clinical assessment	1	15 August 2005
Response to Request for Further Information		27 September 2005
Patient Proforma	1	15 August 2005
Letter of invitation to repeat mailing	1	15 August 2005
Flow Chart	1	15 August 2005
Schematic diagram of the MRC-funded NORSTOP cohorts		15 August 2005
Letter to practice from Rhian Hughes	1	15 August 2005
Other		
Grant Letter		16 August 2005
Postcard	1	15 August 2005
Flowchart	1	15 August 2005
Thank you letter	1	15 August 2005

Research governance approval

The study should not commence at any NHS site until the local Principal Investigator has obtained final research governance approval from the R&D Department for the relevant NHS care organisation.

Statement of compliance

The Committee is constituted in accordance with the Governance Arrangements for Research Ethics Committees (July 2001) and complies fully with the Standard Operating Procedures for Research Ethics Committees in the UK.

05/Q2604/89	**Please quote this number on all correspondence**

With the Committee's best wishes for the success of this project

Yours sincerely

J E Clarke

Miss Nicola Brooks
Chair

Email: Janet.Clarke@northstaffs.nhs.uk

CAS-HA 3-year follow-up

Hereford & Worcester Local Research Ethics Committee
Isaac Maddox House
Shrub Hill Road
Worcester
WR4 9RW

Telephone: 01905 760091
Facsimile: 01905 733249

12 March 2007

Dr Krysia Dziedzic
arc Senior Lecturer in Physiotherapy
Primary Care Musculoskeletal Research Centre
Keele University
Keele
Staffordshire ST5 5BG

Dear Dr Dziedzic

Full title of study: **The course of hand pain and hand problems in community-dwelling adults aged 50 years and over in North Staffordshire; 3-year follow-up of North Staffordshire Osteoarthritis Project 3 (NorStOP3) and the Clinical Assessment Study of the Hand (CAS-HA)**

REC reference number: **06/Q2801/90**

Thank you for your letter of the 1 March 2007, received on the 6 March.

The Chairman, acting on delegated authority of the Research Ethics Committee, has now had an opportunity to review your letter of the 1 March 2007 together with attachments.

Ethical opinion

Favourable opinion following satisfactory response

The Chairman has now given a favourable opinion to the above research following a satisfactory response on the basis described in the application form, protocol and supporting documentation and amendments.

Conditions of approval

The favourable opinion is given provided that you comply with the conditions set out in the attached document. You are advised to study the conditions carefully.

Approved documents

The documents reviewed and approved at the meeting were:

Document	*Version*	*Date*
Application	1	
Covering Letter		01 March 2007
Appendix 2h NORSTOP3 3yr follow up further pain questionnaire	2	12 January 2007
Appendix 2g NORSTOP3 3yr follow up repeat covering letter	2	12 January 2007
Appendix 2f NORSTOP3 3yr follow up reminder postcard	2	12 January 2007
Appendix 2e NORSTOP3 3 yr follow up participant information sheetr	2	12 January 2007

Appendix 2f NORSTOP3 3yr follow up health questionnaire covering letter	2	12 January 2007
Appendix 3f CAS-HA 3yr follow up reminder postcard	2	12 January 2007
Appendix 3e CAS-HA 3yr follow up participant information sheet	2	12 January 2007
Appendix 3c CAS-HA 3yr follow up health questionnaire	2	12 January 2007
Appendix 3b CAS-HA 3yr follow up thank you letter to participants	2	12 January 2007
Appendix 2i NORSTOP3 3yr follow up further pain questionnaire covering letter	2	12 January 2007
Appendix 4b GP proforma for traced participants	2	12 January 2007
Appendix 3l CAS-HA 3yr follow up further pain questionnaire	3	28 February 2007
App 2c NORSTOP3 3 yr follow-up health questionnaire		
App 2b NORSTOP3 & CAS-HA 3 yr follow-up GP letter		
App 2a NORSTOP3: flowchart 3 yr follow-up procedures		
App 2 supporting documentation for NORSTOP3 3yr follow-up		
Appendix 1 schematic diagramme of MRC-funded NORSTOP cohorts		
App 2h - NORSTOP3 3yr follow-up further pain questionnaire		
App 2g NORSTOP3 3 yr follow-up repeat covering letter		
App 2f NORSTOP3 3hr follow-up reminder postcard		
App 2e NORSTOP3 3hr follow- participant information sheet		
App 2d - NORSTOP3 3yr follow-up health questionnaire covering letter		
App 3c CAS-HA 3yr follow up health questionnaire		
App 3b CAS-HA: 3yr follow up thank you letter to participants		
App 3a CAS-HA: flowchart of 3yr follow up procedures		
App 3 supporting documentation for CAS-HA 3yr follow up		
App 2i NORSTOP 3 3yr follow up further pain questionnaire covering letter		
App 3h CAS-HA 3yr follow-up minimum data collection protocol		
App 3g CAS-HA 3yr follow up repeat covering letter		
App 3f CAS-HA: 3yr follow up reminder postcard		
App 3e CAS-HA 3yr follow up participant information sheet		
App 3d CAS-HA: 3 yr follow up health questionnaire covering letter		
App 3m CAS-HA 3yr follow up further pain questionnaire covering letter		
App 3l CAS-HA 3yr follow-up further pain questionnaire		
App 3k CAS-HA 3yr follow up minimum data collection - postal		
App 3j CAS-HA 3yr follow up minimum data collection - postal covering letter		
App 3i CAS-HA 3yr follow up minimum data collection - phone		
CVs for 3 supervisors		
CV for CI		
Research protocol		
App 4b GP proforma for traced participants		
App 4a letter to new GPs for traced participants		
Referees' or other Scientific Critique report x 3		
Letter from funder		
Letter from sponsor and statement of indemnity arrangements		
Appendix 2c - NORSTOP3 3yr follow up health questionnaire	2	12 January 2007

Follow up thank you letter to participants	2	12 January 2007
Health questionnaire		12 January 2007
Questionnaire hand, hip, knee & foot problems		12 January 2007

R&D approval

The study should not commence at any NHS site until the local Principal Investigator has obtained final approval from the R&D office for the relevant NHS care organisation.

Membership of the Committee

The members of the Ethics Committee who were present at the meeting are listed on the attached sheet.

Statement of compliance

The Committee is constituted in accordance with the Governance Arrangements for Research Ethics Committees (July 2001) and complies fully with the Standard Operating Procedures for Research Ethics Committees in the UK.

06/Q2801/90 Please quote this number on all correspondence

With the Committee's best wishes for the success of this project

Yours sincerely

Mr John Barnett
Chair

Email: kath.garrad@sworcs-pct.nhs.uk

Enclosures: *Standard approval conditions [SL-AC1 for CTIMPs, SL-AC2 for other studies]*
Site approval form (SF1)

National Research Ethics Service NHS

28th April 2009

Warwickshire Research Ethics Committee

Dr K. Dziedzic
Arthritis Research Campaign National Primary Care Centre
Primary Care Sciences
Keele University
Keele
Staffs

Lewes House
George Eliot Hospital
College Street
Nuneaton
Warwickshire
CV10 7DJ

Tel: 02476 865244
Fax: 02476 865264
pat.horwell@geh.nhs.uk

Dear Dr Dziedzic

Study title: **The course of hand pain and hand problems in community-dwelling adults aged 50 years and over in North Staffordshire; 3-year follow-up of North Staffordshire Osteoarthritis Project 3 (NorStOP3) and the Clinical Assessment Study of the Hand (CAS-HA)**
REC reference: **06/Q2801/90**
Amendment number: **Amendment 2**
Amendment date: **03 March 2009**

The above amendment was reviewed at the meeting of the Committee held on 29th April 2009.

Ethical opinion

The members of the Committee taking part in the review gave a favourable ethical opinion of the amendment on the basis described in the notice of amendment form and supporting documentation.

Approved documents

The documents reviewed and approved at the meeting were:

Document	Version	Date
Protocol	Amendment 2 V 5	20 October 2008
Letter to new GPS for traced participants	Amendment 2 v2	20 October 2008
Four and a half year screening follow up	Amendment 2 v2	23 October 2008
Present findings	Amendment 2 v2	20 October 2008
Letter of thanks to participants	Amendment 2 v2	20 November 2008
Schematic diagram	Amendment 2 v3	20 October 2008
Notice of Substantial Amendment (non-CTIMPs)	Amendment 2	03 March 2009
Participant Information Sheet	Amendment 2 v4	20 October 2008
letter of invitation for repeat mailing	Amendment	20 October 2008

	2 v4	
Reminder postcard to participants	Amendment 2 v4	20 October 2008
Questionnaire (four and a half years)	Amendment 2	20 October 2008
Flowchart four and a half year follow up procedures	Amendment 2 v3	20 October 2008
Patient Proforma	Amendment 2 v2	20 October 2008
Letter of invitation to participant	Amendment 2 v2	20 October 2008
Questionnaire MDC phone	Amendment 2 v2	20 October 2008
Minimum data set collection	Amendment 2 v2	20 October 2008
Questionnaire MDC postal	Amendment 2 v2	20 October 2008
Postal covering letter	Amendment 2 v2	20 October 2008

Membership of the Committee

The members of the Committee who took part in the review are listed on the attached sheet.

R&D approval

All investigators and research collaborators in the NHS should notify the R&D office for the relevant NHS care organisation of this amendment and check whether it affects R&D approval of the research.

Statement of compliance

The Committee is constituted in accordance with the Governance Arrangements for Research Ethics Committees (July 2001) and complies fully with the Standard Operating Procedures for Research Ethics Committees in the UK.

REC reference number: 06/Q2801/90 **Please quote this number on all correspondence**

Yours sincerely

Pat Horwell
Committee Co-ordinator

Enclosures: List of names and professions of members who took part in the review

Copy to: *R&D office*

Helen Myers

Warwickshire Research Ethics Committee is an advisory committee to West Midlands Strategic Health Authority
The National Research Ethics Service (NRES) represents the NRES Directorate within the National Patient Safety Agency and Research Ethics Committees in England

CAS-HA 6-year follow-up

NHS
National Research Ethics Service
NRES Committee West Midlands - Solihull
REC Offices
Prospect House
Fishing Line Road
Enfield
Redditch
B97 6EW
Telephone: 01527 582532
Facsimile: 01527 582540

Date: 20 July 2011

Dr George Peat
Senior Lecturer in Clinical Epidemiology
Arthritis Research UK Primary Care Centre,
Primary Care Sciences,
Keele University
Keele
Staffordshire ST5 5BG

Dear Dr Peat

Study title: **The course of hand pain, hand problems and hand osteoarthritis in community-dwelling adults aged 50 years and over: The 6-year follow-up of the Clinical Assessment Study of the Hand (CAS-HA).**

REC reference: **11/WM/0196**

The Research Ethics Committee reviewed the above application at the meeting held on 13 July 2011. Thank you for attending to discuss the study.

Discussion at the meeting

1. The time to consent was explained. This is a follow up study so participants are already in the cohort and familiar with the research. The PIS will be sent and those that express an interest will be contacted and invited to attend. The research team will ensure that participants are aware that they can consent to aspects of the study. Non responders get sent a reminder. If there is no response they are sent a health questionnaire. If there is still no response they will receive a reminder postcard. No response to this and they are sent a reminder questionnaire. There is no further chase up. Those that said they do not wish to be contacted again in the previous study will not be contacted at all. Invitations will only be sent to those who gave written consent to contact for follow up in the previous study.

2. You explained that X-rays taken are not being reported on clinically. They are being read for research purposes.

3. The committee commented that GPs are to be paid £50 per 500 participants but that this is not in the PIS.

4. The committee said the reason for X-raying hands and knees is only explained in a further paragraph lower down from the statement saying that

both will be X-rayed. You explained that all participants will have had hands and knees X-rayed for the previous study and will have had the reason for this explained.

5. The committee noted that permission to inform the GP of participation is not being sought. You explained that you are not feeding back any findings this is just a courtesy note to let the GP know that their patient attended the clinic.

6. Consent to link GP records and clinic data was sought in the first part of the study.

7. Appendix 6 – you confirmed that you now have permission to access the NHS tracing service.

Ethical opinion

The members of the Committee present gave a favourable ethical opinion of the above research on the basis described in the application form, protocol and supporting documentation, subject to the conditions specified below.

Ethical review of research sites

NHS Sites

The favourable opinion applies to all NHS sites taking part in the study, subject to management permission being obtained from the NHS/HSC R&D office prior to the start of the study (see "Conditions of the favourable opinion" below).

Conditions of the favourable opinion

The favourable opinion is subject to the following conditions being met prior to the start of the study.

Management permission or approval must be obtained from each host organisation prior to the start of the study at the site concerned.

Management permission ("R&D approval") should be sought from all NHS organisations involved in the study in accordance with NHS research governance arrangements.

Guidance on applying for NHS permission for research is available in the Integrated Research Application System or at http://www.rdforum.nhs.uk.

Where a NHS organisation's role in the study is limited to identifying and referring potential participants to research sites ("participant identification centre"), guidance should be sought from the R&D office on the information it requires to give permission for this activity.

For non-NHS sites, site management permission should be obtained in accordance with the procedures of the relevant host organisation.

Sponsors are not required to notify the Committee of approvals from host organisations

It is responsibility of the sponsor to ensure that all the conditions are complied with before the start of the study or its initiation at a particular site (as applicable).

Approved documents

The documents reviewed and approved at the meeting were:

Document	Version	Date
Covering Letter		07 June 2011
Evidence of insurance or indemnity		28July 2010
GP/Consultant Information Sheets	1	15 April 2011
Investigator CV		21 April 2011
Letter from Sponsor		21 April 2011
Letter of invitation to participant	1	15 April 2011
Other: CV Academic Supervisor Prof van de Windt		15 April 2011
Other: CV Student E Nicholls		20 May 2011
Other: Reminder letter of invitation	1	15 April 2011
Other: Health questionnaire cover letter	1	15 April 2011
Other: DNA invitation letter	1	15 April 2011
Other: Reminder postcard	1	15 April 2011
Other: Health questionnaire reminder invitation letter	1	15 April 2011
Other: Minimum data collection (postal) letter of invitation	1	15 April 2011
Other: CAS-HA 6 years follow up Regional pains questionnaire cover letter	1	15 April 2011
Other: Reminder Regional pains questionnaire cover letter	1	15 April 2011
Other: GP screen letter	1	15 April 2011
Other: GP letter of attendance	1	15 April 2011
Other: GP traced participant cover letter	1	15 April 2011
Other: GP traced participant permission sheet	1	15 April 2011
Other: GP Red flag letter	1	15 April 2011
Other: GP Clinic abnormality letter	1	15 April 2011
Other: GP X-ray abnormality letter	1	15 April 2011
Other: GP X-ray report cover letter	1	15 April 2011
Other: Letter from funder		02 July 2008
Other: Flowcharts of CAS-HA 6 year follow up procedures	1	15 April 2011
Other: Minimum data collection protocol	1	15 April 2011
Other: Tracing protocol	1	15 April 2011
Other: Red flag protocol	1	15 April 2011
Other: Summary of findings so far	1	15 April 2011
Other: Letter of thanks	1	15 April 2011
Other: Confirmation of appointment letter	1	15 April 2011
Other: CAS-HA 6 year follow up Clinic Assessment form	1	15 April 2011
Other: Protocol for taking research x-ray reports	1	15 April 2011
Other: Protocol for significant abnormality	1	15 April 2011
Other: Summary CV for supervisor (student research) Dr Elaine Thomas		06 May 2011
Other: Letter from lead Radiation Protection Advisor		31 May 2011
Other: Letter from lead Clinical Radiation Expert		31 May 2011
Other: Schematic diagram of the North Staffordshire Osteoarthritis Project cohorts	1	15 April 2011

Participant Consent Form: Clinic consent form	1	15 April 2011
Participant Information Sheet: Clinic participant information sheet	1	15 April 2011
Participant Information Sheet: CAS-HA 6 years follow up Health questionnaire participant information sheet	1	15 April 2011
Protocol	1	15 April 2011
Questionnaire: CAS-HA 6 year follow up Health Questionnaire	1	15 April 2011
Questionnaire: CAS-HA 6 year follow up Regional pains questionnaire	1	15 April 2011
Questionnaire: Minimum data collection (postal)	1	15 April 2011
Questionnaire: Minimum data collection (telephone)	1	15 April 2011
REC application	3.1	07 June 2011

Membership of the Committee

The members of the Ethics Committee who were present at the meeting are listed on the attached sheet.

Statement of compliance

The Committee is constituted in accordance with the Governance Arrangements for Research Ethics Committees (July 2001) and complies fully with the Standard Operating Procedures for Research Ethics Committees in the UK.

After ethical review

Reporting requirements

The attached document "After ethical review – guidance for researchers" gives detailed guidance on reporting requirements for studies with a favourable opinion, including:

- Notifying substantial amendments
- Adding new sites and investigators
- Notification of serious breaches of the protocol
- Progress and safety reports
- Notifying the end of the study

The NRES website also provides guidance on these topics, which is updated in the light of changes in reporting requirements or procedures.

Feedback

You are invited to give your view of the service that you have received from the National Research Ethics Service and the application procedure. If you wish to make your views known please use the feedback form available on the website.

Further information is available at National Research Ethics Service website > After Review

11/WM/0196 **Please quote this number on all correspondence**

With the Committee's best wishes for the success of this project

Yours sincerely

Dr Rex J Poloon
Chair

Email: Karen.green@westmidlands.nhs.uk

Enclosures: List of names and professions of members who were present at the meeting and those who submitted written comments

"After ethical review – guidance for researchers" SL-AR2

Copy to: Dr Helen Myers
Keele Universitv
Arthritis Research UK Primary Care
Primary Care Sciences
Keele University
Keele, Staffs
ST5 5BG

Mrs Pamela Deval
NHS Stoke-on-Trent
London House
4th Floor
Hide Street
Stoke-on-Trent
ST4 1NF

NHS

National Research Ethics Service

Central Manchester Research Ethics Committee
Room 181
Gateway House
Piccadilly South
Manchester
M60 7LP

Telephone: 0161-237-2153
Facsimile: 0161-237-2383

29 February 2008

Dr Krysia Dziedzic
arc Senior Lecturer in Physiotherapy
Primary Care Musculoskeletal Research Centre
Keele University
Staffordshire
ST5 5BG

Dear Dr Dziedzic

Full title of study: **Self-management, joint protection education and exercises in hand osteoarthritis (SMOotH): a randomised controlled trial in the community**

REC reference number: **07/H1008/235**

The REC gave a favourable ethical opinion to this study on 21 February 2008.

Further notification has been received from local site assessor following site-specific assessment. On behalf of the Committee, I am pleased to confirm the extension of the favourable opinion to the new site. I attach an updated version of the site approval form, listing all sites with a favourable ethical opinion to conduct the research.

R&D approval

The Chief Investigator or sponsor should inform the local Principal Investigator at each site of the favourable opinion by sending a copy of this letter and the attached form. The research should not commence at any NHS site until approval from the R&D office for the relevant NHS care organisation has been confirmed.

Statement of compliance

The Committee is constituted in accordance with the Governance Arrangements for Research Ethics Committees (July 2001) and complies fully with the Standard Operating Procedures for Research Ethics Committees in the UK.

07/H1008/235	Please quote this number on all correspondence

Yours sincerely

K. Osborne.

K Osborne (Mrs)
Committee Co-ordinator

Email: kath.osborne@northwest.nhs.uk

Enclosure: Site approval form

Copy to: Ms R Hughes
Primary Care Musculoskeletal Research Centre
Keele University
Keele
Staffordshire
ST5 5BG

Appendix 2: Quality assessment tool: QUIPS: QUality In Prognosis Studies - QUIPS) – version adapted from (Hayden et al. 2006) and used in the review

Domain: Study Participation

Goal: To synthesize the 'Considerations' (below) to judge the risk of selection bias.

Considerations
Source Population A **clear description** of **who** the source population is (the source population for studies in this review is older adults in the general population) a), **when** (time period of study), **where** (location), **how** (description of recruitment strategy) and **why** the population was chosen, to allow the reader to determine if the source population was captured. Specific characteristics of the population described will vary according to the study objectives, however a comprehensive description would include characteristics of: **individual** (e.g. age, gender), **characteristics of hand condition** (e.g. history of hand problem, current pain and functioning), and **treatment** (type and extent of care received)
Sampling Frame The **sampling frame** and **procedures** used to sample subjects (e.g. newspaper advertisement, presentation to a health clinic, or captured from a claims database) should not lead to selection of participants that are **systematically different** from eligible non-participants.
Inclusion criteria The inclusion and exclusion criteria used should **define a discreet group of older adults from the general population**. Inclusion/exclusion criteria should not select

participants that are **systematically different** from eligible non-participants.
Baseline Study Population A clear description of **who ends up** in the study, sufficient to allow the reader to judge potential selection bias **relative to the target** population. Specific characteristics of the population will vary according to the study objectives, however a comprehensive description would include characteristics of: **individual** (for example, age, gender), **characteristics of hand condition** (e,g. history of hand problem, current pain and functioning), **and treatment** (type and extent of care received)
Adequate study participation It is important for studies to clearly report the **proportion** of eligible subjects who participate in the study, although it is **not possible to set a criteria** for an 'adequate' (or inadequate) participation rate. Studies should **report factors associated** with non-response, quantify and interpret these associations to determine if it is a selective sample. For example, a low participation raises suspicion that there may be a barrier to participating that may also influence outcomes. Alternative data sources for this information may be necessary to build into the study design.

Study Participation Summary question:

After thorough reflection on all considerations, how would you describe the judgment about the risk of selection bias (i.e. distortion due to relationship between the prognostic factor and outcome being different for participants and eligible non-participants)?

☐ **Low risk of bias** ☐ **Moderate risk of bias** ☐ **High risk of bias**

Low risk of selection bias – due to:

- Complete participation by those eligible to participate
- Incomplete participation, but there is evidence that participation was not likely to be related to the prognostic factor and outcome.

Domain: Study Attrition

Goal: To synthesize the 'Considerations' (below) to judge the risk of attrition bias.

Considerations
<u>Proportion of Population Available for Analysis</u> Proportion of the study population available for follow-up analysis should be **clearly presented**, including those with complete data for prognostic factors, potential confounders and outcomes at time-point(s) of interest. The use of criteria or **cut-points** to assess the adequacy of follow-up less than 100% can give false security and should not be used. Imputed values for missing data should only be used if it is clear that data are **missing at random**, and even then with caution.
<u>Outcome and prognostic factor information on those lost to follow-up</u> For studies with subjects lost to follow-up data, **differences** in **prognostic factors**, **potential confounders**, and **outcomes** between responders and non-responders should be assessed to determine if the data is missing systematically. This should be done by presenting a table comparing responders and non-responders at each follow-up and, if available, using another source of data to capture information on missing outcomes.
<u>Reasons and potential impact of subjects lost to follow-up</u> For studies with subjects lost to follow-up, **reasons** for loss suggest subjects should be

presented and assessed for possible **systematic attrition**. The potential implications of loss to follow-up should be considered with respect to the objectives of the study in question.

Study Attrition Summary question:

After thorough reflection on all considerations, how would you describe the judgment about the risk of attrition bias (i.e. distortion in study results due to relationship between the prognostic factor and outcome being different for completing and non-completing participants)?

☐ Low risk of bias **☐ Moderate risk of bias** **☐ High risk of bias**

Low risk – due to:

- There was no loss to follow-up,
- There is some loss to follow-up, but there is evidence that follow-up was not likely to be related to the prognostic factor and outcome.

List reasons for rating:

Domain: Factor Measurement

For this review, a prognostic factor is defined as any variable tested for its association with the level or progression of hand pain or functional difficulty

Goal: To synthesize the 'Considerations' (below) to judge the risk of measurement bias related to the PF.

Considerations
<u>Definition of the factor</u> The factors of interest **fits with the conceptual framework (is there a clear**

justification for examining the association of this variable with level of hand pain and functional difficulty) . The explicitness of the definition depends on the factor (for example, little description is necessary for factors such as gender or race, however a clear operationalization is required for more complex constructs such as depression or coping).

A comprehensive description should include **dose, level** and **duration of the factor** described adequately to allow easy reproduction of measurements. Continuous variables should be kept as continuous whenever possible; if variables are categorized, the categorization should be based on clear theoretical assumptions.

Valid and Reliable Measurement of factor

Factors are described in a **valid**, **reliable** way that allows you to assess the opportunity for misclassification or mis-reporting bias (e.g are data records accurate and complete, have study questions been understood by participants, if measures have been taken by different observers, has quality control been undertaken?) . Measures that are uncommon or have been modified should provide evidence of reliability and validity. Whenever possible, validated instruments should be used and there should be limited reliance on recall.

Method and Setting of factor Measurement

The measurement **approach**, **timing**, and **setting** of assessment should be standardized across subjects from all prognostic groups, or conducted in a way that limits systematically different measurement. If there are differences, the implications should be considered.

Prognostic Factor Measurement Summary question:

After thorough reflection on all considerations, how would you describe the judgment about the overall risk of measurement bias due to the prognostic factor(s) of interest (i.e. distortion due to differential measurement of the factors related to the value of the outcome)?

☐ Low risk of bias **☐ Moderate risk of bias** **☐ High risk of bias**

Low risk – due to:

- Measurement of the factors are valid, reliable and similar for all subjects
- There are differences or uncertainties in measurement but there is evidence that it is not likely to be related to the outcome.

For studies that contain multiple prognostic factors, where some factors are low risk and others are high risk, such studies are scored as "moderate risk of bias",

Domain: Outcome Measurement

Goal: To synthesize the 'Considerations' (below) to judge the risk of measurement bias related to the outcome.

The outcome for this review is any measure of the level of self-reported hand pain or functional difficulty

Considerations
<u>Definition of the Outcome</u> There is a clear **operationalization** of the outcome of interest, including **how** it is assessed and **when** (time points), related to the conceptual framework. Is the outcome measure assessing the level of self-reported hand pain or functional difficulty?

Valid and Reliable Measurement of Outcome Outcomes are measured in a **valid**, **reliable** way that allows you to assess the opportunity for misclassification of individuals (e.g are data records accurate and complete, have study questions been understood by participants?). Measures that are uncommon or have been modified should provide evidence of reliability and validity. Whenever possible, validated instruments should be used,
Method and Setting of Outcome Measurement The measurement **approach**, **timing**, and **setting** of assessment should be standardized across subjects from all prognostic groups, or conducted in a way that limits systematically different measurement. If there are differences, this should be reported and the implications should be considered.

Outcome Measurement Summary question:

After thorough reflection on all considerations, how would you describe the judgment about the overall risk of measurement bias due to the outcome measure (i.e. distortion due to differential measurement of the outcome related to the value of the PF)?

☐ Low risk of bias **☐ Moderate risk of bias** **☐ High risk of bias**

Low risk – due to:

- Measurement of the outcome is valid, reliable and similar for all subjects
- There are differences or uncertainty in measurement but there is evidence that it is not likely to be related to the PF.

For studies that contain outcome measures, where some outcomes are low risk and others are high risk, such studies are scored as “moderate risk of bias”,

Domain: Study Confounding

Goal: To synthesize the 'Considerations' (below) to judge the risk of bias due to confounding.

Only complete this section for exploratory studies concerning a small number of specific prognostic factors and their association with outcome. Section not relevant for predictive studies looking at combinations of factors to predict outcome

Considerations
Important Confounders Measured Important confounders, **based on a conceptual framework**, are **clearly defined** and all confounders mentioned in the model are **measured**. If a conceptual framework was not included a **reasonably comprehensive** set of factors should be assessed from the domains: individual factors (e.g age, gender) factors related to hand problem e.g length of time with condition, healthcare related factors, and factors related to the social context (e.g. social class). The availability of data on important confounders may be a limitation of the study setting (for example, administrative or medical records). The authors should discuss the implications of missing potential confounders and the reviewer satisfied that bias was unlikely.
Valid and Reliable Measurement of Confounders Confounders are described in a **valid**, **reliable** way that allows you to assess the opportunity for misclassification or mis-reporting (e.g are data records accurate and complete, have study questions been understood by participants, if measures have been taken by different observers, has quality control been undertaken?).. Measures that are uncommon or have been modified should provide evidence of reliability and validity. Whenever possible, validated instruments should be used, with limited reliance on recall.

The measurement approach, timing, and setting of assessment should be **standardized** across subjects from all confounding factor groups, or conducted in a way that **limits systematically different measurement**. If there are differences, this should be reported and the implications should be considered.

Appropriate Accounting for Confounding

Accounting for confounding can be done in the **study design** and/or the **analysis**. In the design, matching for key characteristics, stratification, or initial assembly of comparable groups could be used.

Identification of important confounders in a dataset should be **guided by a conceptual framework**, and **tested systematically**. This includes providing data on potential confounders stratified by the prognostic factor of interest. Crude and adjusted estimates for prognostic factors, and a measure of precision should be presented. If the impact of confounding is systematically investigated without importantly affecting the estimates (they are stable), they are less likely to be confounded.

Multiple regression modelling can be used to **control for confounding**. Stepwise regression techniques do not allow specific testing of confounding and may lead to bias in studies with the objective of causal understanding.

Study Confounding Summary question:

After thorough reflection on all considerations, how would you describe the judgment about the overall risk of bias due to confounding (i.e. the effect of the PF(s) is distorted by another factor that is related to both the PF and the outcome)?

☐ Low risk of bias **☐ Moderate risk of bias** **☐ High risk of bias**

Low risk – due to:

- Inclusion and assessment of confounding was planned by theory, included valid and reliable measures and were appropriately controlled in the design and/or analysis.

For studies that contain outcome measures, where some associations are low risk and others are high risk, such studies are scored as "moderate risk of bias",

Domain: Statistical Analysis and Presentation

Goal: To synthesize the 'Considerations' (below) to judge the risk of bias due to analysis and presentation.

Considerations
Presentation of analytical strategy Sufficient presentation of data to assess the analysis and findings. This includes a clear and exhaustive 'Table 1' providing data on potential confounders stratified by the prognostic factor of interest, including number of subjects available; crude and adjusted estimates for predictive factors, with a measure of precision should be presented with the effect estimates presented consistently in the same direction (i.e. so that an increased risk is presented as a positive coefficient). The clinical importance of the results should be considered and the findings not overstated.
Model development strategy The strategy for model building (i.e. inclusion of variables) is appropriate and based on a conceptual framework or model The selected model is adequate for the design of the study.
Reporting of results Selective reporting of results is avoided.

Statistical Analysis and Presentation Summary:

After thorough reflection on all considerations, how would you describe the judgment about the overall risk of bias due to the statistical analysis?

☐ Low risk of bias **☐ Moderate risk of bias** **☐ High risk of bias**

Low risk – due to:

- Analysis reflects the study's objectives and the design of the study.
- The authors' explanation for analysis employed is sensible and clear. For a study with a causal understanding approach, a conceptual framework guides a thoughtful analysis.

Appendix 3: Potential baseline prognostic factors excluded from analysis

Factor	Exclusion reason
Exclusion – data frequency and reliability	
Race/ethnicity	>90% of participants report their ethnicity to be “White”
Rheumatoid Arthritis (RA)	Low prevalence
Diabetes	Low prevalence
Stroke	Low prevalence
Parkinson’s disease	Low prevalence
Thyroid disease	Low prevalence
Gout	Low prevalence
Severe visual impairment	Low prevalence. All participants had to attend and complete the baseline clinical assessment so severe visual impairment was uncommon - construct largely relevant to hand function only
Cognitive impairment	Highly skewed distribution (as measured by the SIP alertness scale (Bergner et al. 1981)). Majority of participants report no cognitive impairment
Illness perceptions – symptom count	High missing data rate of 29%
Exclusion – clinical opinion	
Consulting a GP or hospital doctor about hand problem	Construct will not be asked in a GP consultation setting
Self-reported diagnosis (OA, RA, other)	GP’s are more likely to be seeking to give a diagnosis than to ask for participants’ perception on the cause of their condition
Taking medication for hand condition	Construct addressed as a future research question. The role of treatment has the potential to be complex as influenced by the type of medication, dosage, and whether the treatment is working well. Medication use could also be susceptible to confounding by indication (Grobbee et al. 1997)) as participants on treatment, although demonstrating fewer symptoms, may have more severe problems

	that indicated them to receive treatment at the outset
Surgery for hand condition	Construct addressed as a future research question as could depend on complex issues e.g. the type of surgery undertaken, length of time since the surgery occurred etc.
Hand/wrist fracture in the last 5-years	Represent this concept as “onset of hand problem following hand injury”. Hand/wrist fracture in the last 5-years are of low prevalence at baseline and this generic question on hand/wrist fracture does not relate the injury to the current hand problem

Footnote: Low prevalence is indicated if <10% of participants in CAS-HA have the factor of interest

Appendix 4: Extended information on methods of measurement (the AUSCAN is excluded from this list as it is described in detail in chapter 4)

Physical and mental component scores of the Short-form 12 (SF-12) (Ware et al. 1996)

The twelve questions below were used to construct the physical and mental component scores for the SF-12. Responses to the questions were summed together to make a total score, with higher weights given to the physical health items in the physical component score and higher weights given to the emotional health items in the mental component score. Details of the weighting scheme and scoring are available from the authors of the tool.

1) In general how would you say your health is?
2) Does your health limit moderate activities e.g. moving a table, pushing a vacuum cleaner, bowling or playing golf?
3) Does your health limit you climbing several flights of stairs?
4) As a result of your physical health have you accomplished less than you would like in the last 4 weeks?
5) As a result of your physical health were you limited in the kind of work or other activities in the last 4 weeks?
6) As a result of any emotional problems have you accomplished less than you would like?
7) As a result of any emotional problems did you not do work or other activities as carefully as usual?
8) During the past 4 weeks, how much did pain interfere with your normal work (including both work outside the home and housework)?
9) How much of the time during the past 4 weeks have you felt calm and peaceful?
10) How much of the time during the past 4 weeks did you have a lot of energy?
11) How much of the time during the past 4 weeks have you felt downhearted and depressed?

12) During the past 4 weeks, how much of the time has your physical health or emotional problems interfered with social activities (like visiting friends, relatives, etc.)?

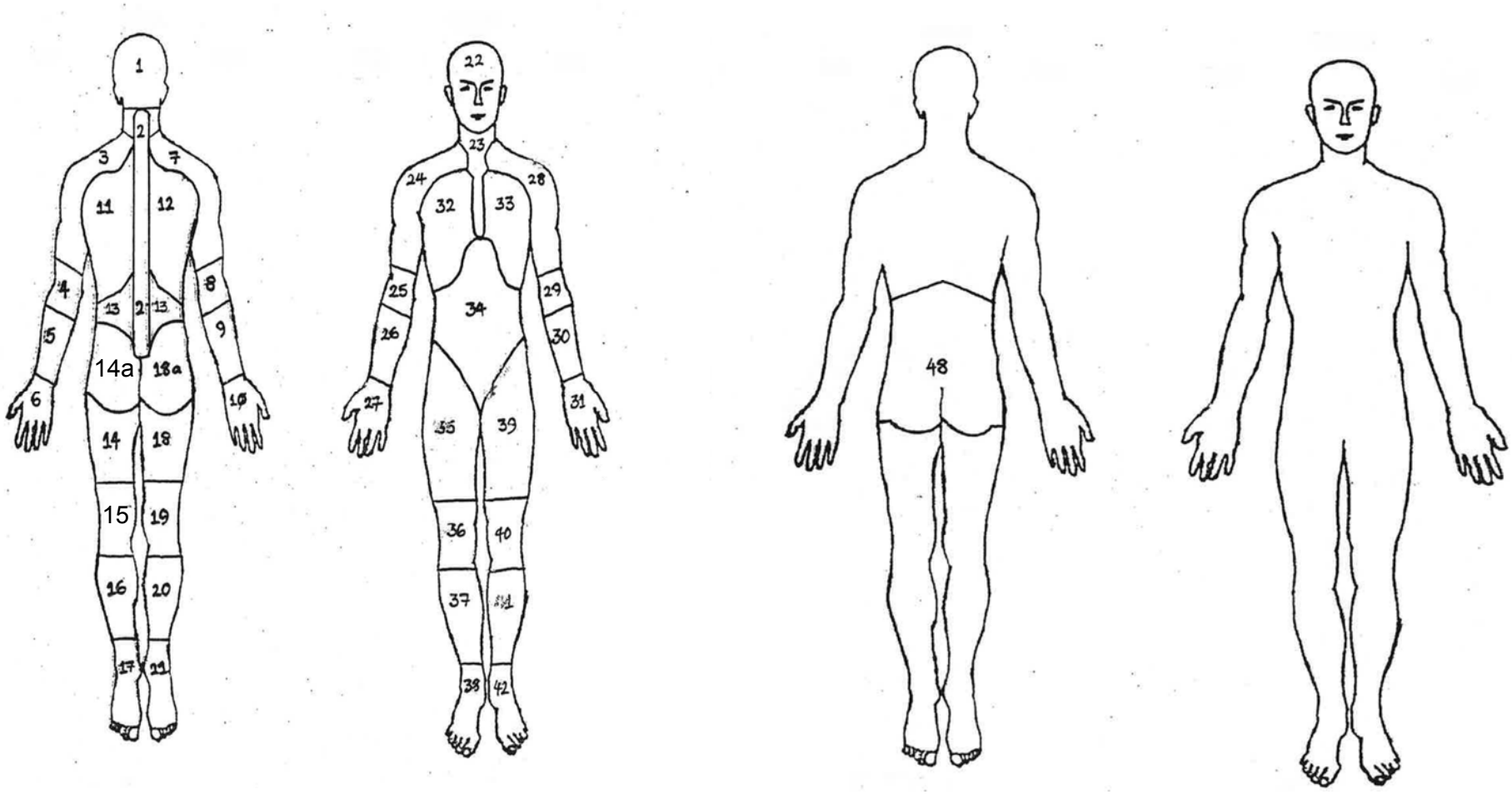
Manchester definition of widespread pain (MacFarlane et al. 1996, Croft et al. 2001, Birrell et al. 2000, Papageorgiou et al. 1995)

The Manchester definition of widespread pain was scored from a body manikin that was included in the questionnaire. Participants were asked to shade on a body manikin any areas where they had experienced pain or aching, lasting for a day or longer, over the last month, but not to include pain due to feverish illness e.g. flu. The body manikin was scored by overlaying a template onto the completed manikin to split the body into distinct regions (see diagram below). Participants were scored with a value of one if they had indicated pain in that region and zero otherwise. It was this data that was used to derive the Manchester definition of widespread pain

Widespread pain was indicated if participants had axial pain (i.e. pain in regions 2 or 48 of the manikin) and pain in at least two regions of two contralateral limbs (i.e. the right upper limb and the left lower limb; or the left upper limb and the right lower limb). The right upper limb was defined as regions 7, 8, 9, 10, 24, 25, 26 and 27 of the body manikin and the right lower limb as regions 18, 19, 20, 21, 35, 36, 37 and 38, with equivalent regions on the left side defined for analysis.

For the analysis, participants were classified into those with widespread pain, those with no pain in their body other than in their hands. Participants not meeting either of these criteria formed a third group which was labelled as “Regional pain”.

Template for the body manikin



Hospital Anxiety and Depression scale (HADS) (Zigmond et al. 1983)

Responses to questions 1, 3, 5, 7, 9, 11 and 13 are totalled to give a score for anxiety

Response to questions 2, 4, 6, 8, 10, 12 and 14 are totalled to give a score for depression

The response options for each question vary but all are measured on a 4 point scale as reported in Zigmond et al. 1983.

1) Past week, I feel tense or wound up
2) Past week, I still enjoy the things I used to enjoy
3) Past week, I get a sort of frightened feeling as if something awful is about to happen
4) Past week, I can laugh and see the funny side of things
5) Past week, worrying thoughts go through my mind
6) Past week, I feel cheerful
7) Past week, I can sit at ease and feel relaxed
8) Past week, I feel as if I am slowed down
9) Past week, I get a sort of frightened feeling like butterflies in my stomach
10) Past week, I have lost interest in my appearance
11) Past week, I feel restless as if I have to be on the move
12) Past week, I look forward with enjoyment to things
13) Past week, I get sudden feelings of panic
14) Past week, I can enjoy a good book or radio or television programme

Illness perceptions questionnaire – revised (IPQ-R) (Moss-Morris et al. 2002)

The questions that form the seven IPQ-R sub-scales used in this thesis are listed below. Each question is rated on a 5-point Likert scale ranging from “Strongly disagree” to “Strongly agree”. Sub-scale scores are constructed by adding together the response options for each question after the items indicate by a * have been reverse coded.

Timeline

1) My hand problem will last a short time*

2) My hand problem is likely to be permanent rather than temporary
3) My hand problem will last for a long time
4) This hand problem will pass quickly*
5) I expect to have this hand problem for the rest of my life
6) My hand problem will improve in time*

Consequences

1) My hand problem is a serious condition
2) My hand problem has major consequences on my life
3) My hand problem does not have much effect on my life*
4) My hand problem strongly affects the way others see me
5) My hand problem has serious financial consequences
6) My hand problem causes difficulties for those who are close to me

Personal control

1) There is a lot I can do to control my hand symptoms
2) What I do can determine whether my hand problem gets better or worse
3) The course of my hand problem depends on me
4) Nothing I do will affect my hand problem*
5) I have the power to influence my hand problem
6) My actions will have no effect on the outcome of my hand problem*

Treatment control

1) There is very little that can be done to improve my hand problem*
2) My treatment will be effective in curing my hand problem
3) The negative effects of my hand problem can be prevented (avoided) by my treatment
4) My treatment can control my hand problem

5) There is nothing that can help my hand problem*

Illness coherence

1) The symptoms of my hand problem are puzzling to me
2) My hand problem is a mystery to me
3) I don't understand my hand problem
4) My hand problem doesn't make any sense to me
5) I have a clear picture or understanding of my hand problem*

Timeline cyclical

1) The symptoms of my hand problem change a great deal from day to day
2) My symptoms come and go in cycles
3) My hand problem is very unpredictable
4) I go through cycles in which my hand problem gets better or worse

Emotional representation

1) I get depressed when I think about my hand problem
2) When I think about my hand problem I get upset
3) My hand problem makes me feel angry
4) My hand problem does not worry me*
5) Having this hand problem makes me feel anxious
6) My hand problem makes me feel afraid

Grip-ability test (GAT) (Dellhag et al. 1995)

The GAT score is calculated as a weighted average of three timed tests (in seconds): putting on a tubigrip, picking up a paper clip, pouring a jug of water (with weights 1.8, 1 and 1.8).

American College of Rheumatology (ACR) criteria for hand osteoarthritis (OA) (Altman et al. 1990)

Osteoarthritis occurs when cartilage covering the ends of the bones gradually roughens and becomes thin and the bone underneath thickens (Arthritis Research UK 2013). For hand OA, this can be painful and result in firm knobbly swellings on the finger joints (Arthritis Research UK 2013). Presence of clinical hand OA was measured in this thesis using the American College of Rheumatology (ACR) criteria for hand OA defined below:

Hand pain, aching, or stiffness$^{\alpha}$ and 3 or 4 of the following features:

- Hard tissue enlargement of 2 or more of 10 selected joints$^{\beta}$
- Hard tissue enlargement of 2 or more DIP joints
- Fewer than 3 swollen MCP joints
- Deformity of at least 1 of 10 selected joints$^{\beta}$

$^{\alpha}$ Hand pain, aching or stiffness was defined as present if it occurred on all or most days in the previous month; $^{\beta}$ The 10 selected joints are the second and third distal interphalangeal (DIP), the second and third proximal interphalangeal, and the first carpometacarpal joints of both hands. MCP = metacarpophalangeal.

Carpal tunnel syndrome (Palmer et al. 2000)

Carpal tunnel syndrome occurs when there is excessive pressure on the median nerve i.e. a nerve in the wrist that allows feeling and movement to parts of the hand (American Accreditation HealthCare Commission 2014a). It can lead to pain, numbness, tingling, weakness or muscle damage in the hand and fingers (American Accreditation HealthCare Commission 2014a). Presence of carpal tunnel syndrome is measured in this thesis using an adapted version of the definition by Palmer et al. 2000. The version was adapted by

excluding both Tinel's and nerve conduction tests as these two components were not measured in the CAS-HA study.

Carpal tunnel definition by Palmer et al. 2000

Pain or paraesthesia or sensory loss in the median nerve distribution and one of:

- Tinel's test positive (excluded as not measured in CAS-HA)
- Phalen's test positive (positive if numbness, tingling or weakness occurred when wrist bent forward)
- nocturnal exacerbation of symptoms
- motor loss with wasting of abductor pollicis brevis (muscle at the base of the thumb)
- abnormal nerve conduction time (excluded as not measured in CAS-HA)

Dupuytren's contracture

Dupuytren's contracture is relatively painless (Adebajo 2010) and occurs through thickening of tissue under the skin in the palm of the hand causing the fingers to stiffen and bend (Medline Plus: Finger injuries and disorders). Unlike other hand conditions, it is more common in men than women (Adebajo 2010). Presence of Dupuytren's contracture was measured in the CAS-HA study by examination of the hand by a clinical assessor. Presence of the condition was recorded dichotomously as present yes/no.

De Quervain's tenosynovitis

De Quervain's tenosynovitis occurs when tendons running from the back of the thumb down the side of the wrist become swollen and irritated (American Accreditation HealthCare Commission 2014b). In the CAS-HA study, presence of De Quervain's tenosynovitis was measured as present if the participant had a positive Finkelstein's test, which was defined as positive if pain increased when the thumb was pushed into the palm of the hand (Adebajo 2010)

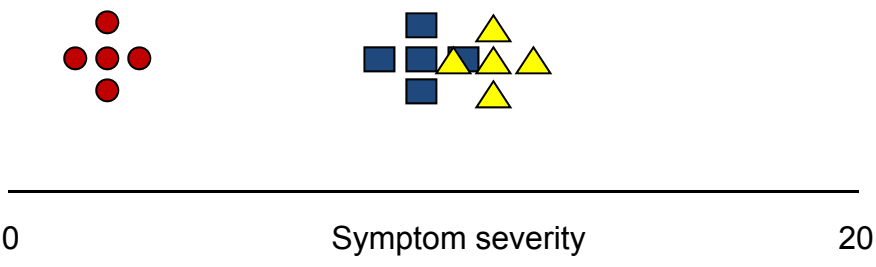
Trigger finger

Trigger finger occurs when a nodule catches at the pulley that overlies the metacarpophalangeal joint in the palm (Adebajo 2010). This causes the affected fingers to lock in a bent position, which may or may not be painful to release (Adebajo 2010). In CAS-HA, participants were defined as having trigger finger if they answered “yes” to both of the following questions: Do your fingers ever lock, trigger or catch? If so, do you have to release them yourself?

Appendix 5: Conceptual differences between reliability and agreement parameters

The diagram below, reproduced from de Vet et al. 2006a[122], is a useful way to illustrate the difference between reliability and agreement parameters. It describes a scenario where three patients (represented by the circles, squares and triangles, respectively) are measured repeatedly on five occasions when no change in symptom state has occurred. Although absolute agreement between all five measures is not obtained, if the study aim is to compare differences between patients, this can be done more reliably for patients that differ greatly in symptom severity (i.e., the patients represented by the circle and squares) than for those with a similar level of symptoms (i.e., the patients represented by the squares and triangles). Information on the reliability of a measurement tool (when derived from similar patient populations to those under study) is therefore informative for studies whose primary aim is to distinguish between patients with different outcomes. In contrast, agreement parameters are useful when the study aim is to evaluate the magnitude of the outcome differences in a patient population, e.g. to assess the precise degree of change in an outcome over time (de Vet et al. 2006a). As the aim of the CAS-HA study includes both these aims, both reliability and agreement parameters were presented.

Descriptive diagram to illustrate differences between reliability and agreement parameters ((reproduced from (de Vet et al. 2006a))



[122] Reprinted from Journal of Clinical Epidemiology, 59 (10), De Vet et al., When to use agreement versus reliability measures, 1033-1039, Copyright (2006), with permission from Elsevier

Appendix 6: The Clinical Assessment Study of the Hand (CAS-HA) – pilot study

A pilot study was conducted prior to the main CAS-HA study to determine the reliability of the clinical and self-report measures used in the main study (Myers et al. 2011). Participants in the pilot study were aged 50 years and over and were purposively sampled from participants in a survey of hand problems in North Staffordshire (Thomas et al. 2004b). Purposive sampling was used so that participants in the pilot study represented the full range of symptom severity likely to present for a clinical assessment in the main study.

A total of 201 participants were sent a postal invite to attend the pilot study and 55 replied, giving consent to attend two research assessments, on average, one-month apart. The time frame between assessments was chosen to be sufficiently long to prevent recall bias but short enough for limited change in symptoms to occur (Streiner 2003). Full ethical approval was given for the study from the North Staffordshire Ethics Committee (ref = 1430).

Data from the CAS-HA pilot study were used to assess the reproducibility of the AUSCAN and to derive a range of possible values for minimum important change. Interpretation of these analyses, however, depended on the amount of variability in the AUSCAN between participants. A descriptive comparison was therefore conducted to compare the key characteristics of participants in the CAS-HA main and pilot studies and to explore whether the amount of variability in the AUSCAN was similar between participants in each study, i.e. to explore whether evidence on the reproducibility of the AUSCAN and minimum important change can generalise from the pilot to the main CAS-HA cohort.

In general, participants in the CAS-HA pilot study were older than those in the main study. They were of better general health and had slightly higher levels of hand pain and functional difficulty. The interquartile ranges of the AUSCAN scores, however, as a marker

of variability, were similar between the two studies, along with the proportion of females that participated. This gives some support that the findings from the CAS-HA pilot study generalise to the main study data although these key differences are present in the data.

Descriptive data to compare the CAS-HA pilot study to the main CAS-HA cohort

	CAS-HA pilot study$^{\alpha}$	CAS-HA main study$^{\beta}$
N	55	623
Age (years)$^{\gamma}$	67 (60, 74)	63 (58, 71)
Gender		
Female	33 (60)	385 (62)
General Health		
Excellent	5 (9)	29 (5)
Very Good	18 (33)	132 (21)
Good	21 (38)	256 (41)
Fair	10 (18)	158 (26)
Poor	1 (2)	44 (7)
AUSCAN$^{\gamma}$		
Hand pain	8.0 (4.0, 11.0)	6.0 (3.0, 9.0)
Hand function	10.0 (3.0, 19.0)	8.0 (3.0, 17.0)

Figures are numbers and percentages unless otherwise stated. α = data taken from the first data collection point in the CAS-HA pilot study, β = data taken from the first data collection point in the main CAS-HA study i.e. baseline, γ = Median (inter-quartile range)

Appendix 7: The Arthritis Impact Measurement Scales 2 (AIMS2)

The AIMS2 is a self-reported questionnaire, developed by Meenan et al. 1992, to measure outcome for patients with arthritis. The questionnaire covers nine domains of outcome (mobility; physical activity; dexterity; household activities; activities of daily living; anxiety; depression; social activity; pain), however, only the specific subscales of 'hand and finger function' and 'hand pain' are used in this thesis.

The AIMS2 questions were only included at baseline and 18-month follow-up in CAS-HA so this measure was not used as the primary measure of hand pain and function in this thesis. The AIMS2 questions for hand pain and function are below along with a description of how they were combined to give total scores for analysis (Meenan et al. 1992).

Hand pain subscale

During the last month….

- How would you describe the hand pain you usually had?
- How often did you have severe pain in your hands?
- How often did you have pain in two or more hand joints at the same time?
- How often did the morning stiffness in your hands last of more than one hour from the time you woke up?
- How often did your hand problems make it difficult for you to sleep?

Hand and finger function subscale

During the last month….

- Could you easily write with a pen or pencil?
- Could you easily button a shirt or blouse?
- Could you easily turn a key in a lock?
- Could you easily tie a knot or a bow?

- Could you easily open a new jar of food?

The items in each subscale were rated using a five point scale: 0 = ‘All days’, 1 = ‘Most days’, 2 = ‘Some days’, 3 = ‘Few days’, 4 = ‘No days’. Items in each subscale were summed together and transformed to a 0-10 scale. Participants with a single item missing on the subscale had their missing value imputed with the mean of the remaining items in the subscale.

Appendix 8: The SMOotH study - Self Management in Osteoarthritis of the Hand

The SMOotH study was a 2x2 factorial clinical trial exploring the clinical and cost effectiveness of hand exercises and joint protection for participants with hand OA (Dziedzic et al. 2011, Dziedzic et al. 2013). The primary aim of the study was to compare changes in hand pain and functional difficulty between the intervention groups. Prior to data collection, the study was given full ethical approval from the North West 7 Research Ethics Committee UK (rec reference: 07/H1008/235).

Participants were recruited to the study using a postal survey mailed to all adults aged 50 years and over registered with five general practices in North Staffordshire and Cheshire. Responders to the postal survey were assessed for trial eligibility and were eligible if they met the following criteria

1) Reported hand pain or problems in the last 12-months
2) Reported hand pain, aching or stiffness on more than a few days in the last month
3) Had an minimum level of hand pain and/or functional difficulty as measured by the AUSCAN (pain score >=5 or function score >=9)
4) Had not seen an OT or physiotherapist in the last 6 months for their hand problem
5) Had not had a joint injection (wrist, fingers or thumbs) in the last 6 months
6) Had not had fractures or hand injury in the last 6-months
7) Another person in the household was not already taking part in the study

Those eligible at the postal survey stage were invited to attend a clinical assessment with a research nurse held at a local hospital or in the participants' GP practice. The aim of the assessment was to further determine trial eligibility. Participants were eligible if they met the following criteria:

1) Gave informed consent to trial
2) Did not have any clinical red flags or abnormalities e.g. swollen joints indicating rheumatoid arthritis

3) Met the American College of Rheumatology Criteria for hand OA (Altman, Alarcon et al. 1990) or had unilateral or bilateral thumb base OA
4) Criteria 5,6, and 7 above had not changed since returning the postal survey
5) Able to attend treatment classes in the 3-month period after randomisation

Individuals eligible at the clinical assessment stage and providing written informed consent were then randomised to one of four treatment arms (leaflet & advice, joint projection, hand exercises or both joint protection and hand exercises). The treatments were delivered by occupational therapists specifically trained to deliver each intervention and participants were followed up by postal questionnaire at 3-, 6- and 12-months post-randomisation.

For the SMOotH study, 12,090 surveys were mailed and 6972 were returned. Of these, 1309 met the eligibility criteria and were invited for a clinical assessment. The clinical assessment was attended by 344 participants, of which, 75% were randomised (N=252). Follow-up rates for the trial were high (N=232; N=218, and N= 219 for 3- 6- and 12-month follow-up respectively).

Participants in the SMOotH study were similar to those in the CAS-HA main study for age, gender, and general health. As expected, given there was a minimal value on the AUSCAN for hand pain and functional difficulty to be eligible for SMOotH, participants in this study had more severe levels of hand pain and functional difficulty than those in the CAS-HA cohort. This is an advantage however of this data set that it can be used to explore whether results from the CAS-HA cohort generalise to other data sets where the profile of participants may differ from the sample they were originally derived, e.g. by symptom severity or by other socio-demographic variables.

Descriptive data to compare the SMOotH study to the main CAS-HA cohort

	SMOotH study[α]	CAS-HA main study[α]
N	257	623
Age (years)[β]	65 (59, 73)	63 (58, 71)
Gender		
Female	170 (66)	385 (62)
General Health		
Excellent	6 (2)	29 (5)
Very Good	50 (20)	132 (21)
Good	128 (50)	256 (41)
Fair	62 (24)	158 (26)
Poor	8 (3)	44 (7)
AUSCAN[β]		
Hand pain	9.5 (7.0, 12.0)	6.0 (3.0, 9.0)
Hand function	15.0 (9.0, 21.0)	8.0 (3.0, 17.0)

Figures are numbers and percentages unless otherwise stated. α = Data collected at baseline, β = Median (inter-quartile range)

Appendix 9: Distribution and correlation of three global assessment of change questions included in the SMOotH clinical trial (N=257)

	Compared to when you were first seen by our research nurse, how is your…..		
	hand problem now?	hand pain now?	ability to use hands?
	N(%)	N(%)	N(%)
Completely recovered	0(0)	0 (0)	0(0)
Much Better	17 (8)	17 (8)	13 (6)
Better	57 (25)	54 (24)	58 (26)
No change	115 (50)	114 (51)	125 (56)
Worse	36 (16)	34 (15)	26 (12)
Much worse	3 (1)	3 (1)	2 (1)
Total	228	222	224
Spearman's rank correlation	0.91	0.81	

Appendix 10: Receiver Operating Characteristic (ROC) curves

ROC curves (Silman 1995) were generated for AUSCAN hand pain and function by plotting (sensitivity) versus (1-specificity) for all possible cut-points on the AUSCAN change score. The values for sensitivity and specificity were calculated using the formulae below when applied to all possible cut-points on the AUSCAN change score.

Calculation of sensitivity and specificity:

The example below demonstrates how sensitivity and specificity values were calculated for the analysis of symptom deterioration if a cut-point of two on the AUSCAN function subscale were chosen to represent minimum important change.

	AUSCAN function (baseline – follow-up)		
Global assessment of change	'Worse' <2	'No change' >=2	Total
No change	a	b	a+b
Worse to some degree	c	d	c+d
Total	a+c	b+d	a+b+c+d

Sensitivity = percentage in the "no change" group correctly identified by the AUSCAN function scale. Sensitivity = b/a+b

Specificity = percentage in the "worse to some degree" group correctly identified by the AUSCAN function scale. Specificity = c/c+d

Appendix 11: Correlation between the global assessment of change question and the AUSCAN pain and function measures at baseline and first follow-up

Data sample	CAS-HA main study		SMOotH clinical trial	
	Baseline	18-month follow-up	Baseline	3-month follow-up
AUSCAN pain	0.21	0.52	0.09	0.37
AUSCAN function	0.24	0.47	0.06	0.28

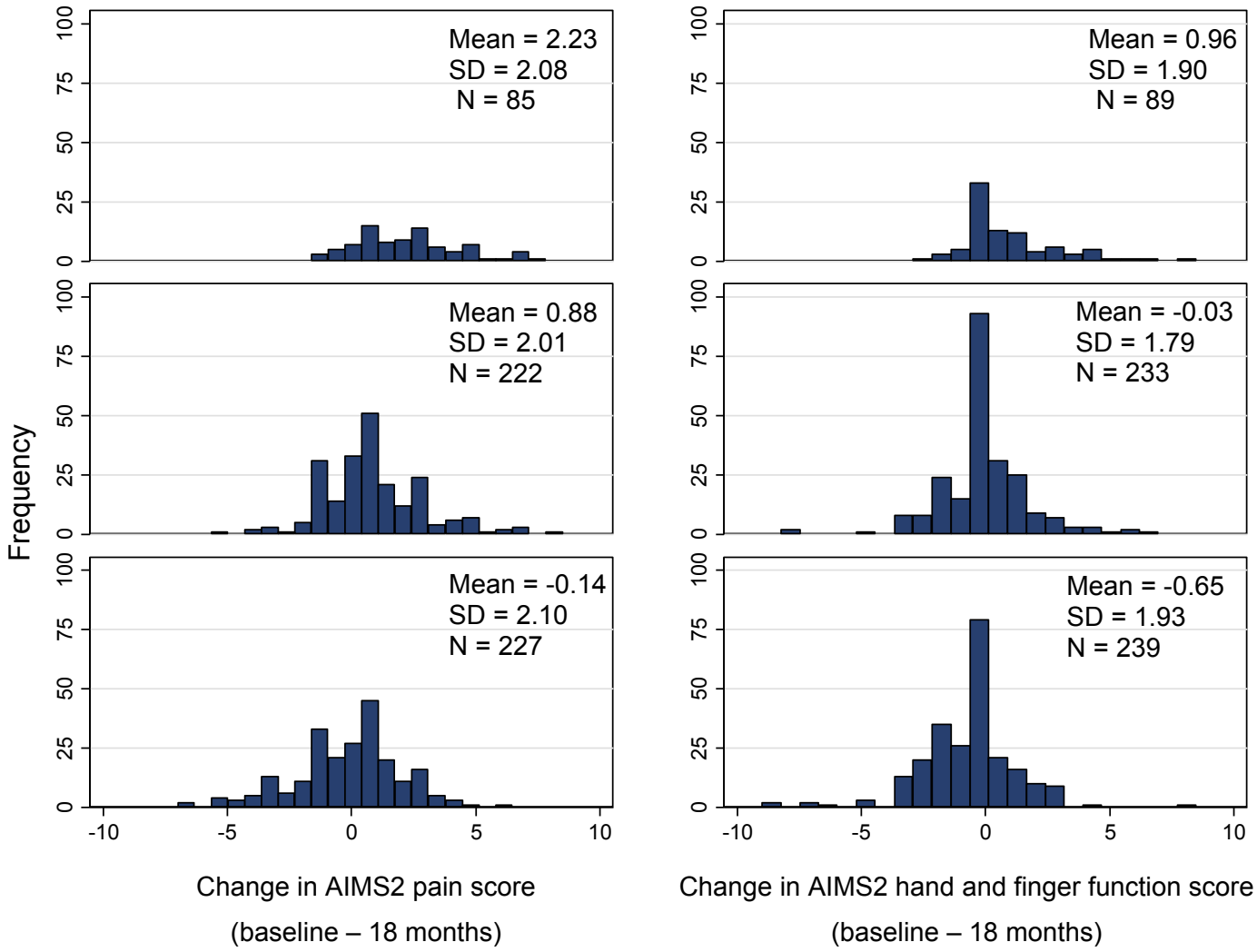
Figures in the table are Spearman's rank correlations

Appendix 12: Sensitivity to change and responsiveness of the AIMS2 and pain numerical rating scale in the main CAS-HA study

Sensitivity to change/responsiveness	All participants (N= 586)			Participants reporting symptom improvement[α] (N= 92)			Participants reporting symptom deterioration[β] (N = 248)		
	AIMS2 pain	AIMS2 hand and finger function	Pain numerical rating[γ]	AIMS2 pain	AIMS2 hand and finger function	Pain numerical rating[γ]	AIMS2 pain	AIMS2 hand and finger function	Pain numerical rating[γ]
Effect Size	0.28	-0.05	0.26	0.94	0.41	0.66	-0.06	-0.25	-0.17
Standardised response mean (SRM)	0.30	-0.06	0.25	1.07	0.51	0.83	-0.07	-0.34	-0.20
Guyatt's responsiveness statistic				1.11	0.54	0.77	-0.07	-0.36	-0.20

Analysis was not completed for the SMOotH study as the AIMS2 was not collected in this study and the time scale for the numerical pain rating was 'in the last the 3 days' rather than 'in the last month' making comparisons between data sets difficult. α = defined as "better", "much better" or "completely recovered" on the global assessment of change question, β = defined as "worse" or "much worse" on the global assessment of change question, γ = defined as "Pain severity in the last month" and measured on a 0-10 numerical rating scale.

Appendix 13: AIMS2 pain and hand and finger function change scores (baseline to 18-month follow-up)



Global assessment of change

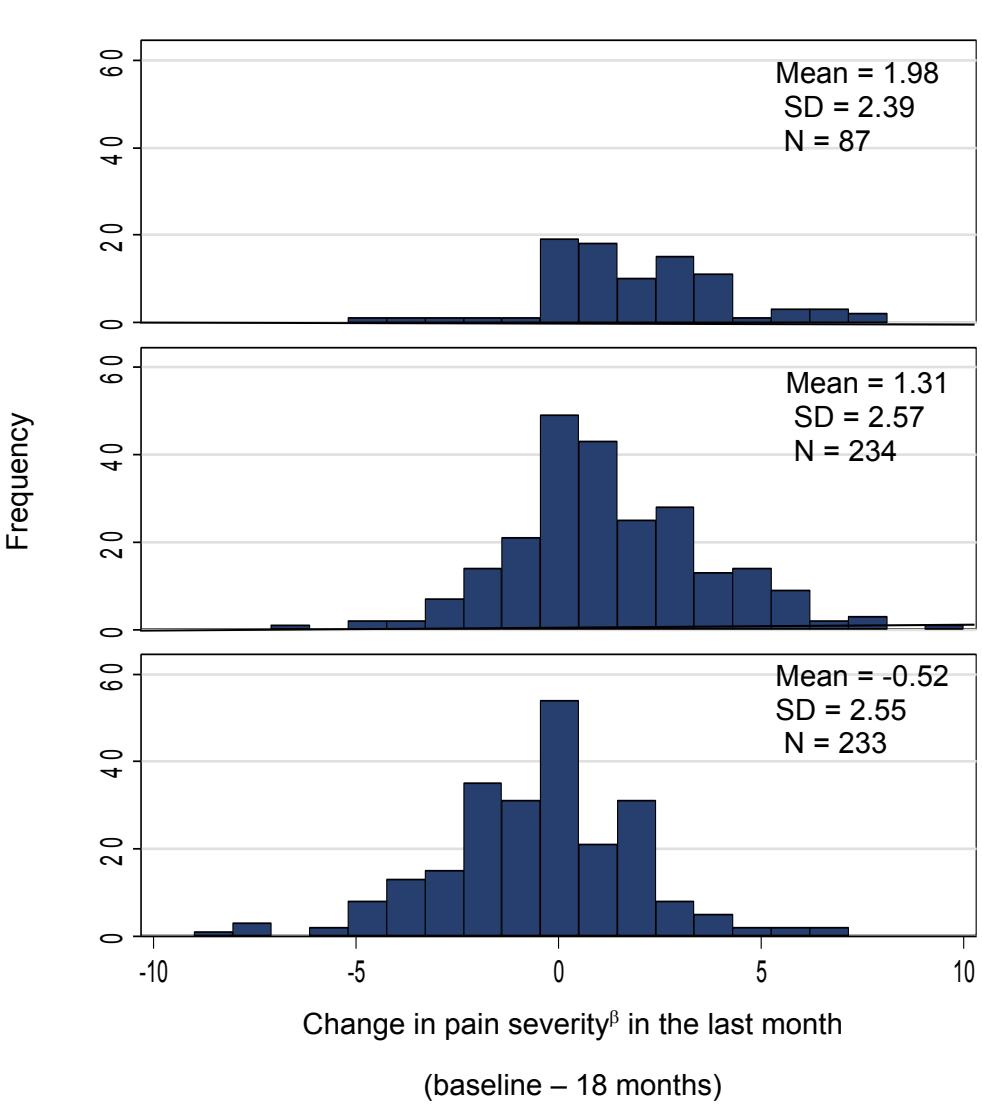
"Better"α
(to include completely recovered, much better or better)

"No change"

"Worse"α
(to include worse or much worse)

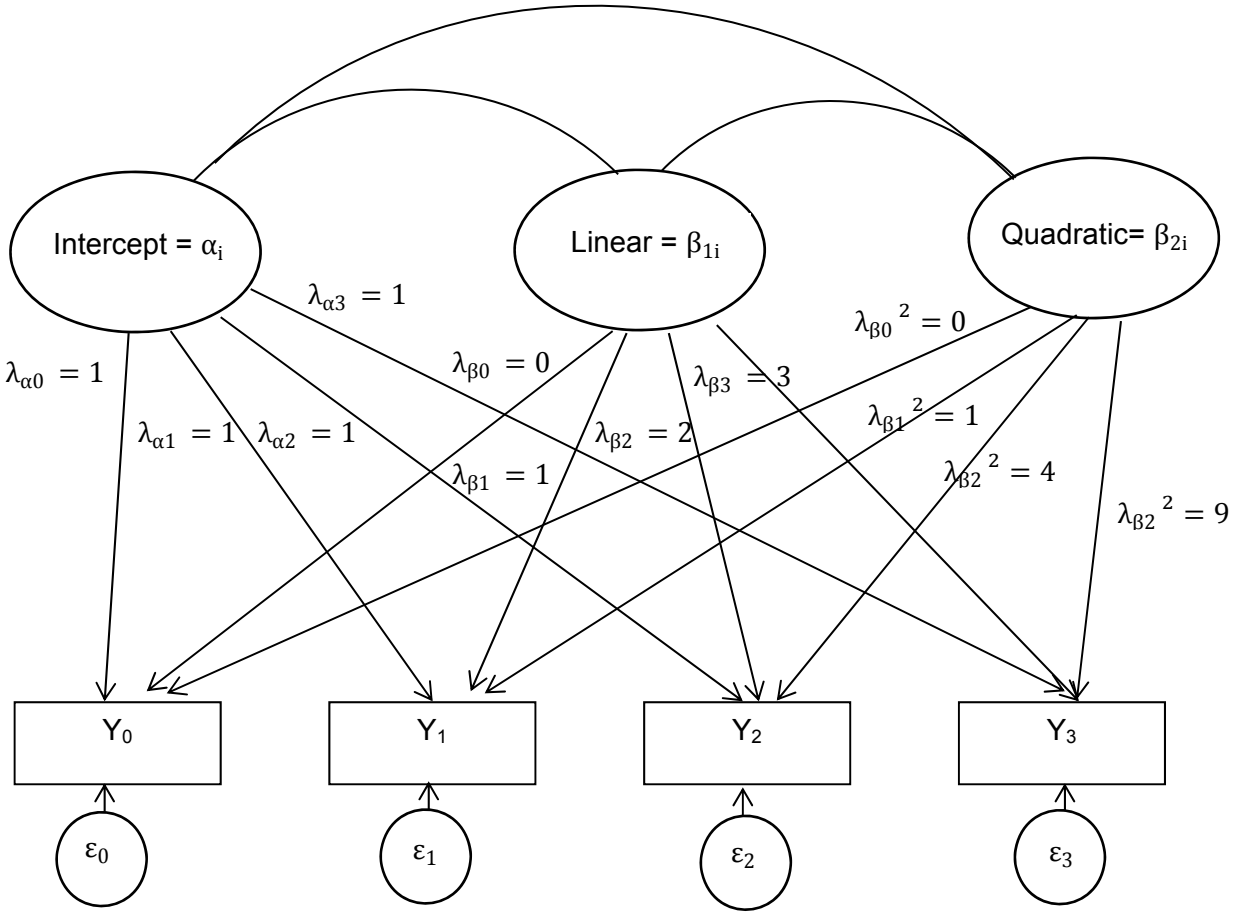
α = Category merging undertaken due to small N. SD = Standard deviation

Appendix 14: Plot of change in pain severity (baseline to 18-month follow-up) stratified by global assessment of change category



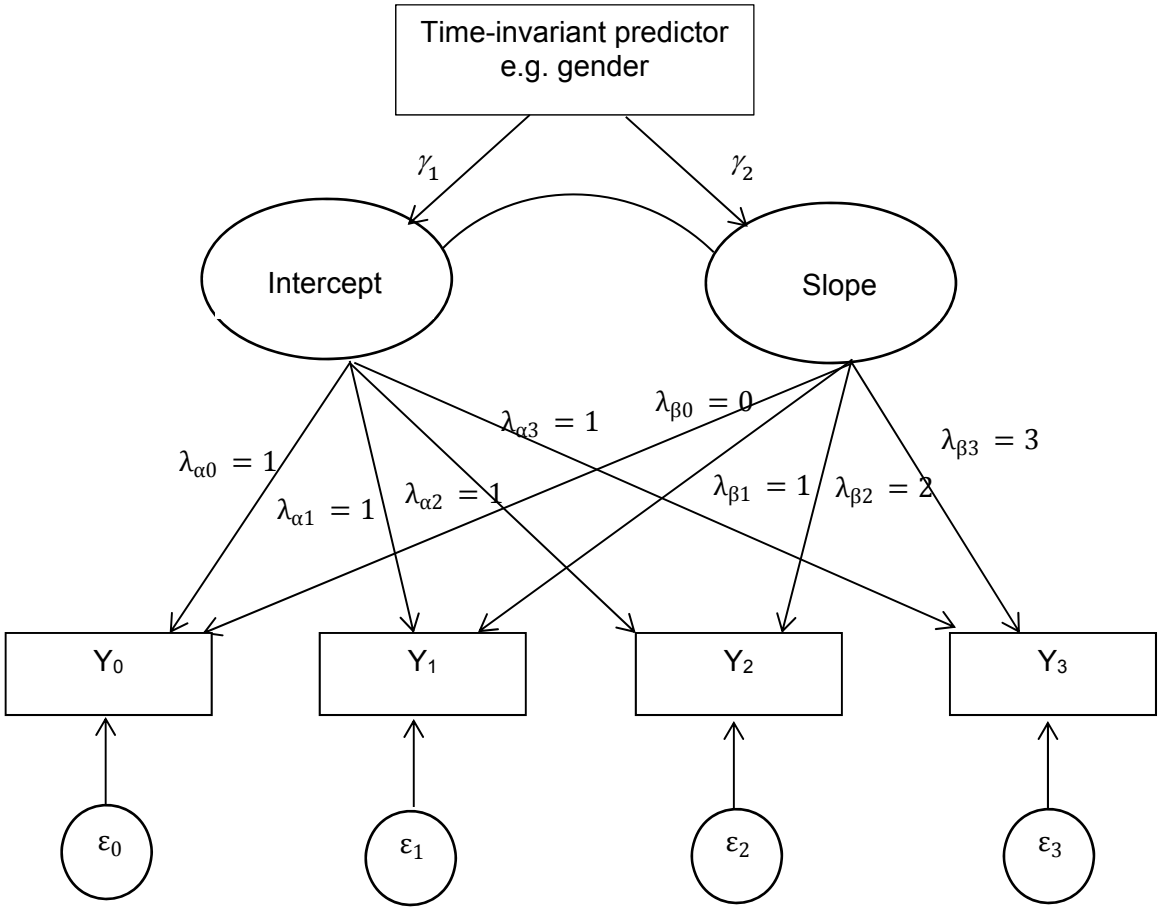
α = Category merging undertaken due to small N.
β = 0-10 numerical rating scale given at each time point to rate pain severity in the previous month
SD = standard deviation

Appendix 15: Quadratic growth model fitted in a structural equation modelling framework for a study with data collected at four time points (Curran et al. 2003)



Footnote: In the SEM framework, circles (or ellipses) represent unobserved factors (latent variables), squares (or rectangles) represent observed variables, single-headed arrows represent the impact of one variable on another, double-headed arrows represent co-variances or correlations between pairs of variables (Byrne & Crombie, 2003). In this diagram, the mean of the latent variables pooled over all participants in the sample represent fixed effects; the random effects are characterised by the variance of each latent factor (Curran & Hussong, 2003).

Appendix 16: Linear growth model fitted in a structural equation modelling framework for a study with data collected at four time points and including one time-invariant predictor of interest (Curran et al. 2003)



Multiple time invariant predictors can be included in the model. Categorical variables can be incorporated by entering dummy variables into the model (Curran & Hussong, 2003)

Appendix 17: Key assumptions for growth modelling

Sampling assumptions

- Participants are selected at random from the population of interest (Byrne et al. 2003)
- Participants have the potential to be sampled at three or more time points so trajectories, over and above change between two time points, can be considered[123] (Byrne et al. 2003)
- No sampling outliers exist that could bias the analysis (Byrne et al. 2003)

Normality assumptions for data to be estimated using maximum likelihood

- Multivariate normality is required for:
    - the repeated measures of interest (Hox et al. 2005)
    - the latent growth factors used to model the trajectory of interest i.e α_i and β_i for the linear model shown in Box 10 (Byrne et al. 2003)
    - the random effects are normally distributed with a mean of 0 and variance as estimated from the data (Singer et al. 2003)

Linearity

- For linear growth, the rate of change per unit of time is assumed to be equal and change in a linear fashion
- Linearity of the relationship between the predictors and each growth factor in the model if a linear relationship is assumed for this

[123] Participants should have the potential to complete at least three time points for the linear model to be over-identified. Similarly, at least four time points for a quadratic model, and so on for higher order polynomials. Over-identification occurs when there is more observed than estimated information (Curran et al. 2003)

Measurement error **(ε_{it})** [124]

Measurement errors are……

- Multi-variately normally distributed with a mean of zero (Curran et al. 2003)
- Independent between the time points (Byrne et al. 2003)
- Independent for time points within an individual i.e. (cov (ε_{it} $\varepsilon_{it'}$) = 0 (Hox et al. 2005)
- Independent of intercept or shape parameters in the model (e.g. cov (ε_{it} α_i) = 0, cov (ε_{it} β_i) = 0 for a linear trajectory over time) (Hox et al. 2005).
- Shown to have homoscedastic variance, i.e. have the same variance at each time point (if time is the only term in the model) (Byrne et al. 2003) and for each level of any categorical predictors included in the model

[124] If these assumptions are not met, the model may need to be redefined (e.g. adding additional predictors or model correlations) to improve model specification (Kahn 2011)

Appendix 18: Types of missing data mechanisms

Three definitions have been proposed in the literature to describe the types of missing data mechanisms that can occur in a study (Little et al. 1987). They describe data that are (a) missing completely at random (MCAR) i.e. the probability of having missing data is independent of observed and unobserved data (b) missing at random (MAR) i.e. the missing data probability is dependent on observed data but independent of unobserved data and (c) missing not at random (MNAR) i.e. missing data is dependent on unobserved data (Twisk et al. 2013). These missing data mechanisms can best be illustrated by examples. For example, if data were missing due to participants moving from the study geographical area, this data would be MCAR if the decision to move did not depend on study outcomes. This data is MCAR as removal of such missing data from the analysis would not lead to bias, just a less efficient analysis (StataCorp, 2013c). In contrast, if outcome data in a clinical trial were more likely to be missing for participants in a particular treatment arm this data would be MAR as it is related to an observed study variable (i.e. treatment arm). This is then contrasted to data that are MNAR that could occur if participants with a particularly high value on the outcome of interest are more likely to drop out of the study than those with a low value on the outcome. This data is MNAR as the probability of missing data is related to a variable that has not been observed for participants with missing data (StataCorp, 2013c). For data that are MAR or MNAR, deletion of such missing data may lead to biased results (StataCorp, 2013c).

Appendix 19: Descriptive statistics of the AUSCAN distributions

	AUSCAN pain			AUSCAN function		
	N	Mean (SD)	Median (IQR)	N	Mean (SD)	Median (IQR)
All participants						
Baseline	589	3.1 (2.2)	3.0 (1.5, 4.5)	593	2.8 (2.4)	2.2 (0.8, 4.7)
18-months	577	3.2 (2.4)	3.5 (0.5, 5.0)	576	2.8 (2.4)	2.5 (0.3, 4.7)
3-years	523	3.3 (2.3)	3.5 (1.5, 5.0)	537	3.1 (2.4)	2.5 (1.1, 5.0)
5-years	417	3.3 (2.2)	3.5 (1.5, 5.0)	418	3.0 (2.4)	2.8 (0.8, 5.0)
7.5-years	384	3.2 (2.2)	3.0 (1.5, 5.0)	384	2.9 (2.3)	2.5 (0.8, 4.7)
Participants with AUSCAN pain and function data at all time-points						
Baseline	301	2.9 (2.1)	2.5 (1.5, 4.5)	301	2.4 (2.1)	1.9 (0.6, 3.9)
18-months	301	2.9 (2.1)	3.0 (0.5, 4.5)	301	2.5 (2.1)	2.5 (0.3, 4.4)
3-years	301	3.0 (2.1)	2.5 (1.5, 4.5)	301	2.6 (2.2)	2.2 (0.8, 4.2)
5-years	301	3.2 (2.2)	3.0 (1.5, 4.5)	301	2.8 (2.3)	2.5 (0.8, 4.4)
7.5-years	301	3.2 (2.3)	3.0 (1.5, 5.0)	301	2.9 (2.3)	2.2 (0.8, 4.7)

SD = standard deviation; IQR = inter-quartile range.

Appendix 20: Growth models - modelling strategy

Growth models were fitted separately for hand pain and function and were progressively increased in complexity according to the stages listed below. As a suitable transformation could not be found to approximate the AUSCAN scores to a normal distribution (transformations tried included the natural log + 1, square root, square and cubic), estimates were calculated using robust standard errors and the Satorra-Bentler Scaled Chi-square Test (SBSCT) was used to test whether the extra level of complexity was needed at each modelling stage.

List of growth models fitted to the hand pain and function data
1) Fixed and random intercept 2) Fixed and random intercept and slope 3) Fixed intercept, slope and quadratic term, but only a random intercept and slope 4) Fixed and random intercept, slope and quadratic term 5) Fixed intercept, slope and quadratic and cubic term, but only a random intercept and slope

Appendix 21: Correlation matrices of AUSCAN hand pain and function by study time point

	AUSCAN pain				
	Baseline	18-months	3-years	5-years	7.5-years
Baseline	1.00 (N=589)				
18-months	0.58 (N=546)	1.00 (N=577)			
3-years	0.64 (N=494)	0.65 (N=509)	1.00 (N=523)		
5-years	0.62 (N=394)	0.56 (N=405)	0.70 (N=402)	1.00 (N=417)	
7.5 years	0.57 (N=362)	0.59 (N=375)	0.64 (N=371)	0.70 (N=350)	1.00 (N=384)
	AUSCAN function				
	Baseline	18-months	3-years	5-years	7.5-years
Baseline	1.00 (N=593)				
18-months	0.74 (N=550)	1.00 (N=576)			
3-years	0.80 (N=510)	0.76 (N=522)	1.00 (N=537)		
5-years	0.78 (N=397)	0.71 (N=407)	0.85 (N=416)	1.00 (N=418)	
7.5 years	0.74 (N=365)	0.71 (N=375)	0.78 (N=381)	0.81 (N=348)	1.00 (N=384)

Appendix 22: Missing data patterns for AUSCAN function

	Missing data pattern (X=data present)					Number of time points with non-missing data
	Baseline	18-months	3-years	5-years	7.5-years	
	N=593	N=576	N=537	N=418	N=384	
N (%)						
322 (52)	X	X	X	X	X	5
78 (13)	X	X	X			3
64 (10)	X	X	X	X		4
49 (8)	X	X				2
33 (5)	X	X	X		X	4
30 (5)	X					1
17 (3)		X	X	X	X	4
6 (<1)	X		X	X	X	4
6 (<1)		X	X			2
3 (<1)	X		X			2
3 (<1)	X		X	X		3
2 (<1)			X	X	X	3
2 (<1)		X	X	X		3
2 (<1)	X	X			X	3
2 (<1)						0
1 (<1)	X		X		X	3
1 (<1)	X	X		X		3
1 (<1)	X	X		X	X	4
1 (<1)		X				1

Footnote: 52%, 19%, 14%, 9% and 5% of participants had data at 5, 4, 3, 2, 1 time points respectively

Appendix 23: Alternative modelling strategies that could have been used to select the predictors of interest in the growth models

Alternative modelling strategies could include those that do not consider the block structure that was imposed on the predictors of interest. For backwards deletion this would involve fitting a model with all predictors in it, so parameter estimates could be unreliable due to data drop-out. For forward selection, this would not have considered that some predictors are less costly and easier to collect than others.

A full stepwise model procedure could have been applied, but this was not practical given that the number of predictors was large and model comparisons were done “by hand”.

Forward or backwards selection could have been used to select the strongest predictors in each block before combining these into a single model. This would be simpler than the modelling strategy chosen as it does not consider the order that the blocks are analysed, but the method is not in keeping with the objective of the analysis to see if predictors with more complex forms of measurement are needed in the model.

Testing whether a factor of importance is a predictor of the model intercept first, prior to then testing whether it is also a predictor of the model slope could have been used, although this strategy may miss predictors that strongly predict the model slope but not the intercept. This method would have proved problematic if the objective of the analysis was to explore factors that predict how the outcomes change over time.

Appendix 24: Illustrative examples of how the coefficients in a growth model can vary depending on other predictors included in the model

Both of the examples below illustrate that the factor coefficients in a growth model should only be interpreted in light of other predictors in the model and, for prediction models that the main emphasis is on model fit and performance of the optimal combination of factors, rather than on the interpretation of the individual predictors per se. To explore the association of individual factors, an explanatory model would need to be built based on clear hypotheses regarding potential confounders, which was not the aim of the models in this thesis.

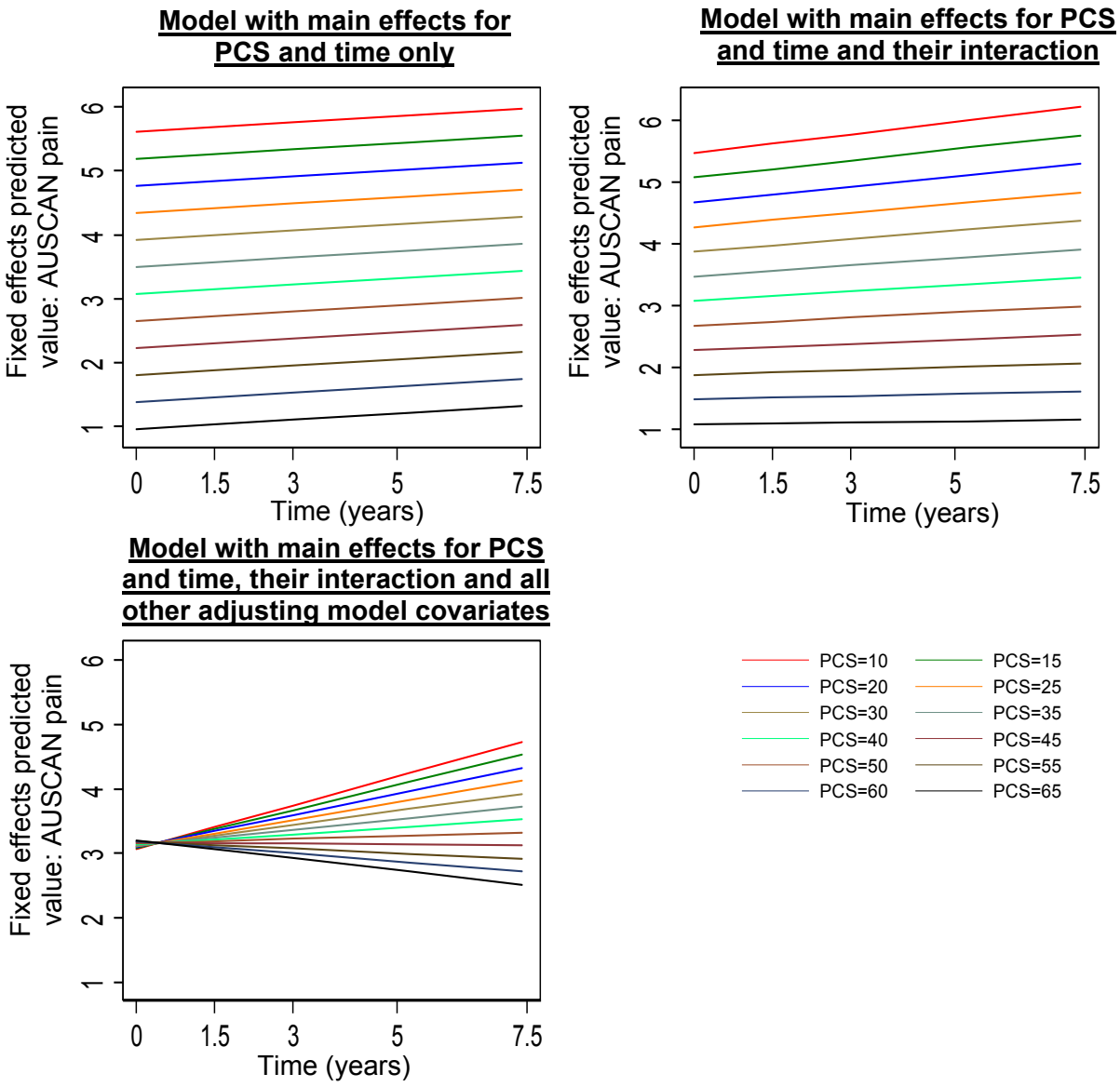
Example 1: Gender

In Chapter 7, the intercept coefficient for gender was positive in the hand pain model and negative in the hand function model. This was further explored, by re-running the model without adjusting for all other predictors in the model. The intercept coefficient for male gender changed to be negative for both hand pain and function (-0.45 for hand pain and -1.14 for hand function with 95% confidence intervals of -0.77, -0.13 and -1.49, -0.79 respectively). These unadjusted models therefore indicate that both hand pain and function are less severe for women than men, which is more in line with previous studies reporting that women tend to experience worse hand pain and functional difficulty than men (Dahaghin et al. 2005b), and illustrates a different conclusion than if the coefficients were interpreted from the combined model.

Example 2: SF-12 Physical function score (PCS)

The figure below shows interaction plots for PCS from the hand pain model. They are plots of the relationship between time and the predicted value of the hand pain outcome for 12 different values of PCS. The values of PCS are equally spaced from 10 to 65, with

such limits chosen to encompass the minimum and maximum values of PCS in the data. It shows that the magnitude of the slope depends on the value of baseline PCS, but that the strength and nature of this relationship depends very much on whether all of the other baseline predictors are included in the model alongside PCS, or not.



Appendix 25: Multinomial logistic regression

Multinomial logistic regression is used to model outcomes with two or more unordered categories. It works by simultaneously estimating a set of logistic regression equations that predict the log of the probability of being in each outcome group relative to a reference group (Biesheuvel et al. 2008). For an outcome with three possible groups (1, 2, and 3), where group 1 is the reference category, and a single (dichotomous or continuous) predictor (X) the logistic regression equations are:

$$log\left(\frac{P(group\ 2)}{P(group\ 1\ )}\right) = \beta_{21} + \beta_{22}X = lp_2$$

$$log\left(\frac{P(group\ 3)}{P(group\ 1\ )}\right) = \beta_{31} + \beta_{32}X = lp_3$$

Only two equations are needed as only two probabilities need to be estimated (the third probability can be calculated by knowing that the sum of all three probabilities is one). The logistic regression equations can be re-written (using algebra) to make the probability values the subject of the regression equations

$$P\ (group\ 2) = \frac{\exp(lp_2)}{1 + exp(lp_2) + exp(lp_3)}$$

$$P(group\ 3) = \frac{\exp(lp_3)}{1 + exp(lp_2) + exp(lp_3)}$$

The probability of being in the reference group is then calculated from the above two probabilities as:

$$P(group\ 1) = 1 - P(group\ 2) - P(group\ 3)$$

A set of three probabilities are then estimated for each individual in the dataset. The model can be extended to outcomes that contain more than three categories and to models with multiple predictors using the principals outlined here.

The regression coefficients from the model (i.e. β_{22} and β_{32}) can be transformed, by taking their exponential, to calculate a relative risk ratio (RRR)[125]. This value is the factor by which the relative risk (i.e. the probability of being in a group relative to the probability of being in the reference category) would change if there was a unit increase in the predictor of interest (holding all other variables in the model constant) (Long et al. 2006).

[125] Some authors refer to this value in the literature as an odds ratio rather than a relative risk ratio (RRR). The term RRR is used here as this is the convention used in STATA (i.e. the software used for analysis). The STATA community highlight that the term odds ratio is misleading (as the denominator in the outcome modelled is not measured as (1 – the probability of being in the group); it is the probability of being in the reference group, which is not the same quantity. They also highlight that the term relative-risk ratio is misleading too. In epidemiology, relative-risk generally refers to the probability of having an outcome if you have the predictor of interest (e.g. an exposed group) compared to not having the predictor of interest (e.g. an unexposed group), however here, the term "relative risk" refers to the probability of being in the outcome group of interest compared to the reference group so it doesn't directly refer to an exposed/unexposed predictor variable until the ratio part is calculated from the model (Gutierrez 2005)

Appendix 26: Calculation of Nagelkerke's pseudo R-square and the Brier score

Nagelkerke's pseudo R-square

The Nagelkerke's pseudo R-square is calculated by comparing the model likelihood for the model without the predictors in it ($(L(0)$), to a model with the predictors in it ($L(\tilde{\theta})$). The numerator in the equation is the Cox and Snell pseudo R-square, which is rescaled by the denominator in the formula to a 0 to 1 scale, to give the Nagelkerke's R-square value (Hu et al. 2006).

$$Nagelkerke's\ pseudo\ R^2 = \frac{1 - \left(\frac{L(0)}{L(\tilde{\theta})}\right)^{\frac{2}{n}}}{1 - \left(L(0)\right)^{\frac{2}{n}}}$$

Brier score

The Brier score is an aggregate measure of disagreement between the observed outcome and a prediction (StataCorp 2013d) and is calculated for a multinomial model using the formula below (Brier 1950). The formula is expressed in the context of the models fitted in this chapter and illustrates that the Brier score is essentially measuring the average squared error difference (StataCorp 2013d) between the predicted and observed data.

$$Brier\ score = \frac{1}{n}\sum_{j=1}^{r}\sum_{i=1}^{n}\left(p_{ij} - E_{ij}\right)^2$$

where i = participant number, j= trajectory group number from the multinomial logistic regression model (ranging from 1 to r), p = predicted probability, and E = 1 if the participant is in the trajectory group from the LCGM, 0 otherwise and n = total number of participants in the sample.

Appendix 27: Using the 'Conditional risk method' to transform the predicted probabilities from a multinomial logistic regression model prior to calculating a set of C-statistics (Van Calster et al. 2012a)

Consider a multinomial model with three outcome categories (A, B, and C). Define A as the reference group and B as the group with which to test discrimination.

Rescale the predicted values to be on a scale of 0 to 1, by defining a new probability variable for each person as:

New probability = P(B)/P(A) +P(B)

This new probability is then used in an ROC curve analysis to compare model discrimination between groups A and B and to calculate a C-statistic for that pairwise comparison.

A similar approach can then be used to compare category A to category C

Appendix 28: The accuracy paradox

This example is taken from http://en.wikipedia.org/wiki/Accuracy_paradox and illustrates that although model 2 has a higher level of accuracy, it is less useful than model 1. In model 2 there is no need for a model to be used as all participants are assigned to having the condition of interest in the sample.

Model 1

	Predict "Yes"	**Predict "No"**	**Total**
True "yes"	9700	150	9850
True "No"	50	100	150

Accuracy = 98%

Model 2

	Predict "Yes"	**Predict "No"**	**Total**
True "yes"	9850	0	9850
True "No"	150	0	150

Accuracy = 98.5%

Appendix 29: Goodness-of-fit for LCGM fitted to AUSCAN Pain (N=311) – complete case analysis

Model type	AIC	BIC	ABIC	Entropy	VLMR LRT	Adjusted LMR LRT	PB LRT	Group N	Average posterior probability
Linear									
1	6821	6847	6825	N/A	N/A	N/A	N/A	311	1.0
2	6244	6281	6250	0.85	p<0.001	p<0.001	p<0.001	124, 187	0.97, 0.95
3	6070	6118	6077	0.83	p<0.001	p<0.001	p<0.001	56, 146, 109	0.94, 0.90, 0.95
4	6031	6090	6040	0.77	p=0.037	p=0.043	p<0.001	89, 72, 43, 107	0.87, 0.88, 0.91, 0.83
5	6020	6091	6031	0.79	p=0.540	p=0.550	p<0.001	109, 40, 9, 70, 83	0.83, 0.94, 0.79, 0.89, 0.86
6	6014	6097	6027	0.81	p=0.510	p=0.516	p=0.013	76, 103, 7, 81, 40, 4	0.89, 0.83, 0.88, 0.86, 0.93, 0.85
7	6011	6105	6025	0.81	p=0.053	p=0.055	p=0.058	86, 10, 68, 5, 39, 102, 1	0.86, 0.80, 0.89, 0.81, 0.90, 0.81, 1.00
Quadratic									
1	6823	6853	6827	N/A	N/A	N/A	N/A	311	1.0
2	6246	6291	6253	0.85	p<0.001	p<0.001	p<0.001	130, 181	0.95, 0.96
3	6072	6132	6081	0.83	p<0.001	p<0.001	p<0.001	108, 147, 56	0.95, 0.90, 0.94
4	6032	6107	6044	0.78	p=0.019	p=0.022	p<0.001	43, 64, 97, 107	0.93, 0.90, 0.87, 0.83
5	6024	6113	6037	0.79	p=0.268	p=0.283	p=0.050	91, 101, 41, 68, 10	0.86, 0.81, 0.94, 0.89, 0.83
6	6014	6118	6030	0.81	p=0.230	p=0.242	p<0.001	92, 62, 39, 102, 6, 10	0.86, 0.91, 0.93, 0.80, 0.77, 0.84
7	6012	6131	6030	0.82	p=0.698	p=0.706	p=0.162	105, 10, 38, 77, 65, 6, 10	0.81, 0.69, 0.94, 0.87, 0.90, 0.79, 0.83
Cubic									
1	6825	6858	6830	N/A	N/A	N/A	N/A	311	1.0
2	6250	6302	6258	0.85	p<0.001	p<0.001	p<0.001	182, 129	0.96, 0.95
3	6076	6147	6086	0.84	p<0.001	p<0.001	p<0.001	56, 108, 147	0.94, 0.95, 0.90
4	6038	6128	6052	0.77	p=0.140	p=0.148	p<0.001	69, 103, 96, 43	0.90, 0.82, 0.85, 0.93
5	6021	6130	6038	0.78	p=0.142	p=0.149	p<0.001	93, 68, 25, 81, 44	0.88, 0.91, 0.72, 0.79, 0.93
6	6013	6140	6033	0.81	p=0.586	p=0.594	p<0.001	9, 42, 91, 62, 20, 87	0.79, 0.95, 0.86, 0.92, 0.79, 0.79
7	6007	6153	6029	0.82	p=0.163	p=0.168	p=0.030	79, 25, 6, 57, 23, 26, 95	0.83, 0.81, 0.89, 0.93, 0.87, 0.79, 0.87

Abbreviations: AIC, Akaike Information Criteria; BIC, Bayesian Information Criteria; ABIC, Sample-size adjusted BIC; VLMR LRT, Vuong-Lo-Mendell-Rubin likelihood ratio test; LMR LRT, Lo-Mendell-Rubin likelihood ratio test; PBLRT, parametric bootstrapped likelihood ratio test; N/A, not applicable; p = p-value. Highlighting indicates models with the lowest AIC, BIC, ABIC values, models with one group less than the model with a non-significant LRT p-value, group frequencies less than 5% of the sample (i.e. N <15) and posterior probabilities <0.7. All models achieved a global solution as in each model the largest log-likelihood was replicated for more than two random starting values.

Appendix 30: Goodness-of-fit for LCGM fitted to AUSCAN Function (N=322) – complete case analysis

Model type	AIC	BIC	ABIC	Entropy	VLMR LRT	Adjusted LMR LRT	PB LRT	Group N	Average posterior probability
Linear									
1	7149	7175	7153	N/A	N/A	N/A	N/A	322	1.0
2	6160	6198	6166	0.92	p<0.001	p<0.001	p<0.001	124, 198	0.96, 0.98
3	5845	5894	5853	0.90	p=0.001	p=0.002	p<0.001	153, 101, 68	0.96, 0.94, 0.95
4	5737	5797	5747	0.88	p=0.027	p=0.031	p<0.001	121, 30, 98, 73	0.95, 0.96, 0.90, 0.94
5	5678	5749	5689	0.86	p=0.035	p=0.041	p<0.001	53, 19, 82, 98, 70	0.94, 0.93, 0.87, 0.91, 0.91
6	5624	5707	5637	0.87	p=0.045	p=0.052	p<0.001	94, 46, 52, 18, 29, 83	0.92, 0.88, 0.95, 0.96, 0.86, 0.89
7	Global solution not obtained (largest log likelihood not replicated in two or more model solutions)								
Quadratic									
1	7150	7180	7155	N/A	N/A	N/A	N/A	322	1.0
2	6163	6208	6170	0.92	p<0.001	p<0.001	p<0.001	198, 124	0.98, 0.96
3	5846	5907	5856	0.90	p=0.001	p=0.001	p<0.001	150, 103, 69	0.97, 0.93, 0.95
4	5739	5814	5751	0.88	p=0.037	p=0.040	p<0.001	99, 120, 73, 30	0.90, 0.96, 0.94, 0.96
5	5682	5772	5696	0.86	p=0.153	p=0.163	p<0.001	19, 70, 53, 98, 82	0.93, 0.91, 0.94, 0.92, 0.86
6	5628	5733	5645	0.87	p=0.169	p=0.178	p<0.001	52, 18, 84, 26, 94, 48	0.95, 0.96, 0.87, 0.88, 0.93, 0.87
7	Global solution not obtained (largest log likelihood not replicated in two or more model solutions)								
Cubic									
1	7152	7186	7158	N/A	N/A	N/A	N/A	322	1.0
2	6166	6219	6175	0.92	p<0.001	p<0.001	p<0.001	198, 124	0.98, 0.96
3	5852	5923	5863	0.90	p=0.002	p=0.002	p<0.001	150, 103, 69	0.97, 0.93, 0.96
4	5743	5834	5758	0.88	p=0.026	p=0.029	p<0.001	99, 120, 73, 30	0.90, 0.96, 0.93, 0.97
5	5682	5791	5699	0.87	p=0.045	p=0.049	p<0.001	82, 97, 56, 18, 69	0.88, 0.93, 0.92, 0.96, 0.92
6	5628	5756	5648	0.87	p=0.191	p=0.201	p<0.001	95, 18, 28, 81, 52, 48	0.93, 0.98, 0.87, 0.89, 0.96, 0.88
7	Global solution not obtained (largest log likelihood not replicated in two or more model solutions)								

Abbreviations: AIC, Akaike Information Criteria; BIC, Bayesian Information Criteria; ABIC, Sample-size adjusted BIC; VLMR LRT, Vuong-Lo-Mendell-Rubin likelihood ratio test; LMR LRT, Lo-Mendell-Rubin likelihood ratio test; PBLRT, parametric bootstrapped likelihood ratio test; N/A, not applicable; p = p-value. Highlighting indicates models with the lowest AIC, BIC, ABIC values, models with one group less than the model with a non-significant LRT p-value, group frequencies less than 5% of the sample (i.e. N <15) and posterior probabilities <0.7.

Appendix 31: Goodness-of-fit Statistics for a LCGM fitted to AUSCAN Pain (N=621[α])

Model type	AIC	BIC	ABIC	Entropy	VLMR LRT	Adjusted LMR LRT	PB LRT	Group N	Average posterior probability
Linear									
1	11125	11156	11134	N/A	N/A	N/A	N/A	621	1.0
2	10231	10276	10244	0.80	p<0.001	p<0.001	p<0.001	343, 278	0.95, 0.94
3	9958	10016	9974	0.78	p<0.001	p<0.001	p<0.001	286, 120, 215	0.88, 0.90, 0.92
4	9885	9956	9905	0.72	p=0.001	p=0.001	p<0.001	165, 72, 187, 197	0.87, 0.88, 0.81, 0.81
5	9866	9950	9890	0.74	p=0.009	p=0.012	p<0.001	189, 85, 7, 155, 185	0.78, 0.83, 0.91, 0.87, 0.80
6	9852	9950	9880	0.73	p=0.315	p=0.330	p<0.001	7, 186, 177, 152, 14, 85	0.92, 0.80, 0.74, 0.85, 0.75, 0.84
7	9846	9956	9877	0.75	p=0.607	p=0.616	p=0.020	173, 88, 10, 182, 7, 154, 7	0.76, 0.83, 0.81, 0.81, 0.92, 0.85, 0.69
Quadratic									
1	11126	11162	11136	N/A	N/A	N/A	N/A	621	1.0
2	10231	10284	10246	0.80	p<0.001	p<0.001	p<0.001	343, 278	0.95, 0.94
3	9954	10025	9974	0.78	p<0.001	p<0.001	p<0.001	116, 288, 217	0.91, 0.87, 0.92
4	9881	9970	9906	0.72	p=0.001	p=0.001	p<0.001	158, 194, 74, 195	0.88, 0.82, 0.88, 0.80
5	9863	9969	9893	0.74	p=0.030	p=0.034	p<0.001	86, 185, 189, 8, 153	0.84, 0.80, 0.79, 0.89, 0.87
6	9842	9966	9877	0.73	p=0.060	p=0.068	p<0.001	8, 158, 167, 27, 87, 174	0.89, 0.86, 0.73, 0.74, 0.84, 0.81
7	9830	9972	9870	0.75	p=0.251	p=0.266	p<0.001	167, 7, 179, 8, 26, 154, 80	0.74, 0.81, 0.81, 0.87, 0.76, 0.86, 0.84
Cubic									
1	11128	11168	11140	N/A	N/A	N/A	N/A	621	1.0
2	10235	10297	10253	0.80	p<0.001	p<0.001	p<0.001	343, 278	0.95, 0.94
3	9958	10042	9982	0.78	p<0.001	p<0.001	p<0.001	119, 287, 215	0.91, 0.88, 0.92
4	9886	9992	9916	0.73	p=0.002	p=0.002	p<0.001	197, 71, 188, 165	0.80, 0.89, 0.82, 0.88
5	9859	9987	9895	0.71	p=0.137	p=0.145	p<0.001	81, 159, 198, 144, 39	0.88, 0.73, 0.83, 0.87, 0.72
6	9835	9985	9877	0.73	p=0.106	p=0.113	p<0.001	140, 90, 8, 49, 198, 136	0.86, 0.85, 0.94, 0.72, 0.80, 0.73
7	9824	9997	9873	0.75	p=0.393	p=0.404	p<0.001	84, 144, 194, 9, 138, 8, 44	0.85, 0.75, 0.80, 0.74, 0.86, 0.93, 0.72

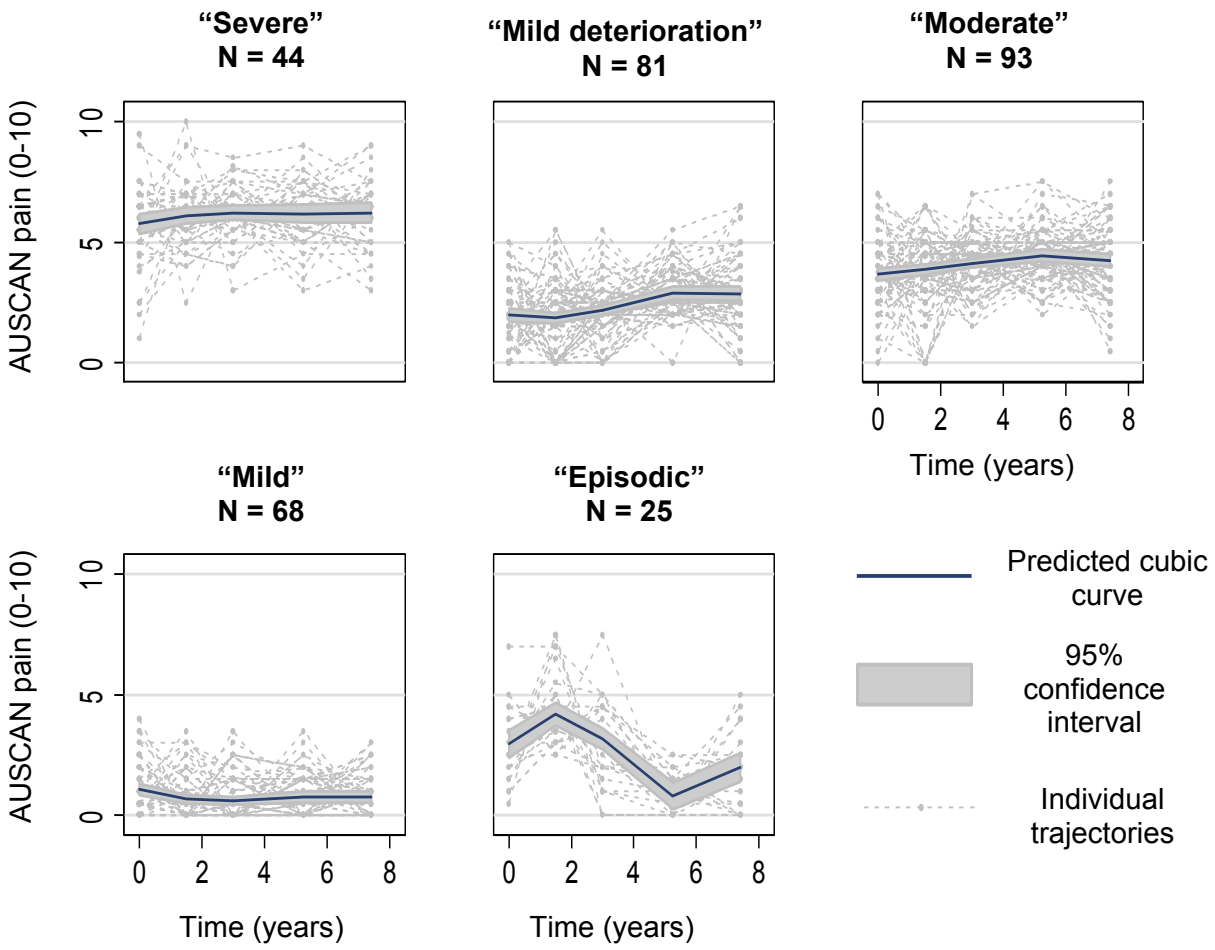
Abbreviations: AIC, Akaike Information Criteria; BIC, Bayesian Information Criteria; ABIC, Sample-size adjusted BIC; VLMR LRT, Vuong-Lo-Mendell-Rubin likelihood ratio test; LMR LRT, Lo-Mendell-Rubin likelihood ratio test; PBLRT, parametric bootstrapped likelihood ratio test; N/A, not applicable; p = p-value. Highlighting indicates models with the lowest AIC, BIC, ABIC values, models with one group less than the model with a non-significant LRT p-value, group frequencies less than 5% of the sample (i.e. N <30) and posterior probabilities <0.7. All models achieved a global solution as in each model the largest log-likelihood was replicated for more than two random starting values [α] Two participants were excluded from the analysis as they had no data at all time-points.

Appendix 32: Goodness-of-fit Statistics for a LCGM fitted to AUSCAN Function (N=621[α])

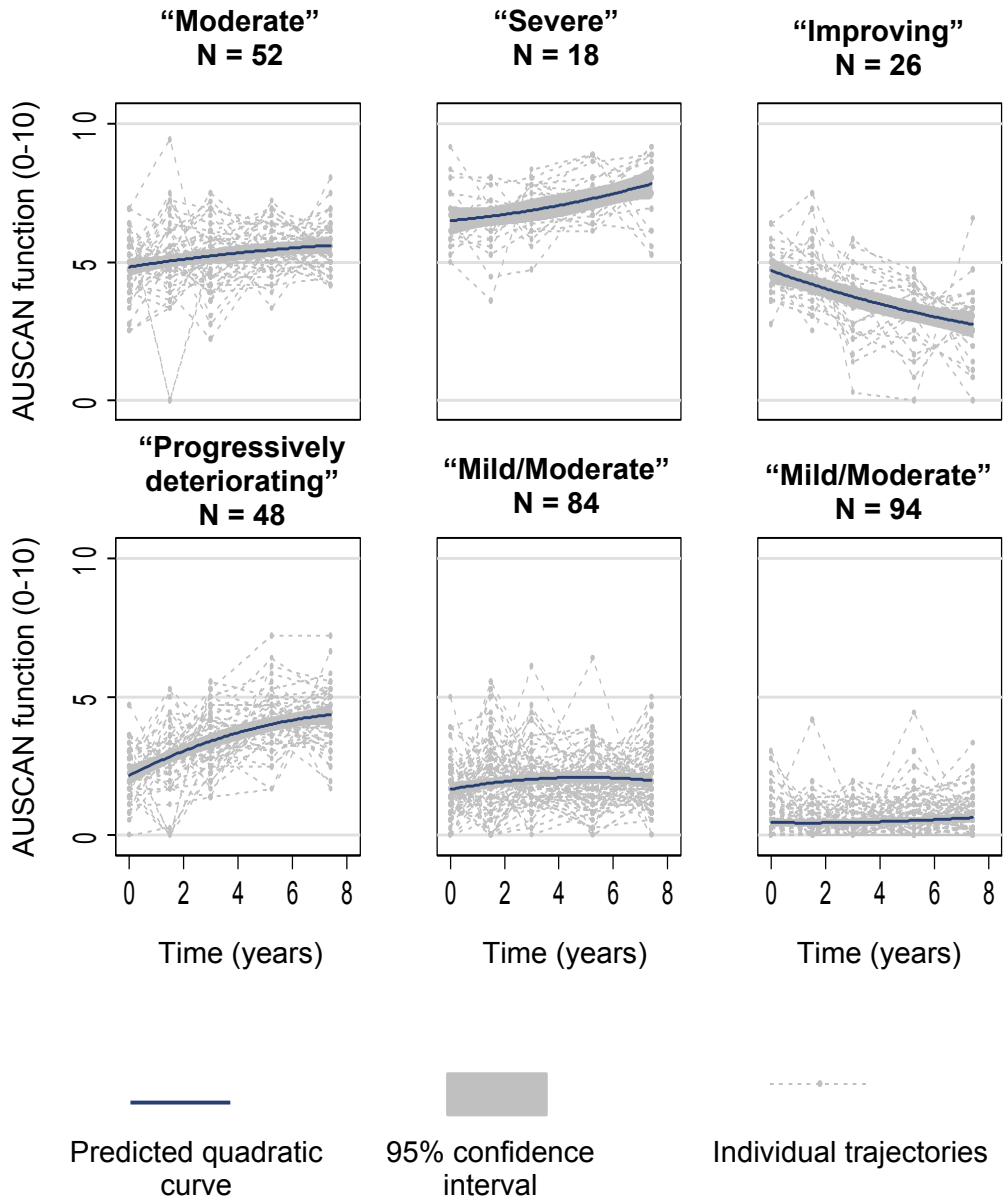
Model type	AIC	BIC	ABIC	Entropy	VLMR LRT	Adjusted LMR LRT	PB LRT	Group N	Average posterior probability
Linear									
1	11519	11550	11528	N/A	N/A	N/A	N/A	621	1.0
2	10059	10104	10072	0.89	p<0.001	p<0.001	p<0.001	247, 374	0.96, 0.97
3	9588	9646	9604	0.86	p=0.077	p=0.084	p<0.001	123, 201, 297	0.94, 0.90, 0.96
4	9415	9486	9435	0.84	p<0.001	P=0.001	p<0.001	258, 171, 128, 64	0.95, 0.87, 0.90, 0.91
5	9346	9430	9370	0.78	p=0.008	p=0.010	p<0.001	129, 60, 173, 117, 142	0.84, 0.91, 0.88, 0.88, 0.78
6	9307	9404	9335	0.77	p=0.243	p=0.258	p<0.001	139, 119, 65, 72, 59, 167	0.78, 0.87, 0.74, 0.73, 0.91, 0.87
7	9278	9388	9309	0.78	p=0.026	p=0.030	p<0.001	70, 67, 114, 6, 59, 138, 167	0.74, 0.74, 0.86, 0.92, 0.88, 0.79, 0.87
Quadratic									
1	11519	11554	11529	N/A	N/A	N/A	N/A	621	1.0
2	10056	10109	10071	0.89	p<0.001	p<0.001	p<0.001	246, 375	0.96, 0.97
3	9585	9656	9605	0.86	p=0.093	p=0.099	p<0.001	291, 129, 201	0.96, 0.93, 0.90
4	9408	9496	9433	0.84	p=0.002	p=0.002	p<0.001	258, 64, 167, 132	0.94, 0.91, 0.88, 0.89
5	9339	9446	9370	0.79	p=0.101	p=0.109	p<0.001	118, 61, 137, 177, 128	0.88, 0.90, 0.79, 0.88, 0.84
6	9302	9427	9338	0.77	p=0.589	p=0.598	p<0.001	121, 58, 70, 64, 138, 170	0.86, 0.91, 0.73, 0.74, 0.78, 0.87
7	9274	9416	9314	0.79	p=0.761	p=0.764	p<0.001	137, 62, 169, 7, 114, 74, 58	0.78, 0.75, 0.88, 0.88, 0.87, 0.74, 0.89
Cubic									
1	11520	11560	11531	N/A	N/A	N/A	N/A	621	1.0
2	10056	10118	10073	0.89	p<0.001	p<0.001	p<0.001	375, 246	0.97, 0.96
3	9586	9670	9610	0.86	p=0.070	p=0.074	p<0.001	202, 293, 126	0.90, 0.96, 0.93
4	9409	9516	9439	0.84	p=0.017	p=0.018	p<0.001	64, 165, 258, 134	0.91, 0.88, 0.94, 0.89
5	9335	9464	9372	0.79	p=0.017	p=0.019	p<0.001	128, 137, 177, 61, 118	0.84, 0.79, 0.88, 0.91, 0.88
6	9299	9450	9342	0.81	p=0.440	p=0.450	p<0.001	131, 118, 60, 175, 124, 13	0.79, 0.84, 0.92, 0.87, 0.87, 0.86
7	9271	9444	9320	0.82	p=0.232	p=0.238	p<0.001	22, 7, 117, 58, 113, 127, 177	0.80, 0.88, 0.88, 0.89, 0.83, 0.80, 0.88

Abbreviations: AIC, Akaike Information Criteria; BIC, Bayesian Information Criteria; ABIC, Sample-size adjusted BIC; VLMR LRT, Vuong-Lo-Mendell-Rubin likelihood ratio test; LMR LRT, Lo-Mendell-Rubin likelihood ratio test; PBLRT, parametric bootstrapped likelihood ratio test; N/A, not applicable; p = p-value. Highlighting indicates models with the lowest AIC, BIC, ABIC values, models with one group less than the model with a non-significant LRT p-value, group frequencies less than 5% of the sample (i.e. N <30) and posterior probabilities <0.7. All models achieved a global solution as in each model the largest log-likelihood was replicated for more than two random starting values [α] Two participants were excluded from the analysis as they had no data at all time-points.

Appendix 33: Trajectory plots for AUSCAN pain based on participants with complete-case data only (N=311)



Appendix 34: Trajectory plots for AUSCAN function based on participants with complete-case data only (N=322)



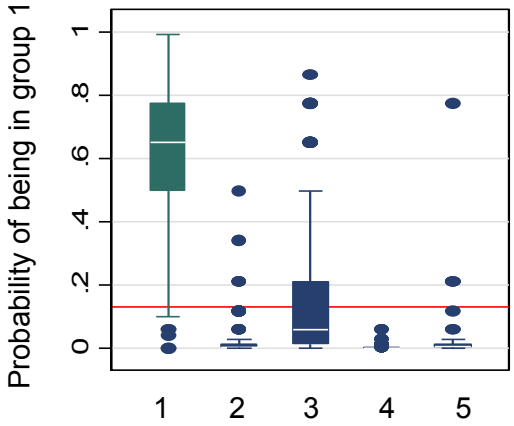
Appendix 35: Model goodness-of-fit and performance statistics for a prediction model of trajectory group membership that only contains the baseline for the outcome of interest as a predictor in the model

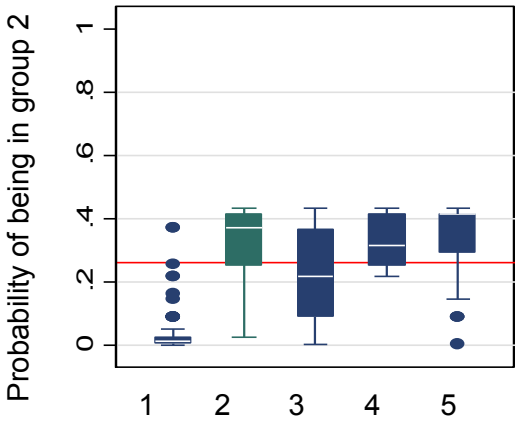
Goodness-of-fit measures and the C-statistic

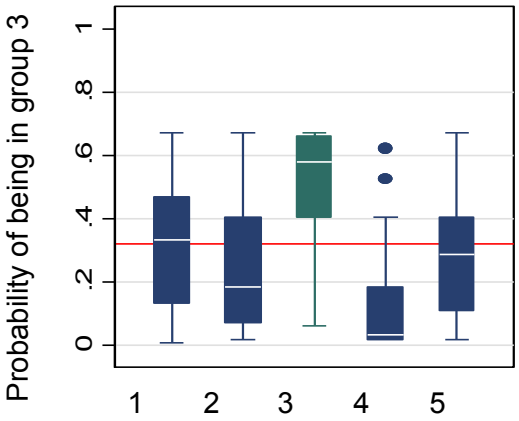
	Nagelkerke's pseudo R-square	Brier Score	C-statistic (95% confidence interval)
Hand pain			
Mild (reference group)			
Severe	0.64	0.55	0.99 (0.97, 1.0)
Mild deterioration			0.72 (0.66, 0.78)
Moderate			0.95 (0.93, 0.97)
Episodic			0.77 (0.69, 0.85)
Hand function			
Mild (reference group)			
Moderate	0.80	0.53	1.00 (0.99, 1.00)
Severe			1.00 (1.00, 1.00)
Improving			0.99 (0.98, 1.00)
Progressively Deteriorating			0.91 (0.87, 0.96)
Mild/Moderate			0.80 (0.74, 0.85)

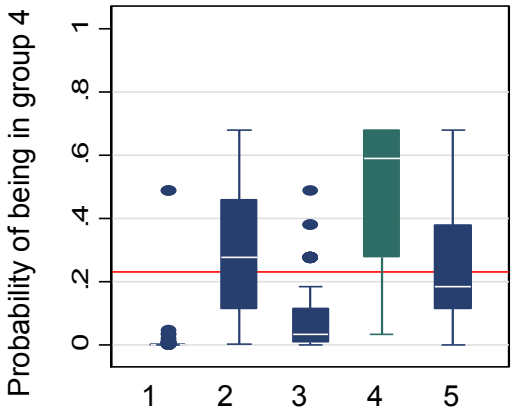
Discrimination plots

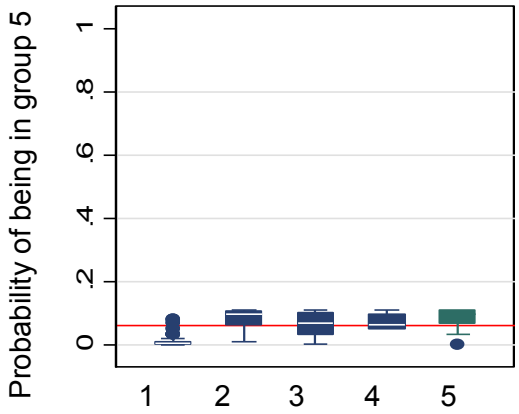
Hand pain









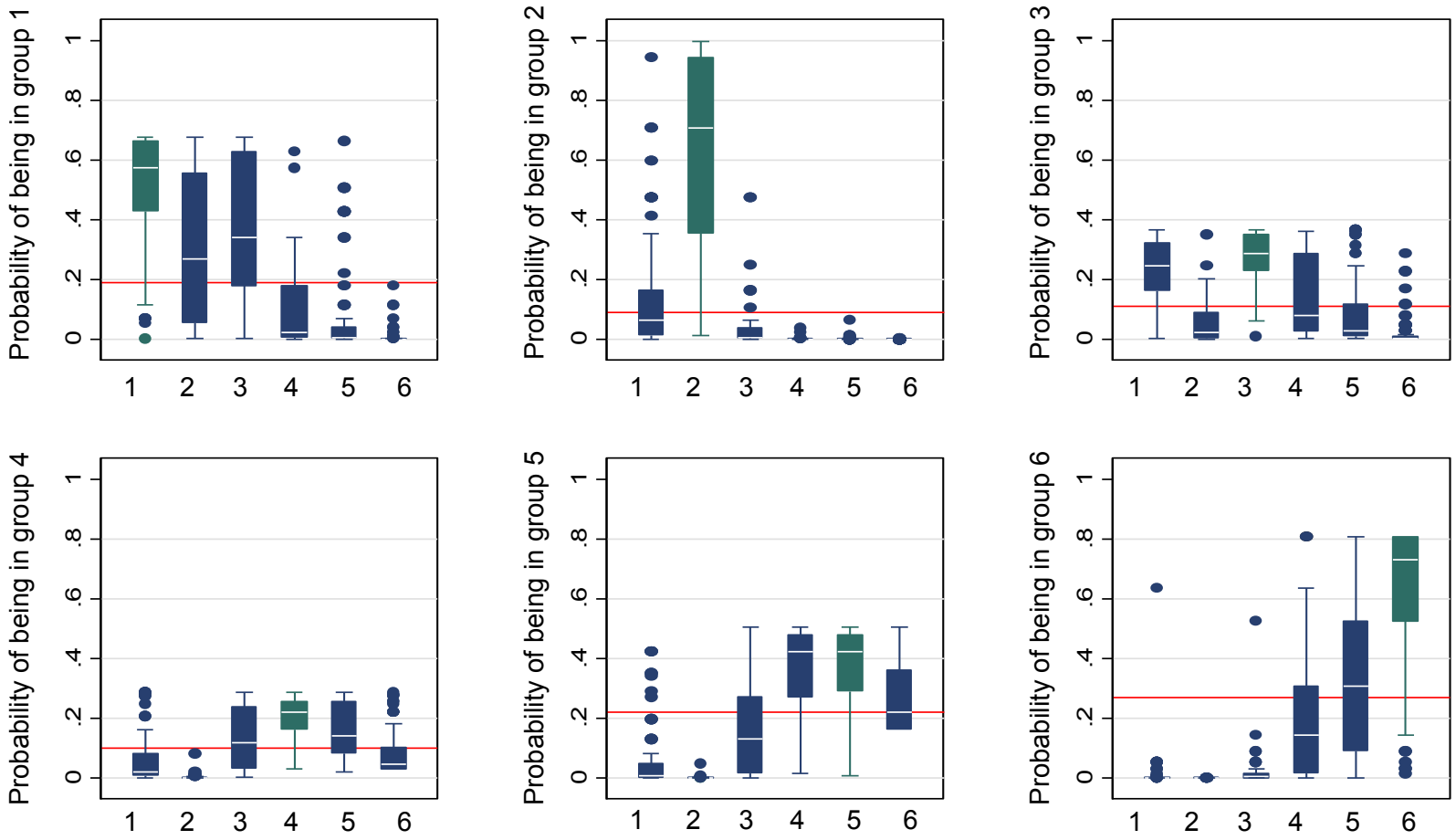


X-axis key (i.e. observed trajectory groups)

1 = “Severe”
2 = “Mild deterioration”
3 = “Moderate”
4 = “Mild”
5 = “Episodic”

Red reference line = group prevalence

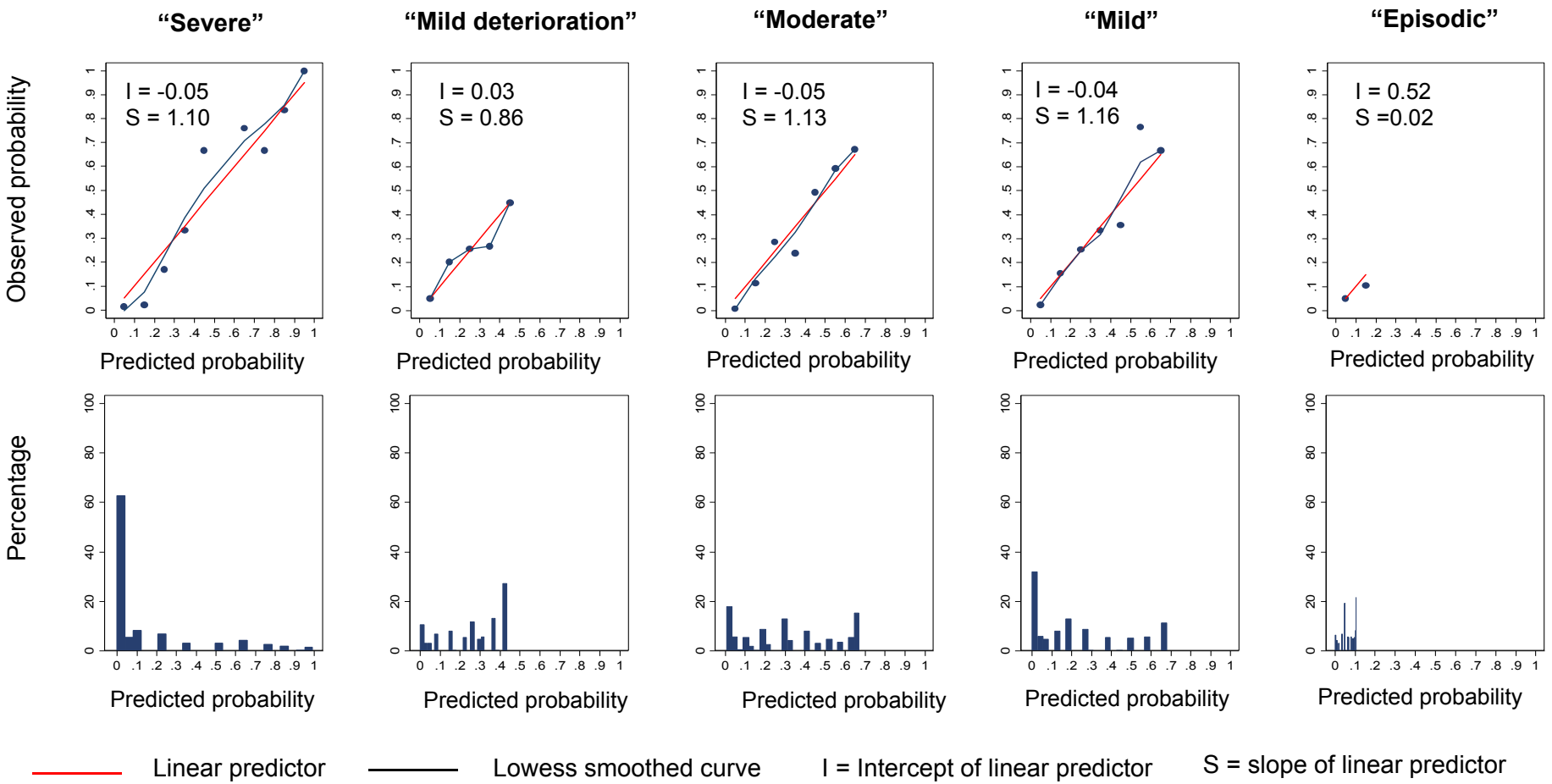
Hand function



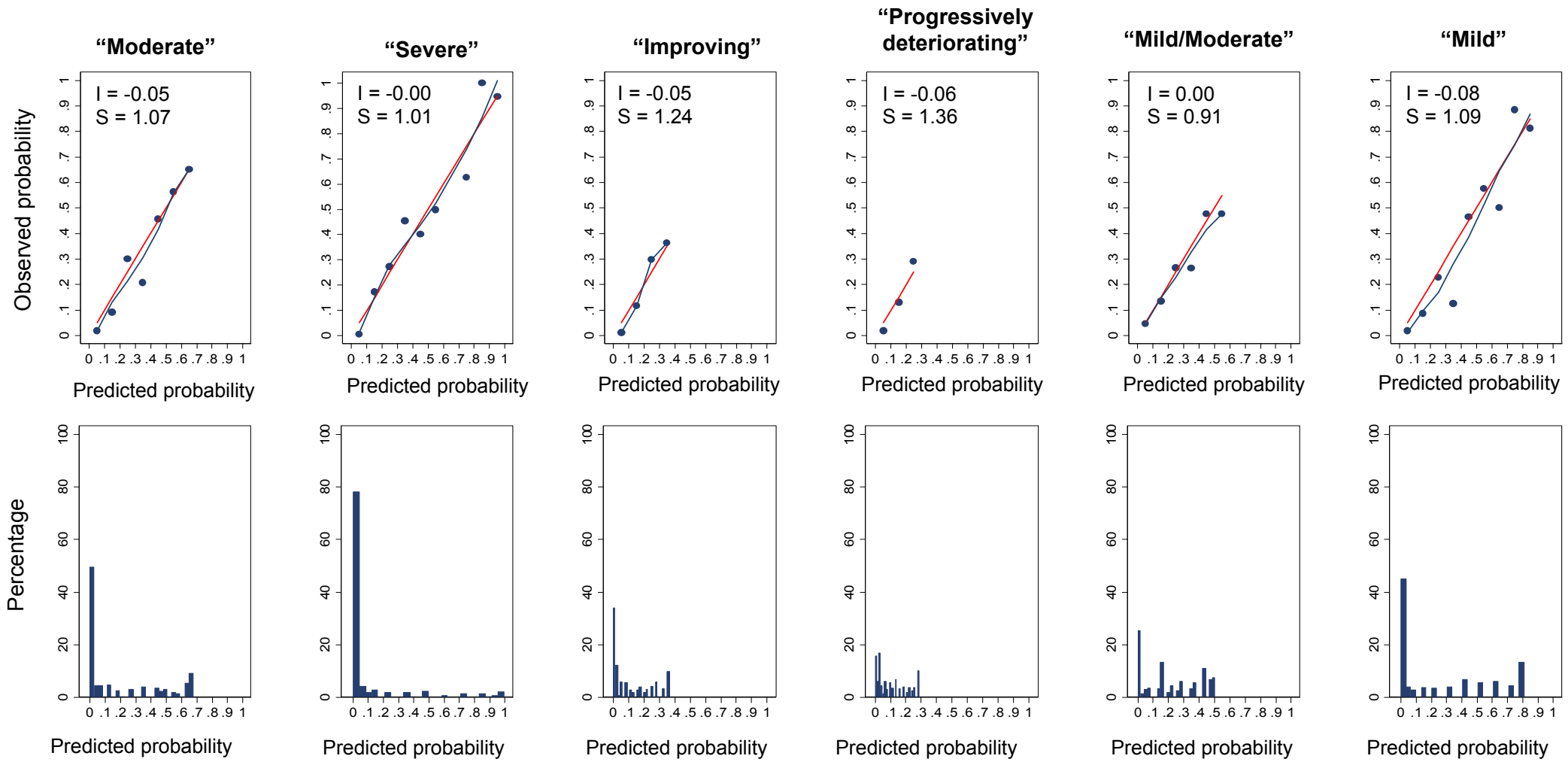
X-axis key (i.e. observed trajectory groups): 1 = “Moderate”, 2 = “Severe”, 3 = “Improving”, 4 = “Progressively deteriorating”, 5 = “Mild/moderate”, 6 = “Mild”. Red reference line = group prevalence

Calibration plots

Hand pain



Hand Function



Model accuracy

Hand pain

Trajectory group from the LCGM	Trajectory group membership predicted from the multinomial logistic regression					Total
	Severe	Mild deterioration	Moderate	Mild	Episodic	
Severe	61	0	18	2	0	81
Mild deterioration	1	72	40	39	0	152
Moderate	17	31	138	1	0	187
Mild	0	37	12	81	0	130
Episodic	1	20	10	8	0	39
Total	80	160	218	131	0	589

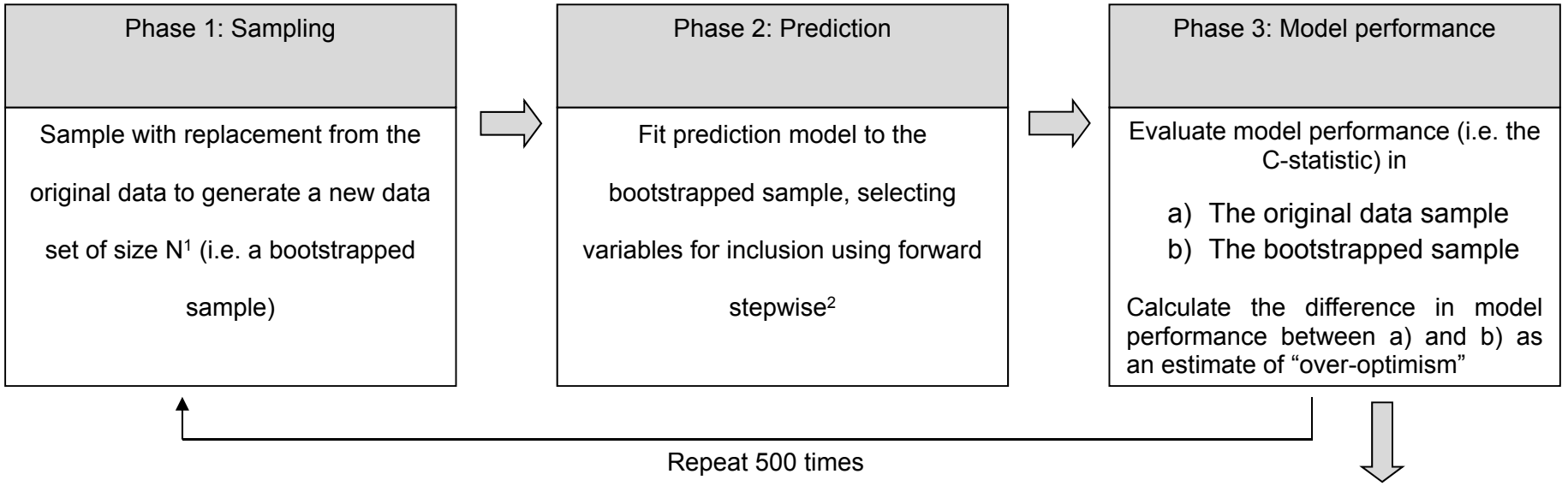
Percentage of participants accurately assigned by the model = 60%

Hand function

Trajectory group from the LCGM	Trajectory group membership predicted from the multinomial logistic regression						Total
	Moderate	Severe	Improving	Progressively Deteriorating	Mild/Moderate	Mild	
Moderate	88	12	7	0	8	1	116
Severe	15	39	0	0	0	0	54
Improving	32	2	21	0	13	1	69
Progressively Deteriorating	2	0	14	0	41	6	63
Mild/Moderate	9	0	5	0	75	42	131
Mild	0	0	1	0	34	125	160
Total	146	53	48	0	171	175	593

Percentage of participants accurately assigned by the model = 59%

Appendix 36: Using bootstrapping to estimate model over-optimism (method suggested by (Steyerberg et al. 2001))



Footnotes

1: N = the total number of people classified into a trajectory group

2: Using forward stepwise selection implies that variables in the prediction model will not necessarily be the same in each bootstrapped sample

Phase 4: Overall evaluation of model performance

a) Calculate an overall estimate of "over optimism" by taking estimates of over optimism at phase 3 and averaging then across all 500 bootstrapped samples

b) Calculate internal validated performance as: Model performance in the original data sample – estimate of "over optimism" from phase 4 stage a)

Appendix 37: Extending a linear parallel process model to include quadratic trajectories over time

Several quadratic models were fitted to the data as it was not clear (initially) which correlation terms between the growth factors should be included in the model. The confusion arose as there was a discrepancy between the correlations included in the model by default in Mplus (when the model was naturally extended to include a quadratic term), to those included in a study with an example of how to include a quadratic term into a parallel process growth model (Cole 2008).

Models one and two (below) describe the nature of the discrepancy, with model one showing the correlations that were automatically included by default in Mplus, and model two, the correlations suggested for inclusion by Cole (Cole 2008). In model two, the correlations in red are the additional correlations that are added to model one, and the correlations highlighted in yellow are those that are excluded from model one to give the model by Cole[126].

A personal communication with the developers of the Mplus software suggested that the choice of which correlations to include in the model should be driven by the research question of interest. This raises two questions:

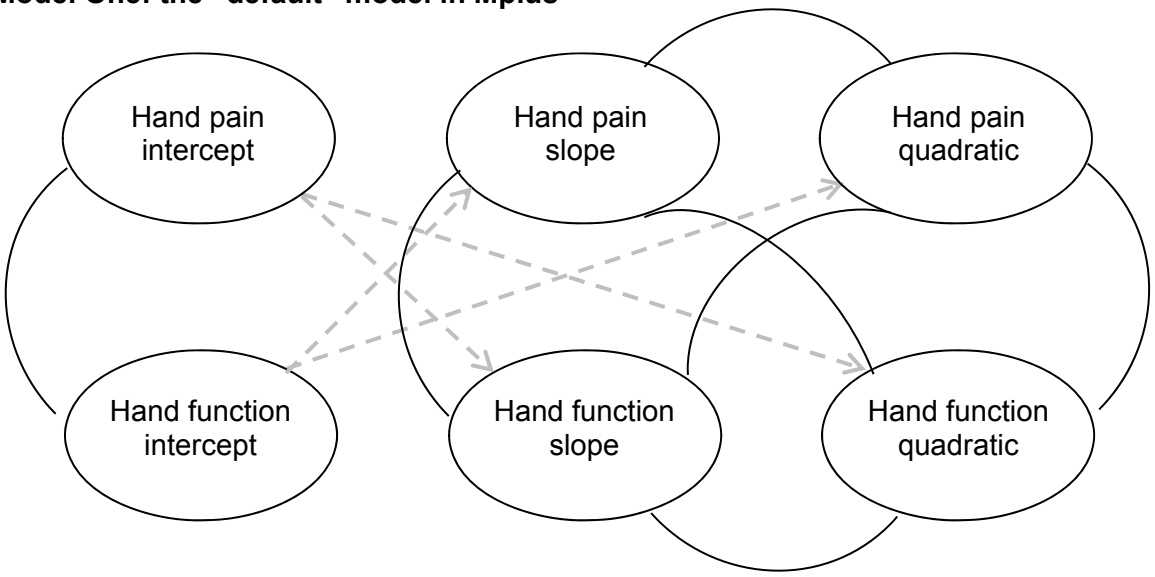
1) Is it plausible that the additional correlations in the model 2 are non-zero in the data? If so, they should be included in the model.
2) Is it plausible that the correlations that have been excluded from model one are non-zero? If so, they should be included in the model (as excluding a correlation term from the model is the same as saying it has a correlation of zero)

[126] The models in this appendix have been simplified to only show the growth factor of interest as there is no ambiguity around how the outcomes of interest should be included in the model – they are included as per the description given in Chapter 5

The initial plan was to test this in the data by producing a third model (model three) that included all of the correlations in models one or two, which could be used to test whether any of the correlations shown in green in model three were statistically significant. It transpired, however, that this approach was not possible as linear dependencies between the latent variables caused the estimated covariance matrix to be non-positive definite and model estimates to be unreliable[127]. The cause of the linear dependency surrounded the inclusion of the additional correlations shown in red in model two. These correlations were therefore removed (from model three) and the model re-run (the correlations highlighted in yellow in model two remained in the model as they were not the cause of the linear dependency between the latent variables)[128]. This model produced plausible estimates and is "the quadratic model" reported in chapter 9 (it is identical to the default model in Mplus and the correlations highlighted in yellow in model two are included to explore whether these correlations are statistically significant in the data)[129].

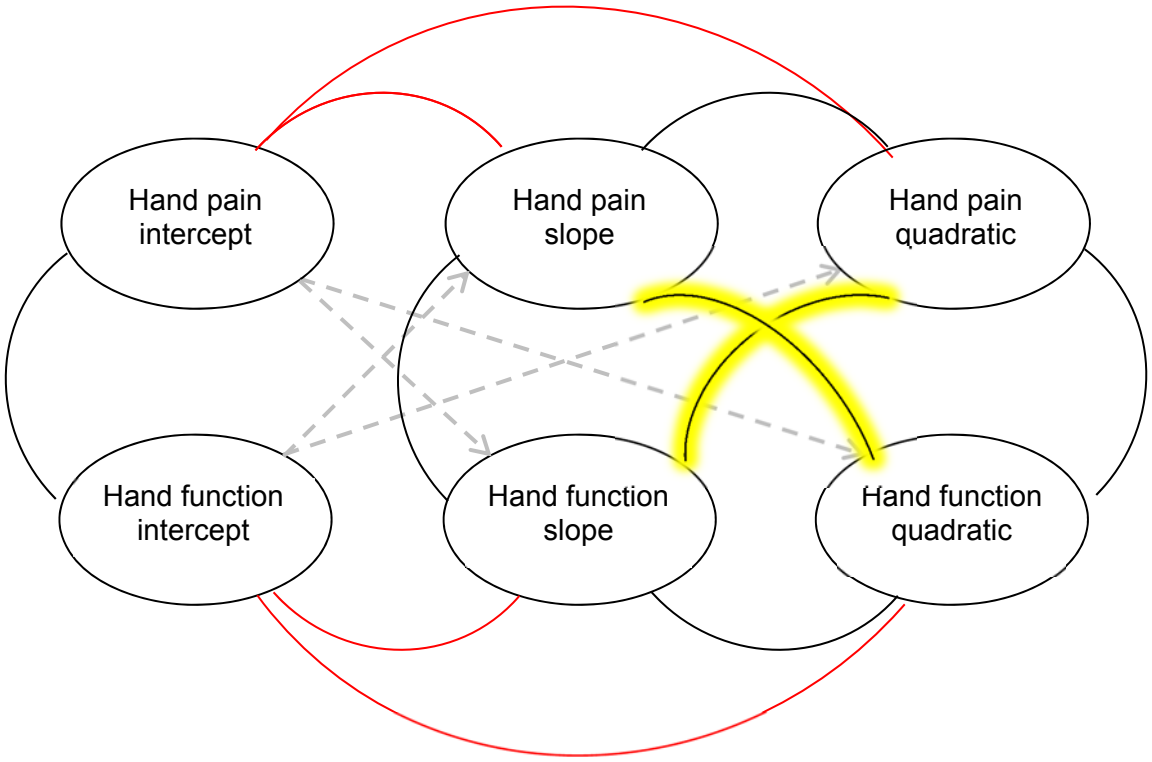
[127] A non-positive definite covariance matrix means that at least one eigenvalue in the covariance matrix is <=0. This causes estimation problems for the model (for further details see (Rigdon 1997))
[128] Linear dependencies still occurred between the latent variables if only the correlations highlighted in yellow in model two were removed from model three
[129] It was possible to eradicate the linear dependency between the latent growth factors by removing only one of the correlations shown in red in model 2, but this caused other estimates problems, namely correlations that were greater than one. All correlations in red in model 2 were therefore removed from the model, as it had previously been shown that the correlation between the random intercept and the random growth factors were not strong when the outcomes were modelled separately, thus supporting the removal of these correlations from the analysis

Model One: the "default" model in Mplus



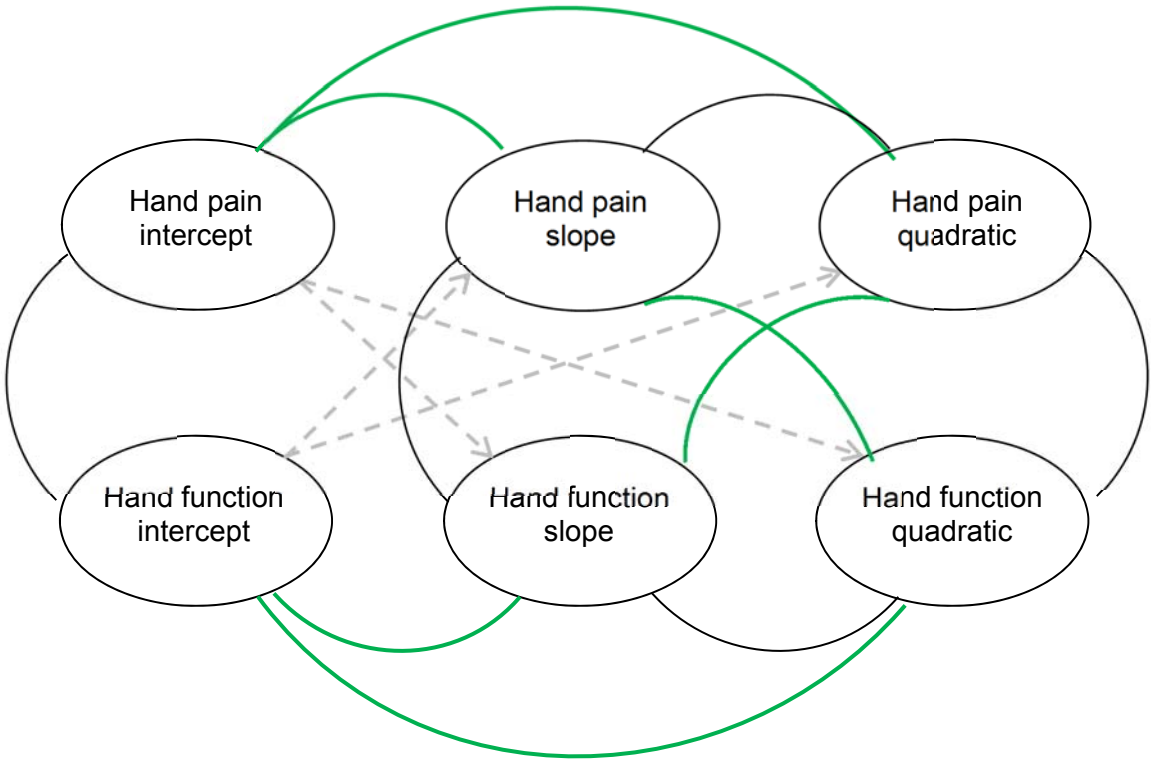
Curved arrows indicate correlations, dashed lines represent predictive relationships. The predictive relationships remain the same across models one to three

Model Two: the model by Cole (Cole 2008)



Curved arrows indicate correlations, dashed lines represent predictive relationships. The predictive relationships remain the same across models one to three

Model Three: The combined model



Curved arrows indicate correlations, dashed lines represent predictive relationships. The predictive relationships remain the same across models one to three

Appendix 38: Combinations of clinical conditions

Combination of conditions	OA – meets the ACR criteria	Carpal tunnel	Dupuytren's contractuture	De Quervain's tenosynovitis	Trigger finger	Number of participants
1	No	No	No	No	No	127
2	No	Yes	No	No	No	68
3	No	No	Yes	No	No	42
4	Yes	Yes	No	No	No	38
5	Yes	No	No	No	No	32
6	No	Yes	No	Yes	No	26
7	No	Yes	Yes	No	No	25
8	No	No	No	Yes	No	24
9	No	No	No	No	Yes	24
10	Yes	Yes	No	Yes	No	19
11	Yes	Yes	No	No	Yes	17
12	Yes	No	No	Yes	No	16
13	No	Yes	No	No	Yes	13
14	No	No	Yes	No	Yes	12
15	No	Yes	Yes	Yes	No	10
16	No	No	Yes	Yes	No	8
17	Yes	No	Yes	No	No	8
18	No	Yes	Yes	No	Yes	8
19	Yes	Yes	Yes	Yes	No	8
20	Yes	No	No	No	Yes	6
21	Yes	Yes	Yes	No	No	6
22	Yes	Yes	Yes	No	Yes	6
23	No	Yes	No	Yes	Yes	5
24	No	Yes	No	Missing	No	5
25	No	Missing	No	No	No	4
26	Yes	Yes	No	Missing	No	4
27	Yes	No	Yes	Yes	No	4

28	Yes	Yes	No	Yes	Yes	4
29	Yes	Yes	Yes	Missing	Yes	4
30	Yes	Missing	No	No	No	3
31	No	No	No	Missing	No	3
32	Yes	No	Yes	No	Yes	3
33	No	No	No	Yes	Yes	3
34	No	Missing	No	Missing	No	2
35	No	Yes	No	Missing	Yes	2
36	Yes	No	Yes	Missing	Yes	2
37	No	Missing	Yes	No	Yes	2
38	No	Missing	No	Yes	No	2
39	Yes	Missing	Yes	Missing	No	2
40	No	Missing	Yes	Missing	No	2
41	Yes	No	No	Missing	No	2
42	No	Missing	Yes	No	No	2
43	No	No	Yes	Yes	Yes	2
44	No	No	Yes	Missing	No	2
45	Yes	Yes	Yes	Yes	Yes	1
46	No	Missing	Yes	Yes	Yes	1
47	No	Missing	Yes	Yes	No	1
48	Missing	Yes	No	Missing	No	1
49	Yes	Yes	No	Missing	Yes	1
50	No	Yes	Yes	Yes	Yes	1
51	Yes	Missing	No	Yes	Yes	1
52	Yes	Yes	Yes	Missing	No	1
53	No	No	No	No	Missing	1
54	Yes	No	No	Missing	Yes	1
55	Yes	Missing	No	Missing	No	1
56	Yes	No	No	Yes	Yes	1
57	No	Missing	No	Missing	Yes	1
58	No	Yes	Yes	Missing	No	1
59	No	Yes	Yes	Missing	Yes	1
60	No	No	No	Missing	Yes	1

Appendix 39: Proportion of participants with each clinical condition stratified by presence of radiographic hand OA

Clinical condition	No joints with radiographic OA[α] N(%)	At least one joint with radiographic OA[α] N(%)
Carpal tunnel syndrome	43 (43)	216 (46)
Dupuytren's contracture	23 (22)	130 (27)
De Quervain's tenosynovitis	19 (18)	113 (24)
Trigger finger	19 (18)	95 (20)

α = Radiographic hand OA defined if at least one hand joint has a Kellgren and Lawrence (KL) grade >=2