Improving the assessment of fatigue in Axial Spondyloarthritis: improving patient outcomes

Nathan Ashley Pearson

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy in Health Science

Warwick Medical School, University of Warwick

October 2019
# Contents

List of tables ......................................................................................... viii
List of figures ......................................................................................... x
List of abbreviations ............................................................................... xi
Publications during thesis ..................................................................... xiii
Acknowledgements ................................................................................ xiv
Abstract ............................................................................................... xv

1. Chapter 1: Introduction to health outcome measurement in Axial Spondyloarthritis ......................................................................................................................... 1
   1.1. Introduction ...................................................................................... 1
   1.2. Axial Spondyloarthritis .................................................................. 2
       1.2.1. Diagnosis ................................................................................. 3
       1.2.2. Epidemiology .......................................................................... 4
       1.2.3. Aetiology .................................................................................. 5
       1.2.4. Symptoms ............................................................................... 5
   1.3. Measuring health outcomes ............................................................. 8
       1.3.1. PROMs taxonomy ................................................................... 10
       1.3.2. Application of PROMs in healthcare ........................................... 12
   1.4. Understanding and developing patient-reported outcome measures ...... 13
       1.4.1. Establishing the need for a new measure ................................. 14
       1.4.2. Development of a conceptual (measurement) framework .......... 15
       1.4.3. Crafting the first version of the PROM ....................................... 17
       1.4.4. Reviewing and modifying the PROM ........................................ 18
       1.4.5. Patient and public involvement in research ............................... 21
   1.5. Purpose of an axSpA-fatigue specific PROM ..................................... 23
       1.5.1. Aims and objectives ................................................................. 24
       1.5.2. Project design .......................................................................... 24

2. Chapter 2: Systematic review of the quality and acceptability of patient reported measures of fatigue used in axSpA ................................................................. 27
   2.1. Introduction .................................................................................... 27
2.2. Measuring fatigue in axSpA .................................................................27
2.3. Systematic reviews of PROMs .............................................................28
   2.1.1 Methodological quality (COSMIN) and measurement properties ......30
   2.1.2 Existing reviews of measures of fatigue ........................................38
   2.1.3 Review aims and objectives .........................................................40
2.4. Methods .............................................................................................41
   2.4.1. Search strategy ...........................................................................41
   2.4.2. Study selection ...........................................................................41
   2.4.3. Data extraction ...........................................................................42
   2.4.4. Assessment of study methodological quality .................................43
   2.4.5. Assessment of PROM quality ......................................................43
   2.4.6. Data synthesis ............................................................................43
   2.4.7. Comparative item content appraisal ............................................44
   2.4.8. PROM recommendations ............................................................44
   2.4.9. Sensitivity analysis .....................................................................45
2.5. Results ..................................................................................................45
   2.5.1. Identification of studies and PROMs ............................................45
   2.5.2. Study and sample characteristics ...............................................46
   2.5.3. Measurement properties and methodological quality ....................46
   2.5.4. Fatigue conceptualisation and patient involvement .......................46
   2.5.5. Multidimensional fatigue-specific PROMs .................................59
   2.5.6. Unidimensional fatigue PROMs ..................................................61
   2.5.7. Single-item fatigue PROMs .........................................................64
   2.5.8. Sensitivity analysis .....................................................................66
2.6. Discussion ............................................................................................67
2.7. Next step .............................................................................................71
3. Chapter 3: The lived experience of fatigue and energy in patients with axial
   spondyloarthritis ....................................................................................72
   3.1. Introduction .....................................................................................72
3.2. Background ..............................................................................................................72
3.2.1. Aims and objectives ..........................................................................................75
3.3. Methodology ...........................................................................................................81
3.3.1. Phenomenology .................................................................................................81
3.3.2. Hermeneutics .....................................................................................................83
3.3.3. Idiography principle ..........................................................................................83
3.3.4. Other methodological approaches .....................................................................84
3.4. Method .....................................................................................................................86
3.4.1. Recruitment ........................................................................................................86
3.4.2. Patient research partner (PRP) group involvement ...........................................87
3.4.3. Methods and data collection ..............................................................................88
3.4.4. Data analysis of lived experience ......................................................................88
3.4.5. Data analysis to develop a measurement framework .......................................89
3.4.6. Rigour ................................................................................................................89
3.5. Results .....................................................................................................................90
3.5.1. Participants .........................................................................................................90
3.5.2. Findings .............................................................................................................90
3.5.3. The patient lived experience of axSpA and fatigue .........................................91
3.5.4. Patient research partner (PRP) – review of study findings .............................109
3.5.5. Developing a measurement framework ..............................................................110
3.6. Discussion ...............................................................................................................116
3.7. Next step ................................................................................................................121
4. Chapter 4: Item generation and PROM refinement ..................................................122
4.1. Introduction ............................................................................................................122
4.2. Background ............................................................................................................122
4.2.1. Aims and objectives ..........................................................................................123
4.3. Item generation, selection and scale development ...............................................124
4.3.1. Item generation from existing measures of fatigue (‘PROM mapping’) ..........124
4.4. Overview of methods and results .................................................. 127

4.5. Method 1: reviewing the axSpA-fatigue and energy measurement
     framework and items ........................................................................ 127
     4.5.1. Aims and objectives ............................................................... 127
     4.5.2. Justification for using focus groups ........................................ 128
     4.5.3. Participants ........................................................................... 129
     4.5.4. Methods and data collection ................................................... 130
     4.5.5. Recording and transcription .................................................. 132
     4.5.6. Data analysis ......................................................................... 133
     4.5.7. Patient research partners ....................................................... 134

4.6. Results 1: reviewing the axSpA-fatigue and energy measurement
     framework ......................................................................................... 134
     4.6.1. Participants ........................................................................... 134
     4.6.2. Findings ................................................................................ 135
     4.6.3. Part one: Thematic analysis findings ....................................... 140
     4.6.4. Part two: Action points for PROM development ..................... 142
     4.6.5. Patient research partner group meeting ................................... 144

4.7. Method 2: pretesting the WASTEd ................................................. 150
     4.7.1. Aims and objectives ............................................................... 150
     4.7.2. Justification for approach ....................................................... 150
     4.7.3. Techniques in pretesting interviews ....................................... 151
     4.7.4. Cognitive interviews .............................................................. 153
     4.7.5. Three-step test interview (TSTI) ............................................ 153
     4.7.6. Chosen pretesting approach in this study ............................... 154
     4.7.7. Sampling .............................................................................. 155
     4.7.8. Participants .......................................................................... 155
     4.7.9. Methods and data collection ............................................... 155
     4.7.10. Data analysis and modifying the PROM ............................... 157
     4.7.11. Review and refinement meetings ........................................ 160

4.8. Results 2: pretesting the WASTEd ................................................. 161
5.4.5. Research team review ................................................................. 222
5.4.6. PRP group review ................................................................. 223

5.5. Discussion ................................................................................. 230

5.6. Next step (outside of the PhD) .................................................. 234

6. Chapter 6: Overall discussion of the thesis ................................... 235

6.1. Overview of thesis ................................................................... 235

6.2. Strengths and contributions of this PhD .................................... 239

6.2.1. Incorporating active involvement of a patient research partner group 239

6.2.2. Summary of knowledge contributions per phase of study ......... 243

6.3. Limitations ............................................................................. 245

6.4. Future research initiatives .......................................................... 246

6.4.1. Comprehensive psychometric evaluation of the WASTEd ........ 246

6.4.2. Cross-cultural translation studies .......................................... 247

6.4.3. Implications for clinical practice ............................................ 248

6.4.4. Challenges with mental health and emotional wellbeing .......... 249

6.4.5. Implications for research and PROM development .................. 250

6.4.6. Implications for intervention studies for axSpA fatigue .......... 251

6.5. Conclusion .............................................................................. 252

References ..................................................................................... 253

Appendix 1A: flowchart of project stages ........................................ 294

Appendix 2A: search strategy ............................................................ 295

Appendix 2B: characteristics of included studies (n=23) ................. 299

Appendix 3A: scoping review search strategy for qualitative studies of axSpA and fatigue .......................................................... 307

Appendix 3B: lived experience interview schedule ........................... 308

Appendix 3C: example extract of exploratory comments and initial emergent themes from IPA analysis for participant 4 .................. 311

Appendix 3D: example table of superordinate and subordinate themes for participant 5 ............................... 313
Appendix 3E: completed COREQ checklist .................................................................316
Appendix 4A: Results of PROM mapping ..................................................................319
Appendix 4B: Thematic analysis findings from focus groups (Chapter 4) ...............324
Appendix 4C: interview schedule for cognitive interviews .........................................329
Appendix 4D: interview schedule for TSTI ..................................................................332
Appendix 5A: factor structure for model 2 .................................................................336
Appendix 5B: item-total correlations and Cronbach’s α per model.........................339
Appendix 5C: the final short form WASTEd following psychometric evaluation....341
**List of tables**

Table 1.1: Types of spondyloarthritis ................................................................. 2
Table 1.2: ASAS axSpA classification criteria adapted from Rudwaleit et al. (7) .... 4
Table 2.1: Quality assessment criteria for measurement properties taken from Pearson et al. (146) .................................................................................. 31
Table 2.2: Eligibility criteria .................................................................................. 42
Table 2.3: Quality and data synthesis criteria with definitions .............................. 43
Table 2.4: Methodological quality (COSMIN) per study (n=23) per PROM (n=9) and investigated measurement properties adapted from Pearson et al. (146) ........................................................................ 48
Table 2.5: Data synthesis, levels of evidence and overall quality of reviewed PROMs (n=9) adapted from Pearson et al. (146) ................................................................. 54
Table 2.6: PROM content appraisal of reviewed single- and multi-item fatigue measures (n=9) utilising Hewlett’s model of RA fatigue, adapted from Pearson et al. (146) .................................................................................................. 56
Table 3.1: Methodologies and summated, non-synthesised findings of identified qualitative studies in axSpA and fatigue ................................................................. 76
Table 3.2: Sampling matrix showing the distribution of participants across key variables .................................................................................................................. 91
Table 3.3: Master table of superordinate and subordinate themes with definitions and key quotes from patients .................................................................................. 93
Table 3.4: Domains and subdomains identified and extracted for the purpose of underpinning future measurement ........................................................................ 111
Table 4.1: Extract of nominal group topic guide groups 1-3 .................................. 131
Table 4.2: List of themes and subthemes with supporting quotes from thematic analysis .................................................................................................................. 136
Table 4.3: Item list for pretesting interviews with all verbal probes ...................... 148
Table 4.4: Advantages and disadvantages of using ‘think aloud’ and ‘verbal probing’ (130) ...................................................................................................................... 152
Table 4.5: Coding framework for analysis pretesting data adapted from QAS-99 and including an FDA recommendation (89,292) ................................................................ 159
Table 4.6: Characteristics of pretesting interview participants ............................ 161
Table 4.7: Reference 32-item list tested in round 1 cognitive interviews .............. 161
Table 4.8: Summary of identified changes from cognitive interviews (n=5), research team review and PRP group review .............................................................. 167
Table 4.9: Summary of identified changes from TSTIs (n=5), research team review and PRP group review ...................................................................................... 175
Table 4.10: Final item list for field-test evaluation ................................................ 178
Table 5.1: General advantages and disadvantages of postal and online survey methods (299,300) ........................................................................................................................................ 186
Table 5.2: Questionnaire type, name and number of items in order of presentation ........................................................................................................................................ 194
Table 5.3: Category scheme to identify good or poor performing items ........................................................................................................................................ 195
Table 5.4: Person separation index – statistical strength and number of distinct groups adapted from (328) ........................................................................................................................................ 200
Table 5.5: Summary of statistical tests and applied quality criteria ........................................................................................................................................ 202
Table 5.6: Demographic details of respondents .................................................................................................................................................................................. 205
Table 5.7: Extracted factors for Model 1 ............................................................................................................................................................................................... 207
Table 5.8: Items removed in model 1 and 2 with brief rationale ........................................................................................................................................ 210
Table 5.9: Model 3 item factor loadings (bold numbers represent major factor loadings) ........................................................................................................................................ 211
Table 5.10: Summary of energy factor items identified as problematic following RMT analysis ........................................................................................................................................ 215
Table 5.11: Summary of fatigue factor items identified as problematic following RMT analysis ........................................................................................................................................ 218
Table 5.12: Summary of symptom factor items identified as problematic following RMT analysis ........................................................................................................................................ 221
Table 5.13: Summary of control factor items identified as problematic following RMT analysis ........................................................................................................................................ 222
List of figures

Figure 1.1: Conceptual model of RA-fatigue, taken from Hewlett et al. (50) .............8
Figure 1.2: WHO-ICF conceptual framework (115)..................................................16
Figure 1.3: An illustration of the PROM development process taken from Haywood et al. (59) .........................................................................................................................20
Figure 1.4: Photo of the PRP group and facilitator during a meeting ......................23
Figure 2.1: PRISMA flow chart of article identification and selection for this study adapted from Pearson et al. (146) .........................................................................................47
Figure 4.1: Working measurement blueprint for the WASTEd..............................147
Figure 5.1: Person-item threshold location distributions for the energy factor ......214
Figure 5.2: Example disordered threshold for item E6 (response option 0) ..........215
Figure 5.3: Person-item threshold location distributions for the fatigue factor ......217
Figure 5.4: Person-item threshold location distributions for the symptom factor....220
Figure 5.5: Substantial disordered threshold where the probability of selecting the middle (moderate) response option is significant regardless of underlying energy levels .........................................................................................................................220
Figure 5.6: Final measurement blueprint of axSpA fatigue and energy ..............229
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Term</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACP</td>
<td>Active Cancer Patient</td>
</tr>
<tr>
<td>AS</td>
<td>Ankylosing Spondylitis</td>
</tr>
<tr>
<td>ASAS</td>
<td>Assessment of Spondyloarthritis International Society</td>
</tr>
<tr>
<td>ASSIGNnw</td>
<td>Ankylosing Spondylitis Special Interest Group North West</td>
</tr>
<tr>
<td>AUC</td>
<td>Area Under the Curve</td>
</tr>
<tr>
<td>axSpA</td>
<td>Axial Spondyloarthritis</td>
</tr>
<tr>
<td>BASDAI</td>
<td>Bath Ankylosing Spondylitis Disease Activity Index</td>
</tr>
<tr>
<td>BASFI</td>
<td>Bath Ankylosing Spondylitis Functionality Index</td>
</tr>
<tr>
<td>BFI</td>
<td>Brief Fatigue Inventory</td>
</tr>
<tr>
<td>BRAF-MDQ</td>
<td>Bristol Rheumatoid Arthritis Fatigue Multi-Dimensional Questionnaire</td>
</tr>
<tr>
<td>CDAQ</td>
<td>Coeliac Disease Assessment Questionnaire</td>
</tr>
<tr>
<td>CFA</td>
<td>Confirmatory Factor Analysis</td>
</tr>
<tr>
<td>CFS</td>
<td>Chronic Fatigue Syndrome</td>
</tr>
<tr>
<td>CIRF</td>
<td>Cognitive Interview Reporting Framework</td>
</tr>
<tr>
<td>CIS</td>
<td>Checklist Individual Strength</td>
</tr>
<tr>
<td>COREQ</td>
<td>CONSolidated criteria for REporting Qualitative research</td>
</tr>
<tr>
<td>COSMIN</td>
<td>CONSensus-based Standards for the selection of health Measurement INstruments</td>
</tr>
<tr>
<td>CS</td>
<td>Cancer Survivor</td>
</tr>
<tr>
<td>CTT</td>
<td>Classic Test Theory</td>
</tr>
<tr>
<td>DA</td>
<td>Discourse Analysis</td>
</tr>
<tr>
<td>DEMQOL</td>
<td>Dementia Health-Related Quality of Life</td>
</tr>
<tr>
<td>DEMQOL</td>
<td>Dementia Health Related Quality of Life</td>
</tr>
<tr>
<td>DIF</td>
<td>Differential Item Functioning</td>
</tr>
<tr>
<td>EASi-QoL</td>
<td>Evaluation of Ankylosing Spondylitis Quality of Life</td>
</tr>
<tr>
<td>EFA</td>
<td>Exploratory Factor Analysis</td>
</tr>
<tr>
<td>EQ-5D</td>
<td>EuroQoL</td>
</tr>
<tr>
<td>ES</td>
<td>Effect Size</td>
</tr>
<tr>
<td>ESSG</td>
<td>European Spondyloarthopathy Study Group</td>
</tr>
<tr>
<td>FA</td>
<td>Factor Analysis</td>
</tr>
<tr>
<td>FACIT</td>
<td>Functional Assessment of Chronic Illness Therapy</td>
</tr>
<tr>
<td>FDA</td>
<td>US Food and Drug Administration</td>
</tr>
<tr>
<td>FSS</td>
<td>Fatigue Severity Scale</td>
</tr>
<tr>
<td>GF</td>
<td>General Fatigue</td>
</tr>
<tr>
<td>GT</td>
<td>Grounded Theory</td>
</tr>
<tr>
<td>HADS</td>
<td>Hospital Anxiety and Depression Scale</td>
</tr>
<tr>
<td>HIV</td>
<td>Human Immunodeficiency Virus</td>
</tr>
<tr>
<td>HLA</td>
<td>Human Leucocyte Antigen</td>
</tr>
<tr>
<td>HRQoL</td>
<td>Health-Related Quality of Life</td>
</tr>
<tr>
<td>ICC</td>
<td>Intra-class Correlation Coefficient</td>
</tr>
<tr>
<td>ICF</td>
<td>International Classification of Functioning, Disability and Health</td>
</tr>
<tr>
<td>IPA</td>
<td>Interpretative Phenomenological Analysis</td>
</tr>
<tr>
<td>IRT</td>
<td>Item Response Theory</td>
</tr>
<tr>
<td>ISPOR</td>
<td>International Society for Pharmacoeconomics and Outcomes Research</td>
</tr>
<tr>
<td>ITC</td>
<td>Item-Total Correlation</td>
</tr>
<tr>
<td>JM</td>
<td>Jane Martindale</td>
</tr>
<tr>
<td>JP</td>
<td>Jonathan Packham</td>
</tr>
<tr>
<td>KH</td>
<td>Kirstie Haywood</td>
</tr>
<tr>
<td>LT</td>
<td>Elizabeth Tutton</td>
</tr>
<tr>
<td>MAF</td>
<td>Multi-dimensional Assessment of Fatigue</td>
</tr>
<tr>
<td>MFI-20</td>
<td>Multi-dimensional Fatigue Inventory</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>MFSI-SF</td>
<td>Multi-dimensional Fatigue Symptom Inventory–Short Form</td>
</tr>
<tr>
<td>mNY</td>
<td>Modified New York criteria</td>
</tr>
<tr>
<td>MRI</td>
<td>Magnetic Resonance Imaging</td>
</tr>
<tr>
<td>MS</td>
<td>Multiple Sclerosis</td>
</tr>
<tr>
<td>NASS</td>
<td>National Ankylosing Spondylitis Society</td>
</tr>
<tr>
<td>NGT</td>
<td>Nominal Group Technique</td>
</tr>
<tr>
<td>NHS</td>
<td>National Health Service</td>
</tr>
<tr>
<td>nr-axSpA</td>
<td>Non-radiographic Axial Spondyloarthritis</td>
</tr>
<tr>
<td>NRS</td>
<td>Numeric Rating Scale</td>
</tr>
<tr>
<td>NSAID</td>
<td>Non-steroidal Anti-Inflammatory Drugs</td>
</tr>
<tr>
<td>OMERACT</td>
<td>Outcome Measures in Rheumatology</td>
</tr>
<tr>
<td>PCM</td>
<td>Partial Credit Model</td>
</tr>
<tr>
<td>PFS</td>
<td>Piper Fatigue Scale</td>
</tr>
<tr>
<td>PPI</td>
<td>Patient and Public Involvement in Research</td>
</tr>
<tr>
<td>PRISMA</td>
<td>Preferred Reporting Items for Systematic Reviews and Meta-Analyses</td>
</tr>
<tr>
<td>PRO</td>
<td>Patient-Reported Outcome</td>
</tr>
<tr>
<td>PROM</td>
<td>Patient-Reported Outcome Measure</td>
</tr>
<tr>
<td>PROMIS</td>
<td>Patient-Reported Outcomes Measurement Information System</td>
</tr>
<tr>
<td>PRP</td>
<td>Patient Research Partner</td>
</tr>
<tr>
<td>PS</td>
<td>Physical Sensations</td>
</tr>
<tr>
<td>PsA</td>
<td>Psoriatic Arthritis</td>
</tr>
<tr>
<td>PsAID</td>
<td>Psoriatic Arthritis Impact of Disease questionnaire</td>
</tr>
<tr>
<td>PSI</td>
<td>Person Separation Index</td>
</tr>
<tr>
<td>QAS</td>
<td>Questionnaire Appraisal System</td>
</tr>
<tr>
<td>QoL</td>
<td>Quality of Life</td>
</tr>
<tr>
<td>RA</td>
<td>Rheumatoid Arthritis</td>
</tr>
<tr>
<td>Ra</td>
<td>Reduced Activity</td>
</tr>
<tr>
<td>REC</td>
<td>Research Ethics Committee</td>
</tr>
<tr>
<td>RM</td>
<td>Reduced Motivation</td>
</tr>
<tr>
<td>RMT</td>
<td>Rasch Measurement Theory</td>
</tr>
<tr>
<td>RSM</td>
<td>Rating Scale Model</td>
</tr>
<tr>
<td>SD</td>
<td>Standard Deviation</td>
</tr>
<tr>
<td>SEM</td>
<td>Standard Error of Measurement</td>
</tr>
<tr>
<td>SF-36 VT</td>
<td>Short-Form Health Status Survey (36-item) Vitality Subscale</td>
</tr>
<tr>
<td>SpA</td>
<td>Spondyloarthritis</td>
</tr>
<tr>
<td>SRM</td>
<td>Standardised Response Mean</td>
</tr>
<tr>
<td>SRQR</td>
<td>Standards for Reporting Qualitative Research</td>
</tr>
<tr>
<td>TNF</td>
<td>Tumour Necrosis Factor</td>
</tr>
<tr>
<td>TSTI</td>
<td>Three-Step Test Interview</td>
</tr>
<tr>
<td>VAS</td>
<td>Visual Analogue Scale</td>
</tr>
<tr>
<td>WASTEd</td>
<td>Warwick Axial Spondyloarthritis measure of Fatigue and Energy</td>
</tr>
<tr>
<td>WF-NRS</td>
<td>Worst-Fatigue Numeric Rating Scale</td>
</tr>
<tr>
<td>WHO</td>
<td>World Health Organisation</td>
</tr>
</tbody>
</table>
Publications during thesis


Acknowledgements

This research was completed as part of a funding grant by the National Ankylosing Spondylitis Society (NASS) who have been incredibly supportive and encouraging throughout the completion of my degree.

I feel incredibly privileged to have had the opportunity to work with a such fantastic research team which includes my academic supervisors and patient research partners. I am thankful to my supervisors, Drs Kirstie Haywood, Elizabeth Tutton, Helen Parsons, Jane Martindale and Jon Packham for their unwavering support throughout my PhD. I have been incredibly fortunate to have a supervisory team who I could to turn to for support during both academic and personal challenges.

I would like to extend my thanks to the patient research partner group for their interest, support and enthusiasm for our project. Particularly, I am extremely grateful for the relationship we have been able to develop over the last 4 years and their warm welcome and engagement during and after our meetings. I would also like to thank all the patients and healthcare professionals who participated in my research studies.

Finally, I would like to dedicate my thesis to my two grandfathers who unfortunately passed away at the start, and end of my PhD experience. They were two of my biggest supporters and always encouraged me to keep going. Their loss will always be felt, but their energy, excitement and enthusiasm for me and what I do will stay with me forever.
Abstract

Background: Fatigue is an important symptom for people with axial spondyloarthritis (axSpA). Current measurement relies on a single item visual analogue scale of fatigue severity that provides a limited assessment of fatigue. This thesis sought to improve the understanding and measurement of axSpA-fatigue and capture what matters to patients. Coupled with current evidence and patient and clinician involvement, I developed a novel patient-reported outcome measure (PROM) for use in both clinic and research.

Methods: The four-phase mixed-methods research process was: i) a systematic review to identify whether there was a need for a new PROM; ii) phenomenology-based semi-structured interviews to develop a lived experience and measurement framework; iii) item (question) generation and refinement; and iv) a psychometric evaluation of the new PROM. A patient research partner group was incorporated throughout the project as co-producers of the PROM.

Results: The systematic review identified 9 PROMs of limited quality and relevance to axSpA-fatigue. Interviews with patients identified ‘achieving balance’ as an active process of integrating axSpA symptoms and fatigue into daily life through three themes: having energy, engaging in everyday life and making sense of axSpA. Energy emerged as a related but distinct component in axSpA-fatigue experience. The measurement framework of axSpA-fatigue consisted of four domains (symptoms, impact, psychological/emotional wellbeing, self-management) which underpinned the content of a 30-item PROM. Psychometric evaluation showed that the PROM is good quality, producing a four domain (fatigue, energy, symptoms, control) 18-item short form version – the Warwick Axial Spondyloarthritis measure of faTigue and Energy (WASTEd).

Conclusion: The WASTEd is a co-produced, axSpA-fatigue specific PROM. Its use will help make visible the work patients do to balance their axSpA and fatigue symptoms and support the development of appropriate interventions. Further psychometric testing of the WASTEd is required to determine evidence of its reliability (test-retest) and responsiveness.
1. **Chapter 1: Introduction to health outcome measurement in Axial Spondyloarthritis**

1.1. **Introduction**

This project originated from the findings of a patient survey conducted by the National Ankylosing Spondylitis Society (NASS) which identified understanding the impact of fatigue as one of the top three priorities for axial spondyloarthritis (axSpA) patients (1). Current assessment guidance stipulates that fatigue should be measured using a single item visual analogue scale (VAS) on fatigue severity (2), however, this is a limited assessment of a complex symptom and it is unclear what (about fatigue experience) the single item really captures. This chapter provides an overview of axSpA, fatigue and health outcome measurement to provide the context for this thesis.

Section 1.2 explores the epidemiology, pathophysiology and the symptoms of axSpA alongside an introduction to the significance of fatigue in inflammatory conditions. Section 1.3 describes the changing landscape in health outcome measurement from a clinician-led approach to a model of health that is both patient-derived and patient-centric. Section 1.4 explores patient-reported outcomes and how they can be operationalised into patient-reported outcome measures to address specific health issues. Finally, Section 1.5 outlines the necessity of the empirical research work completed as part of this thesis, the study aims and the project design.

Subsequent chapters describe the iterative process of developing the Warwick Axial Spondyloarthritis measure of Fatigue and Energy (WASTEd) from understanding what fatigue in axSpA is and what is important for measurement, through to how best to measure axSpA-fatigue. A brief description of each chapter is presented below and illustrated in a flowchart provided in Appendix 1A.

*Phase 1 – Identifying a need (evidence synthesis):* Chapter 2 describes a systematic review of the quality and relevance of existing single and multi-item measures used to assess fatigue in axSpA. This chapter identifies the need for a new measure through highlighting the limited relevance and poor quality of existing measures.

*Phase 2 – Conceptualising fatigue in axSpA:* Chapter 3 describes qualitative interviews with patients to explore their lived experience of axSpA and fatigue. Two
frameworks were derived from the data: (1) a lived-experience framework of axSpA and fatigue, and; (2) a measurement framework to underpin the content of the WASTEd.

**Phase 3 – Developing and pretesting the WASTEd:** Chapter 4 details a three-stage qualitative process in development and refinement of the WASTEd. This begins with item generation, drawing on the measurement framework developed in chapter 3 and existing measures of fatigue. Then, a review of the measurement framework and generated items with both patients and professionals is described. Finally, the chapter closes with pretesting interviews as the final stage of qualitative refinement of the WASTEd.

**Phase 4 – Psychometric evaluation of the WASTEd:** Chapter 5 describes a national postal survey of axSpA patients to psychometrically evaluate and refine the WASTEd. Using classic and modern approaches to psychometric testing, this study determines the internal structure of the WASTEd and provides statistical arguments for item reduction, producing a short form version of the questionnaire.

Finally, Chapter 6 draws the findings of chapters 2-5 together to summarise the findings of the thesis and sets out the implications for clinical practice and research in axSpA fatigue. Further, this chapter suggests future research initiatives for PROM and intervention development and evaluation.

### 1.2. Axial Spondyloarthritis

Spondyloarthritis (SpA) is an umbrella term that categorises a group of inflammatory rheumatological conditions that have some shared features but have distinctive manifestations (3). Two key diagnostic categories within SpA exist whereby it may manifest as predominantly axial (axSpA) or predominantly peripheral SpA, although it should be noted that axSpA may also present features of peripheral SpA and vice versa (Table 1.1).

**Table 1.1: Types of spondyloarthritis**

<table>
<thead>
<tr>
<th>Spondyloarthritis</th>
<th>Predominantly Axial SpA</th>
<th>Predominantly Peripheral SpA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Radiographic axSpA (Ankylosing Spondylitis)</td>
<td>Psoriatic Arthritis</td>
<td></td>
</tr>
</tbody>
</table>

1.2.
<table>
<thead>
<tr>
<th>Non-radiographic axSpA</th>
<th>Reactive Arthritis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Enteropathic arthritis (comorbid inflammatory bowel disease)</td>
</tr>
<tr>
<td></td>
<td>Undifferentiated SpA</td>
</tr>
</tbody>
</table>

Axial Spondyloarthritis (axSpA) is a chronic inflammatory condition characterised by frequent, wide-spread joint pain and stiffness (4) that primarily manifests in the sacroiliac joints (SI) and spine (4). Patients typically report experiencing pain, stiffness, reduced mobility, impaired physical function and fatigue (5).

1.2.1. Diagnosis

Historically, diagnosis of axSpA in the absence of classical plain x-ray changes in the pelvis was challenging. Non-radiographic axSpA (nr-axSpA) was not recognised as a disease type meaning that only patients showing radiographic sacroiliitis received a diagnosis of AS. However, as the availability of magnetic resonance imaging (MRI) increased, the recognition (and subsequent classification) of AS evolved. Now, nr-axSpA and AS are recognised as two classes of the same disease (4–6). Improved understanding of the development and progression of axSpA has highlighted key diagnostic indicators and informed current criteria proposed by the Assessment of Spondyloarthritis International Society (ASAS) (7). Other criteria, pre-dating the ASAS classification criteria exist for the diagnosis of SpA: the European Spondyloarthropathy Study Group (ESSG) and Amor criteria, and for the diagnosis of AS: the New York criteria and the Rome criteria. These pre-existing criteria were developed prior to the widespread availability of MRI scanners and do not incorporate the concept of MRI proven sacroiliitis (or x-ray negative/MRI negative but clinically classifiable axSpA).

For AS to be diagnosed, SI joint lesions must be observable via x-ray and meet the modified New York (mNY) criteria for AS (8). In nr-axSpA, sacroiliac lesions are not apparent on plain x-ray, but may be visible on MRI. These MRI changes, combined with pre-defined clinical criteria, may be used to diagnose nr-axSpA (9). ASAS classification criteria allows three avenues to axSpA diagnosis: sacroiliitis on x-ray imaging (radiographic group), sacroiliitis on MRI and at least one clinical feature of SpA (non-radiographic group), or possession of the gene HLA-B27 and at least two SpA features (non-radiographic clinical group) (Table 1.2) (7).
From symptom onset to diagnosis, an average 8.5-year delay to diagnosis has been reported with only 15% of individuals receiving a diagnosis within three months of presenting symptoms (10). A range of diagnostic delays between 8 and 10 years from symptom onset and diagnosis has been reported (10,11) and evidence suggests this delay can negatively impact on the patients’ psychological health (12). In part, difficulty with diagnosis may be due to the slow but progressive nature of axSpA, coupled with the difficulties in identifying individuals with inflammatory spinal pain (13), from the much more commonly occurring non-inflammatory mechanical back pain in the general population; 60-80% of people in the general population report back pain at some stage in their lives (14).

Table 1.2: ASAS axSpA classification criteria adapted from Rudwaleit et al. (7)

ASAS classification criteria for axSpA
(in patients with back pain ≥ 3 months and onset age < 45 years)

<table>
<thead>
<tr>
<th>Sacroiliitis on imaging</th>
<th>or</th>
<th>HLA-B27</th>
<th>plus</th>
<th>≥ 2 other SpA features</th>
</tr>
</thead>
<tbody>
<tr>
<td>plus</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥ 1 SpA feature</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

SpA features:
- Inflammatory back pain
- Arthritis
- Enthesitis (heel)
- Uveitis
- Dactylitis
- Psoriasis
- Crohn’s disease/ulcerative colitis
- Good response to NSAIDs
- Family history for SpA
- HLA-B27
- Elevated CRP

Sacroiliitis on imaging:
- Active (acute) inflammation on MRI highly suggestive of sacroiliitis associated with SpA or
- Definite radiographic sacroiliitis according to mNY criteria

1.2.2. Epidemiology
True prevalence rates of axSpA are unknown; however, a recent systematic review suggested AS prevalence rates within Europe of 0.24% and 0.17% in Asia (15).
Females are more commonly diagnosed with nr-axSpA due to the slower progression of disease and less radiographic sacroiliitis change; consequently, leading to an under-representation of females with SpA when using the modified New York criteria. Accordingly, the prevalence of nr-axSpA in females is likely to be greater as many do not develop radiographically observable symptoms.

In a large cross-sectional cohort study in a UK primary care setting, Hamilton et al applied three commonly used criteria for classifying axSpA and AS to estimate the prevalence of axSpA (16). Prevalence rates were estimated to be 1.2% (ESSG criteria), 0.3% (ASAS criteria) and 0.15% (modified NY criteria). These findings were reportedly in keeping with European estimates when applying the ESSG criteria and ASAS criteria (16). The prevalence rates of axSpA and RA are thought to be similar (17), but the trajectory for both conditions is very different with the incidence of RA declining, whilst incidence of axSpA is increasing (possibly related to improvements in screening and the introduction of nr-axSpA classification) (18).

1.2.3. Aetiology

A defining characteristic of AS is the formation of new bone which leads to ankylosis of the SI joints and the development of bony growths in spinal ligaments which permanently impair spinal mobility (19). The current aetiology of axSpA is unknown, however it is thought that a combination of genetic, environmental (20) and immunogenic (19) factors together produce clinical disease with genetic evidence suggesting a strong association between AS and HLA-B27 (21).

Growing genetic evidence suggests a strong link with first-degree family members of individuals diagnosed with AS at high risk of developing AS (75.5) compared to third-degree family members at significantly lower risk (3.5) (22). Human leucocyte antigen (HLA)-B27 is thought to be the most significant genetic risk factor (21,23) with between 70-90% of AS patients possessing this gene (19). However, it is recognised that a large number of genes in addition to HLA-B27 also increase the risk of developing AS (19). Multiple hypotheses have been proposed to explain how this may occur and are discussed in detail in a recent review (19).

1.2.4. Symptoms

There is currently no cure for axSpA, it is managed through life-long physical and drug therapies tailored to the patients’ needs (24–26). Patients are recommended to
engage in an active exercise programme designed to maintain range of motion and posture, irrespective of whether their axSpA is active or stable (27). Over the last decade, effective drug therapies for axSpA (primarily for AS) have been successfully developed, relying on timely, sustained administering of anti-inflammatory medication to minimise structural damage to the spine (19). Non-steroidal anti-inflammatory drugs (NSAIDs) are considered the primary treatment choice for pain and stiffness and can be provided continuously at adjusted doses (as per the patients’ needs) (27). Up to 60% of patients show sustained, adequate responses to TNF inhibitors (27) and anti-IL-17 medications. However, if treatment fails, or there are adverse reactions to drug therapy, there are currently no further alternatives available (5). Drug therapies are often effective in dealing with the physical symptoms of disease such as pain management and stiffness (28,29). These drug therapies may co-incidentally improve fatigue, but improvement is variable and there are currently no established approaches for dealing with fatigue.

Fatigue is a common symptom of illness and qualitative research in a range of conditions has highlighted the multifaceted and wide-ranging impact of fatigue on patients, including in: RA (30–32); axSpA (33–36) and cancer (37–39) amongst many other chronic illnesses. A national survey of patients with axSpA conducted by NASS in 2013 highlighted the importance of fatigue to axSpA patients, achieving a rating higher than pain (1). Current ASAS international advice on fatigue measurement in axSpA suggests that a single item measurement (taken from the BASDAI questionnaire; question 1) is sufficient to reflect fatigue in axSpA (2). In a UK cohort of 612 axSpA patients, the importance of measuring fatigue frequency and severity was explored (40). This study reported the following detection rates: 75% of patients reported both frequent and severe fatigue, 15% reported frequent but not severe fatigue, and 10% reported severe but not frequent fatigue (40). Patients reporting frequent and severe fatigue had significantly worse outcomes across disease-specific (BASDAI, BASFI) and health-related QoL outcomes (SF-36) compared to the just frequent or severe groups. Further, those with just frequent and just severe fatigue were similarly affected across a range of outcomes (40). This study highlights two key issues: firstly, the insufficient information a single item on fatigue severity provides about fatigue impact; and secondly, the significant and detrimental impact fatigue can have on the quality of life of axSpA patients.

This limitations of using a single item to measure fatigue experience has also been demonstrated in a matched cohort study of active cancer patients (ACPs), cancer survivors (CS) and the general population (41). In this study, participants were
asked to complete a fatigue intensity numeric rating scale (NRS), and the Multi-dimensional Fatigue Inventory (MFI-20) (42) so differences in fatigue experiences between the groups could be explored. Findings highlighted the differential experience of fatigue between ACPs and CSs, with the former describing more intense physical fatigue; this was supported by a statistically significant difference between ACPs – who reported much higher physical fatigue – compared with CSs and controls (41). Moreover, the fatigue intensity NRS scores were predicted by only two domains of the MFI (physical and mental fatigue) but not the remaining three domains. The researchers also noted that variance explained by the NRS was lower for ACPs than CSs, suggesting other factors were influencing how ACPs scored on the NRS (41). This study further demonstrates the limitations of relying on a single item to measure fatigue and highlights the valuable contribution multi-item measures can make to understanding the experiences and impact of fatigue on patients.

Fatigue has been identified as a complex, multi-faceted and naturally occurring phenomenon which whilst a widely used concept, is sometimes challenging to incorporate into clinical care due to a lack of a universally agreed definition (43). It can affect those in good or poor health (44–46) but is commonly observed in inflammatory conditions (45). Crucially, there appears to be a distinction between disease-related fatigue and healthy fatigue, which manifest differently and subsequently present different challenges for the individual.

In healthy adults, fatigue tends to occur in response to great stress, physical exertion and/or mental exertion (47). Aaronson and colleagues, using semi-structured interviews, investigated the experience of fatigue among healthy working adults (48). Their findings highlighted an experience of fatigue that was temporary but overwhelming and attributed to the worker’s life roles and stress which the worker could actively address through restorative measures (48). In healthy individuals, the experience of fatigue is typically predictable and with suitable rest, even mitigated or diminished (47,48). Conversely, disease-related fatigue can be experienced in the absence of exertion, is often unpredictable and is not receptive to sleep.

Recent years has seen a growing focus on the importance of fatigue in the rheumatic diseases, with fatigue increasingly understood to be an important cognitive and physical symptom (30,36,49). Patients with RA have described disease-related fatigue as unpredictable, cannot be controlled and overwhelming.
A patient-derived conceptualisation of fatigue in rheumatoid arthritis (RA) described the complex interactions between disease, cognitive/behavioural and personal factors. This conceptualisation is the first detailed insight, within rheumatology, into what the experience and impact of fatigue is for patients (Figure 1.1). However, recognising that fatigue may disease-specific, it is unclear whether and to what extent the RA-fatigue model is an adequate reflection of fatigue experience in axSpA.

Figure 1.1: Conceptual model of RA-fatigue, taken from Hewlett et al. (50)

1.3. Measuring health outcomes
Historically, health outcome measurement was seen through a biomedical prism whereby the predominant view was disease-based. In practice, this creates a dichotomy for judging patient health: those with disease and those without disease or ‘cured’. This view of health does not recognise any lasting impact of illness on health, omitting chronic illness, and is thus irreconcilable with a solely disease-based view of health. This reductionist view fails to grasp the complexity of health and draws the focus away from illness manifestations and impacts relevant to
patients, professionals and researchers. Growing recognition of health status as representing more than simply ‘being cured’ of disease challenged the prevailing biomedical model approach to health outcome assessment and highlighted the importance of psychological, social and behavioural factors when considering patient health outcomes (52). This broadened perspective was reflected in the World Health Organisation’s (WHO) definition of health (52) as:

‘a state of complete physical, mental and social well-being, not merely the absence of disease and infirmity’

This move from a disease-centred view of patient health toward a patient-centric approach originated with the idea of a “normative standard” (53). The notion is that an individual’s personal experience is implicit and is thus inaccessible to others. With clinician-led measurement of health outcome, interpretation of the patients’ capabilities would, therefore, be based on the clinician’s own normative standard thus failing to reflect the patients’ experience. This failure to reflect the patients experience is evidenced in studies which highlight the discrepancies between patients and healthcare professionals when assessing health outcomes (54,55). For example, in a study investigating which criteria patients and clinicians use to just AS disease activity, a clear disparity emerged between both (56). Clinicians judged disease activity through measures of disease severity and inflammation whilst patients made this judgement through their AS-associated complaints (56). This highlights how priorities and perspectives between clinicians and patients can vary, and the likelihood that aspects of health, that are important to patients, could be overlooked.

The shift to reflecting the patient perspective and capturing subjective experiences of health is captured by the International Classification of Functioning, Disability and Health (ICF) framework on which measures of functioning may be based, and:

‘provides a standard language and a conceptual basis for the definition and measurement of health and disability’ (57)

Capturing a patient’s report about the psychosocial impact of illness can complement traditional health approaches and afford a richer understanding of illness manifestation and impact (58). A patient-reported outcome (PROs) (59) has been defined as (60):
‘any report coming directly from patients, without interpretation by physicians or others, about how they function or feel in relation to a health condition and its therapy’

To afford measurement of these often non-observable, subjective concepts, PROs are operationalised into PRO measures (PROMs), which seek to capture how people feel, function and/or live their lives as a consequence of their health and associated healthcare (59). PROMs may, therefore, assess constructs ranging from pain and fatigue to health-related quality of life. PROM questionnaires consist of single or multiple items which provide a structured assessment of health from the patient’s perspective. They may be unidimensional, measuring a single concept – for example, the Functional Assessment of Chronic Illness Therapy (FACIT) fatigue (61) considers fatigue as a single over-riding construct; or multidimensional, measuring multiple domains relating to a central concept – for example, the Evaluation of Ankylosing Spondylitis Quality of Life (EASi-QoL) which measures four domains of health-related quality of life (HRQoL) in axSpA patients: physical function, disease activity, emotional well-being and social participation (62).

The preference is that PROMs are patient self-completed, however, interview and proxy-based (by a family member or caregiver) completion may also be recommended where, for example, age or ill-health limit self-completion (59). For example, in response to the cognitive and functional decline associated with dementia, both interview and proxy-completed versions of the Dementia health-related Quality of Life (DEMQOL) questionnaire were developed (63).

1.3.1. PROMs taxonomy

Three broad PROM taxonomies can be described: generic measures which provide broad summaries of health; specific measures which relate to specific conditions, aspects of health or patient groups; and individualised approaches which support responders in identifying aspects of health that are personally relevant.

Generic PROMs have a broad application, allowing them to be used within the general population and across different patient populations irrespective of underlying health condition(s), providing general and comparable information about respondent’s health-related quality of life (64,65). Generic measures may be classified as health profile or utility measures. Health profile measures are considered preferable to index measures as they can provide scores separately for domains, supporting data interpretation, or produce a summary score. One of the
The most widely used generic profile measures is the SF-36, a 36-item questionnaire providing an assessment across eight broad health status domains including physical and mental health, bodily pain and social functioning (66). Using the SF-36, physical and mental component summary scores can be generated, and a summary score can also be calculated (66). Population norms have been generated across several counties (67–70), further supporting data interpretation and comparison of health status. In a recent review of PROM performance in primary and community care settings, the SF-36 (66) was identified as the best performing generic measure in terms of its quality (measurement properties) and acceptability (71).

Utility measures provide scores ranging from 0 (death) and 1 (perfect health), representing a value placed by patients on their own health (65). This information can be used to calculate quality-adjusted life years used to determine cost-utility of medical interventions (72). The most widely used generic index measure is the EuroQoL 5 dimensions (EQ-5D), a five-item preference-based utility measure covering five dimensions: mobility, self-care, usual activities, pain/discomfort and anxiety/depression (73). The EQ-5D is used to compare the benefit and cost of interventions and healthcare programs (74). Currently, it is the preferred measure for cost-effectiveness evaluations and is used by the department of health PROMs programme evaluating the cost-effectiveness of interventions for four surgical outcomes (75). Generic measures are considered most suitable for the purpose of conducting economic evaluations of value-for-money, productivity and performance (58). However, as these measures are general, they can fail to adequately capture or reflect important aspects of health status (58).

In contrast, specific measures can be focused on a condition (e.g. axSpA), a domain of health (e.g. fatigue), a population (e.g. children) or an intervention (e.g. total hip replacement), providing insight into the patient’s own assessment of their health (58). For example, in axSpA, several disease-specific measures exist, which are further specific to aspects of health such as disease activity (BASDAI) (76), functional impairment (BASFI) (77) and quality of life (EASi-QoL) (62)).

The specificity of content of specific measures enhances their clinical appeal and, when well developed with appropriate patient engagement (59), their relevance to patients (59). This should not only improve their acceptability (for example, enhanced completion rates), but their responsiveness to small but important changes in health (78,79). As such, specific measures are generally more responsive to change than generic measures (79) and tend to have better face
validity and credibility compared to generic measures (64). However, facilitated by the broader content, generic measures may better capture treatment side-effects and co-morbid features. Due to the narrow focus and specificity of content, specific PROMs tend to outperform generic measures through their ability to capture greater detail (58).

Where generic and specific PROMs typically utilise a fixed and standardised set of items, individualised PROMs are flexible measures without fixed domains and/or weightings, allowing responders to select content that is relevant to them (80). For example, nominating the areas of their life that they consider most important. An example is the Patient Generated Index for Ankylosing Spondylitis (PGI-AS) (81). Whilst such approaches have high levels of content validity, are well-received by patients and are more responsive than fully standardised measures (82,83), they can be time-consuming to complete and score. Similarly, the tailored nature of individualised PROMs mean they lack standardisation of content (79), reducing their utility in comparative evaluation. Due to their relative strengths, the combined use of generic and specific PROMs is advocated (58,79,84,85).

1.3.2. Application of PROMs in healthcare

The ability of PROMs to provide insight into the patient’s perspective means they have a wide range of potential applications, including in clinical trials as primary or secondary endpoints, routine practice to support tailored healthcare provision and in clinical audits to improve patient care such as the national health service (NHS) PROMs programme (75).

Within clinical trials, incorporating PROs and PROMs as trial endpoints ensures a more comprehensive assessment of the trial and its impact on the patient, from the patient perspective (86,87). Alternatively, PROs may be used as secondary endpoints to support interpretations of primary endpoints within the trial or be collected to support the generation of hypotheses for future study (87). A review of Phase III oncology trials explored the benefits of including PROs as endpoints in clinical trials, highlighting three benefits: (i) to assist clinicians and patients to select the best treatment relative to their costs and benefits; (ii) enriching understanding of the patient experience on subjective domains of health such as pain and fatigue, and; (iii) improving future clinical trial methods (88). This review highlights the value that can be gained through inclusion of PROMs in clinical trial research through better informed decision making and enriching understanding. Increasing value is
being given to the patient’s perspective in research and clinical trials (89) and thus underscores the need for high quality PROMs that are well developed and therefore, capturing the patient perspective.

Routine practice presents an important context for PROM application. The collection of PRO data can be useful for screening patients or monitoring them over time, including to evaluate the impact of treatment (90). Increasingly, PROMs provide a method to identify functional and psychological health problems that may be missed in usual routine practice and provide a means for regular, standardised monitoring which can complement clinical information (91). Moreover, there is evidence from cancer research that routine PRO collection can support better patient-centred care (92). Evidence also suggests that PROMs improve communication and decision-making between patients and doctors whilst also improving patient satisfaction (92–94), and this data can be aggregated to support decision-making and assess care quality (95,96).

Since 2009, four surgical interventions: knee, hip, groin hernia and varicose vein repair, have been assessed using a combination of generic and disease-specific PROMs (72). These PROMs are being administered prior to, and after surgery to determine the effectiveness of surgical interventions on outcomes relevant to the patient’s condition and health-related quality of life, rather than simply whether surgical outcomes were achieved. For example, to determine whether knee surgery has improved function, mobility and reduced pain amongst other outcomes, data from the Oxford Knee Score PROM is used in conjunction with clinical data (72). This data is then analysed to determine which providers’ outcomes are significantly different than what would be expected.

1.4. Understanding and developing patient-reported outcome measures

Growing recognition of the importance of including the patient’s perspective in measurement has resulted in an exponential growth in the availability of PROM measures (19). However, the quality and acceptability of these measures is variable (97–99). In response, in 2006 the US Food and Drug Administration produced draft guidance for the development and evaluation of PROMs necessary for them to support medicine product labelling claims (100). This was developed following engagements with the international health measurement community (for example, the International Society for Quality of Life Research; the International Society for Pharmacoconomics and Outcomes Research (ISPOR) and the pharmaceutical
industry), informing the development of their final guidance document which was made available in 2009 (89).

A bibliographic review in 2002 assessed the growth of QoL measures and reported exponential growth in the development and evaluation of QoL measures, with most of this research activity occurring within musculoskeletal and rheumatology disciplines (101). This exponential growth means there is a substantive amount of choice when selecting measures, highlighting the necessity for structured reviews to support the identification and selection of appropriate outcome measures. Structured reviews of measurement properties are increasingly conducted to evaluate the quality and relevance of existing PROMs and guidance has been developed to support this type of review (102–104). However, the reporting of the development process for legacy measures is often scant, limited and lacks transparency – for example, the BASDAI has no reports on how items were generated or qualitative research to underpin its development (76). In addition, many legacy measures failed to engage with, or incorporate the patient voice – whether as participants in qualitative work, collaboratively, or as research partners – meaning it is unclear whether these measures are relevant to patients or have good face and content validity.

In response to the lack of transparency in PROM development, and the claims underpinned by measures with unknown development processes, the US Food and Drug Administration (FDA) in 2009 published guidance for industry detailing the criteria that PROMs are required to meet for treatment related QoL claims, in clinical trials, to be accepted (89). The five-stage process includes three development stages: hypothesising a conceptual framework, adjusting the framework and draft instrument and confirming the conceptual framework and assessing other measurement properties (89). The penultimate and final stages involve collecting, analysing and interpreting data and modifying the PROM through rigorous psychometric evaluation and refinement (89). This guidance has underpinned a new generation of PROMs which puts meaningfully capturing the patient perspective at the fore of each stage of PROM development, with great emphasis on the importance of face and content validity (105–108).

1.4.1. Establishing the need for a new measure
The first step in PROM development is to establish whether there is a need for a new measure: this involves identifying whether there is an existing measure
available. If there are PROMs available, the next step is to then determine their adequacy. This can be determined by considering the quality of their development, the relevance of the measure to the intended users, their acceptability for use and evidence of their psychometric quality. A systematic review of evaluative studies is a commonly used approach to establish PROM quality (97–99,109) and more recently, their relevance and acceptability (110), and whether patients were involved in their development (109). For example, a recent systematic review of PROMs used for inflammatory bowel disease found that of 20 PROMs currently available, only two had evidence of patient involvement in their development, and only one of those two measures had used patient interviews as part of the PROM development process (111).

1.4.2. Development of a conceptual (measurement) framework
The purpose of a PROM should be well defined, providing an explicit indication of what the PROM is seeking to capture. The generation of a conceptual framework to underpin a PROM is recommended in current guidance, providing transparency and clarity on ‘what’ the PROM is intended to measure (89,112). This should describe the important conceptual domains, any relevant subdomains and identify the PROs important for the measurement target.

Existing conceptual models of health may be useful when considering the breadth of factors essential for measuring patient health. For example, the WHO developed the ICF framework which describes how illness may impact on an individuals’ functioning and disability (113). In their overview, the WHO-ICF distinguishes between two parts: 1) functioning and disability; and 2) contextual factors and considers these across five categories: 1) components; 2) domains; 3) constructs; 4) positive aspects; and 5) negative aspects (Figure 1.2). This conceptual model, demonstrates the nature and direction of associations, was developed through international consensus and is internationally accepted (114).
The WHO-ICF model moves beyond simply recognising the consequence of disease, and instead provides a more holistic representation of how illness may affect a person biologically, socially and psychologically – representing the move toward the biopsychosocial model of health (115). This holistic model can support explanations about the role of other factors (environmental and personal characteristics) and how they may modulate ability to engage in daily life, where illness alone does not provide the whole answer (116). However, whilst the WHO-ICF is an important contribution to supporting our understanding of the wider picture of ill health and its associated factors, a recent review in rheumatoid arthritis (RA) has identified some shortcomings in PROM development (117). The review considered 42 PROMs and established that personal factors which were considered important to patients – and an important component of the WHO-ICF conceptual model – were only minimally covered in half of the PROMs identified in the review (117). This review highlights the value of conceptualising a phenomenon to underpin measurement – ensuring relevant and important components of illness experience are reflected in the PROMs content.

Researchers investigating fatigue experience in RA have developed a conceptual framework of RA-fatigue which describes the complex interrelationships between disease-specific, cognitive/behavioural and personal factors (50). For example, over
and under-activity is thought to influence and be influenced by cognitions, illness beliefs and stress; both of which interact with fatigue experience (50). This type of domain-specific conceptual framework provides a transparent, conceptual underpinning for the development of a PROM with a clear indication of what is important for measurement and therefore what types of questions should be asked. Unfortunately, in the case of RA-fatigue, the conceptual model came after the development of an RA-fatigue specific questionnaire, the Bristol Rheumatoid Arthritis Fatigue Multidimensional Questionnaire (BRAF-MDQ) (106).

1.4.3. Crafting the first version of the PROM

The content of a new PROM can be informed from multiple sources including (but not limited to): existing, related PROMs; existing literature in the field and empirical qualitative research. All item content decisions should be underpinned by the conceptual framework generated as part of the PROM development process.

Given that existing PROMs have undergone a form of development – even if not to the standard of current practice guidance – they provide useful sources of content, having already undergone some psychometric testing, which may help to inform the generation of an initial item set (59,118). By mapping existing item content to the conceptual framework, an item-set that reflects the model can be established.

The Patient-Reported Outcomes Measurement Information System (PROMIS) develops PROs to evaluate and monitor mental, physical and social health that can be used in clinical research and practice (119). In addition to this, PROMIS have developed item-banks for a range of issues such as social participation (120), pain (121) and depression (122). These are standardised measures which can be used with the general population and chronic illnesses and have been developed using item response theory (IRT; modern psychometric method), meaning the PROMIS network can provide a useful source of potential item content for a PROM.

Qualitative research using semi-structured interviews, focus groups or Delphi methods should be conducted to understand the patients' lived experience and perspectives on the concept under investigation, ensuring both relevance and comprehensiveness to the target population (59,112,123). Current guidance recommends developing the conceptual framework, conceptual domains and items using the language of participants. An example of this is the recent development of the coeliac disease assessment questionnaire (CDAQ) – a measure designed to assess quality of life (124). The development of the PROM began with in-depth
qualitative interviews with 23 patients which were conducted and analysed iteratively, using thematic analysis to derive a coding framework (125). The framework consisted of six domains covering: (i) symptoms; (ii) gluten-free diet, (iii) emotional health; (iv) impact on activity; (v) relationships, and (vi) financial issues. The qualitative analysis identified the significance patients placed on holidaying, yet due to infrequency of holidaying between individuals, the study reconciled inclusion of such an important aspect of QoL by including it within a subdomain on travel (125). This study demonstrates the importance of knowing what is important to patients to support the development of a relevant and meaningful measure. However, this study did not report any patient and public involvement at any stage of the qualitative process.

1.4.4. Reviewing and modifying the PROM
Two processes of review and refinement are frequently used when developing PROMs: pretesting interviews – which are advocated for in current guidance (60,89,126) and psychometric evaluation (89).

Questionnaire pretesting is a qualitative endeavour that seeks to explore practical properties of the developing item set and PROM, particularly in relation to its comprehensiveness, relevance, acceptability and comprehension in the target population (126–128). Questionnaire pretesting typically uses a semi-structured interviewing format within which the techniques ‘think aloud’ and ‘verbal probing’ are used to appraise the PROMs practical properties. Pretesting interviews are underpinned by existing models of cognitive processing; the most commonly used being the four-stage model of survey response processes (129). This model describes four processes that occur from reading an item through to giving a response (129):

1. **Comprehension:** is the respondent able to make sense of key words, phrases and the overall question?
2. **Retrieval:** what information do respondents need to recall, and how do they do that?
3. **Judgement:** is the respondent using sufficient mental resources to give a meaningful answer, and is there any element of social desirability in their response?
4. **Response mapping:** is the respondent able to match their internal answer to the response categories available?
PROM development studies frequently report using cognitive interviewing as their pretesting method. For example, in the development of a stiffness PROM for RA – cognitive interviews were conducted with 11 patients across two rounds of five interviews, where think aloud and verbal probing techniques were used to assess items (107). Using this method, the study was able to identify problems with retrieval, understanding, judgement and response to improve item clarity.

However, cognitive interviewing as a method for questionnaire appraisal was originally developed for interviewer-administered questionnaires (130). This raises some potential methodological issues with arbitrarily applying such a method to self-report questionnaires – particularly because cognitive interviews were designed for verbally-administered questionnaires, yet PROMs are self-report measures. Hak et al have since developed the three-step test interview (TSTI) – a three stage interview process specifically for pretesting self-report questionnaires (131). Both cognitive interviews and TSTI have been used to pretest developing PROMs (described in detail in Chapter 4).

**Psychometric evaluation**

Current guidance specifies that once the content validity of a new PROM is established, evidence of its validity, reliability and responsiveness needs to be determined (89). A preliminary psychometric evaluation may be conducted to reduce the PROM from a long-form version to a short-form. This preliminary psychometric evaluation should use both classic test theory (89,118) and modern psychometric methods; Rasch Measurement Theory or Item Response Theory (118). The purpose of this preliminary psychometric evaluation is to determine the internal structure of the developing PROM, and identify which items are contributing to measurement (59).

The refined short-form PROM can then undergo a more comprehensive evaluation to determine evidence of its essential measurement properties (validity, reliability), longitudinal properties (responsiveness) and interpretability (59). These measurement properties and concepts are defined and described in Chapters 2 and 5. An illustration of the PROM development process is provided in Figure 1.3.
Figure 1.3: An illustration of the PROM development process taken from Haywood et al. (59)

1. Establish need for new PROM
2. Establish Core Research Team – responsible for day to day running of the initiative
3. Develop conceptual framework or ‘blueprint’
   - Conceptualisation of health - potential items and domains
4. PROM Crafting I: concept elicitation, item generation and selection
   - Domain refinement
   - Item identification / development
5. PROM Crafting II: cognitive interviews
   - Elaboration of items
6. Content validation and further refinement
7. PROM evaluation: item reduction and refinement in the target audience
8. Psychometric evaluation of the final PROM in the target audience
   - Expert Reference Group: Content validation and further refinement
   - Opinion on finalised PROM
   - Second face to face meeting of Advisory Group
   - Opinion on domains, items and wording
   - Third face to face meeting of Advisory Group
1.4.5. *Patient and public involvement in research*

Patient and public involvement in research (PPI) is increasingly advocated for in research studies, including PROM development (132–134); however, PPI is not currently included in PROM development guidance (89,112,126). INVOLVE (invo.org.uk) is a UK-wide organisation that encourages and supports patient and public involvement (PPI) in healthcare research and the NHS. INVOLVE propose a shift in how patients are involved in research, moving away from completing research ‘for’, ‘to’ or ‘about’ patients; rather, research should be conducted ‘with’ or ‘by’ the public (135).

INVOLVE distinguish between three types of PPI which can be adopted at all stages of a research study (136):

1. **Consultation**: asking members of the public for their views to inform decision making.
2. **Collaboration**: an ongoing partnership between the researchers and members of the public they are working with.
3. **User controlled**: research that is user-led; managed, controlled and directed by service users.

More recently, INVOLVE has produced guidance regarding co-production which they define as (137):

> ‘an approach in which researchers, practitioners and the public work together, sharing power and responsibility from the start to the end of the project, including the generation of knowledge’

INVOLVE propose that co-production is not a set of fixed processes to be followed, but instead about key principles of sharing power, including all perspectives and skills, respecting and valuing everyone and their knowledge, reciprocity and relationships (137). This type of approach to PPI represents a shift from more formalised roles toward a more integrated model of research.

Patients and the public can be involved in and contribute to the full research process from identifying a need for a study through to interpreting data and dissemination (138), and their involvement can improve the clinical relevance of research (139,140). It is recognised that engaging with patients as research partners introduces the patient perspective to the research process, ensuring the focus is on what is important to patients (138,141,142). It has been argued that PPI in PROM development is an evolution for the field (138), however, the lack of
guidance for PPI in PROM development accompanies a limited and disparate evidence-base in the field. Recently developed PROMs describe varying approaches and levels of PPI engagement in research. For example, in the development of the CDAQ (124,125) in 2018, there was no reported PPI in the development of the PROM, whilst an RA stiffness PROM did include PPI in qualitative data analysis (reading two transcripts and highlighting points from their perspective) (108). However, patients were not involved in item generation or refinement decisions (107). This contrasts with the development of the BRAF-MDQ which included a patient research partner within the research team, and whom contributed to the wording of a final item list, labelling domains identified in a factor analysis and whom was consulted, with a second patient partner to help identify the best order to present domains and items within the questionnaire (143).

These three examples demonstrate multiple issues regarding PPI involvement in PROM development studies. Firstly, PROM development studies still fail to include PPI, or where PPI is included, the approach lacks justification and can be poorly reported. Secondly, the approach to including PPI can vary substantively from working with a patient research partner to incorporating them within a research team. Finally, involvement can be very limited, and in the examples above PPI was restricted to a specific stage in the research process or to quite inconsequential aspects of research such as simply labelling domains.

To support greater transparency in the reporting of PPI in research, the Guidance for Reporting Involvement of Patients and the Public was developed, seeking to address inconsistent reporting of PPI in research and improve consistency, quality and transparency (144). This checklist was developed using evidence from systematic literature reviews (144) and was later updated following a three-round Delphi survey with a range of participants including researchers and patients (145). Two versions of the checklist were generated, a long-form and short-form – the latter of which was designed for use in studies where PPI is not the focus of, but a component within, the study.

In my thesis I describe working with a PPI group, based at Wrightington Hospital (Wigan) which primarily falls under the description of ‘co-production’ (http://www ww nhs uk Specialities Patient Research Advisory Group aspx). PPI members receive two and a half days training before they officially participate in meetings. I met and communicated with the group over the course of 4 years and sought to maximise their input into the project as much as possible. I used the
GRIPP2 short-form checklist as a guide to assist me with conducting my PPI meetings, and to assist my reporting of PPI involvement throughout this thesis. Where the PPI group were involved in the thesis, a separate header is used in the method and results. A deeper discussion of PPI within the study (including my reflections) is provided in Chapter 6. A photo of the team during a meeting is shown in Figure 1.4.

Figure 1.4: Photo of the PRP group and facilitator during a meeting

1.5. Purpose of an axSpA-fatigue specific PROM

Whilst evidence suggests that a single item measure of fatigue severity provides an inadequate assessment of axSpA fatigue, a review of the quality and acceptability of alternative fatigue assessments in axSpA is not available. Additionally, an empirically derived conceptualisation of axSpA-fatigue is not available, and it is unclear if the RA conceptualisation also applies to axSpA. Although related rheumatological conditions, axSpA is typically active and symptomatic at a young age and primarily affects the spine and pelvis of patients, with no clear treatment route. Further, the evidence indicates that fatigue may have a disease-specific element – that is, its perceived causes, impact and how it is experienced may vary between conditions. Therefore, it cannot be assumed that the RA-fatigue model is representative of axSpA-fatigue experience and this warrants investigation.
**Research question:** What is the experience of fatigue in patients with axSpA and how should it be assessed to inform clinical research and routine practice?

1.5.1. **Aims and objectives**
To explore the experience of fatigue in axSpA patients to inform the co-production of a high quality and relevant axSpA-fatigue specific PROM that is suitable for use in routine clinical settings and research.

Objectives:

1. To review the quality and acceptability of single and multi-item PROMs used in the assessment of fatigue in people with axSpA.
2. To explore the lived experiences of patients with axSpA with a specific focus on the experience and impact of fatigue.
3. To co-produce a new, patient-derived and patient-reported measure of axSpA-specific fatigue for use in clinical research and routine practice.
4. To assess the psychometric properties of the new PROM to develop initial evidence of quality and statistically refine the PROM into a short-form measure.

1.5.2. **Project design**
This project was completed in four phases which are described below. A patient research partner (PRP) group based at Wrightington, Wigan and Leigh Partnership NHS Trust were involved in the project from conception and throughout.

**Phase 1: Systematic review of the quality and acceptability of fatigue PROMs used in axSpA-fatigue assessment**

Objective: To systematically appraise, compare and synthesise published evidence of the quality and acceptability of clearly defined single and multi-item PROMs used to assess fatigue in axSpA patients, to inform PROM recommendation.

Method: Systematic literature review incorporating information of study methodological quality with evidence of evaluation, per measurement property, to produce a robust evidence synthesis to inform identification of a ‘best’ fatigue
PROM for use with axSpA patients. An additional content-validity assessment using an existing conceptual framework of fatigue from rheumatoid arthritis to determine PROM content relevance.

**Phase 2: Exploration of the patients' lived experience of axSpA and fatigue**

Aim: To explore the patients' lived experience of axSpA and fatigue to inform the development of a lived experience framework of axSpA and fatigue, and an axSpA-fatigue specific measurement framework.

Method: Semi-structured interviews with patients (drawing on interpretative phenomenological analysis) were conducted to explore and understand the patients' lived experience of axSpA and fatigue.

**Phase 3: Item generation and pretesting**

*Step 1: item generation*

Aim: To operationalise the measurement framework developed in Phase 2 and craft a first iteration axSpA-fatigue specific item set to be reviewed and confirmed alongside the measurement framework.

Method: Through a process of ‘PROM mapping’, items from existing measures were mapped to the concepts identified in the measurement framework. Findings from Phase 2 were also used in the generation of item content.

*Step 2: focus groups*

Aim: To review and actively appraise the measurement framework of axSpA-fatigue and developing item set.

Method: Separate focus groups with patients and healthcare professionals were held to evaluate the measurement framework of axSpA-fatigue. Involvement of both patients and professionals ensures both patient and clinical relevance of the framework and developing item set, resulting in a confirmed ‘blue-print’ of axSpA and fatigue and revised item set to inform development of a new axSpA-fatigue specific PROM.
Step 3: pretesting interviews

Aim: To assess the comprehensiveness, relevance, acceptability and comprehension of the developing PROM and support the refinement of the item set to inform the development of a long-form PROM for psychometric evaluation.

Method: Two pretesting methods for interviewing were utilised: (i) cognitive interviews to focus on conceptual issues within the questionnaire; (ii) three-step test interviews to support a holistic appraisal of the questionnaire. This supported item modification and refinement.

Phase 4: Preliminary psychometric evaluation of the long-form PROM

Aim: To elucidate the internal structure of the new PROM and establish evidence of its key measurement properties (quality), supporting item refinement decisions to produce a short-form measure.

Method: A cross-sectional psychometric evaluation in an England-wide survey of axSpA patients. Analysis used traditional and modern psychometric methods to complete a robust psychometric evaluation of the new PROMs quality, and support item refinement decisions.
2. Chapter 2: Systematic review of the quality and acceptability of patient reported measures of fatigue used in axSpA

2.1. Introduction
This chapter presents a systematic literature review of the quality, relevance and acceptability of single- and multi-item PROMs used to assess fatigue in axSpA patients between 1980 and August 2017. The current status of fatigue assessment in axSpA is described in section 2.2 and the role of systematic reviews of PROMs is discussed in section 2.3. The methods for performing a systematic review and explicit evaluation of PROMs is described in section 2.4. The results of the review are described in section 2.5. The chapter closes with a discussion (2.6). This review was published in 2018 (146).

2.2. Measuring fatigue in axSpA
Current assessment guidance recommends the assessment of fatigue severity in axSpA with a single-item visual analogue scale (VAS) or numerical rating scale (NRS) (2), taken from the Bath Ankylosing Spondylitis Disease Activity Index (BASDAI) (76). The BASDAI is a six-item, clinician-derived measure of axSpA disease-activity, including single items about pain, stiffness and fatigue. To improve data quality, in 2002, NRS’s were recommended as alternative response options to the original 0-100 mm VAS scales (147). However, there was little involvement of patients in the development of the measure, with little transparency in the conceptualisation of disease activity. Moreover, this recommendation was not informed by a review of the quality and acceptability of a single-item measure of fatigue in this population.

Given the known complexity of fatigue, accurate and relevant assessment therefore requires a measure that captures the components of fatigue that matter to patients and will detect change in fatigue over time. Currently, it is unclear what the single-item VAS on fatigue severity captures about fatigue experience, and by virtue of being a single item it is not possible to deduce how fatigue is affecting patients and their ability to function or live their daily lives. This is an inherent problem of single-item measures which may oversimplify and provide an incomplete assessment of the target concept (148).
Whilst single items offer simplicity in measurement, they are susceptible to random measurement errors and bias from misunderstanding or unintended interpretations, meaning it is unclear what exactly a single item really captures. This challenge with single items was exemplified in the matched cohort study described in Chapter 1 (41). However, well developed multi-item measures can overcome some of these limitations and provide a much more robust, thorough understanding of what and how patients are affected by complex issues. Moreover, multi-item questionnaires are better able to measure complex constructs reliably (149).

Despite recommendations to assess fatigue using a single item, numerous methods of fatigue assessment are available (150–152) and many have been applied across axSpA populations. The multiplicity of methods of assessment may reflect dissatisfaction with the recommended single item assessment of fatigue. There is a need for evidence-based guidance to improve the standardisation of fatigue assessment in axSpA that takes into consideration both (i) the psychometric properties of measures and (ii) the relevance and acceptability of measures to patients. Such guidance will reduce research waste – removing any unnecessary burden to patients, clinicians and researchers from the completion of irrelevant and unhelpful measures – and promote improved data synthesis.

To date, there has been no systematic review exploring the quality, relevance or acceptability of measures used to assess fatigue in people diagnosed with axSpA. This information is essential to inform the selection of the best quality, most relevant measure to adequately assess fatigue in this population. Section 2.3 explains the benefits of such systematic reviews, introduces essential measurement properties (including quality criteria) and ends with the review aim and objectives.

### 2.3. Systematic reviews of PROMs

A systematic review identifies and summaries the findings of all available studies on a given topic. The Cochrane Library (153) defines a systematic review as a process:

> ‘to identify, appraise and synthesise all empirical evidence that meets pre-specific eligibility criteria to answer a specific research question’

Systematic reviews involve following a clear, structured process which begins with the formulation of a review question, through to analysis and dissemination. Uman describes eight stages for conducting systematic reviews or meta analyses (154).
1. Formulating a review question.
2. Defining eligibility criteria to sort through identified articles.
3. Developing a comprehensive search strategy and identifying studies (extracting a list of abstracts).
4. Reviewing abstracts for inclusion or exclusion (as per eligibility criteria).
5. Data extraction using extraction tables (by at least two reviewers to minimise data entry errors).
6. Assessment of study quality using guidance where available.
7. Analysis and interpretation of results.
8. Disseminating findings.

Systematic reviews of PROM quality, relevance and acceptability provide robust evidence-based guidance for PROM selection (155). Such reviews involve identifying and synthesising evaluative evidence for all available measurement properties, with each measurement property being considered a review in itself (156). The aim of PROM reviews is to support identification and selection of the best quality, most suitable PROM for use in the target population for a given construct. In addition, this type of review can highlight where evidence of PROM quality is weak or missing, identifying possible future research avenues to address gaps in knowledge (156).

To support measurement selection in rheumatology, the Outcome Measures in Rheumatology (OMERACT) group proposed a filter (157) that simplifies more complex measurement concepts into three questions:

- Truth relates to validity (face, content, construct and criterion): does the PROM measure what is intended and are results relevant and unbiased?
- Discrimination related to reliability and responsiveness: are responses consistent when the respondent’s state has not changed, and is the measure sensitive enough to detect change in health.
- Feasibility explores the practical elements of measurement: e.g. ease of administration within financial and time constraints; includes interpretability.

The OMERACT filter does this by simplifying the complex nomenclature around measurement into simple, easy to understand terms that are more widely accessible (157) – thus, enabling non-measurement experts to identify and select appropriate outcome measures for practice or research. However, whilst this provides accessibility and identifies important aspects of measurement quality, it provides no
guidance or criteria for determining the quality of each measurement property, or how to adequately synthesise this information to inform PROM selection.

The COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) initiative has developed extensive guidance to support systematic reviews of PROM quality, including: (i) a protocol for conducting such reviews (available since 2011; http://www.cosmin.nl) (158); (ii) a checklist to evaluate study methodological quality (103) including supporting user manuals, and; (iii) a taxonomy of measurement properties (102). The COSMIN guidance is the result of an international, multiple-stakeholder initiative to provide consensus-based guidance on the essential attributes required of studies which seek to evaluate the measurement properties (psychometric, clinimetric) of methods of health status assessment – including PROMs (103). The resulting guidance supports the assessment of study quality. The COSMIN taxonomy is composed of three key measurement properties: validity, reliability and responsiveness (102).

Unlike the OMERACT filter, the COSMIN guidance outlines the minimum information necessary, per measurement property, to evaluate study methodological quality (103). This provides a structured, transparent approach to evaluating study methodological quality, increasing confidence in the output of PROM reviews. For these reasons, I chose to use the COSMIN checklist in my systematic review.

2.1.1  Methodological quality (COSMIN) and measurement properties

The COSMIN guidance includes a four-point checklist which allows each property to be rated for its quality from poor to excellent (103,159). The key measurement properties and components identified in the taxonomy are captured in the checklist. Unfortunately, the checklist contains no quality criteria to appraise content or face validity.

A summary of measurement properties, quality criteria and quality ratings are provided in Table 2.1.
Table 2.1: Quality assessment criteria for measurement properties taken from Pearson et al. (146)

<table>
<thead>
<tr>
<th>Measurement properties</th>
<th>Rating</th>
<th>Quality criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Validity</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Content validity</td>
<td>+</td>
<td>Authors provide a clear description of the measurement aim, target population, concept(s) measured and process of item selection. Members of the target population and experts in the field were clearly identified as being involved in development. For measures applied for the first time in a new population, evidence that the views of members of the target population (and experts in the field) have been sought to determine relevance, comprehension and comprehensiveness.</td>
</tr>
<tr>
<td></td>
<td>?</td>
<td>Insufficient evidence available</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>No detail re measurement aim, target population, concept(s) measured, process of item selection; members of the target population or experts were not specifically involved in development. For measures applied for the first time in a new population, evidence whereby the relevance and acceptability of the measure with members of the target audience or experts was not provided.</td>
</tr>
<tr>
<td>Construct validity – Structural validity</td>
<td>+</td>
<td>Factors should explain at least 50% of the variance</td>
</tr>
<tr>
<td></td>
<td>?</td>
<td>Explained variance not stated</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>Factors explain &lt;50% of the variance</td>
</tr>
</tbody>
</table>

1 Table informed by criteria and recommendations of other studies (97,163)
<p>| Construct validity – Hypothesis testing | + | Correlations with measures of the same construct should be &gt;0.50 OR at least 75% of the results in accordance with hypothesised associations AND correlations with related constructs should be higher than with those reported with unrelated constructs |
|                                         | ? | Only report correlations with unrelated constructs OR the extent to which between group differences are expected is not described /justified |
|                                         | - | Correlations with measures of the same construct are &lt;0.50 OR &lt; 75% of the results in accordance with hypothesized associations OR correlations with related constructs are lower than those reported with unrelated constructs |
| Construct validity – Known-group validity (not part of the COSMIN checklist) | + | Hypothesised between group differences are supported (or can be assumed) AND between group differences are statistically significant |
|                                         | ? | Between group differences are poorly hypothesized, but between group differences are statistically significant |
|                                         | - | Expected between group difference poorly defined or justified AND the statistical significance of between group differences not reported |
| Reliability                             | + | Cronbach’s alpha(s) ≥ 0.70 |
|                                         | ? | Cronbach’s alpha not determined or dimensionality unknown |
|                                         | - | Cronbach’s alpha(as) &lt; 0.70 |
|                                         | + | Intra-class Correlation Coefficient (ICC)/ weighted Kappa ≥0.70 OR Pearson’s r ≥0.80 |</p>
<table>
<thead>
<tr>
<th>Reliability (test-retest / inter-rater / intra-rater)</th>
<th>?</th>
<th>Neither ICC/ weighted Kappa, nor Pearson's r established</th>
</tr>
</thead>
<tbody>
<tr>
<td>-</td>
<td>ICC/ weighted Kappa &lt;0.70 OR Pearson’s r &lt;0.80</td>
<td></td>
</tr>
</tbody>
</table>

| Reliability – measurement error | N/A | Descriptive (not rated) |

| Responsiveness | + | Change-score correlations with measures of the same construct are >0.50 OR at least 75% of the results are in accordance with hypothesized associations OR the Area Under the Curve (AUC) is >0.70 AND change-score correlations with measures of related constructs are higher than those reported with unrelated constructs |
| ? | Solely correlations with unrelated constructs |

| - | Change-score correlations with measure of the same construct <0.50 OR < 75% of the results are in accordance with hypothesized associations OR AUC is <0.70 AND change-score correlations with related constructs are lower than those reported with unrelated constructs |

| Interpretability | N/A | Descriptive (not rated) - requires evidence that the minimal important (within-person) change and/or minimal importance (between group) difference exceeds evidence of the smallest detectable difference. Supported by evidence of acceptable data quality (score distribution, absence of end effects (floor/ceiling)) |
**Validity**

Validity is defined as the extent to which a PROM measures the construct it intends to measure (118,160). It is not a general characteristic of a measure and never absolutely determined; rather, it is evaluated as part of an ongoing process. There are three types of validity: content, construct and criterion (160).

**Content validity:**

Content validity does not require statistical analysis; rather, a qualitative appreciation of the content of a PROM. Determines if PROM content is relevant to, and reflective of, the target construct (160). Two component of content validity are informed by a subjective appraisal:

- **Face validity:** Overall, does it capture relevant concepts? For example, if more than fifty percent of items in a measure of depression focused on anxiety, the measure may be deemed to lack ‘face validity’.
- **Content validity:** Are individual items relevant and comprehensive?

Relevance may be addressed by considering the following questions (160):

- Is it relevant to the target population (age, sex, disease characteristics)?
- Is it comprehensive enough?
- Are the items relevant to the purpose of the measurement’s application?

Comprehensiveness requires knowledge of the target construct. Item comprehensiveness may, for example, be compared to an empirically derived conceptual model of the target construct.

Developers often omit to evaluate face or content validity meaning little is often know about the relevance of items to the target population, or to the construct under measurement. Additionally, there is often poor theoretical underpinnings and no conceptualisation of the construct, meaning it is difficult to deduce whether the measure contains a reflective item-set of the construct.

**Construct validity**

Three sources of evidence can be collectively considered to demonstrate or challenge the construct validity of a measure: structural validity, hypothesis testing and cross-cultural validity.
**Structural validity**

Dimensionality is the number of domains that a measure assesses and is determined using a factor analysis (FA) (161); a statistical method that evaluates the underlying structure of a multi-item PROM through the identification of different dimensions, or components, into which questions or items may group (162). There are two approaches to conducting FA: exploratory factor analysis (EFA) and confirmatory factor analysis (CFA).

EFA is an approach best utilised when there is no quantitative evidence about dimensionality. This is usually conducted during the development phase of a measure to identify the number of domains within a measure, and the items that are contributing to measurement.

CFA is used to test whether a hypothesised domain structure of a measure fits with the data, and therefore is a hypothesis-driven evaluation. Multiple alternative models can also be tested at this stage to ascertain whether there is a better fit model than the hypothesised or proposed model (160).

**Hypothesis testing**

Hypothesis testing involves crafting a priori hypotheses about the relationship between scores on the measure under evaluation and other measures of similar, or dissimilar constructs (163). For example, in a study comparing the Functional Assessment of Chronic Illness Therapy (FACIT) fatigue against three other measures of fatigue, it was expected that due to theoretical similarities between the measures i.e. all purporting to measure fatigue, there would be significant association with the FACIT-fatigue (164). This was assessed using correlations, however, the hypothesised relationship only stated that a relationship was expected, with no magnitude specified and the direction (a positive relationship) was implied. The study found moderate to strong correlations ranging between the three measures and the FACIT-fatigue ranging from 0.52 and 0.79 (164). Current criteria indicate that correlations with measures of the same construct should be >0.50 (97) to be sufficient evidence of convergent validity.

Evidence of convergent and/or divergent validity adds to the evidence base of whether the measure is capturing the intended construct. However, generating hypotheses a priori is important to reduce bias; moreover, hypotheses should be clear, directional (positive or negative) and indicate the magnitude of the anticipated relationship with justification.
A second approach can be to examine known groups validity – that is, administering the measure to two groups known to differ on the underlying construct. Here, researchers may hypothesise the extent of difference expected between scores, for patients, on a target construct. For example, in a longitudinal evaluation of the arthritis-specific work productivity survey, it was hypothesised that axSpA patients with worse health status would have higher losses of pay and lower household work productivity; a finding which was confirmed in the study (165).

**Cross-cultural validity**

Cross-cultural validity addresses the issues relating to item relevance and performance that may arise from the translation process or cultural differences. Mokkink et al. (102) define cross-cultural validity as:

> ‘the degree to which the performance of the items on a translated or culturally adapted PRO instrument are an adequate reflection of the performance of items in the original version of the instrument’

Therefore, equivalence between the scores of those completing the new translated measure is sought with a sample completing the original measure (160). The new translated measure can then be evaluated in the same way as the original: evaluating structural validity, internal consistency and so forth.

**Reliability**

Reliability refers to score consistency, not score validity (118) and measures should: (i) produce consistent scores in individuals who have not changed on the target construct, and (ii) be as free from measurement error as possible (102,160). In addition, items within a given measure should be sufficiently correlated with one another, indicating they are measuring the same construct. There are three types of reliability: test-retest, internal consistency and measurement error.

**Test-retest reliability**

Test-retest reliability is a measure of the consistency of the PROM across two time-points where the respondent is stable i.e. there has been no change in their status on the target construct (118,160,166). In health outcome measurement, determining the stability of a respondent’s health status can be done using anchor-based
Internal consistency

Internal consistency is a measure of interrelatedness amongst items within a given measure (166). Items can be assessed at a scale level using Cronbach’s alpha or at item level using item-total correlation (ITC) or corrected ITC.

Item-total correlation (ITC) provides information about the extent to which scores on a given item correlate with the total score. To produce an ITC, the item score is correlated against the total score of the scale. However, this calculation means the total score includes the item score – that is, the item is being correlated against itself which inflates the ITC. Corrected ITC seeks to address this by finding ways to ‘remove’ the item score from the total score of the scale. Ways to deal with this include simply removing the item during analysis, mathematically omitting the item score or adjusting the score (167) – all of which produce a corrected ITC.

Inclusion of multiple items about a clearly defined domain generates more information about the construct, and if they relate to one another, reliable information about the domain can be obtained. Where items do not relate with one another, error may be introduced. It has been proposed that ITC’s below 0.3 (166) indicate a poor relationship between the item score and total score, thus highlighting the item for potential revision or deletion from the measure.

Cronbach’s Alpha assesses the relationship between the items within a scale (166). Such an assessment is important to identify which items are measuring the ‘same’ concept within a scale and therefore contributing its measurement, and those which are not. Values between 0.7 and 0.9 are generally accepted as evidence that the items are measuring something different about a construct (163,168). Some researchers accept alpha values up to 0.95 as being indicative of strong internal consistency, with anything >0.95 thought to highlight probable redundancy within the measure (i.e. some items are likely measuring the same thing) (161).

Measurement error

Measurement error is defined as “The systematic and random error of a patient’s score that is not attributed to true changes in the construct to be measured” (102).
Random error may have no identifiable reason for occurring and is unbiased, with its influence on scores increasingly diluted across repeated measurements. Systematic error can also occur and is a significant challenge that introduces bias to scores. Sources of systematic error can include the respondent or the PROM itself which contaminates scores over time and subsequently undermines the measures reliability. Additionally, systematic measurement error can lead to inappropriate score assessments and risk rendering the measure ineffective. There are a group of statistical approaches available to investigate measurement error including standard error of measurement (SEM) and limits of agreement (118,160).

**Responsiveness**

Responsiveness is defined as “the ability of an instrument to detect change over time in the construct to be measured” (102). As change should be detectable on the underlying construct of the measure, if the measure truly measures what it purports to, then responsiveness can be seen as an aspect of validity (‘longitudinal validity’). However, an important distinction is that responsiveness is estimated from two assessment points (therefore longitudinal), whilst validity is usually estimated in a cross-sectional assessment. There are two approaches to estimating responsiveness: criterion and construct approaches.

A criterion approach uses a ‘gold standard’ (where available) to assess the extent changes in scores on the measure of interest reflect changes in scores on the ‘gold standard’ (160).

A construct approach relies on hypothesis testing and is frequently used when a gold standard is not available. Hypotheses are usually generated a priori to minimise the risk of bias and may postulate expected differences between change scores between the target measure and other measures known to be responsive. Alternatively, change on measure (A) may be anticipated to correlate with change on measure (B) due to similarity between the constructs (160).

### 2.1.2 Existing reviews of measures of fatigue

Although no review has been conducted to evaluate the quality, relevance and acceptability of measures used to assess fatigue in axSpA, three reviews were identified that evaluated the quality of fatigue measures: one in neurological
conditions (169), one in chronic fatigue syndrome (CFS) (99), and one in long-term chronic health conditions (152).

In a review of the quality of measures used in the assessment of fatigue in central nervous system disorders, 17 measures were identified with evidence of psychometric evaluation (169). The review found that, despite measures previously being endorsed as having good psychometric evidence, this was only when evaluated using classic test theory methods. Evaluation studies that used modern psychometric methods (item response theory; IRT) highlighted inadequacies in the underlying construct, scoring structure and highlighted item redundancy within measures used in this population (169). However, this review did not evaluate study methodological quality, nor did it provide detail or transparency regarding its data extraction processes. In fact, the authors made arbitrary judgements about measurement property quality, drawing on existing criteria taken from another study and review (170,171).

In contrast, the review into the quality and acceptability of PROMs used in CFS was substantively more structured, drawing on recommendations for PROM development and evaluation and existing reviews to support transparent data extraction, and appraisal and utilising the COSMIN checklist to evaluate study methodological quality (102,103). This review identified 77 measures – 11 CFS-specific, 55 domain-specific and 11 generic – but found almost all the measures had limited evidence of their quality (99). In addition, the study reported whether measures had evidence of content validity to support their use in CFS – a measurement property seldom reported and often overlooked.

Finally, a third review which reviewed measures of fatigue used in chronic conditions described 39 measures: six established measures had good evidence of their psychometric properties, but no measure had a full range of psychometric evidence to support use (152). 17 measures were newly developed without any essential evidence of key measurement properties (152).

Of the three fatigue reviews identified, only one used the COSMIN checklist to support logical and transparent assessment of both study and PROM quality, considered the content validity of the identified measures and involved patient research partners (99).

Selection of an appropriate PROM should consider how robust the development process was (including relevant patient involvement as research partners), its intended purpose, target sample, conceptual underpinnings and evidence of its
psychometric properties (59,172). To further advance the ways in which PROM reviews are conducted, my systematic review will report the development process of all identified PROMs (including whether patient partners were involved), and empirically evaluate content validity through an item-content appraisal using an existing conceptual framework of fatigue.

2.1.3 Review aims and objectives

The aim of the review was:

1. To systematically appraise, compare and synthesise evaluative evidence of single- and multi-item fatigue-specific PROMs used with axSpA patients to determine their quality and acceptability.

2. To provide evidence-based recommendations on the ‘best’ measure of fatigue for use in axSpA research and clinical practice.

There are two potential outcomes from the review which will dictate the next research steps.

a. A measure emerges that is deemed to have both evidence of relevance and sufficient psychometric properties to support its uptake and use. If multiple measures are found to fit these criteria, a “best measure” will be identified.

b. No measure emerges with evidence of adequate relevance and psychometric properties leading to one of two possible avenues for pursuit: conceptualising fatigue in axSpA to compare with the existing RA-fatigue model:

i. If the models are identical or have significant alignment, then modification of the BRAF-MDQ may be sufficient to fulfil measurement needs in axSpA-fatigue.

ii. If there is little or no similarity between the models, a new measure specific to the needs of axSpA patients in relation to fatigue may need to be developed.
2.4. **Methods**

This systematic review was registered on the International prospective register of systematic reviews (Reference: CRD42016042271) in advance of it being conducted, and a protocol paper was written and submitted prior to the analysis and completion of the review (173).

2.4.1. **Search strategy**

To maximise the searches ability to capture evaluation studies of PROMs, a two-phase search strategy was conducted: 1) search to identify fatigue PROM evaluation studies; 2) PROM-specific searches (see Appendix 2A). Searches were conducted from the beginning of January 1980 until the end of August 2017 across five databases: Medline (OVID), Embase (OVID), PsycINFO (OVID), Cumulative Index of Nursing and Allied Health Literature and Web of Science. Medical subject headings and free text terms were identified and grouped to form filters specific to: (i) population (axSpA); (ii) construct (fatigue); (iii) assessment type (PROMs); and (iv) evaluation evidence (measurement and practical properties). Development of filter iii (assessment type) was informed by a filter developed by a PROM group and the knowledge centre of Oxford University, within the department of Public Health (174). Filter iv (evaluation evidence) was developed through modifying an existing filter designed specifically to identify studies of PROM evaluation in PubMed (175). This filter has high, demonstrable levels of sensitivity and precision with terms covering a breadth of measurement terms (175). In phase 2 searches, assessment type was replaced with named measures identified as a result of the initial phase 1 search. Reference lists of included studies and existing reviews were checked (150–152).

2.4.2. **Study selection**

I conducted all searches and assessed the identified titles and abstracts. A 10% subset of titles and abstracts were independently checked by a second author (KLH) to ensure consistency in the application of eligibility criteria. A third author (JCP), who is a consultant rheumatologist, double-assessed all abstracts that related to psoriatic arthritis (PsA) to ensure only data relating to patients clinically recognised as having axSpA were extracted. Any disagreements were resolved through discussion.
Table 2.2: Eligibility criteria

<table>
<thead>
<tr>
<th>Inclusion</th>
<th>Study-specific criteria</th>
<th>PROM-specific criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Contains a clearly identifiable and reproducible patient-reported assessment of fatigue</td>
<td>PROM is fatigue-specific, assessed fatigue within a multidomain measure or were single/multi-item assessments</td>
</tr>
<tr>
<td></td>
<td>Reports evidence of PROM development and/or evaluation following completion by axSpA patients</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Written or available in English</td>
<td></td>
</tr>
<tr>
<td>Exclusion</td>
<td>Available only as an abstract</td>
<td>Assessments were clinician-reported</td>
</tr>
<tr>
<td></td>
<td>Fatigue assessment was not patient-reported or clearly identifiable or reproducible</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Describes PROM application only</td>
<td></td>
</tr>
</tbody>
</table>

2.4.3. Data extraction

Development of data extraction sheets was informed by previous published reviews (98,99,163,176) and the COnsensus-based Standards for the selection of health status Measurement INstruments (COSMIN) checklist (102,103,159), to ensure a standardised and consistent extraction of data. Study and PROM-specific information was extracted as per the study eligibility criteria (see Table 2.2). Measurement properties and quality criteria are reported in Table 2.1. Practical properties were also sought and included evidence of feasibility (administration time and scoring) and acceptability (relevance to patients) which was determined through consideration of the original development paper for the PROM and a content appraisal of PROM items against an existing conceptual model of fatigue from a related condition – rheumatoid arthritis (50). I extracted data for all studies into the data extraction sheets; a second author (KLH) selected a 10% subset and independently double-extracted into separate data extraction sheets. Agreement was checked by comparing the data identified in each article and extracted, followed by a discussion during a face-to-face meeting.
2.4.4. Assessment of study methodological quality

COSMIN uses a four-point checklist to inform assessment of study methodological quality per reported measurement property. The checklist contains 119 items to appraise methodological quality in relation to 10 categories. Each measurement property can achieve one of four ratings: poor, fair, good or excellent (102,103,159). The overall score is informed by the lowest item rating for the given measurement property. A definition for all ratings for each stage of the analysis is provided in Table 2.3.

2.4.5. Assessment of PROM quality

PROM quality was determined using a synthesis of recommendations described in other studies (97,98,159,163) which facilitate transparency in PROM quality appraisal. Each measurement property per study (where available) was appraised and rated as either: unclear (?); conflicting (±); inadequate (-); adequate (+) in accordance with the definitions outlined in Table 2.3.

2.4.6. Data synthesis

Data synthesis was informed by four factors: (i) study methodological quality (COSMIN) – determining the methodological quality of the evidence available: (ii) the number of studies reporting evidence – the strength of the evidence base: (iii) ratings for measurement/practical properties per measure – consideration of the strength of evidence of a given property; and (iv) consistency of the results between studies – confidence that the finding is accurate (97,98,163,176). Together with the evidence of PROM quality, the final data synthesis reflects: (i) the quality of each measurement property and, (ii) the overall level of evidence per measurement property (97,98).

Table 2.3: Quality and data synthesis criteria with definitions

<table>
<thead>
<tr>
<th>Analysis stage</th>
<th>Criteria</th>
<th>Definitions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study quality²</td>
<td>Study methodological quality criteria</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Poor</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Fair</td>
<td>-</td>
</tr>
</tbody>
</table>

² specific criteria for judging study quality are specified, per measurement property, in the COSMIN checklist (103)
### Data synthesis

| Good | - |
| Excellent | - |

**PROM quality criteria (97,98,163,176)**

| Unclear (?) | Not possible to determine |
| Conflicting (±) | Conflicting findings |
| Inadequate (-) | Findings do not meet minimum accepted standards for the measurement property |
| Adequate (+) | Findings meet or exceed minimum accepted standards for the measurement property |

**Strength of evidence (97,98)**

| Strong | Consistent findings in multiple studies of good methodological quality OR in one study of excellent quality |
| Moderate | Consistent findings in multiple studies of fair methodological quality OR in one study of good quality |
| Limited | One study of fair methodological quality |
| Conflicting | Conflicting findings |
| Unknown | Only studies of poor methodological quality |

2.4.7. **Comparative item content appraisal**

The conceptual model of RA-fatigue is made up of three overarching domains describing eight subdomains: disease specific with no sub domains; cognitive/behavioural with subdomains behaviour, cognition and emotion; personal with subdomains support, health, environment and responsibilities, and symptoms with no sub domain (see Chapter 1, Figure 1.1) (50). For the purpose of conducting a comparative item content appraisal, each item within each PROM was assigned to a domain within the RA-fatigue model based on its content (e.g. a question about low mood would be assigned to the emotion domain). The extent to which each PROM captures the domains of the RA-fatigue model is discussed in the results.

2.4.8. **PROM recommendations**

There is no current guidance on how to select or recommend a 'best' PROM in this type of review – therefore, I generated the following criteria which takes into
consideration key minimally important measurement properties, and evidence from both the data synthesis and comparative item-content appraisal (146):

1. Content validity – How closely do the key domains of fatigue that were identified in the RA-fatigue model resemble those in the PROM?
2. Is there adequate evidence of both structural and construct validity, and reliability (internal consistency and test–retest)?
3. Is there at least moderate quality evidence to support confidence in the findings?

2.4.9. Sensitivity analysis
The COSMIN checklist introduces significant rigour to the evaluation of methodological quality for studies of measurement properties; however, the stringent criteria of COSMIN may be overly conservative and significantly reduce the evidence base that inferences can be drawn from. This may be particularly noticeable for legacy measures whereby such standards did not exist. As this systematic literature review was only able to identify 23 eligible studies for inclusion, a sensitivity analysis is prudent to maximise the potential to draw inferences from the data, albeit with caution. This sensitivity analysis will provide a synthesis of data from all studies irrespective of their methodological quality and consider the adequacy and consistency of the study findings.

2.5. Results
2.5.1. Identification of studies and PROMs
The review process and search results are summarised using a Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flowchart (see Figure 2.1). Evidence for nine fatigue-specific PROMs was extracted from twenty-three articles. The PROMs were grouped into three categories: multi-dimensional, unidimensional and single-items.

_Multidimensional measures:_ Multi-dimensional Fatigue Symptom Inventory–Short Form (MFSI-SF) (177), Multi-dimensional Assessment of Fatigue (MAF) (178) and Multi-dimensional Fatigue Inventory (MFI-20) (42).

_Unidimensional measures:_ Functional Assessment of Chronic Illness Therapy (Fatigue) (FACIT-fatigue) (61), Fatigue Severity Scale (FSS) (179) and the vitality subscale (VT) of the Short-Form 36-item Health Status Survey (SF-36) (66).
Single-item measures: Worst-Fatigue Numeric Rating Scale (WF-NRS) from the Brief Fatigue Inventory (BFI) (180), the 10 cm fatigue severity VAS (from the BASDAI) (76) and a modified 10 cm VAS with the original descriptor ‘none’ changed to ‘no problem’ (181).

2.5.2. Study and sample characteristics
Participants in all studies were adults with a primary diagnosis of axSpA (including AS) and were aged between 18 and 72. Sample sizes in the studies ranged between 40 and 812. Almost all studies were cross-sectional investigations of fatigue prevalence or exploration of its association with other variables. Some data was also extracted from trials (see Appendix 2B). An overview of PROM content is provided in Table 2.6.

2.5.3. Measurement properties and methodological quality
Study methodological quality was assessed and recorded, per PROM, and is detailed in Table 2.4. A synthesis of evidence is presented in Table 2.5. There was no evidence identified for the following measurement properties: measurement error, content validity, structural validity, criterion-based responsiveness, measurement acceptability or feasibility of completion.

2.5.4. Fatigue conceptualisation and patient involvement
Following review of PROM development articles, there is very limited evidence of fatigue conceptualisation for four PROMs (MFI-20, MFSI-SF, SF-36 and BFI). The process of item generation, selection and refinement was either not reported or poorly reported and lacked transparency. Only one measure – the single-item fatigue-severity VAS was specifically developed for use with axSpA patients; however, this measure did not provide any conceptualisation of fatigue. Patient involvement in the studies was as research participants only and no study involved patients as partners in measurement evaluation. The item-content appraisal, per PROM, against the Hewlett RA-fatigue model is presented in Table 2.6.
Figure 2.1: PRISMA flow chart of article identification and selection for this study adapted from Pearson et al. (146)
Table 2.4: Methodological quality (COSMIN) per study (n=23) per PROM (n=9) and investigated measurement properties adapted from Pearson et al. (146)

<table>
<thead>
<tr>
<th>PROM / Study</th>
<th>Country (language)</th>
<th>(n)</th>
<th>Reliability</th>
<th>Validity</th>
<th>Responsiveness</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Internal consistency</td>
<td>Measurement error</td>
<td>Content validity</td>
</tr>
<tr>
<td><strong>Multidimensional fatigue measures (3/9)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>MAF</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aissaoui et al. (182)</td>
<td>Morocco Arabic</td>
<td>110</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Durmus et al. (183)</td>
<td>Turkey Turkish</td>
<td>43</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ibn Yacoub et al. (184)</td>
<td>Morocco Arabic</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stebbings et al. (185)</td>
<td>New Zealand English</td>
<td>67</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Turan et al. (186)</td>
<td>Turkey Turkish</td>
<td>68</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Akkan et al. (187)</td>
<td>Turkey Turkish</td>
<td>110</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>MFI-20</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3 COSMIN provides a grading system for four possible rating outcomes: Excellent / Good / Fair / Poor (103,159).

4 Measurement property quality has four possible rating outcomes (97,163): (1) adequate (+) - it fulfils the assessment criteria; (2) inadequate (-) - it has failed to meet the assessment criteria; (3) conflicting (+/-) – the evidence is conflicting and therefore difficult to interpret; and (4) unclear (?) – the results are unclear.
<table>
<thead>
<tr>
<th>Study</th>
<th>Country</th>
<th>Language</th>
<th>Sample Size</th>
<th>Quality</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Da Costa et al. (188)</td>
<td>Canada</td>
<td>English</td>
<td>125</td>
<td>Fair</td>
<td>Mean (SD), ES, SRM, Guyatt</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(US)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>van Tubergen et al. (43)</td>
<td>The Netherlands</td>
<td>Dutch</td>
<td>40</td>
<td>Fair</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Sub-group – Arm 1</td>
</tr>
<tr>
<td></td>
<td>The Netherlands</td>
<td>Dutch</td>
<td>812</td>
<td>Poor</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Whole population</td>
</tr>
<tr>
<td></td>
<td>The Netherlands</td>
<td>Dutch</td>
<td>776</td>
<td>Poor</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Patients with BASDAI fatigue VAS score =5 excluded</td>
</tr>
<tr>
<td><strong>MFSI-SF</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gunaydin et al. (189)</td>
<td>Turkey</td>
<td>Turkish</td>
<td>63</td>
<td>Poor</td>
<td>Poop</td>
</tr>
</tbody>
</table>

**Unidimensional fatigue measures (2/9)**

**FACIT-fatigue**
<table>
<thead>
<tr>
<th>Study</th>
<th>Country/Region</th>
<th>Language</th>
<th>Sample Size</th>
<th>Quality</th>
<th>Measure/Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Revicki et al. (190)</td>
<td>Canada/US/Europe/Canada</td>
<td>English</td>
<td>82</td>
<td>Poor</td>
<td></td>
</tr>
<tr>
<td>Maksymowych et al. (191)</td>
<td>Canada (US)</td>
<td>English</td>
<td>302</td>
<td>Poor</td>
<td></td>
</tr>
<tr>
<td>FSS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bedaiwi et al. (192)</td>
<td>Canada (US)</td>
<td>English</td>
<td>457</td>
<td>Poor</td>
<td></td>
</tr>
<tr>
<td>Schneeberger et al. (193)</td>
<td>Argentina</td>
<td>Spanish</td>
<td>159</td>
<td>Poor</td>
<td>Poor</td>
</tr>
<tr>
<td>Wanders et al. (194)</td>
<td>The Netherlands</td>
<td>Dutch</td>
<td>40</td>
<td>-</td>
<td>Mean (SD), mean change, ES, SRM, T-statistic</td>
</tr>
<tr>
<td>van Tubergen et al. (43)</td>
<td>The Netherlands</td>
<td>Dutch</td>
<td>40</td>
<td>Fair</td>
<td></td>
</tr>
<tr>
<td></td>
<td>The Netherlands</td>
<td>Dutch</td>
<td>812</td>
<td>Poor</td>
<td></td>
</tr>
</tbody>
</table>

Single-item fatigue measures (3/9) – WF-NRS separately appraised (qualitative study) (WF-NRS a single item taken from BFI, but practical properties reviewed ONLY)
<table>
<thead>
<tr>
<th>Study</th>
<th>Country</th>
<th>Language</th>
<th>Sample Size</th>
<th>Quality</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aissaoui et al. (182)</td>
<td>Morocco</td>
<td>Arabic</td>
<td>110</td>
<td>Poor</td>
<td></td>
</tr>
<tr>
<td>Dernis-Labous et al. (195)</td>
<td>France</td>
<td>French</td>
<td>639</td>
<td>Poor</td>
<td>SRM only</td>
</tr>
<tr>
<td>Fallahi et al. (196)</td>
<td>Iran</td>
<td>Persian</td>
<td>163</td>
<td>Poor</td>
<td></td>
</tr>
<tr>
<td>Gunaydin et al. (189)</td>
<td>Turkey</td>
<td>Turkish</td>
<td>63</td>
<td>Poor</td>
<td></td>
</tr>
<tr>
<td>Ibn Yacoub et al. (184)</td>
<td>Morocco</td>
<td>Moroccan</td>
<td>100</td>
<td>Poor</td>
<td></td>
</tr>
<tr>
<td>Park et al. (197)</td>
<td>Korea</td>
<td>Korean</td>
<td>50</td>
<td>Poor</td>
<td>Poor</td>
</tr>
<tr>
<td>Revicki et al. (190)</td>
<td>US, Europe</td>
<td>English</td>
<td>397</td>
<td>Good</td>
<td></td>
</tr>
<tr>
<td>Schneeberger et al. (193)</td>
<td>Argentina</td>
<td>Spanish</td>
<td>159</td>
<td>Poor</td>
<td></td>
</tr>
<tr>
<td>Stebbings et al. (185)</td>
<td>New Zealand</td>
<td>English</td>
<td>67</td>
<td>Poor</td>
<td></td>
</tr>
<tr>
<td>Yilmaz et al. (198)</td>
<td>Turkey</td>
<td>Turkish</td>
<td>74</td>
<td>Poor</td>
<td>Poor</td>
</tr>
<tr>
<td>Alkan et al. (187)</td>
<td>Turkey</td>
<td>Turkish</td>
<td>110</td>
<td>Poor</td>
<td></td>
</tr>
<tr>
<td>Dagfinrud et al. (44)</td>
<td>Norway</td>
<td>Norwegian</td>
<td>302</td>
<td>Fair</td>
<td></td>
</tr>
<tr>
<td>van Tubergen et al. (43)</td>
<td>Austria</td>
<td></td>
<td>40</td>
<td></td>
<td>Mean (SD), ES, SRM, Guyatt</td>
</tr>
<tr>
<td>Study</td>
<td>Country(s)</td>
<td>Language(s)</td>
<td>Sample Size</td>
<td>Quality</td>
<td>Fatigue Measure</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>-----------------</td>
<td>-------------</td>
<td>-------------</td>
<td>---------</td>
<td>-----------------</td>
</tr>
<tr>
<td>Fernandez-Sueiro et al. (199)</td>
<td>Spain</td>
<td>Spanish</td>
<td>103</td>
<td>Poor</td>
<td>Modified 10cm VAS</td>
</tr>
<tr>
<td>Wheaton et al. (181)</td>
<td>Canada (US)</td>
<td>English</td>
<td>140</td>
<td>Poor</td>
<td>Fatigue-specific PROM subscale (1/9)</td>
</tr>
<tr>
<td>Revicki et al. (190)</td>
<td>US and Europe</td>
<td>English</td>
<td>397</td>
<td>Good</td>
<td>SF-36 vitality subscale</td>
</tr>
<tr>
<td>Bodur et al. (200)</td>
<td>Turkey</td>
<td>Turkish</td>
<td>962</td>
<td>Poor</td>
<td>SF-36 vitality subscale</td>
</tr>
<tr>
<td>Durmus et al. (183)</td>
<td>Turkey</td>
<td>Turkish</td>
<td>43</td>
<td>Poor</td>
<td></td>
</tr>
<tr>
<td>Turan et al. (186)</td>
<td>Turkey</td>
<td>Turkish</td>
<td>68</td>
<td>Poor</td>
<td></td>
</tr>
<tr>
<td>van Tubergen et al. (43)</td>
<td>The Netherlands</td>
<td>Dutch</td>
<td>812</td>
<td>Poor</td>
<td></td>
</tr>
<tr>
<td></td>
<td>The Netherlands</td>
<td>Dutch</td>
<td>776</td>
<td>Poor</td>
<td></td>
</tr>
<tr>
<td>Yilmaz et al. (198)</td>
<td>Turkey</td>
<td>Turkish</td>
<td>74</td>
<td>Poor</td>
<td></td>
</tr>
</tbody>
</table>

*Modified 10cm VAS*
<table>
<thead>
<tr>
<th>Study</th>
<th>Country</th>
<th>Language</th>
<th>Sample Size</th>
<th>Effect Size</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Akan et al. (187)</td>
<td>Turkey</td>
<td>Turkish</td>
<td>110</td>
<td>Poor</td>
<td></td>
</tr>
<tr>
<td>Wanders et al. (194)</td>
<td>Netherlands</td>
<td>Dutch</td>
<td>40</td>
<td>Mean (SD),</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>mean change, ES, SRM</td>
<td></td>
</tr>
</tbody>
</table>
Table 2.5: Data synthesis\(^5\), levels of evidence and overall quality of reviewed PROMs (n=9) adapted from Pearson et al. (146)

<table>
<thead>
<tr>
<th>PROM / Study</th>
<th>Number of evaluation studies</th>
<th>Reliability</th>
<th>Validity</th>
<th>Construct Validity</th>
<th>Responsiveness</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Internal consistency</td>
<td>Reliability</td>
<td>Measurement error</td>
<td>Content validity</td>
<td>Structural validity</td>
</tr>
<tr>
<td>Multidimensional fatigue measures (3/9)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MAF</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td>?</td>
<td></td>
</tr>
<tr>
<td>MFI-20</td>
<td>2</td>
<td>+ Limited</td>
<td>+ Limited</td>
<td></td>
<td>?</td>
<td>Unknown</td>
</tr>
<tr>
<td>MFSI-SF</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>?</td>
<td>Unknown</td>
</tr>
<tr>
<td>Unidimensional fatigue measures (2/9)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^5\) Data synthesis: synthesis of the data determined the quality and acceptability of each reviewed PROM. The synthesis considered four factors:

1) Study methodological quality (COSMIN scores)
2) Number of studies reporting evidence of measurement properties (per PROM)
3) Results for each measurement property (per PROM)
4) Consistency of results between studies.

The results of data synthesis include two ratings:

1) Overall measurement property quality: adequate (+), not adequate (-), conflicting (+/-), or unclear (?).
2) Levels of evidence for the overall quality per measurement property. Five outcomes were defined:

- `strong` – consistent findings in multiple studies of good methodological quality OR in one study of excellent quality; `moderate` – consistent findings in multiple studies of fair methodological quality OR in one study of good methodological quality; `limited` – one study of fair methodological quality; `conflicting` – conflicting findings; or `unknown` evidence – only studies of poor methodological quality (97,98).
<table>
<thead>
<tr>
<th>Measure</th>
<th>Responses</th>
<th>Practical Properties</th>
<th>Reliability</th>
<th>Validity</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>FACIT-fatigue</td>
<td>2</td>
<td>Unknown</td>
<td>+ Moderate</td>
<td>Unknown</td>
<td>ES, SRM</td>
</tr>
<tr>
<td>FSS</td>
<td>3</td>
<td>Unknown</td>
<td>Unknown</td>
<td>Unknown</td>
<td></td>
</tr>
<tr>
<td><strong>Single-item fatigue measures (3/9)</strong></td>
<td></td>
<td>Review of practical properties only</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BFI WF NRS</td>
<td>1</td>
<td></td>
<td></td>
<td>Unknown</td>
<td></td>
</tr>
<tr>
<td>10cm VAS (BASDAI)</td>
<td>15</td>
<td>- Limited</td>
<td>+ Moderate</td>
<td>Unknown</td>
<td>ES, SRM, Guyatt</td>
</tr>
<tr>
<td>Modified 10cm VAS</td>
<td>1</td>
<td></td>
<td>+ Limited</td>
<td>Unknown</td>
<td></td>
</tr>
<tr>
<td><strong>Fatigue-specific PROM subscale (1/9)</strong></td>
<td></td>
<td></td>
<td></td>
<td>Unknown</td>
<td></td>
</tr>
<tr>
<td>SF-36 (vitality domain)</td>
<td>8</td>
<td>+ Moderate</td>
<td>+ Limited</td>
<td>Unknown</td>
<td>ES, SRM</td>
</tr>
</tbody>
</table>
Table 2.6: PROM content appraisal of reviewed single- and multi-item fatigue measures (n=9) utilising Hewlett’s model of RA fatigue, adapted from Pearson et al. (146)

<table>
<thead>
<tr>
<th>PROM</th>
<th>Concepts of RA Fatigue</th>
<th>Symptoms</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Disease-specific Fatigue</td>
<td>Cognitive, Behavioural</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Behaviour</td>
</tr>
<tr>
<td>MAF</td>
<td>Multi-dimensional fatigue measures (3/9)</td>
<td>Physical impact (2) – ability to do chores in house, exercise.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Physical leisure activities (2) – sex, recreational activities.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Physical effect on daily activities (5) – cook, bathe, dress, walk, shop/errands.</td>
</tr>
<tr>
<td>MFI-20</td>
<td></td>
<td>Limitation (1) – physically feel I can only do a little.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Capability (1) – physically I feel I can take a lot on.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Activity level - (1) - I get little done</td>
</tr>
<tr>
<td>MFSI-SF</td>
<td>Memory (2) – trouble</td>
<td>-</td>
</tr>
<tr>
<td>FACIT-fatigue</td>
<td>-</td>
<td>Ability (1) – able to do usual activities.</td>
</tr>
<tr>
<td>-</td>
<td>Impairment (1) – I need help to do my usual activities.</td>
<td></td>
</tr>
<tr>
<td>-</td>
<td>Social activity (1) – have to limit because tired.</td>
<td></td>
</tr>
<tr>
<td>-</td>
<td>Motivation (2) – trouble starting, trouble finishing.</td>
<td></td>
</tr>
<tr>
<td>-</td>
<td>Mood (1) – frustrated by being too tired.</td>
<td></td>
</tr>
<tr>
<td>-</td>
<td>Impact (2) – need to sleep, too tired to eat.</td>
<td></td>
</tr>
<tr>
<td>-</td>
<td>Symptom manifestation (4) – fatigued, weak, listless, tired.</td>
<td></td>
</tr>
<tr>
<td>-</td>
<td>Energy (1) – I have energy.</td>
<td></td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>-</td>
<td>Interference (1) – work, family or social life.</td>
<td></td>
</tr>
<tr>
<td>-</td>
<td>Generic (1) – easily fatigued?</td>
<td></td>
</tr>
<tr>
<td>-</td>
<td>Global perception (1) – what number best reflects global fatigue</td>
<td></td>
</tr>
<tr>
<td>-</td>
<td>Frequency (1) – causes frequency problems</td>
<td></td>
</tr>
<tr>
<td>-</td>
<td>Causes of fatigue (1) -</td>
<td></td>
</tr>
<tr>
<td></td>
<td>BFI WF-NRS</td>
<td>10cm VAS</td>
</tr>
<tr>
<td>------------------------------</td>
<td>------------</td>
<td>----------</td>
</tr>
<tr>
<td><strong>Single-item fatigue measures (3/9)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BFI WF-NRS</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>10cm VAS</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Modified 10cm VAS</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>Fatigue-specific PROM subscale (1/9)</strong></td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>SF-36 VT (vitality subscale)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Sense of energy/fatigue (4)</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

- full of pep?
- A lot of energy?
- Feel worn out?
- Feel tired
2.5.5. **Multidimensional fatigue-specific PROMs**

**Multi-dimensional Assessment of Fatigue (MAF)** (178)

*Development:* The MAF was developed following revision of the Piper Fatigue Scale (PFS) (201) to measure fatigue in adult RA patients. It purports to measure three domains of fatigue: temporal fatigue, fatigue severity and sensory fatigue. The PFS is a 41-item measure, however, this was reduced to 16-items for the MAF which was created to measure four domains of fatigue: severity, distress, impact and timing. The ways in which items were selected, retained or generated was not described.

*Content appraisal:* There was no evidence of patient involvement as research partners during development and it is unknown whether the measure reflects what matters to axSpA-fatigue. Comparative appraisal indicates that of the nine subdomains, only four were addressed by the PROM: behaviour, emotion, responsibilities and symptoms. Items were asking about physical impact, physical leisure activities, physical effect on daily activities, emotional impact of fatigue, interference with social and work life and the severity, frequency and general fatigue experience.

*Measurement evaluation:* There was limited evidence of construct validity (correlations and known-groups validity) from six studies of poor methodological quality. All correlations lacked a priori hypothesised associations. Small to moderate associations were reported between the MAF total and AS-specific Bath measures (range 0.23-0.73), the MAF subscales and the SF-36 VT (range 0.3-0.53) and 10 cm single-item fatigue severity VAS (taken from the BASDAI; range 0.39-0.53) (182–187).

**Multi-dimensional Fatigue Inventory (MFI-20)** (42)

*Development:* The MFI-20 was developed to measure fatigue in cancer patients without somatic items. Initial PROM development was informed by both authors and previous research which culminated in five proposed domains of fatigue: general fatigue (GF), physical sensations (PS), mental fatigue (MF), reduced motivation (RM) and reduced activity (Ra).
Domains general, PS and MF were theoretically supported for inclusion based on other studies which identified these domains in their own studies (see Smets et al. (42) for more detail). The final two components were RM and Ra, however, it is unclear whether Ra was considered as a consequence of fatigue experience. The item development process was not described. Initial PROM evaluation was conducted with patients diagnosed with cancer, chronic fatigue syndrome and a healthy group who may experience physical fatigue through known causes (e.g. military personnel), or mental fatigue (e.g. newly qualified doctors).

**Content appraisal:** There was no evidence of patient involvement as research partners during PROM development. The cognitive/behavioural domain had items addressing all three subdomains (behaviour, cognition, emotion). Items for these domains explored limitation, capability, activity levels, cognition, forethinking, motivation and anxiety. However, only one other subdomain (health) was addressed with items on self-perceptions of health, leaving 5 subdomains unaddressed.

**Measurement evaluation:** Limited evidence was available of construct validity from one poor-quality study (188). Acceptable evidence of internal consistency (Cronbach’s α from 0.68 [RM subscale] to 0.86 [Ra subscale]) and construct validity (moderate to strong associations between subscales [general fatigue with PF 0.69/Ra 0.52/RM 0.45/MF 0.45; MF with PF 0.40/Ra 0.42/RM 0.48; RM with PF 0.51/Ra 0.54]) supporting assumed a priori hypothesised associations was found from a fair-quality study (43). In another study of fair methodological quality (188) there was limited evidence of 1-week test-retest reliability for patients after completion of an ‘overall perceived health’ visual analogue scale, taken from the EuroQoL (EQ-5D) (ICC range: PF 0.57–0.75 RM/MF) with values for the GF, PF and Ra subscales reported <0.70. The study (a spa therapy trial) also provided distribution-based measures of responsiveness (effect size statistics and standardised response means were calculated from the trial data without a priori hypotheses. After spa therapy concluded, a 3-month follow-up was conducted: small values were observed (<0.3) for domains reflecting reduced activity to large (>0.82) for domains reflecting general fatigue and PF (Effect Size (ES): GF 0.82/PF 0.81/RA 0.28/RM 0.54/MF 0.38; Standardised Response Mean (SRM): GF 0.70/PF 0.82/RA 0.23/RM 0.51/MF 0.49; Guyatt statistics: GF 0.86/PF 0.96/RA 0.30/RM 0.50/MF 0.57).
Multi-dimensional Fatigue Symptom Inventory—Short Form (MFSI-SF) (177)

Development: The MFSI-SF is a 30-item measure derived from a previous iteration of the MFSI – an 83-item measure which was developed using evidence from fatigue literature, expert discussions (with those who treat patients with fatigue) and a review of existing fatigue measures and is targeted at cancer patients. Factor analysis derived five domains: global fatigue and somatic, cognitive, affective and behavioural symptoms. The MSFI-SF is made up of the five empirically derived subscales which purport to measure general fatigue, physical fatigue, emotional fatigue, mental fatigue and vigor. Each domain consists of 6 items.

Content appraisal: There was no evidence of patient involvement as research partners, only as research participants. Only three subdomains were captured by the measure: cognition – with items on memory and concentration; emotion – with items on mood, and health – with items on symptom manifestation and self-perceived health.

Measurement evaluation: Available evidence was limited to only one study of poor methodological quality which had limited evidence of construct validity (189). Associations ranging from weak to strong were reported between the MFSI-SF subscales and the 10cm fatigue severity VAS taken from the BASDAI (10 cm VAS with GF 0.71/PF 0.74/emotional fatigue 0.56/MF 0.45/Vigor −0.32) following completion by 62 AS patients. A priori hypothesised associations were not reported, however, association between variables could be assumed.

2.5.6. Unidimensional fatigue PROMs

Functional Assessment of Chronic Illness Therapy – fatigue (FACIT-fatigue) (61)

Development: FACIT-fatigue is a 13-item fatigue-specific measure developed to measure cancer-related fatigue (61). Items ask about fatigue and its impact on daily activities. Two development phases were described: phase 1 – item generation using semi-structured interviews with cancer patients and medical experts; phase 2 – item reduction performed by presenting the generated items to a second group of medical experts for their review. No conceptualisation of fatigue was reported. A
preliminary psychometric evaluation in a sample of cancer patients demonstrated good evidence of validity and reliability.

Content appraisal: There were no patients involved as research partners, only participants. The cognitive/behavioural domain was reflected in the measure’s items on ability, impairment and social activity for the behaviour subdomain; items on motivation for the cognition subdomain and one item on mood for the emotion subdomain. Only one other subdomain – health, was addressed by the items which explored the impact on health and symptom manifestation.

Measurement evaluation: One study provided poor quality evidence of good internal consistency (Cronbach’s α 0.82/0.86) and item-level performance (corrected ITC: 0.56/0.88) (190). The same study also provided good-quality evidence of construct validity, reporting strong associations between FACIT-fatigue and the SF-36 VT subscale (range r = 0.74-0.82) and the 10 cm VAS (r = -0.69), and moderate associations with the BASFI (r = -0.56) and the BASDAI index score (r = -0.47) (190), thus confirming a priori hypotheses that variables would be related.

Fatigue Severity Scale (FSS) (179)

Development: FSS is a 9-item measure developed to measure fatigue in patients diagnosed with multiple sclerosis and systemic lupus erythematosus. The initial long-form was a 28-item measure which was reduced using the results of a factor analysis, item analysis alongside theoretical considerations – however, the theoretical considerations were not clearly outlined. There was no conceptual model of fatigue reported. A group of five sorted items (without labels) into domains to form the questionnaire structure.

Content appraisal: There were no patients involved as research partners. The symptom domain was addressed with items asking about generic fatigue experience, global reflection of fatigue, frequency and causes. A further three subdomains were addressed by the measure: behaviour, cognition and responsibilities. Behaviour items explored impact of fatigue, emotion items covered
motivation and a reflection of the respondents most disabling symptom, responsibilities covered inference with social, family and work.

**Measurement evaluation:** Two studies of poor methodological quality reported both strong (0.77) (192) and moderate (0.53) (193) associations between the FSS and 10 cm fatigue-severity VAS. In a placebo-controlled trial of etanercept, participants in both the placebo and intervention arms reported small ES at 28 days (ES 0.15/ -0.23; SRM 0.22/ 0.22) (194).

**Short Form 36-item Health Survey (SF-36) Vitality Subscale (VT). (66)**

**Development:** The SF-36 is a 36-item generic measure of health status that measures eight health domains: physical functioning, role limitations due to physical health or emotional problems, energy, pain, emotional wellbeing, social role functioning and general health (66). Many items included in the measure were selected from established measures. The content of these were reviewed in order to assign the content to the pre-defined domains. A previous iteration of the short form health survey provided the data for this (SF-20).

**Content appraisal:** Patients were not involved as research partners in the development process. The four-item vitality subscale addressed the health subdomain with items on the participants’ sense of energy and fatigue.

**Measurement evaluation:** One study of fair methodological quality provided acceptable evidence of construct validity (190), reporting a strong association between the SF-36 VT subscale and the FACIT-fatigue (r = 0.74; r = 0.82), a moderate association with the 10 cm VAS (r = -0.49) and a weak association with the BASFI (r = -0.33). Acceptable evidence of internal consistency and item-level performance was provided by one study of good methodological quality (Cronbach’s α 0.78/ 0.88; item-total correlation 0.57/ 0.64) (190). In one study examining the effects of a 25mg dose of etanercept twice weekly reported, moderate to large ES statistics at 28 days (ES = 0.54; SRM = 0.83) and 112 days (ES = 0.69; SRM = 0.75) (194).
2.5.7. **Single-item fatigue PROMs**

10 cm fatigue-severity VAS (taken from the BASDAI) (76)

*Development:* The 10 cm fatigue-severity VAS is the first question of a 6-item questionnaire that is specific to disease activity in AS – the Bath Ankylosing Spondylitis Disease Activity Index (BASDAI) (76). Questionnaire development was undertaken by a team of physiotherapists, research associates and rheumatologists. Patients completed a pilot version of the measure but were not involved as research partners. No conceptualisation of fatigue was reported.

*Content appraisal:* Patients diagnosed with AS participated in research activities related to development of the BASDAI but were not partners in the research process or item generation. The item only addressed fatigue severity, which reflects the symptoms domain.

*Measurement evaluation:* One study of good methodological quality provided acceptable evidence of construct validity, showing a strong association between the single-item VAS and the FACIT-fatigue (r = -0.69), and a moderate association with the SF-36 VT subscale (r = -0.49) following completion by AS patients in a double-blind, placebo-controlled clinical trial which supported a priori hypothesised associations (190). In another study of fair quality, some evidence of test-retest reliability was reported after a 6-week test-retest period in a group of patients defined as stable using the EuroQoL EQ-VAS (general health) but was judged to be below accepted standards for group analysis (ICC = 0.60) (43). Estimates for test-retest reliability were below accepted thresholds for use with groups (0.70) or individuals (0.90) (168). Comparison between recipients of a placebo or non-steroidal anti-inflammatory drugs (NSAIDs) showed small ES statistics (ES -0.35), however, at 6 weeks large ES statistics were reported for participants receiving active, spa therapy intervention (ES = 0.89; SRM = 0.89; Guyatt statistics 0.92) (43).

**Modified 10 cm fatigue-severity VAS** (181)

*Development:* This is a modified version of the 10 cm fatigue-severity VAS (see 10 cm fatigue-severity VAS for its development). In this modified version, the low-score
anchor has been modified from 'none' to 'no problem', thus modifying the meaning of the response scale.

*Content appraisal:* see 10cm fatigue-severity VAS.

*Measurement evaluation:* One study which had a poor methodological quality rating provided inadequate, poor quality interpretative guidance (181).

**Brief Fatigue Inventory (BFI) Worst-Fatigue Numeric Rating Scale (WF-NRS) (180)**

*Development:* The BFI, based on the Brief Pain Inventory, is a 10-item measure developed to assess the impact of fatigue in cancer patients over the last day (180). There are two subscales: fatigue severity consisting of 4-items, and fatigue impact consisting of 6-items. The content of items investigates severity of fatigue, interference with function, factors that exacerbate fatigue and factors contributing to increased fatigue. Questionnaire items were selected and modified from a revised fatigue questionnaire completed by cancer patients and healthy controls, in a previous study. Only limited conceptualisation was evidenced during the explanation of item revision.

*Content appraisal:* see 10cm fatigue-severity VAS.

*Practical considerations:* No evidence of measurement evaluation was available. One study provided a qualitative exploration of the WF-NRS’s relevance and acceptability (202). The item was judged to be relevant, however, the use of ‘best’ and ‘worst’ in the same item was considered confusing (“what best describes your worst fatigue”). Recommendations also included increasing the recall period beyond 24 hours as it was felt the short recall period failed to capture variability in fatigue experience.
2.5.8. Sensitivity analysis

The sensitivity analysis synthesised all data from all included studies as per the original data synthesis, irrespective of its methodological quality rating. As some data from poor studies was interpreted as part of the original analysis (due to no other evidence of higher quality being available), only measures with poor quality evidence that was not interpreted were included in the sensitivity analysis.

18 sources of evidence rated as ‘poor’ in the original data analysis were interpreted across five measurement properties for three PROMs: MFI-20, FACIT-fatigue and 10cm fatigue severity VAS.

**MFI-20:** evidence of construct validity testing and known groups validity was available from one study (43). Negative correlations were observed between each subscale of the MFI-20 with SF-36 VT (GF -0.73/ PF -0.65/ RA -0.59/ RM -0.53/ MF -0.42), however, no a priori hypotheses were reported and could not be assumed. Known groups evidence in the form of mean (standard deviation; SD) per subscale of the MFI-20 were reported for individuals identified as high fatigue (10cm fatigue severity VAS >5) and low fatigue (<5), however, there was overlap between groups across all subscales.

**FACIT-fatigue:** evidence of internal consistency reliability was available from one study (190). Cronbach’s alpha (α) and item-total correlations were reported at two time points: baseline (Cronbach’s α = 0.82/ ITC = 0.56) and week 12 (Cronbach’s α = 0.86/ ITC = 0.88). One study provided a receiver operating characteristic curve, however, this analysis was substantively flawed: the analysis controls for a variable included in the analysis (191). Therefore, findings from this analysis are not being considered or reported in this sensitivity analysis.

**10cm fatigue-severity VAS:** eleven studies provided evidence of construct validity of which six provided evidence of associations with other measures of disease or health-related quality of life (43,182,184,187,189,193,195–198). As it is unclear how fatigue measurement might be associated with measures such as disease activity, joint pain or duration of morning stiffness, the evidence from six studies is difficult to
interpret meaningfully without a priori hypotheses (182,184,195–198). Two studies provided evidence of a positive, moderate ($r=0.512$) (187) and positive, strong ($r=0.710$) (185) association with the MAF total score. One study provided evidence of a positive, moderate association with the FSS ($r=0.53$) (193). One study provided evidence of a negative, moderate association ($r=-0.64$) with the SF-36 VT subscale (43). One study provided evidence for associations of the 10cm VAS with each of the MFSI-SF subscales (GF 0.711/ PF 0.741/ EF 0.559/ MF 0.447/ Vigor -0.315) (189). These associations corroborate the established finding from the original synthesis that there is adequate, moderate evidence of construct validity for the 10cm single-item fatigue severity VAS – as determined from the better-quality studies in the data synthesis. Two studies provided partial evidence of interpretability showing a floor-ceiling effect of 30.1% - 1% (199) and 2% - 2% (197). However, no information about score distribution, missing values, change scores or a minimal important difference were provided.

Overall, the inclusion of these data produced very little change to ratings for the measurement properties. Only the internal consistency rating for the FACIT-fatigue would change from ‘unknown’ to ‘limited and adequate’ in the evidence synthesis. This is limited because the finding is from a single study, and adequate because the Cronbach's Alpha’s are within accepted standards. To surmise the implications of this sensitivity analysis: the findings would not lead to any material change in the overall judgements made as a result of the evidence synthesis.

2.6. Discussion
AxSpA patients have identified that understanding the impact of fatigue is an important priority (1), however, current assessment guidance is limited to a single-item measure of fatigue severity (2). Current guidance therefore fails to recognise and reflect the often significant and widespread impact of fatigue on an individual’s life. Only three of nine reviewed measures were multidimensional, containing items on different aspects of fatigue. Moreover, there were no measures that were specific to the experience of fatigue in axSpA, and no fatigue measures had been evaluated for relevance to axSpA patients. Evidence of reliability and construct validity was limited and the evidence mostly poor quality. There was no interpretative guidance or evidence of measurement error, content validity or structural validity for the reviewed measures. Evidence of responsiveness was confined to the reported
effect size statistics, which do not provide an accurate evaluation of how well a measure can detect meaningful change in health (163). Subsequently, the paucity of evidence available for even minimally important measurement properties (validity, reliability) means no assessment recommendations can be made.

This is the first systematic literature review of existing fatigue measures to determine their quality and acceptability following completion by axSpA patients. This review is strengthened by its pluralistic approach to evaluation, considering both study (102,103) and PROM quality (97,98,163,176) alongside an appraisal of PROM content using the RA-fatigue conceptual model as a framework for a comparative item-content appraisal. However, one explanation for why these studies scored poorly is that the rigorous COSMIN criteria were applied to data where PROM evaluation was not the primary focus of the study. Although I assessed all titles and abstracts for review eligibility, a second reviewer (Dr Kirstie Haywood) assessed a sub-set of titles and abstracts and reliability was checked.

Since completion of the review, two new checklists have emerged for determining study methodological quality: an updated COSMIN checklist called the “risk of bias checklist” (104) and a new checklist by Francis and colleagues (203). The updated COSMIN checklist now contains items on PROM design (including its development process and conceptual underpinnings) and content validity which are substantive changes (104). However, my review already considered the development pedigree of PROMs and examined content validity in relation to an existing framework of fatigue (50). The new contribution by Francis is a composite checklist which was generated through a systematic review and refined using cognitive interviews with the intended audience: clinicians and researchers with survey-based research expertise (203). The checklist uses critical aspects of the COSMIN checklist whilst drawing upon other existing literature with the aim to reduce complexity and improve ease of engagement for clinician’s and researchers who may lack expertise in measurement theory (203). A key strength of this checklist is its explicit recognition of the need for an underlying conceptual model for measurement and practical properties of the PROM – something which was omitted in the COSMIN checklist. However, the meticulous approach used in the development of the original COSMIN checklist coupled with the involvement of a group of international experts raises concerns about the use of a simplified checklist. Efforts to reach international consensus by an interdisciplinary group of measurement experts on the quality criteria necessary to have confidence in measurement was an important step forward for PROM evaluation, therefore, to distil this simply for ease and
accessibility risks diminishing the quality of future systematic reviews. An immediate consequence of this may be misinformed decision-making for the selection of appropriate outcome measures toward PROMs that may not be good quality, raising the question and risk of unethical research waste.

Application of the RA-fatigue conceptual model as a reference framework for axSpA-fatigue has highlighted the poor content validity of the reviewed measures, with no PROM providing a full reflection of the model. The MAF and FSS both had content reflecting two important components of axSpA-fatigue: frequency and severity (40). However, only two PROMs: the MFI-20 and FACIT-fatigue had items that assessed the cognitive/behavioural (and emotional) impact of fatigue, with 10/20 and 6/13 items respectively. The MAF, MFSI-SF and FSS only had items reflecting two of the cognitive/behavioural domains. Despite the MFI-20 having adequate evidence of internal consistency and reliability, it was unclear whether the PROM was sensitive enough to detect change, or whether its content reflects what is important about fatigue for axSpA patients. There was limited but acceptable evidence of a strong association between the FACIT-fatigue and SF-36 VT, enhancing confidence that the FACIT-fatigue is a measure of fatigue in this population, however, evidence of reliability and responsiveness was lacking. Therefore, whilst the item content of these measures (although not complete) was acceptable, adequate evidence of essential psychometric properties was lacking, meaning their application cannot be recommended for axSpA-fatigue assessment. A thorough, robust assessment of axSpA-fatigue is essential to both detect and detail the nuances of fatigue experience necessary to deliver tailored, individualised healthcare to axSpA patients.

Qualitative research in axSpA-fatigue has highlighted a significant impact of fatigue on social life, psychological wellbeing, relationships, engagement with usual activities of daily living (36) and reliance on self-management (35). This demonstrates the need for measures to be reflective of the wider spectrum of fatigue experience and impact and further reveals how insufficient the single-item VAS on fatigue severity is (36). Although current qualitative evidence in axSpA is limited, the currently available qualitative evidence indicates that there is similarity between RA and axSpA-fatigue experience, therefore justifying use of the RA-fatigue model (50) as a comparative framework with which to appraise the relevance of PROM item content against (30,36). However, growing evidence suggests that fatigue experience is a complex, dynamic and multifaceted experience that is predominantly disease-specific. This was exemplified in a review
of similarities and differences in fatigue experience between both related and unrelated conditions (fibromyalgia, multiple sclerosis, AS and stroke) (47) and in another study comparing patients between two stages of illness: active cancer patients and cancer survivors (41). Therefore, despite the appropriateness of the RA-fatigue model as a comparative appraisal framework for measures used in axSpA fatigue assessment, this should be confirmed through the development of an axSpA-fatigue conceptual model. This will allow for the nuances of fatigue experience and impact in axSpA to be identified, better informing measurement selection, modification or creation.

A review of fatigue PROM quality used in a range of chronic illnesses identified a lack of evidence of essential measurement properties which impeded recommendations (152). However, there was a lack of transparency regarding how measurement quality was judged in this study, with study methodological quality not reported. International guidance highlights the importance of transparency in the assessment of measurement quality and acceptability (97,163). Use of the COSMIN checklist, as in the present review, enables ratings of study methodological quality to be incorporated into data synthesis and contribute to a more robust judgement of PROM quality (102,103,159).

PROMs that are both patient-derived and follow a strong development process are more robust and relevant to the experience of patients and better able to capture outcomes that really matter (138,204). However, the content of many legacy measures was clinician driven and therefore may lack relevance to patients (138,204). PROM failure to capture meaningful outcomes that matter to patients (58,60,89,205) diminishes the contribution that PROMs can make to patient-centred care and shared decision making. This was a driver for the co-development of a new, patient-derived RA-fatigue measure – the Bristol RA Fatigue Multi-dimensional Questionnaire (BRAF-MDQ) (106,143). Despite this review identifying nine PROMs used in axSpA fatigue assessment, only four provided a limited conceptualisation of fatigue, predominantly informed by literature reviews and clinical experts. Only the FACIT-fatigue used qualitative research (semi-structured interviews) in its development, but this measure did not provide a conceptualisation of fatigue. Qualitative research can afford greater insight into the leading health issues affecting patients, including insight into their experience and impact, subsequently improving the relevance and acceptability of PROM content. This can support targeted health-care efforts to address the unmet needs of patients by measuring what really matters.
The content of MFI-20 and FACIT-fatigue provided the most comprehensive assessment of fatigue (50), however, current available evidence of their psychometric quality in the axSpA population is limited.

A small number of fatigue-specific PROMs have undergone evaluation for their quality and acceptability for use in axSpA fatigue assessment. These measures have been evaluated in studies of often poor methodological quality which only evaluate a few measurement or practical properties, undermining the ability of this review to provide any recommendations. Further, the lack of robust evaluative evidence suggests that any data produced from their application in clinical practice or research should be interpreted with caution. The comparative item-content appraisal identifies the MFI-20 and FACIT-fatigue as providing the most comprehensive assessment of fatigue, capturing the impact of fatigue on cognition and behaviour. Despite this, further exploration of relevance and acceptability of the PROMs identified in this review for use in axSpA-fatigue assessment is encouraged. Further, PROMs identified as having acceptable content validity require urgent evaluation of essential measurement properties (validity, reliability, responsiveness and interpretation) to enable greater confidence in their selection and application.

2.7. Next step
Despite extensive review, only two measures (MFI-20 and FACIT-fatigue) contained items more broadly reflective of the RA-model of fatigue. However, there was no strong or convincing evidence of their quality or acceptability for use in axSpA fatigue assessment. There is a clear need to establish an understanding of what fatigue in axSpA is and determine the extent to which a conceptualisation of axSpA-fatigue corroborates with the RA-fatigue model. Should there be significant alignment between the two conceptual models then modification of the RA-fatigue specific questionnaire – the BRAF-MDQ – may be sufficient to measure axSpA-fatigue. However, if there is little or no alignment between the two conceptual models then the development of a new axSpA-fatigue specific measure may be warranted. Chapter 3 describes a qualitative exploration of the lived experience of fatigue and axSpA to inform the development of: (i) a lived experience framework, and; (ii) an axSpA-fatigue specific measurement framework to underpin a future PROM.
3. Chapter 3: The lived experience of fatigue and energy in patients with axial spondyloarthritis

3.1. Introduction
This chapter presents a qualitative exploration of patients' lived experience of axial spondyloarthritis and fatigue to produce a measurement framework to underpin the content and item generation of the new PROM (Chapter 4). The use of qualitative research in axSpA and fatigue is described in section 3.2 and the qualitative methodology underpinning this study is discussed in section 3.3. The methods for exploring patients' lived experiences is described in section 3.4. The results of the study are described in section 3.5. The chapter closes with a discussion (section 3.6).

3.2. Background
Fatigue is known to be associated with poor health-related quality of life in axSpA (40), yet there is limited understanding about the experience of fatigue in axSpA from a patient perspective. A scoping review of qualitative studies (interviews and focus groups) conducted with axSpA patients to explore their experiences of axSpA and fatigue was undertaken to inform this research (see Appendix 3A for search strategy). This scoping review identified just two, UK-based, qualitative studies which sought to understand the association between axSpA and fatigue: one interview study (36) and one focus group study (35). The remaining evidence comes from two interview studies conducted with axSpA patients in Norway (33,206), one in Egypt (34), and one comparative review of fatigue experience in chronic illnesses (including axSpA) (47). A synthesis of the studies is provided below to establish what is currently known about the experience of axSpA and fatigue. Study methodologies and a summary of their findings are provided in Table 3.1.

Across these studies there is growing evidence that axSpA-fatigue is a complex symptom. The experience included experiencing both physical and mental symptoms (36), that fatigue is distinct from normal tiredness (33) and that fatigue is unpredictable (35). One study challenged these findings and suggested that fatigue was predominantly a physical symptom (34), which was explained as a cultural difference in fatigue experience. Fatigue symptoms also differentially affected participants, which could be influenced by changes in the weather, seasons or overexertion (206). AxSpA symptoms such as pain could lead to a lack of sleep,
subsequently exacerbating fatigue (35). However, pain is a symptom of axSpA and therefore this study begins to highlight the potential influence of disease symptoms on fatigue experience.

The studies described the impact of fatigue on the social life of participants whereby they changed their social activities (34,36), felt that they lacked social engagement (36) and managed their energy expenditure within the demands of social occasions (36). It is within the context of exercise, social and leisure activity that the term energy begins to appear in some axSpA-fatigue studies (33,36). These studies referred to energy and used the term interchangeably with fatigue, sometimes as a synonym and other times as an antonym. However, none of the studies provided any detail or information about the nature of energy in axSpA, whether it was a new concept or simply part of the vernacular patients use when discussing or describing their fatigue. The concept of energy has previously been described in human immunodeficiency virus (HIV) research as a related, but distinct component of fatigue (207,208). This research described energy as something individuals used to manage fatigue: when energy is available, individuals were able to function; in contrast, when energy was lacking, function was affected (208). Moreover, energy was needed to be physically and mentally active (209). The lack of clarity around energy and interchangeable use of fatigue and energy in existing axSpA-fatigue studies could be essential to understanding the lived experience and impact of fatigue in axSpA.

Just two studies explore patients’ approaches to fatigue management in axSpA (35,36), which were: (i) seeking support and help from others (36); and (ii) self-management strategies (35,36). A small limited study using diaries and interviews (36) suggested professional support was rare and there was reliance on family or spouses for support with daily tasks such as housework or childcare (36). A deeper understanding of support and the link to fatigue or axSpA symptoms was not evident. This aspect warrants further investigation to understand the types of support required and the impact it has on patients.

Self-management strategies were explored in two studies (35,36). A range of approaches including adopting a positive attitude, self-medicating with painkillers (e.g. co-codamol), taking naps, staying active, complementary therapies (acupuncture), yoga and Pilates were identified (35). Some parts of the results specified that participants achieved pain relief (e.g. when using acupuncture) but did not state whether it helped or worsened fatigue. Consequently, it is not clear which
management strategies relate to axSpA symptoms, and which strategies specifically relate to the self-management of fatigue. Overall the study methodology and methods were poorly described, and further clarification of the findings is required.

Farren’s study (36) described self-management approaches taken by participants which included seeking medical help and undergoing diagnostic tests (e.g. blood test), using prescription medication (amitriptyline) to aid sleep and using complementary therapy for pain relief which in this study, reportedly had a secondary effect of reducing fatigue. In addition to these approaches, participants described using pacing techniques (targets, prioritising, budgeting energy and rest), modifying behaviour to reduce mental fatigue, adopting a positive attitude and exercise (36). However, this study also provided very little depth about how these approaches helped with managing fatigue. The lack of distinction between axSpA symptoms (particularly pain) and fatigue means these issues are conflated in both studies (35,36) and it is unclear what patients are doing for, or how they are coping with or managing their fatigue. A richer and more detailed report described three different states (“life conditions”): ordinary, slowed-down and disrupted (206). These states were influenced by symptom experience and intensity, and the effectiveness of self-management strategies. The more challenging the symptoms, the greater the difficulty experienced by patients to incorporate these into daily life and to self-manage. However, there was little reference to tiredness and fatigue.

In managing daily life with axSpA two states (“conditions of fatigue”) were identified: life strain-related tiredness and unfamiliar and unmanageable illness-related fatigue. Life-strain related tiredness was influenced by axSpA symptoms like stiffness and pain, which could increase the effort needed by participants to function in daily life, leading to tiredness. This type of tiredness was associated with daily life situations, whereas illness-related fatigue was not understandable, had no identifiable cause and could not be relieved with self-management strategies (33). This study begins to explore fatigue and distinguishes its characteristics (unpredictable, no identifiable cause) from normal tiredness, however, the study was descriptive and alluded to a broader picture of fatigue in axSpA, describing symptoms such as pain, stiffness and inflammation rather than focusing on the experience of fatigue.

Overall there is a paucity of qualitative studies and existing studies tend to be small with poorly described methodology and methods. Only one study explicitly stated its epistemological position (36). The findings identify the impact of fatigue on the lives of patients’ with axSpA across multiple dimensions: impaired physical function,
reduced social life and leisure activities and has a negative impact on mood (35,36). The term ‘energy’ is used frequently but there has been no exploration of whether it is conceptually important within the fatigue experience. Further, a variety of self-management strategies are described by patients (35,36), however, whilst identified they are poorly developed and lack detail. Consequently, there are gaps in our understanding of how people live with fatigue and how fatigue is influenced by fluctuations in axSpA. To address these limitations, a qualitative study informed by interpretative phenomenological analysis using semi-structured interviews has been undertaken to explore the lived experience of fatigue in axSpA. From this, a measurement framework can be derived to underpin the content and development of items for the fatigue PROM described in chapters 4 and 5.

3.2.1. Aims and objectives
The overall aim of this study was to explore patients’ lived experiences of axSpA and fatigue.

Objectives:

1. Developing an understanding of the patients’ lived experience of axSpA and fatigue (lived experience framework).
2. Producing a measurement framework of what really matters to patients to underpin a future axSpA fatigue PROM.
Table 3.1: Methodologies and summated, non-synthesised findings of identified qualitative studies in axSpA and fatigue

<table>
<thead>
<tr>
<th>Study</th>
<th>Sample/ aim</th>
<th>Method/ methodology</th>
<th>Findings⁶</th>
</tr>
</thead>
<tbody>
<tr>
<td>Farren 2013</td>
<td>Sample: 10 ankylosing spondylitis patients (UK)</td>
<td>Methodology: framework analysis, adhering to subtle realism. Epistemology: interpretivism</td>
<td>Three fatigue themes emerged: i) daily and seasonal patterns; ii) consequences and iii) management.</td>
</tr>
<tr>
<td></td>
<td>Aim: “to explore the perceived causes, consequences and management of fatigue” (36)</td>
<td>Method: semi-structured interviews (restricted to 30-minutes) and patient diary’s (recorded over the 7-days prior to interview). Interview topic guide developed from previous research and evolved as study progressed.</td>
<td>i) Daily and seasonal patterns: varies between participants. Physical and mental fatigue is described; overall fatigue is distinct from normal tiredness. Fatigue is worse in winter for some. Disturbed sleep is problematic for fatigue; pain could affect sleep. Disease activity affected fatigue for some.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ii) Consequences: fatigue affects their ability to cope with activity - but two groups emerged: a) those who were able to push through physical activity; and b) those who were unable to cope with increases in activity. Impacts on work - some described early retirement. Full-time workers limited their social and leisure activities because of the energy expenditure of working. Non-working participants found leisure and social activities difficult to participate in. Fatigue impacted on concentration, enthusiasm, self-esteem, memory and led to low social engagement.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>iii) Management: Two approaches to management were described: a) strategies for self-management and b) seeking help and support from others.</td>
</tr>
</tbody>
</table>

⁶ Only findings attributed to axSpA are reported
When seeking help, professional support for fatigue was rare. Spouses and family were common sources of support.

| Mengshoel 2008 | Sample: 12 ankylosing spondylitis patients (Norway) | Methodology: no explicit methodology stated. Thematic analysis used coding which informed theme generation. Method: unstructured interviews (lasted between 50-90 minutes). Two open questions: how participants lived before AS and how they lived during good and bad periods of illness. | Three states (“life conditions”) were described.

i) Ordinary life: stiffness when staying still – movement prevents and deals with stiffness. Sleep is disturbed. Integrating movement into daily habits at home and work (including adjusting the environment). Avoiding overexertion/strain on back, neck and arms as the body can take revenge for overexertion. Nutrition is important – benefits of supplements and a good diet described; but some foods can induce feelings of stiffness. Want to carry on with normal life including working – be a contributor and not burden on others. Changing exercise to more appropriate choices (compared to those before illness) or stopping sport. Heavy or strenuous tasks reallocated to close family members. Unpredictability of AS affected scheduling of meetings with friends so invited them to their home instead of meeting outside the house. Changing and developing of new friendships.

ii) Slowed down life: their body requires greater attention. Stiffness more problematic, fatigue and acute symptoms could be reversed with 20-30 minutes of rest (lay, rest, stretch). More intense fatigue noticed upon waking. Causes of fatigue thought to be bad sleep, changing weather and overdoing it the previous day. Participants reported feeling less tolerant and had to slow down to recover.

iii) Disrupted life: intense, painful sensations that could not be reversed. Lots of uncertainty. Occurs rarely (2-3 times per year) for no known reason. Movement could worsen pain, but so did resting for long periods. Pain could be localised or distributed throughout body. Personal care and concentration became difficult. Recovery occurs over several weeks. |

| Mengshoel 2010 | Sample: 12 ankylosing spondylitis patients (Norway) | Methodology: inductive coding. Interrelationships between codes identified. Coding patterns used to inform | Two themes were described: i) comprehensible, manageable, life strain-related tiredness with a second theme of life-strain related tiredness as a sign of overload, and; ii) unfamiliar and unmanageable illness-related fatigue with a second theme on illness-related fatigue as a sign of sickness. |
Aim: "to examine the nature of fatigue and how it is managed in daily life situations by individuals with ankylosing spondylitis" (33)

| generation of categories. Data analysis informed by previous (2008) study findings. Method: Semi-structured interviews. Topic guide developed to be open and became more focused as the study progressed. i) Comprehensive, manageable, life strain-related tiredness: AS makes life harder for participants who make adjustments to live similarly to how they did before AS. Some had to make significant adjustments/ develop new ways of living. Immobility causes stiffness. Overcoming stiffness required movement but that demanded a lot of energy. Many participants were woken by stiffness, disturbing their sleep. Participants often felt tired in the morning and found it hard to rest during the day. Work used their energy, so they had nothing left to exercise. For some, exercise gave some new energy. Life-strain related tiredness was a sign of overload: participants must manage their energy. Participants prioritised activities, slowed down and avoided overstraining themselves. Tiredness was a signal that they were low on energy and needed to recharge. Overdoing it can lead to punishment (feeling worn out) the next day. Easy tasks became the focus with more demanding tasks being delayed. Being social is a strain. Recovery was about taking it easy: listening to music, reading, seeking peace and calm. Participants feel they can tolerate less now than they used to be able to. ii) Unfamiliar and unmanageable illness-related fatigue: unbearable pain, heaviness and fatigue is now unmanageable. No relief for stiffness anymore. Ordinary tasks become impossible to perform. Pain experience either localised or throughout the body. Great need for sleep but some could not sleep or calm down. Felt worn out and even after slowing down or sleeping they still felt tiredness. Illness-related fatigue as a sign of sickness: associated with flare onset (inflammation). Ankylosing spondylitis is an unpredictable disease. Deep and intense exhaustion and weakness. Powerless and helpless – cannot predict or relieve it. |
| Eilertsen 2015 | Sample: 95 patients: 12 ankylosing spondylitis patients (Norway) 25 fibromyalgia patients (Sweden) 10 multiple sclerosis patients (Sweden) 48 stroke patients (32 from Denmark, 16 from Norway) | Methodology: secondary qualitative thematic analysis on original transcripts to identify (sub) themes of fatigue. Systematic comparative analysis across (sub) themes focusing on similarities and differences between the different chronic illnesses. Methods: all studies used semi-structured thematic interviews. | Characteristics of fatigue experience: weariness, unpredictable, uncontrollable, invisible, difficult to describe. Fatigue experience varies in ankylosing spondylitis. Mutual reinforcement with pain. Impact on daily life: greater need to sleep, rest, lack of energy to continue former activities or roles (before illness), poor understanding from others, difficult to make decisions or plans, difficulties with memory and concentration. Trying to manage: Resting and sleeping, adjusting activity levels and avoiding stress. Creating meaningful activities and relief by warmth and light. Attitude adjustment: hoping for alleviation, normalising, trying to accept and account for symptoms |

| Mortada 2015 | Sample: 60 Egyptian patients: 20 with rheumatoid arthritis 20 with fibromyalgia (FM) 20 with axSpA | Methodology: Prospective monocentric study using content analysis, inductive reasoning and grounded theory. Method: semi-structured interviews using an interview guide. No information about how the topic guide was developed, content informed or whether it was (i) piloted; (ii) AxSpA-specific findings were not separately reported and were reported within and alongside findings for RA and FM. AxSpA-specific findings were extracted, where possible, per theme and are reported below. Description and causes of fatigue: overwhelming physical tiredness, difficulty moving. AxSpA patients described fatigue as “a sense of freezing of the body”. Mental fatigue not reported by axSpA patients. Daily patterns of fatigue: tiredness is most prominent in the morning and can subside after an hour. Associated with pain. |
characteristics among Egyptian, Muslim patients with rheumatic diseases” (34) evolved during interview process

Consequences of fatigue: participants did the minimum physically necessary. Difficulty getting to work. Reduced sexual activity (males).
Coping strategies: dividing big tasks into multiple, smaller tasks.

| Davies 2013 | Sample: 14 ankylosing spondylitis patients (UK)  
Aim: “to explore from the patients’ perspective (i) the effect of fatigue in ankylosing spondylitis (ii) ways of self-managing fatigue as part of their everyday lives, and (iii) to identify potential interventions” (35) | Method: exploratory focus groups on the patient perspectives of managing fatigue in AS including views on potential interventions. Two focus groups.  
Data analysis: thematic analysis – aimed to gain theoretical sensitivity which puts emphasis onto the frame of mind of the participant. | Three themes were identified: i) pervasive fatigue; ii) current limitations of self-management, and; iii) a new direction for future interventions.

i) Pervasive fatigue: influences which contributed to, or exacerbated fatigue were described, including age, poor sleep, pain at night, low-mood or depression, poor concentration, side-effects of medication, work responsibilities and unpredictability. Some participants were unsure whether it is fatigue (due to disease) or age-related. Pain causes poor sleep, leading to fatigue for some participants. Medication contributes to or worsens feelings of fatigue for some participants. Some found that work was a motivator, whilst others could cope better after finishing work. Difficulty concentrating was an identified problem.

ii) Current limitations (of self-management): approaches used by participants included adopting a positive attitude, self-medication and keeping active. Hydrotherapy was popular but expensive for participants. Short naps were effective for fatigue. Some avoided medication due to reported side-effects. Alternative complementary therapies like meditation worked for one participant but not during flares. Acupuncture and other therapies provided temporary pain relief. Exercises like Yoga or Pilates were avoided due to concerns about overexertion.

iii) A new direction (for future interventions): Participants were shown a video of mindfulness-based stress reduction psychological intervention, which the majority liked. Most were enthusiastic about the intervention however; one person was more sceptical about whether it would be successful. Suggestion of extending intervention to including significant others and carers. Some participants expressed a preference for group delivery, others expressed preference for online or distance intervention delivery. |
3.3. **Methodology**

This study draws on Interpretative Phenomenological Analysis (IPA) as its methodological framework (210). IPA was developed to provide methodological direction for research that focuses on understanding the lived experience of others. Smith et al. (2009) expand on this stating that IPA is (210):

> ‘concerned with understanding personal lived experience and thus with exploring persons’ relatedness to, or involvement in, a particular event or process (phenomenon)’

The theoretical underpinning of IPA draws on three philosophical principles: phenomenology (a philosophical approach to the study of experience), hermeneutics (the theory of interpretation) and idiography (concerned with the particular – that is, to focus on the individual within their context) (210). Through these principles, it enables participants to say what is important to them within their personal, social and historical context whilst encouraging the researcher to be aware and reflective of their own role within the interpretation of the participant’s experience.

IPA seeks to develop an interpretative account of the participants’ experience within the context of that experience. To do this, the researcher develops an ‘insider perspective’ (211) by focusing on what the participant is saying, and how they have understood their experience (210). This is achieved through a process of double hermeneutics (two layers of interpretation). Double hermeneutics describes a process whereby the researcher is making sense of the participant, who is making sense of their experience of a phenomenon (212). It has been argued that IPA is appropriate for research investigating unique experiences, the meaning of experiences and how these meanings manifest within the individual (213). Therefore, IPA is well-suited to address the aim of this study.

3.3.1. **Phenomenology**

A person’s experience of a phenomenon is a blend of personal meaning to the individual and meaning for others who interact with the individual (214). Phenomenology originated in the work of Husserl and is guided by the principle that experiences should be considered within their context, in the way that they occurred (210). Husserl’s conceptualisation of phenomenology focused on the intricate examination of experience in the person’s consciousness and how individuals come
to accurately ‘know’ their experiences. He reasoned that experiences may possess distinct, essential features which may be separated from – or transcend – the circumstances of their appearance (210). This could then provide insight into a given experience for other individuals. However, Husserl recognised that we hold predisposed views and perspectives which he referred to as our ‘natural attitude’ – that we ‘bracket’ (step outside of) to examine everyday experiences. Instead, he favoured adopting a ‘phenomenological attitude’ which serves to reorient our attention away from, for example, an everyday environmental object and toward introspection and how we perceive that object.

The key contribution of Husserl to phenomenology was to draw focus to the individual’s personal understanding of their experience of a phenomenon. This idea was further developed by others, recognising that living within the world may influence an individual’s perspective on their experience (215–217).

Heidegger deviated from Husserl’s phenomenology, instead emphasising the importance of hermeneutics and the existential (210). Heidegger was focused on how our interactions with the world (activity, relationships) and the appearance of the world are made meaningful to us (215). Heidegger posits that people exist within their context and to examine this, we should use a process of “intersubjectivity” (215) – that is the:

‘Shared, overlapping and relational nature of our engagement in the world’

This supported further evolution of the philosophy of phenomenology which recognised that individuals are immersed within a complex world of interactions which includes objects, relationships, culture and language (215,217). Together, this evolving philosophy of phenomenology recognises that experiences are lived processes whereby meaning, and an individual’s perspective, is influenced and shaped by their relationship with the world.

Recognising the importance of understanding subjective experience is a key focus of phenomenology. Experiences are lived processes (210) and recognising this can enable researchers and healthcare professionals to better understand the implications of health conditions or symptoms on a patients’ life and subsequently support the provision of tailored, targeted healthcare. This approach fits with current recommendations for PROM development (112,123).
3.3.2. Hermeneutics

Cohen, Kahn and Steeves (218) define hermeneutic phenomenology as:

‘the tradition of looking at a phenomenon, a single kind of human experience, rather than a social process or structure or a culture’

Heidegger introduced the concept of hermeneutics to interpret experience or phenomena (214). Gadamer expanded on this, describing the relationship between past experience or fore-understanding with new experiences or phenomenon (219). Therefore, a key premise of hermeneutics is that the whole is the sum of its parts, thus, to understand the whole you must understand its parts (210,219).

Hermeneutics is a circular process which moves between the parts and the whole of a given experience or phenomena. Notably, this describes the process of an individual moving between experiencing a phenomenon and understanding the phenomenon. In IPA, it is recognised that the researcher will transition between the part and the whole in their interpretive process – a circular process termed ‘double hermeneutics’. Prior exposure to a participant’s transcript can influence new readings of it which may lead to new or different interpretations (210).

As a researcher, this is an essential component of the IPA process whereby the researcher attempts to make sense of the participant, who themselves are trying to make sense of their experience. This recognises that the interpretative account of the researcher is an outcome of their relationship with the participant and their encounter (212,220). Therefore, it is essential that the researcher is aware of any held pre-conceived views or notions and tries to conduct the analysis without being steered by these views (210). Engaging in a process of reflexivity can help provide transparency to the analysis process with regard to the views and beliefs of the researcher.

3.3.3. Idiography principle

IPA is committed to idiography in two ways: detail and depth of analysis, and understanding experience from the perspective of particular individuals, and within their particular context (210). Therefore, idiography is focused on the individual and their story, rather than more descriptive group-level claims (idiographic vs nomothetic research). IPA studies are interested in rich, detailed and in-depth data about an experience of a phenomenon that recognises the importance of the participants phenomenology, and the process of double hermeneutics in making
sense of the experience. This lends itself to a two-stage analysis process: firstly, each individual interview is analysed; and secondly, upon conclusion of the final interview and its analysis, similarities and differences between reported experiences can be explored and reported.

The proposed research seeks to provide a detailed understanding of what it is like to live with axSpA and fatigue and to generate a measurement framework with which to underpin future measurement of fatigue. IPA is, therefore, an appropriate methodological approach.

In summary, IPA draws on phenomenology, double-hermeneutics and idiographic principles to enable a thorough examination of the lived experience of a phenomenon – in this case, to examine the patients' lived experience of axSpA and fatigue.

3.3.4. Other methodological approaches

For the purpose of PROM development, qualitative methods (interviews, focus groups, pretesting interviews) are frequently used to inform development of a measurement framework, item generation and item modification despite there being no consensus on the best approaches to take. Moreover, there is little consideration for the methodological underpinnings that influence data collection and analysis. Qualitative analysis approaches are adopted for specific purposes such as generating practicable, translatable recommendations (thematic analysis) (221) or generating theory (grounded theory) (222,223). To ensure the most appropriate qualitative approach was adopted in my study, I considered a variety of methodological approaches which I briefly describe below and accompany with reasons why they were not chosen for this study.

Discourse analysis

Discourse analysis (DA) relates to “language in use”, the premise being that words and language are tools of communication that alone, are meaningless (224). However, it is “through the shared, mutually agreed-on use of language that meaning is created” (224). Therefore, language is used as a means to construct our understanding of reality (224) and enables people to express their identity (225,226). Gee (2005) outlined seven “building tasks” which, when analysed, can provide insight into how social norms are created and maintained and how personal
and group identities are constructed (227). Variants of discourse analysis have been
developed and are in use – however, DA is focused on interaction. Due to this focus
on interaction, discourse analysis was not chosen for this study as the preferred
focus was an individual’s lived experience of phenomena within their social and
historical context.

Ethnography

Ethnography is the study of social interactions – which includes behaviours and
perceptions – of individuals or groups within their natural setting. Ethnographic
studies seek to “get inside” (228) the experience of individuals or groups, and
develop a rich insight into their behaviours, perspectives and practices (229). This is
exploratory research that lends itself to unstructured observation that aims for depth
and detail. Within a healthcare context, ethnography can be used to generate
detailed insights into relationships between healthcare professionals, professionals
and patients, care delivery and the care experience of patients (229). However, the
focus of ethnography is on social interactions as opposed to lived experiences of
specific illness symptoms. Consequently, whilst ethnography would be appropriate
to examine care provision and patient-professional relationships, it is not as suited
to explore individual experiences and the meaning individuals ascribe to
experiences like fatigue.

Grounded theory

Initially developed by Glaser and Strauss in 1967 (230) to provide a sequential and
systematic guide to conducting qualitative fieldwork and analysis (210), grounded
theory (GT) is an inductive, qualitative approach that is highly structured with the
aim to generate or discover theory by transitioning from data to theory (231,232).
The generated theory is ‘grounded’ in data that has been systematically collected
and analysed iteratively (233). Since its conception, Glaser and Strauss have
deviated from one another on how grounded theory should be conducted (234).
Glaser adopted a more data-centric view of the analysis process whereby theory
and its conceptualisation emerges from the data (231,232). In contrast, Strauss
proposed a more structured, prescriptive approach to conducting grounded theory
analysis (235).
More recently, constructivist grounded theory (223) has been proposed which offers more flexibility to the process with a more clear epistemological position. Charmaz considers that there is an underlying assumption that data is produced through the interactions between the participant and the researcher, which the researcher observes and defines (236). The ability of GT to produce rich and detailed theory suggests it would be a suitable methodological approach for this study; to conceptualise fatigue in axSpA. However, generated theory typically extends beyond lived experience to produce a high tier account of the phenomenon under investigation in the target population (210,237). Its tendency toward an explanatory conceptualisation of phenomenon contrasts with IPA which is focused on lived experience, meaning and the individual (210). Therefore, IPA and its focus on the individual and their lived experience is better suited to the aims of this study.

3.4. Method

3.4.1. Recruitment
Three rheumatology teams from three secondary care, National Health Service rheumatology departments in England supported patient recruitment. A nominated team member identified potential patients from the axSpA database. Patients were purposefully sampled for age, sex and disease duration and the sampling matrix was updated after each interview and relayed to the local recruitment teams. Patients were contacted by the local rheumatology nurse who provided them with a study cover letter, information sheet and a consent form to express interest and give permission for me to contact them. This form was returned to the staff member who made the approach in person or by postage-paid envelope and this was then transferred to me in person (if I was local) or sent to me in a password protected file via email.

Those who expressed interest and returned consent forms were contacted by me by phone or email – whichever their stated preference. I began by asking whether the patient had received the information sheet and cover letter, and whether they had read them. Those who had not received them were provided with copies by me. I explained that the study involved participating in an audio-recorded interview that could be conducted at their home, their local rheumatology clinic or by phone. I explained that the purpose of the interviews was to explore their experiences of living with fatigue and axSpA and explained that their confidentiality and anonymity would be maintained, and they had the right to withdraw without reason without it
affecting their medical care. Participants were not contacted again until at least 24 hours had transpired to arrange interviews.

Those who agreed to participate were asked to sign a consent form, prior to the interview, confirming they understood and received the information sheet, have the right to withdraw, that data would be recorded, transcribed and viewed by the research team and confidentially stored. Assessments of disease-activity (BASDAI) (76), functional limitations (BASFI) (77) and mood (Hospital Anxiety and Depression scale; HADS) (238) were taken to ensure that, within the sampling frame, a range of participants were recruited across the spectrum of disease activity, function, anxiety and depression. I administered these assessments upon the interview's conclusion to avoid influencing the responses of participants.

The study was granted ethical approval by a Research Ethics Committee (REC reference: 16/WM/0147) and the Health Research Authority (approval dated: April 22, 2016).

3.4.2. Patient research partner (PRP) group involvement

Two face-to-face meetings with the PRP group were held: one prior to the study beginning (described below) and one as part of the analysis of study findings (described in the results section). The aim of PPI in this interview study was to co-produce study materials to assist with conducting interviews, and analyse data to co-produce the lived experience framework (145).

The first meeting was held with 5 members lasting 2.5 hours. The interview schedule was discussed, with the PRP group members, sensitising the language of the questions, and me to the issues and challenges around axSpA and fatigue. Their ideas, comments and suggestions predominantly related to the language used in the interview schedule (for example, clarifying the term clinician to interviewees) and together we modified the topic guide.

The second meeting was held toward the latter stages of the analysis to discuss the emerging lived experience framework, with particular focus on the concept of energy. I shared a summary of interview findings with the PRP members and a discussion followed where they provided their views of the framework. Copies of an abbreviated analysis were taken to the meeting to ensure the developing framework could be reviewed in full. The abbreviated analysis contained the main interpretations of the data and exemplar patient quotes. Members were encouraged
to read the abbreviated analysis, make notes, comments or annotations on the paperwork. All PRP comments and feedback were recorded and facilitated my reflection on the process of analysis.

3.4.3. Methods and data collection
Willing participants took part in semi-structured interviews during which they explored their experiences of living with axSpA and fatigue. For patient convenience and to minimise the burden of travel, interviews were arranged to coincide with hospital appointments and conducted in a private room within the local rheumatology department. Written informed consent was provided prior to the interview. With the permission of participants, all interviews were digitally recorded on an encrypted recorder and transcribed verbatim by me.

Development of the interview schedule was led by the methodology, focusing on the patients' lived experience. Schedules from other relevant studies (30,33) were also considered to ensure thoroughness. Further development involved working together with an established group of PRPs (http://www.wwl.nhs.uk/Specialities/Patient_Research_Advisory_Group.aspx) to ensure that the interview schedule was relevant and used language that was accessible. Members also highlighted the importance of appropriate seating and recommended providing chairs with arms to provide support for patients when sitting and standing (see Appendix 3B for final interview schedule).

3.4.4. Data analysis of lived experience
Data analysis followed the IPA six-step process (210):

1. Immersion in each patient’s lived experience was achieved through taking part in the interviews, transcribing the interviews, reading and re-reading the transcripts.
2. Exploratory notes were made, examining semantic content and language.
3. From the exploratory notes emergent themes were identified.
4. Key concepts, relationships between them and patterns in the data were identified.
5. Repeat these processes (1-4) for each participant.
6. Patterns between participants were sought.
NVIVO10 software was used to manage the study data (239) (see Appendix 3C and 3D for examples of key analysis stages).

I undertook the analysis and early in the process, met with Drs Elizabeth Tutton and Kirstie Haywood who had independently analysed my first three interview transcripts. In this meeting we discussed my interview style, my questions and the emerging findings. Following this, I met with and had regular discussions with Dr Elizabeth Tutton who independently coded one third of the interviews and, together with the PPI group and research team, I developed the resulting framework of fatigue and energy. In terms of my positionality, I am a white male who has experience of living with a chronic health condition for almost 10 years. I have a background in psychology and cognitive neuroscience, with prior knowledge – but only limited experience of – conducting IPA interviews. I had also participated in a qualitative research training course at Oxford University prior to beginning this study.

3.4.5. Data analysis to develop a measurement framework
In addition to the primary analysis a second analysis (on the same data) was undertaken to generate a measurement framework, derived from patients’ lived experience, to capture what is important about axSpA fatigue for measurement. This analysis considered the impact triad described for rheumatic disease to ensure the model captured fatigue impact (240). The triad describes the importance of assessing severity, importance and self-management to determine the extent to which an individual is impacted by the concept under investigation (240). The research question for this analytical process was: what are the key elements of axSpA-fatigue experience and what is important and meaningful for measurement?

3.4.6. Rigour
The four criteria proposed by (241) for quality in qualitative research guided this study. The researcher was committed to each individual and the significance of their story through reading and re-reading their transcripts which also ensured sensitivity to their context. Rigour was demonstrated through a detailed audit trail of quotes and themes (see Table 3.3 for a master table), demonstrating theme derivation. To ensure transparency, the sampling matrix detailing distribution across key variables is presented in Table 3.2.
The researcher reflected on their role as an interviewer and during analysis to help illuminate ways in which the researchers’ positionality could be influencing their interpretations (see Chapter 3 discussion).

The COnsolidated criteria for REporting Qualitative research (COREQ) 32-item checklist (242) was completed to ensure clear and thorough reporting of study method, context and analysis. The checklist consists of 3-domains: (i) research team and reflexivity; (ii) study design, and; (iii) analysis and findings (Appendix 3E).

3.5. Results

3.5.1. Participants
From an original 21 patients who agreed to participate, seventeen (9 male; mean age 45; median age 46; range 22 – 72 years) with diagnosed axSpA (mean diagnosis duration 14.4 years; range 1 – 42 years) took part in semi-structured face-to-face interviews (47.4 minutes, median 48.4 minutes, range 22.1 – 85.49 minutes). Two participants declined due to changed circumstances; two females declined due to childcare difficulties (one subsequently participated in a focus group (Chapter 4)).

Each participant was allocated a number (1-17) to anonymise their identity. The distribution of patients across demographics and health status scores are reported in Table 3.2.

3.5.2. Findings
The findings of the IPA analysis are described below, and the secondary analysis to produce the measurement framework of axSpA fatigue is reported in section 3.5.3.
**Table 3.2: Sampling matrix showing the distribution of participants across key variables**

<table>
<thead>
<tr>
<th>Variables</th>
<th>Proportions (%)</th>
<th>Number of patients (Total n=17)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>52.94</td>
<td>9</td>
</tr>
<tr>
<td>Female</td>
<td>47.05</td>
<td>8</td>
</tr>
<tr>
<td><strong>Age</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18 to 30</td>
<td>23.53</td>
<td>4</td>
</tr>
<tr>
<td>31 to 40</td>
<td>11.76</td>
<td>2</td>
</tr>
<tr>
<td>41 to 50</td>
<td>29.41</td>
<td>5</td>
</tr>
<tr>
<td>51 to 60</td>
<td>11.76</td>
<td>2</td>
</tr>
<tr>
<td>61+</td>
<td>23.53</td>
<td>4</td>
</tr>
<tr>
<td><strong>BASDAI</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt; 4</td>
<td>41.18</td>
<td>7</td>
</tr>
<tr>
<td>≤ 4</td>
<td>58.82</td>
<td>10</td>
</tr>
<tr>
<td><strong>BASFI</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt; 4</td>
<td>52.94</td>
<td>9</td>
</tr>
<tr>
<td>≤ 4</td>
<td>47.05</td>
<td>8</td>
</tr>
<tr>
<td><strong>HADS Anxiety</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt; 7</td>
<td>47.05</td>
<td>8</td>
</tr>
<tr>
<td>≤ 7</td>
<td>52.94</td>
<td>9</td>
</tr>
<tr>
<td><strong>HADS Depression</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt; 7</td>
<td>35.29</td>
<td>6</td>
</tr>
<tr>
<td>≤ 7</td>
<td>67.71</td>
<td>11</td>
</tr>
</tbody>
</table>

3.5.3. *The patient lived experience of axSpA and fatigue*

The findings identify the central concept of achieving balance as an active process of integrating axSpA symptoms and fatigue into their world, working with and not against their condition, to lead the best possible life within the confines of their illness. This is conveyed through the three superordinate themes: having energy, engaging in everyday life and living with axSpA (Table 3.3).

---

7 BASDAI 6 items reflecting disease activity; score range 0 to 10, where 0 is no disease activity and 10 is maximum; scores less than 4 represent low disease activity. BASFI 10 items reflecting axSpA function; score range 0 to 10, where 0 is no functional impairment and 10 is maximum; scores less than 4 represent low functional impairment. HADS 14 items, seven per subscale (anxiety and depression); subscale score ranges 0 to 21; subscale scores less than 7 represent normal anxiety or depression, 8-11 mild, 12-14 moderate and 15+ severe.
Achieving balance is dependent upon the participant having energy to be active in daily life. When participants do not have energy, the balance may shift, negatively affecting their daily life and their control over their condition. Low energy can lead to a worsening of axSpA symptoms, greater feelings of being restricted in daily life and reduced activity levels. This can occur alongside fatigue making participants feel low and exhausted. Reduced activity can lead to worsening stiffness and functional impairment creating a vicious circle for participants. Participants could achieve balance through preserving and re-energising, using plans and routines to manage daily activity and through a process of experiential learning, usually through trial and error, to improve efforts of self-management. The three superordinate themes and their respective subordinate themes are presented using quotes from the participants to illustrate key points.
Table 3.3: Master table of superordinate and subordinate themes with definitions and key quotes from patients

<table>
<thead>
<tr>
<th>Central concept</th>
<th>Superordinate themes</th>
<th>Subordinate themes with definitions</th>
<th>Example patient quotes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Achieving balance</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Lacking in energy</strong></td>
<td></td>
</tr>
</tbody>
</table>
|                           |                       | The experience of being drained and having trouble beginning or continuing with activity | Being drained  
“If you imagine a dementor sucking away the happiness, right… that is how it feels with your energy. It sort of like suckssss all the… [pause]… the get up and go out of ya” (Participant 12)  
“I just went off like somebody had taken my batteries out” (Participant 13) |
|                           |                       | **Getting and preserving enough energy** |                        |
|                           |                       | “Yeah it can, yes it can do depending what, depending what it is and what level of energy it will… use.” (Participant 7)  
“Umm, everything took energy. And sometimes it just seemed almost… [pause] yah, that it was, it was going to be too much to even try to do just the simplest thing. Getting out of bed, just took so much effort.” (Participant 16)  
“I probably drink about 12-15 energy drinks a day now.” (Participant 3) |
|                           |                       | **Struggle to find energy to live everyday life** |                      |
|                           |                       | An active process of finding ways to preserve stamina and to re-energise |                        |
|                           |                       | **Being active**                    |                        |
|                           |                       | “I think, mentally it makes me better because it, it proves to meself (myself) that I can I can beat it. So, I think that by keeping active, makes me think that well ‘I'm beating this, I'm not gonna let it beat me’.” (Participant 17)  
“I can do my AS exercises, but I don't go to the gym anymore” (Participant 7) |
<p>|                           |                       | <strong>Being organised</strong>                 |                        |
| Engaging in everyday life |                       |                                     |                        |
|                           |                       | <strong>Maintaining daily life</strong>          |                        |
|                           |                       | A process whereby participants maintain engagement by adapting their approach to preserve their body and energy without having to reduce activity |                        |</p>
<table>
<thead>
<tr>
<th>Changed ways of being</th>
<th>“I have to plan the week as like a timetable. And I have to build in down time.” (Participant 12)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>“I am conscious of it because I know sometimes if I’m planning something I think “right when, if I need a sleep when can I have a sleep?” so it is always there…” (Participant 17)</td>
</tr>
<tr>
<td></td>
<td>Knowing through experience</td>
</tr>
<tr>
<td></td>
<td>“My tiredness is just tiredness. And I understand my tiredness” (Participant 3)</td>
</tr>
<tr>
<td></td>
<td>“Even though I’ve been through it thousands of times it’s still you know hard to understand” (Participant 2)</td>
</tr>
<tr>
<td></td>
<td>Learning to self-manage symptoms</td>
</tr>
<tr>
<td></td>
<td>“[with fatigue] I feel like I just have to sit and do nothing, just rest your body” (Participant 7)</td>
</tr>
<tr>
<td></td>
<td>“I reduced the dosage and by the end of November, I’d managed to stop taking all of them which I was quite pleased with” (Participant 8)</td>
</tr>
<tr>
<td></td>
<td>Being accepting</td>
</tr>
<tr>
<td></td>
<td>“Just accept the fact you’re going to get it and it’s going to happen. Equally accept the fact it’s going to last for days. It’s not completely debilitating but don’t push yourself at that particular moment when you feel more than tired.” (Participant 10)</td>
</tr>
<tr>
<td></td>
<td>“You either deal with it, accept it or you go down the other route” (Participant 9)</td>
</tr>
</tbody>
</table>

Identifies how participants adapt to incorporate their symptoms and fatigue into their life through experiential learning, sensitising and accepting.
<table>
<thead>
<tr>
<th>Making sense of axSpA</th>
<th>Living with axSpA</th>
<th>Living with unpredictability</th>
</tr>
</thead>
<tbody>
<tr>
<td>An active process of developing understanding and recognising change</td>
<td>Captures the challenges that living with chronic illness poses to living a normal life</td>
<td>“The problem is the unpredictability, and the unpredictability is the biggest problem because, I know, I know I’m going away in 10 days right... but, I have had to build into (the plan), while we’re away, the fact that I know that I will probably manage x amount of time, x amount of distance, then I will have to have a sleep or have to have a lie down.” (Participant 12)</td>
</tr>
</tbody>
</table>

| | | I can say if I do that my back will hurt, but I cannot tell you if I do that I can be fatigued at the end of it.” (Participant 10) |

| | | Searching for a cause |
| | | “The pain does make the fatigue worse because you’re looking for an escape, from the pain” (Participant 1) |

| | | “But I think that could be contributing to the fatigue as well not eating but I think it was the fatigue that wasn’t making me eat as well cos I didn’t have the energy to get up and make food.” (Participant 4) |

| | | Struggling to maintain normal life |
| | | The physical and emotional effort required to keep life moving forward demonstrated in living with restrictions, the impact on psychological wellbeing and putting on a positive front |

| | | Living with restriction |
| | | “You have to fight really hard to do simple things because you, you feel tired, and you feel worn out and you think “I’ve really done nothing” and [pause 2 seconds] and that’s about all really.” (Participant 1) |

| | | “If it’s something I really, really, really! want to do... then I’m fairly confident that if I push myself I’ll do it, but I am going to pay the penalty the next day” (Participant 12) |

<p>| | | Impact on psychological wellbeing |
| | | “I do feel that I’m a little bit depressed and I think the fatigue has altered my moods, and because I’m constantly in a low mood.” (Participant 3) |</p>
<table>
<thead>
<tr>
<th>Feeling supported</th>
<th>Being able to connect with others</th>
</tr>
</thead>
<tbody>
<tr>
<td>Captures the importance of caring relationships demonstrated in being able to connect with, and trust others</td>
<td>&quot;You know, if that means that you can offload with the person, you know even if they can't even understand what you're going through, you know having that available to you I would say is, yeah. Is really good.&quot; (Participant 16)</td>
</tr>
<tr>
<td></td>
<td>&quot;When they say to be “Oh you look tired”, its normally like, “Did you have a late night?” “No I was in bed probably four hours before you were but this is how I feel after a full night's sleep. So I think it's misunderstanding, absolutely.” (Participant 3)</td>
</tr>
<tr>
<td></td>
<td>Being able to trust</td>
</tr>
<tr>
<td></td>
<td>&quot;I wouldn't ask for help if I can. I won't ask for help.” (Participant 13)</td>
</tr>
<tr>
<td></td>
<td>&quot;You know I, had to put my pride you know aside and learn to take others offers of help (laughs).” (Participant 16)</td>
</tr>
</tbody>
</table>
3.5.3.1. **Superordinate theme 1: Having energy**

Having energy is defined as the challenge of retaining enough stamina to do things in daily life. Challenges with managing stamina can lead to participants struggling to function in everyday life. This is expressed through two categories: lacking in energy and the struggle to find energy to live everyday life.

**Lacking in energy**

**Lacking in energy** is expressed as the experience of being drained and having trouble beginning or continuing with activity.

*Being drained.* Feeling drained could be mental or physical and create difficulty for participants with initiating or continuing activity. Although difficult to articulate, being physically drained followed a sudden loss of energy “like somebody had taken my batteries out” *(Participant 13)* and often interfered with participants’ thoughts and plans. Some participants discussed being drained in terms of motivation whereby it can be “really hard to push yourself…” *(Participant 13)* to push through.

> “I can’t be bothered doing owt (to do anything). Or anything that involves getting up and moving around. Anything remotely active.”
> 
> *Participant 10*

Efforts to self-motivate included task prioritisation and finding a balance between essential tasks and current capability. However, pain could amplify this challenge and act as a barrier to generating motivation, leaving participants feeling restricted and consequently negatively influencing exercise behaviour.

> “Exercise is the best thing that can be done (to) help AS but … sometimes you’re in no position whatsoever to move… you’re not mobile enough or have the energy to do that. So, you’re stuck in a bit of a vicious circle really because if you had the energy to, then you could help yourself a bit more.” *Participant 7*

**The struggle to find energy to live everyday life**

**The struggle to find energy to live everyday life** is an active process of finding ways to preserve stamina and to re-energise.
Getting and preserving enough energy. Careful use of energy resources and preserving stamina is important for participants to get through the day, but they often lacked the ability to do this “mentally don’t have the capacity to do that anymore because I don’t have the energy to” (Participant 7). Low mental or physical stamina can limit how long participants can function on a given day, however, active appraisal of task demands can support preservation efforts through task avoidance or modification.

“Yeah it can, yes it can do depending (on) what, depending what it is and what level of energy it will... use.”

Participant 7

Energy reflected the participants’ ability to be physically and mentally active. Failing to preserve energy could interfere with task execution, requiring extra effort from participants and further draining their energy levels. Increasing energy expenditure forms part of a cyclical process of insufficient energy leading to greater effort and energy usage.

Finding effective ways to re-energise were a counterbalance to this challenge and included rest breaks, pursuing a healthier diet or consuming high sugar or caffeinated drinks to quickly re-energise.

“I probably drink about 12-15 energy drinks a day now.”

Participant 3

Although the effectiveness of these approaches was relatively unreliable, they enabled participants to self-motivate. Conversely, recommendations from others to rest or take naps when experiencing fatigue were difficult to reconcile with commitments such as work, meaning they were often unworkable.

Maintaining stamina can introduce new challenges for participants, however, it can enable participants to engage in everyday life and soften the impact of fluctuating energy levels.

3.5.3.2. Superordinate theme 2: Engaging in everyday life

Engaging in everyday life captures two processes: i) maintaining daily life through being active and organised, and ii) a changed way of being developed through experiential learning and being accepting that circumstances are dynamic and may change.
Maintaining daily life

Maintaining daily life is a process whereby participants maintain engagement by adapting their approach to preserve their body and energy without having to reduce activity.

Being active. Participants maintained daily life through the determination to keep moving, changing patterns and finding a compromise between activity and symptoms whilst incorporating activity into daily life. The desire to “beat” (Participant 1) their condition was encouraging for participants.

“Mentally it makes me better because it, it proves to meself (myself) that … I can beat it. So, I think that by keeping active, makes me think that well, ‘I'm beating this…’."

Participant 17

Compromising enabled participants to balance between their exercise needs to alleviate symptoms and maintaining energy levels. Where compromise was not possible, the possibility of worsening axSpA symptoms or over-expending energy increased. Here, experiential knowledge of how inactivity can impact on axSpA-symptoms and fatigue can act as an important impetus for participants to find a compromise.

“It’s that compromise between doing enough not to seize up completely and not doing that much that it hurts.”

Participant 10

“I find that the less I do I seem to have more fatigue.”

Participant 6

Fatigue raised a unique challenge for some participants who described an altered perception of time: ‘a day’ meant something different and reflected a few hours within the day which activity was restricted to. This perception seemed to be sensitive to the severity of fatigue, with more severe fatigue feeling more restrictive.

“If I wasn’t as fatigued I’d know I would be out doing more and just going about my day but fatigue … restricts that to only a few hours a day maybe at most.”

Participant 4

Through being organised, participants can limit the impact of restriction by incorporating structure into daily life and prioritising tasks.
Being organised. Planning can make it easier to achieve more complicated activities for participants. Incorporating more rigid methods of planning when arranging key social events such as holidays enabled rest breaks to be scheduled into plans and a balance reached between doing too little or too much. However, plans could be jeopardised by fatigue or a flare (a period of increased symptom intensity (243)) with axSpA symptoms; both of which are unpredictable, difficult, uncontrollable and require participants to be responsive. Here, going with the flow and adopting a flexible approach can be effective, although some participants were defiant and pushed through, irrespective of any payback they might experience as a result.

“I’ll plan particularly carefully that I’m not doing too much. I’ll particularly plan I’m not doing anything too little. If I’m moving around I will move around and stretch but I will not push it too far anymore, very much in terms of my daily life. It’s within limits.”

Participant 10

Routine was a way of incorporating activities into daily life in a systematic way, introducing structure and ensuring prioritisation of essential tasks. Establishing routines that are effective at maintaining energy levels can positively impact on the participant “I can get up and feel full of energy” (Participant 9). Notably, routine included dietary behaviour which can help to mitigate feelings of fatigue “when I’m on a healthy diet it does help with the fatigue” (Participant 13).

Changed ways of being

Changed ways of being identifies how participants adapt to incorporate their symptoms and fatigue into their life through experiential learning, sensitising and accepting. It represents tacit knowledge (not explicit) and is expressed as knowing through experience and learning to self-manage symptoms.

Knowing through experience. Experiential learning is an essential process for participants to develop their knowledge to become more adept at distinguishing between fatigue and normal tiredness, knowing the type of fatigue (mental or physical) and pattern of fatigue. The distinction between normal tiredness and fatigue was an important piece of tacit knowledge whereby participants could ‘feel’ the difference and understand that fatigue is something more. Conversely, participants found it difficult to develop their explicit knowledge and understanding of
fatigue which undermined their ability to react directly to it “even though I’ve been through it thousands of times it’s still you know hard to understand” (Participant 2).

“But there’s no doubt that sometimes it’s fatigue and sometimes it’s just general tiredness that everyone suffers with, and I, you know the difference between the two I know the difference.”

Participant 7

Physical fatigue was described as a sensation throughout the “whole body” (Participant 2) leaving limbs feeling heavy “like iron” (Participant 6). Mental fatigue was described as having “brain fog” (Participant 12) and being “drunk with tiredness” (Participant 7). Importantly, these metaphors provide insights into the hazing effect that mental fatigue can have on cognitive function. Participants reported challenges with concentrating, staying focused on tasks and slurring speech during social interactions.

**Learning to self-manage symptoms.** Through a process of trial-and-error, complimented by increased self-awareness and ‘knowing’ their body, participants learnt how to best self-manage their symptoms. Self-management needs and decisions vary depending on symptoms, difficulties with energy levels and fatigue. Whilst participants discussed using medication to manage their pain, there are no established approaches for their energy levels or fatigue.

“But with fatigue I don’t know what to do, don’t know what best to do.”

Participant 2

Developing approaches through trial-and-error to deal with fatigue and low energy was important for participants, leading to the development of a ‘learnt expertise’. Over time, this learnt expertise is demonstrated as greater proficiency in identifying, developing and refining effective strategies and can be complemented by the participants’ sensitivity to their body and cues “I know, before I can feel it…” (Participant 12). Common self-management approaches included sleeps, short rests/breaks, or re-energising with food or drink. Some tried alternative therapy methods such as acupuncture, however, this method was of uncertain effectiveness, inflexible and not easily incorporated into daily life.

Participants demonstrated learnt expertise when experimenting with their medication, self-regulating doses to learn ‘how much’ to take at a given time.
Confidence grew with expertise. Others described a different relationship between medication and fatigue which lead them to reduce or eliminate medication they felt was responsible for worsening fatigue. This feeds into the challenge of unmanageable fatigue whereby how participants feel, and what they need to do, become misaligned thus resonating as a feeling of losing control “it was like it controlled me? And my body” (Participant 16) and contributing to the vicious circle.

**Being accepting.** Being accepting is a sense of becoming comfortable with the condition and working with rather than against it by keeping it in perspective. Some adopt a positive outlook “glass half full” (Participant 5) and accept that they will experience their symptoms or fatigue and are comfortable with modifying plans and going with the flow.

> “Just accept the fact you’re going to get it and its going to happen. Equally accept the fact it’s going to last for days. It’s not completely debilitating but don’t push yourself at that particular moment when you feel more than tired.”
> Participant 10

This can help participants to cope with the impact of axSpA, fatigue and low energy on parts of their lives such as sleep “I’ve just got used to knowing that I don’t sleep well and I just kind of adapt, I just deal with it now” (Participant 9). Acceptance seemed to be one route in a crossroad, enabling participants to deal with illness-related challenges and to preserve their mental health “you either deal with it, accept it or you go down the other route” (Participant 9). However, the characteristics of the participants and their personal circumstances means some participants may be better able to cope with their illness than others.

> “My character is to embrace something and move forward. I’m not about sitting and... kicking my feet and waving my arms kind of thing.”
> Participant 16

**3.5.3.3. Superordinate theme 3: Making sense of axSpA**

Making sense of axSpA is an active process of developing understanding and recognising change. This is conveyed through living with axSpA with categories living with unpredictability and searching for a cause; the struggle to maintain
normal life with categories living with restriction, impact on psychological wellbeing and putting on a positive front; and feeling supported with categories being able to connect with others and being able to trust others.

Living with axSpA

Living with axSpA captures the challenges that living with chronic illness poses to living a normal life.

Living with unpredictability. Symptoms of axSpA and fatigue are dynamic, fluctuate and consequently introduce new challenges to daily life for participants. Aside from normal fluctuation, participants experience flares which are periods characterised by worsened pain, stiffness and fatigue. Due to the unpredictable nature of axSpA and fatigue, it is difficult for participants to mitigate any impairing effects they may have on them or the negative impact it may have on them. This creates challenges to creating stability day-to-day.

In a journey to develop a sense of ‘predictability’ over their symptoms, participants discussed looking for associations between changes in their symptom state with specific events such as changing seasons, which could feel comforting for participants, even if the association turns out to be unreliable. Despite these efforts, the lack of predictability means participants are often living and making decisions without knowing the extent of the impact (if any) on their illness and symptoms.

“I’ve not got any predictability over triggers for it where you can think ‘Well, if you do that, that will happen’. I can say if I do that my back will hurt, but I cannot tell you if I do that I can be fatigued at the end of it.”

Participant 10

Searching for a cause. Participants attributed fatigue to a variety of possible causes: age, the weather or illness-specific explanations (e.g. pain). Older participants tended to feel that their fatigue worsened as they got older and became more central in their fatigue experience over time. However, these participants did not consider whether greater fatigue was due to worsening axSpA symptoms and loss of mobility which occur over the course of illness, consequently leading to a reduction in activity levels. For others, changing weather or specific seasons (e.g.
winter) were identified as the reason for changes in their fatigue. Pain was highlighted as a possible causal factor in fatigue for many participants whilst others described pain as an amplifier of fatigue, rather than a cause.

“The pain does make the fatigue worse because you’re looking for an escape, from the pain.”

Participant 1

Changes in dietary patterns due to fatigue and low energy were reported with participants describing skipping meals or choosing quicker but less nutritious meals. Low energy was heavily influential on choices, with poorer dietary decisions negatively impacting on their fatigue and contributing to a vicious circle.

“But I think that could be contributing to the fatigue as well not eating but I think it was the fatigue that wasn’t making me eat as well cos I didn’t have the energy to get up and make food.”

Participant 4

The challenge of unpredictability coupled with the difficulty identifying causal factors highlights the challenges participants face with trying to understand, make sense of and live with axSpA.

Struggle to maintain normal life

Struggling to maintain normal life was the physical and emotional effort required to keep life moving forward demonstrated in living with restrictions, the impact on psychological wellbeing and putting on a positive front.

Living with restriction. Restrictions that result from axSpA and fatigue may require extra effort to complete even basic movements. Normal movements may require an additional two or three actions, which can quickly culminate in a much greater demand on the participant’s energy levels. This can further diminish already low energy levels and lead to overexertion and payback. Overexertion may leave participants feeling drained, however, they can respond to this by appraising the energy cost associated with tasks or social events (described in ‘Struggle to find energy to live everyday life’).

“I still do try and go out with my friends but it, it depends what it is, what energy levels it will require.”

Participant 7
Conversely, payback is distinct, occurring after overexertion or even after minor activities such as meeting a friend for a coffee “I pay for it you know” (Participant 2) and represents a prolonged recovery period during which participants report feeling worse than normal. This is different to a flare as it has a predictable cause, worsening feelings of fatigue and leaving energy levels less responsive to normal efforts to re-energise.

Living with restriction demonstrates how participants can be obstructed from doing what they want to, interfere with physical and mental functioning and diminish home, work and social life with some participants feeling that axSpA and fatigue have “shrunk” their “world” (Participant 12).

**Impact on psychological wellbeing.** Diminished psychological wellbeing may require effort from participants in the form of coping, introducing new challenges to living a normal home, work or social life. Participants disclosed feelings of anxiety, guilty and downturns in their mood attributable to fatigue. This influenced the temperament of some participants who felt more irritable and short-tempered with others.

“I do feel that I’m a little bit depressed and I think the fatigue has altered my moods, and because I’m constantly in a low mood.”

Participant 3

Others shared feelings of guilt for two reasons: putting their own fatigue-needs ahead of the perceived needs of others (e.g. children), and when pulling out of plans last minute meaning the participant and others are affected. Participants described feeling “stressed” (Participant 7) because of fatigue which may contribute a flare “Yes, it can. I think, if you’re stressed, I can definitely kick off an AS attack if I’m very stressed” (Participant 13). This exacerbation of axSpA symptoms can contribute to a vicious cycle for the participant with increased pain, poorer sleep quality and duration, greater feelings of tiredness and fatigue which can further contribute to stress. Here, participants may feel overwhelmed and no longer be able to action their self-management strategies.

“Erm, I still cyclically get horribly tired or I will get horrifically
erm... when I get flare ups, flare ups are awful.”

Participant 12
Putting on a positive front. Some participants described ‘putting on’ a positive front which seemed to occlude how they really felt. This allowed them to keep non-observable symptoms of their condition private and remove them from dialogue, ensuring they are not treated differently by others. In some instances, there was an element of denial and avoidance involved which, through putting on a front, allowed participants to disconnect from their condition and move it into the background.

“I try not to think about it actually and not many people know that I've actually got it (AS). I think it only hits me that I've got it when I have a real bad spell. I know when I see [physiotherapist's name] at me (my) appointments, I forget that.. it's my AS that's causing it. 

Does that make sense?”

Participant 17

Unfortunately, this front also introduced a barrier, stopping others from developing their understanding and sensitivity to the participants’ experience of axSpA and fatigue. This was juxtaposed by a perceived benefit by the participant for those around them who, through not knowing, were protected from the negative aspect of their illness.

“If I’m tired or fatigued or if I’m in a lot of pain, I just put a smile on and crack on you know (laughs), because why should you bother other people with your ailments.”

Participant 1

Feeling supported

Feeling supported captures the importance of caring relationships demonstrated in being able to connect with, and trust others.

Being able to connect with others. Participants drew on the help of others, whether with physical tasks or to talk to. This support was frequently local to them and readily accessible (family, neighbours, friends) and could be effective for balancing chores and tasks, protecting participants from overexertion whilst allowing engagement within daily life. Similarly, confiding with others can help participants deal with any emotional build-up.

“If that means that you can offload with the person … even if they can't even understand what you're going through… having that
Although some participants could no longer do what they used to, support enabled them to integrate into normal life within the household.

Understanding within the support network enabled participants to seek specific support for their needs. A spectrum of understanding was described, which varied between participants. Misunderstanding was a source of difficulty for most and acted as a barrier to them seeking support consequently leading to feelings of irritability and more frequent clashes with significant others.

“My wife doesn’t really understand it. She sees me attempting to continue to do my stuff daily, my job daily, working, cleaning up doing everything on its own just trying, she sees me doing all that, so she probably doesn’t understand how I actually feel.”

Participant 3

Participants frequently reported frustration at advice from friends, family or work colleagues rooted in poor understanding, for example, others implying that the participant may be responsible for feeling tired, rather than that they are experiencing axSpA-related fatigue.

“When they say to be “Oh you look tired”, it’s normally like, “Did you have a late night?” “No I was in bed probably four hours before you were but this is how I feel after a full night’s sleep. So I think it’s misunderstanding, absolutely.”

Participant 3

Being able to trust others. Developing trusting relationships with friends, family and professionals was key to participants being able to share their experiences of illness and utilise support. Where trust is absent, shared understanding may not develop and thus become the source of tension within the relationship. However, developing trust requires participants to have a good understanding of their condition, knowing fatigue is part of that condition, and to be transparent and open about life with their illness and fatigue. Where the participant is not aware of a symptom such as fatigue, it may become occluded and therefore not be addressed or receive the attention it needs.
“I’ve never really mentioned (the fatigue). Just thought it was part of the parcel, something that you had to deal with.”

Participant 2

Equally it requires others, such as the clinical professionals working with patients, to ask key questions about the illness and its impact on patients, address fatigue and if fatigue is raised as a key issue, respond to it seriously and not brush over it.

“If I brought it up and wanted it addressed, and they just brushed over it. As in ‘oh that tired thing, let’s move on’.”

Participant 3

In relation to fatigue, some patients may not raise the topic in clinic due to concerns about its relevance, making fatigue a tentative topic “am I wasting their time?” (Participant 1), although this was not the case for all participants, some of whom spoke positively about healthcare professionals in relation to how they discussed their fatigue “She’s always really thorough” (Participant 11). Despite this, challenges remain in delivering quality support for fatigue.

“I know I can say stuff to [physio's name]... and she'll listen and give good advice... much more than most other people will”

Participant 12

Wanting to remain or appear independent to others can lead participants into creating a barrier that can hinder the development of a reciprocal relationship with others.

“You know I, had to put my pride you know aside and learn to take others offers of help (laughs).”

Participant 16

Participants who felt empowered in their relationship with professionals were able to disclose issues with their fatigue and seek help. Receptive and proactive professionals were reassuring and were referred to positively by participants. Trust, therefore, may help to establish a framework between the participant and professional that is conducive to openness and empowers the patient to raise issues affecting them.

“I feel I can talk to him about it, so I try to pass off as much detail as I can cos then I know that if there is something that can be
3.5.4. *Patient research partner (PRP) – review of study findings*

The meeting was held with 5 PRP members for approximately 3 hours. A summary of the findings was shared with patient partners and the group discussed the lived experience framework, often describing how it resonated with them and highlighting that it really captures their experiences. It was during this meeting that the learnt expertise aspect of learning to live with and manage fatigue arose. However, the concept of energy garnered most of the members’ attention with members agreeing it was important but unable to reach a consensus about whether it was distinct from fatigue, or more intertwined. Members were divided and proposed exploring the concept further. They also specified that they disliked the label ‘energy capacity’ to capture energy limits and felt ‘stamina’ was a more acceptable and meaningful label. Some member also disliked language about ‘causes’ of fatigue and felt that its cause could never be identified, and instead things like the weather were ways for them to justify the way they feel because of fatigue. No consensus was reached on this language change, therefore, I agreed to revisit the analysis to determine the most appropriate language as used by the participants.

The discussions, comments and feedback from PRP members resulted in changes to the lived experience framework. This included smaller changes to language and revisiting the analysis to determine the importance of concepts like learnt expertise – this was developed and incorporated into the analysis.

3.5.4.1. *Summary of lived experience findings*

Achieving balance was expressed through three key concepts of ‘being’ that related to living with axSpA and fatigue. These concepts were: having energy, engaging in everyday life and making sense of axSpA. Having energy captured the challenges of being proactive. Engaging in everyday life identified the proactive approaches participants take to daily life. Making sense of axSpA demonstrated the challenges of living with illness, understanding it and how participants were affected by their axSpA and fatigue.

The importance of energy as a resource that participants draw on to initiate and maintain physical and mental activity is an important, emerging finding from the
Energy was experienced alongside the challenge of having an invisible illness, whether that was in reference to the participants’ axSpA, fatigue or both. Achieving a balance was crucial to how participants maintained their lives, the process of learning to live with their axSpA and fatigue including the complications that having an invisible illness can bring.

3.5.5. Developing a measurement framework
Using the lived experience framework derived from the patient interviews (Table 3.3) and guided by the three components of the impact triad (severity, importance, self-management) (240), the following measurement framework was derived to underpin the questionnaire (Table 3.4). Five domains were identified: symptoms, impact on the patient, function, psychological and emotional wellbeing and self-management. These measurement domains were extracted as part of the analysis from the first version lived experience framework of axSpA fatigue and energy and were for reviewed and appraised in the focus groups (chapter 4).
**Table 3.4: Domains and subdomains identified and extracted for the purpose of underpinning future measurement**

<table>
<thead>
<tr>
<th>Domain</th>
<th>Subdomain</th>
<th>Component and supporting quotes from interviews</th>
</tr>
</thead>
</table>
| Symptoms  | Fatigue     | Frequency: “It’s just fatigue and it’s like that every day. I don’t get one day where I don’t get any fatigue and then the next day I’m okay.” (Participant 6)  
Severity: “I am that tired, … I feel sick with it sometimes ….” (Participant 3)  
Duration: “Mine tends to be between 2 and 4 in the afternoon I could just lie down anywhere and that’s it …” (Participant 5) |
|           | Low energy  | Frequency: “I just felt drained and had no energy, no motivation” (Participant 4)  
Severity: “I feel completely drained, just got no energy left.” (Participant 6)  
Duration: “It’s constantly draining. Constantly draining, I feel drained now. I feel like I could just roll over there and go to sleep” (Participant 5) |
| Impact on self | Cognitive   | Reading: “… You can’t read books or newspapers all the print gets into one line, like bar codes and with my eyes they are bar codes” (Participant 1)  
Talking: “You’re so light-headed and dizzy with tiredness and you, you just, I can’t even hold a conversation …” (Participant 7)  
Memory: “The memory, you know the mind and it and all just sheer tiredness of forgetting what I’m doing you know” (Participant 8)  
Concentration: “When I’m tired and feeling like it’s flared up I won’t play that because the concentration side of things” (Participant 9) |
<table>
<thead>
<tr>
<th>Physical Self-care/eating:</th>
<th>“I didn’t have the energy to get up and make food … It affected me preparing food and it sort of affected my appetite as well because I felt tired and if I was a little bit hungry I would just sort of choose sleep over making food …” (Participant 4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exercise:</td>
<td>“The thing I find really difficult is that obviously exercise is the best thing that can be done to help AS but … if you had the energy to then you could help yourself a bit more but it’s so draining” (Participant 7)</td>
</tr>
<tr>
<td>Social Usual work/job:</td>
<td>“It makes my work life very difficult, makes my home life very difficult, and makes me feel like shit. Day in, day out.” (Participant 3)</td>
</tr>
<tr>
<td>Leisure activities:</td>
<td>“I wouldn’t play football I’ll just sort of sit and watch. Whereas before, I’d be like the first person on the pitch …” (Participant 4)</td>
</tr>
<tr>
<td>Socialising:</td>
<td>“I don’t really have many friends but the ones that I do have are married so, I hardly ever see them anyway, but when they want to see me sometimes I think “actually I just cannot be bothered” and I’ll come up with some excuse!” (Participant 11)</td>
</tr>
<tr>
<td>Function Difficulty sleeping:</td>
<td>“I’m tossing and turning all night but I don’t go into that really deep sleep again” (Participant 13)</td>
</tr>
<tr>
<td>Disturbed sleep:</td>
<td>“I’ll wake up two or three times a night” (Participant 4)</td>
</tr>
<tr>
<td>Feeling when awake:</td>
<td>“I can sleep for the dead, wake up, my heads really bright but my body’s saying “sod off”.” (Participant 12)</td>
</tr>
<tr>
<td>Psychological and emotional wellbeing Mood Depression (low mood, feeling down):</td>
<td>“I do feel that I’m a little bit depressed and I think it’s the fatigue has altered my moods” (Participant 3)</td>
</tr>
<tr>
<td>Losing control:</td>
<td>“I was definitely more frustrated.. I was definitely more emotional.. erm [pause] I, I think because I felt so out of control …” (Participant 16)</td>
</tr>
</tbody>
</table>
| Anxiety and worrying | **Worry:** “I do, I do, worry” (Participant 17)  
**Dread:** “I only ever sort of go to things if I absolutely have to because I dread… you know, even things like it’s my nieces 4th birthday party” (Participant 7)  
**Frustration (stress):** “It can be frustrating, sometimes cus you can be, busy doing something and then all of a sudden you just hit that brick wall…” (Participant 17)  
**Unpredictability:** “It’s so unpredictable and you just don’t know. You can’t plan anything, you can’t…” (Participant 2) |
| --- | --- |
| Sense of self | **Guilt:** “… That spoils it for me then, and it spoils it for her, because we’ve wasted it. But I didn’t know it was, I was going to be that bad on that day” (Participant 1)  
**Embarrassment:** “I was talking like I am with you now and I just went off like somebody had taken my batteries out … I felt so embarrassed cos I’d just fallen asleep but I mean, we laughed about it but, [laughing]” (Participant 13)  
**Helplessness:** “I think it’s more tiring for me than a normal human being” (Participant 10) |
| (Self)-isolation | **Pushing others away:** “I didn’t really want to speak to anyone. It would be short answers if people did speak to me. I just sort of shut myself off put my headphones in and just sort of lock myself off away from the world …” (Participant 4)  
**Hiding fatigue:** “I’m probably a bit more isolated than I feel I really am” (Participant 9)  
**Invisible illness:** “So they look at you and think “oh you look pretty healthy” you know and they don’t understand what's going on inside.” (Participant 13) |
| Self-management | **Achieving balance:** “I’ll plan particularly carefully that I’m not doing too much. I’ll particularly plan I’m not doing anything too little… so if I plan to do something and I don’t feel particularly good on that day, I will not o it I’ll wait for another day” (Participant 10) |
| Energy          | Coping and adapting: “I’ll just make changes and adapt differently but at the end of the day I’ll just crack on” (Participant 9)  
Learnt expertise: “I now feel brave enough most of the time to turn round and say... with the people that are close to me I can’t do any more, I have to stop, even if I don’t really want to stop, I know that it’s a law of diminishing returns” (Participant 12)  
Flexible approach: “When you’re feeling well, get the jobs done. When you don’t, feel too grand, you leave them.” (Participant 1)  
Feeling lost: “But the fatigue, mentally and physically that goes with it I can’t seem to get a grip on it or manage it at all, it drives me mad” (Participant 7)  
| Difficulty starting things: “I can’t be bothered doing owt (to do anything). Or anything that involves getting up and moving around. Anything remotely active.” (Participant 10)  
Re-energising: “I have energy drinks” (Participant 3)  
Declining events (avoiding wipe-out): “You know, say someone wants to go for a coffee if I’m not feeling 100% I won’t put myself in that situation.” (Participant 2)  
Feeling drained: “… That drained quite a bit of my energy cos I couldn’t take like a certain amount of step I had to walk in a certain way when I have the massive flare ups I had to sort of swing my hip out and around” (Participant 4)  
| Support          | Delegating responsibilities: “If I’m struggling, obviously hubby is first port of call, but there are other people who can help me, when I need it” (Participant 12)  
Friends, family and colleagues: “It would really make me sad to know that there’s someone who suffers the way I suffered and doesn’t have anybody, to be there, to support them” (Participant 16)  |
“Luckily my partner’s really, really understanding he wasn’t at first it’s taken him ages to get used to it, erm, but now he’ll be like ”go and you know, go and have your nap, go to sleep” (Participant 7)
3.6. Discussion

The findings identify the central concept of achieving balance and the hidden work undertaken by patients to attain a balance. Where balance was achieved, participants were able to self-manage their condition, cope with their fatigue and its effects, and maintain their energy levels. This study contributes to the understanding of living with axSpA and fatigue by the identification of: (i) the key role of energy and the activities undertaken by participants to support their energy levels; (ii) the overarching concept of achieving balance and ways in which expertise was acquired, and (iii) the hidden challenges and work within participant’s daily life. A detailed discussion and reflection on PRP are reported in Chapter 6.

Having energy emerged as a distinct, but related concept to fatigue that to date, has not previously been highlighted in axSpA. Participants described energy as a short-term, daily resource that can be drawn on and replenished in order to be physically or mentally active. Whilst other researchers in axSpA (33,36,244), multiple sclerosis (MS) (245) and cancer (246) have also referred to energy, they did not conceptualise what energy means to patients, and it was often used as a synonym or antonym of fatigue. Having energy was essential to initiate, continue and finish activities and despite sudden draining of energy, participants could recover and replenish energy levels to continue with daily life. This is juxtaposed by fatigue which was a difficult symptom that permeated all spheres of the participants' life and remained mostly unresponsive to rest or sleep. Where energy was sparse or difficult to maintain, participants reported being socially withdrawn and less active – a similar finding of reduced social activity or withdrawal has been reported in other axSpA studies (33,36). Evidence from research in HIV has posited a distinction between fatigue and energy and a working definition of energy emerged - “Energy is the individual’s potential to perform physical and mental activity” (209) – a proposal which corroborates inferences made in this study. Similarly, energy was reported to be more sensitive to diurnal change than fatigue and was not simply antithetical to fatigue (207). In my study, energy is defined as an internal resource that participants draw on to be physically and mentally active, enabling them to be engaged in, and live their lives.

Participants in my study proposed that managing fatigue and energy levels required the ability to identify, develop and appraise different approaches to...
ease or better self-manage their symptoms. Many participants described a trial-and-error approach towards developing a learnt expertise of their condition, symptoms and effective ways to deal with them. A similar finding of trial-and-error has been reported by others (36,206), however, the additional aspect of learnt expertise is a new contribution from my study. Learnt expertise was acquired over time and enabled participants to respond to their fatigue and energy needs using strategies that they know have been effective in the past. However, a wider challenge around self-management in axSpA emerged and related to a lack of understanding, professional support and guidance. Most participants stated that they did not understand their fatigue, and some did not know it was associated with their axSpA. Further, only some participants reported having excellent support from professionals – most did not. This highlights two potentially significant deficits for patients: firstly, a lack of patient knowledge and understanding about what it means to have axSpA and what it entails and; secondly, a disparity in the provision of support from healthcare professionals.

Living with an invisible illness was identified as a difficult aspect of having axSpA and fatigue. They often felt they were judged as lazy and their challenges in daily life were not recognised by others. The invisibility of fatigue has been highlighted in a review of fatigue experiences amongst people living with a range of chronic illnesses (including axSpA) (47). This seemed to be rooted in a poor understanding of axSpA and fatigue by patients and those around them, including professionals. Evidence from RA has found that rheumatology nurses had a lack of knowledge and communicated poorly with patients with regards to fatigue; seldom referring patients to other services despite recognising fatigue as a multidisciplinary diagnosis (247). It has been reported that professionals do not discuss fatigue with the patient and their family (30), despite there being evidence that it can significantly affect both the patient and their families (248). This demonstrates a need for better education about fatigue for patients, professionals and family alike; supporting more timely access to support and help patients to achieve balance and confidently self-manage their symptoms.

Putting on a ‘positive front’ was one-way participants coped; which inevitably acted as a barrier to others being exposed to the challenges of
their axSpA and fatigue, jeopardising participants’ access to practical and emotional support from others. This type of coping method has not been reported in other axSpA studies. Participants who felt understood by their friends and family were able to turn to them for help with their responsibilities which reduced the burden on the participant, therefore better enabling them to cope with their condition. In contrast, those who described feeling unsupported described being snappy, irritable and quick to temper which affected their relationships. Similar detrimental effects of fatigue on social interactions have been highlighted in a recent cross-sectional survey (249); however, there are currently no studies specifically exploring the impact of axSpA fatigue on social interaction and family life.

Although living with axSpA and fatigue can bring new challenges, participants were determined to find ways to self-manage. Participants described a range of processes such as acceptance, a changed way of being and making practical changes including prioritising important tasks and being active. The adoption of similar processes has been reported in other interview studies (33,35,36) and chronic illnesses (47), although no study identified the importance of being organised to create structure and routine in the patients’ lives. Whilst these studies identified similar processes, they were poorly explored with limited explanation of how these strategies were implemented by participants – information that is essential to understand how participants integrate self-management strategies into daily life.

This study has used current best practice guidance to develop the measurement framework, such as conducting qualitative interviews to establish the patient perspective of the target concept (89,112,123), establishing a phenomenological framework (123) and producing a table of identified domains with supporting quotes (89). However, guidance has concentrated on preferred methods to develop a PROM (e.g. qualitative research interviews) and provides little guidance on how to analyse data to produce a measurement framework. In this study, I have utilised a developing theoretical framework for extracting meaningful PRO concepts relating to impact through the use of the impact triad (240). This framework provides the foundations upon which a future PROM has been developed, articulating the PROs important for measurement (250). Whilst this is an emerging theoretical framework for the purpose of measurement, it
captures and is reflective of the multidimensional nature of axSpA-fatigue and energy conceptualised in my study. The absence of guidance on how to best extract a measurement framework, from qualitative data, to underpin measurement, means there is a gap in current guidance recommendations and the impact triad is a step toward addressing this. Moreover, I produced a table detailing the key domains of axSpA fatigue with accompanying patient quotes which provide transparency, as per FDA guidance (89).

Method and analysis reported in this study uses current guidance from the COREQ 32-item checklist (242), strengthening the rigour of the study through transparent and robust reporting. Other checklists such as the 21-item Standards for Reporting Qualitative Research (SRQR) (251) were considered, however, these tended to offer a more generic reporting framework for a full write-up and less in terms of reporting a full methodology – for example, there is little guidance in the SRQR on what to report regarding the underlying theoretical framework of a study (251). The COREQ checklist was specifically designed for interview and focus group studies, increasing its suitability for use in my study. I flexibly applied the checklist to ensure fit with the theoretical underpinnings of my study, and its purpose. Such a flexible approach has been argued for in a Delphi study exploring potential reporting guidelines for qualitative research which recognised the need to adapt criteria in line with the approach taken (252).

This study used purposive sampling, open questions and unrestricted interview durations to maximise the discussion and ensure a range of views were expressed in interviews. The sample included a wider range of disease durations than in other identified qualitative axSpA studies (33–36,206), however, the average disease duration of participants in my study was up to 10 years lower (35,36). The lower average disease duration is explained by the wider range of ages represented in my study compared to the two other UK studies (35,36). Moreover, the balance of sexes in my study was almost 50-50, similar to one UK focus group study (35) – meaning the female perspective was equally represented in my study. This contrasts with qualitative axSpA studies which have reported male sampling biases between 60% (36) in the UK to 90% in Egypt (34), and a 60% female bias has been reported in Norway (33,206). The prevalence of axSpA between males and female is almost equal (253), therefore
supporting the representativeness of my study sample. One third of respondents in my study showed evidence of mild depression, and over half showed evidence of mild anxiety. Moreover, one third of participants who reported anxiety met the criteria for moderate levels of anxiety (data not reported). This fits within estimates for anxiety and depression in axSpA (254–259).

Although study participants were given the option of being interviewed at home or by telephone, all interviews were conducted in a quiet, private room as has been recommended (260), at the participants’ local rheumatology department. This accommodated patient’s schedules as most were already attending the hospital (e.g. personal appointments, weekly AS group meets, hydrotherapy), thus reducing participant burden. Interview location may influence responses – for example, if asked about health professional support, participants may limit their views when interviews are held within the same hospital. Conversely, however, conducting interviews in familiar surroundings can enable participants to be more confident and open about their experiences (261). During the interviews in this study, participants did discuss their thoughts and feelings about local care including sharing their negative experiences.

An earlier study of 60 Egyptian patients, 20 with rheumatoid arthritis, fibromyalgia and axSpA proposed that in axSpA, fatigue was predominantly related to the physical impact of disease, with few issues that spoke to the mental impact of fatigue (34). In contrast my study highlights the multifaceted nature of fatigue, describing physical, cognitive, emotional components alongside the social impact. Additionally, my study highlighted that fatigue experience can fluctuate independently of disease activity; this raises an interesting question about using a summated score from the BASDAI which could be inflated by fatigue scores, potentially masking real change in underlying disease activity from treatments or interventions. The patients in both studies were mostly similar and whilst the only substantive difference was the limited number of females included in Mortada’s study (n=2), this is unlikely the reason for any difference. However, the questions used in data collection directly referred to the effects of fatigue on physical activity, but no items addressed the impact of fatigue on mental activity or its psychological and emotional implications, thus indicating a possible methodological reason for the difference in findings.
3.7. Next step
The next step in the research process was to operationalise the measurement framework developed in this chapter. However, the views of a wider stakeholder group were necessary to ensure the measurement framework was reflective of what matters, and acceptable to all end-users; that is, healthcare professionals as well as axSpA patients. Chapter 4 describes the crafting, review and pretesting of an item set (underpinned by the measurement framework) to produce an axSpA fatigue specific PROM ready for psychometric evaluation.
4. Chapter 4: Item generation and PROM refinement

4.1. Introduction
This chapter presents the process of item generation, refinement and pretesting. Items were generated to reflect the domains and PRO concepts identified in the measurement framework (chapter 3) to form the initial item set. The measurement framework and newly generated items were reviewed by both patient and professional end-users in focus groups, followed by pretesting interviews to qualitatively appraise the comprehensiveness, relevance, acceptability and comprehension of the item set. This process produced an axSpA-fatigue specific long form PROM for statistical evaluation in a field test; this field test evaluation is described in chapter 5. A short background to item generation and pretesting is presented in section 4.2. The process of item generation is described in section 4.3. Section 4.4 provides an overview of the methods are results for focus groups and pretesting interviews. Section 4.5 describes the focus group methods, followed by the results of the focus groups in section 4.6. Pretesting interview methods are described in section 4.7 and the results of these pretesting interviews are described in section 4.8. The chapter closes with a discussion (section 4.9).

4.2. Background
To ensure the development of PROMs that are relevant and meaningful to patients, it is essential to establish evidence of content validity (126,262). FDA guidance highlights the necessity of content validity, which it defines as (89):

‘the items and domains of an instrument are appropriate and comprehensive relative to its intended measurement concept, population, and use’

The development of a measurement framework in chapter 3 provides a foundation to underpin item content relevant to fatigue experience in axSpA. However, current guidance does not explain how best to translate the findings of qualitative research into meaningful items and instead is limited to issues of rigour, such as transparency and producing a clear audit
trail in the item development process (89,112). Approaches to item
development include translating concepts that occur frequently (112,126),
generating an exhaustive list of PROs that are important to patients and
reflective of the research (263) and examining relevant existing measures
as source material to supplement or provide item content for a new
measure (59). The latter step is commonly used in PROM development has
important benefits including the fact that these published items have usually
undergone some form of psychometric evaluation (118).

In addition to developing PROM content, it is important to confirm that the
measurement framework is a comprehensive reflection of the construct and
that the generated item set is fit for purpose. Different approaches have
been taken in PROM development with some using focus groups that
specifically incorporate item development and refinement into the session.
For example, in the development of a PROM for recovery from drug and
alcohol dependence, a short list of potential items was generated following
Delphi groups with service providers to be reviewed in focus groups (264).
Participants were invited to judge items for their importance, wording and
relevance (264). Substantive changes were made to items following the
focus group reviews, demonstrating their value in supporting content
development and refinement. Conversely, a hybrid meeting that involved
reviewing a measurement framework and developing item set was used in
the development of an RA stiffness PROM (107).

Once an item set has been crafted to reflect the measurement framework, it
is then recommended that pretesting interviews are conducted to
qualitatively determine the comprehensiveness, relevance, acceptability
and comprehension of the developing PROM (126). Moreover, ISPOR
advocate using pretesting interviews to explore respondent understanding
of items, instructions and response options (126). Item generation to craft
an item set, focus group reviews of the measurement framework and item
set, and pretesting interview methods are described in this chapter.

4.2.1. Aims and objectives
To develop a long-form PROM of axSpA fatigue ready for psychometric
evaluation in a national sample of axSpA patients.
Objectives:

1. To develop an initial item set following current best practice guidance.
2. To review and appraise the measurement framework and initial item set, particularly considering relevance, acceptability and comprehensiveness.
3. To pretest items to assess their comprehensiveness, relevance, acceptability and comprehension; supporting item modification and refinement to produce a long-form axSpA-fatigue specific PROM.

4.3. **Item generation, selection and scale development**

An initial item pool was developed to reflect the domains and PROs identified in the measurement framework, using sources of both primary data and secondary information:

a. Qualitative data from axSpA patients (Chapter 3).

b. Existing single and multi-item measures of fatigue used in axSpA (Chapter 2).

Additional measures of fatigue were identified from existing reviews of fatigue PROMs (150–152). Items were developed or extracted and reviewed in focus groups with patients and healthcare professionals (Method 1).

The results of PROM mapping are described below, followed by decisions for recall period, response scales and polarity. Where an existing item could not be identified, new items were crafted by me, using the measurement framework informed by the findings from the qualitative study (Chapter 3).

4.3.1. **Item generation from existing measures of fatigue (‘PROM mapping’)**

A process of ‘PROM mapping’ was conducted – that is, identifying appropriate existing PROMs and extracting items considered to be relevant to the domains and PROs documented in the measurement framework.

There is only a finite number of ways a question can be posited (118),
therefore, using existing items that have some development and
psychometric credentials can help with the development of a new measure,
identifying items crafted specifically to capture intended PRO concepts.

To support extraction and transparency, a table was developed that
included details of: (i) the identified conceptual domains within the
measurement framework; (ii) subdomains; (iii) PRO concepts, per
subdomain, identified for measurement; (iv) existing fatigue PROMs
consulted, and; (v) the extracted item(s). It was essential that the list was
not too long or onerous so that the focus groups could manage the amount
of data. PROMs were reviewed for content until a fatigue item list was
populated.

4.3.2. PROM mapping results
The item list was mostly populated using items from existing fatigue
PROMs (see Appendix 4A for the results of PROM mapping). Nine PROMs
were identified from the review conducted as part of this thesis (Chapter 2),
and an additional 15 fatigue measures were identified from a review of
fatigue measures used in chronic illness (152). One further measure was
identified from two other reviews (150,151) – the checklist individual
strength (CIS20R and CIS8R) (151). The BRAF-MDQ for RA fatigue was
also considered for potential items (106). No reviews of energy measures
could be identified, and energy items identified within existing measures
were generated to measure fatigue except for the SF-36 which asked two
questions about energy and vitality (66).

44 items were extracted during PROM mapping to produce an initial item-
set, with 15 new items being crafted by me to address PROs where no
relevant items were available. This item set was reviewed by a content
expert (KH), a clinician (JP) and me to assess its comprehensiveness (in
relation to the measurement framework), relevance (patient and clinical),
item structure, language and to ensure all PRO concepts were captured.
Recognising that the phraseology of items was disjointed (due to being
extracted from an array of measures), we standardised the wording.
Following discussions, 19 items were removed with repetition being the
most common reason for item removal, and some items being removed for
using unsuitable language. For example, an item taken from FACIT-Fatigue
asking about feeling ‘listless’ was rejected as this term was not used by patients in their interviews (61). This item list was reviewed by a focus group with both patients and professionals to maximise engagement in item development and refinement.

**Recall period**

Of the measures identified in the systematic review (Chapter 2), a range of recall periods were used in PROMs with one week being the most frequently used across single and multi-item measures (61,76,177–179) through to one month in a generic measure of general health status (66). The worst-fatigue NRS taken from the BFI specified over the past 24 hours (180) and one measure did not specify a time period (42). Current best practice guidance highlights the complex interplay between recall period and item content and recommends using as short a recall period as possible, recognising the potential for recall bias and potential burden on the participant (126). In addition, guidance recommends considering variability and frequency of what is being measured, with higher variability and frequency recommended for more frequent, short-recall assessment (126). Therefore, as the evidence from chapter 3 suggested fatigue was unpredictable, difficult to manage and variable, a one-week recall period was selected for the items and was subsequently discussed in the focus groups.

**Response scales**

A study that replaced the VAS’s in the BASDAI with 11-point NRS’s found that 87% of patients preferred to answer the study questionnaires using a NRS or Likert scale, with only 9% preferring the VAS (147). In addition, a review of pain intensity measurement found that an 11-point NRS was slightly more preferable to patients due to it placing slightly lower cognitive demand on them (265). Therefore, only the NRS and Likert scale response options were selected for review in the focus groups.
Polarity

A unipolar scale was selected for use in this study. This is because the measurement framework identified intensities, severities, frequencies and interference as important components of fatigue and energy for measurement. This means that the scale begins from 0 or ‘not at all’ and goes to an extreme as per other measures of fatigue (61,106,177).

4.4. Overview of methods and results
Method 1 describes the focus groups used as part of the review process to (i) confirm a working measurement blueprint of axSpA fatigue and energy, (ii) to identify important outcomes for measurement and (iii) to generate, modify and refine the developing item set whilst checking for breadth and relevance. Method 2 describes the pretesting assessment of the new axSpA fatigue and energy questionnaire as part of its development and refinement into a long-form questionnaire for psychometric evaluation.

4.5. Method 1: reviewing the axSpA-fatigue and energy measurement framework and items
NHS ethical approval was granted following review by a research ethics committee (REC reference: 16/WM/0147).

4.5.1. Aims and objectives
To review and further refine the measurement framework of axSpA fatigue and energy by collaborating with both patients and healthcare professionals with experience of fatigue or working or researching this field to reach a consensus on a working measurement blueprint of axSpA-fatigue and energy.

Objectives:

1. To identify important and relevant outcomes for patients and healthcare professionals for measuring fatigue and energy in axSpA.
2. To explore and refine the measurement framework of axSpA fatigue and energy into a working measurement blueprint to underpin a new PROM.

4.5.2. Justification for using focus groups

PROM development is an iterative process and current guidance recommends the use of both interviews and focus groups as part of the development process (89). Focus groups provide an opportunity for the target population and other relevant stakeholders (e.g. healthcare professionals) to review and have direct input into the developing measurement framework. Specifically, focus groups provide an opportunity for new issues and topics to emerge through sharing and challenge, and to check for any important concepts that may be missing (266). Focus groups have been used in the development of item banks (267–269) and PROMs (107,270). In PROM development studies, the focus groups tended to be structured into exploring the groups’ experience of a phenomena, followed by a review of PROs or the new questionnaire (107,270). In the development of an RA-stiffness PROM, a structured style of focus groups was used with part A exploring a previously derived qualitative conceptualisation, and part B exploring the groups views on how RA-stiffness should be measured (107). A similar approach was taken in the development of a PROM for real-time symptom assessment for individuals with functional urologic complaints, however, the second part of these focus groups involved reviewing a pre-defined questionnaire (270). Through using focus groups, both studies were able to identify potential issues, such as a question on stiffness duration being too difficult to recall or quantify (107), informing refinement decisions in anticipation of pretesting interviews.

Nominal group technique (NGT) is a group process that involves three key stages: (1) identification of a problem; (2) developing a solution, and; (3) decision making (271). However, as the problem has already been identified (fatigue in axSpA) and a solution has been proposed (the measurement framework of axSpA-fatigue), the group will instead follow a modified NGT process: (1) a detailed exploration and appraisal of the measurement framework; (2) identifying and ranking what are the most
important components for measurement, particularly regarding impact (240), and; (3) deciding on how best to measure axSpA-fatigue by reviewing, generating and modifying items. For ease of reading, I refer to the modified NGT used in my study as focus groups herein.

4.5.3. Participants

4.5.3.1. Patients participants

Patients were approached by a member of their local rheumatology unit and provided with a study cover letter, information sheet and a consent form for agreement that I could contact them. Contact was made up to one month before the focus group was due to begin. The local unit clinicians at two participating sites (Stoke-on-Trent and Wrightington), along with their unit colleagues approached patients in three ways: after routine appointments, after weekly AS meetings (Stoke only) and by telephone. Participants who confirmed their willingness to participate were given a letter specifying the full details of the study: purpose, time, date and location. Reminder calls were made locally, and I sent reminder emails to participants a few days prior to the group being held. A minimum of 12 patients were invited to each scheduled focus group.

Patients were approached (and included) if they had a primary diagnosis of Axial Spondyloarthritis, were aged 18 years or older and were able to read, speak and write in English.

4.5.3.2. Healthcare professionals

Clinicians at two participating sites were asked to share the study with eligible colleagues in person or via email. Their invitation to participate included a participant information sheet and study invitation letter outlining the purpose of the focus group. Professionals were also approached through the Ankylosing Spondylitis Special Interest Group North West (ASSIGNw) network (http://www.research.lancs.ac.uk/portal/en/publications/the-ankylosing-spondylitis-special-interest-group-north-west-assignw(50b7e6c8-674e-49b3-a221-5b8e3c741ae).html). Those who returned a consent to be contacted form were then directly invited by me and a group date and time
was arranged. Unfortunately, response by professionals was limited across both sites which was thought to be due to a lack of experience with axSpA-fatigue.

Healthcare professionals were included if they had experience of working with axSpA patients and or experience working with patients with fatigue in rheumatology.

**Sample size**

Focus group sampling guidance ranges between 4 and 12 participants per group (272–274), but typically groups contain between 6 and 8 participants (275). To maximise likelihood of achieving at least 4 participants per group, the research nurses responsible for patient approach were asked to approach and invite at least 15 participants per focus group. This was informed by an existing study exploring fatigue self-management in a UK sample and achieved a 46% expression of interest rate (35). Close contact via telephone and email was maintained by me with the nurses responsible for recruitment to ensure enough participants were being approached for the focus groups.

4.5.4. **Methods and data collection**

Four separate focus groups: two with patients and two healthcare professionals were held between October 20, 2017 and November 28, 2017. I facilitated one professional group in Stoke-on-Trent independently, two patient groups in Stoke-on-Trent were co-facilitated by JP and one professional group in Wrightington was co-facilitated by JM. All focus groups were conducted in the local rheumatology clinic of patients and professionals to facilitate maximum participation through its locality and familiarity.

Before beginning the focus group, participants were asked to complete consent forms and to provide some basic demographic information. For patients, this included age, sex, disease duration and symptom duration. For healthcare professionals, this included age, sex, their current position (e.g. nurse, physiotherapist etc) and years of experience working in axSpA. Verbal probes were used throughout groups to encourage further
discussion on key points such as asking participants 'why', 'how' and inviting them to expand on points raised. Co-facilitators were also able to prompt to maximise discussion between group members. Asking for examples allowed for points raised to be contextualised into a real-life scenario for patient participants and related back to clinical experience and settings for healthcare professionals.

Groups 1-3

A topic guide was developed collaboratively with the research team (KH, ET, JP) and consisted of three parts (Table 4.1): Part A was a detailed discussion about the participants’ experience of, or working with, axSpA-fatigue followed by a review of the measurement framework that emerged from the patient interviews. Part B involved a ranking task where participants were invited to individually generate lists of what is important to them about axSpA-fatigue. Finally, Part C was a group review and appraisal of the relevance and acceptability of a group of questions generated to address key concepts that emerged in the measurement framework. Each item, per domain was reviewed. All focus groups opened with ground rules (respecting anonymity, confidentiality and right to withdraw) and ended with a closing statement thanking attendees for participating and asking if there were any questions.

Table 4.1: Extract of nominal group topic guide groups 1-3

<table>
<thead>
<tr>
<th>Part A: Presentation of results from stage 1 interviews and initial discussion – Experience of fatigue</th>
</tr>
</thead>
<tbody>
<tr>
<td>- What is your personal experience of fatigue or tiredness?</td>
</tr>
<tr>
<td>- Are there similarities with the experiences of the patients we have interviewed?</td>
</tr>
<tr>
<td>- Are there any differences?</td>
</tr>
<tr>
<td>- Have we included the most important things about fatigue and tiredness?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Part B: Ranking what matters</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Individual activity: write down your own ‘importance’ list</td>
</tr>
<tr>
<td>- Group activity: sharing your list with the group (write up all lists on Flip Charts)</td>
</tr>
</tbody>
</table>
- Group activity: wider group discussion about the lists and any discrepancies that appear
- Group activity: is it possible to reach agreement on top 6 most important things?

**Part C: Exploring the relevance and acceptability of new items**
- Which questions do you think work well? Why?
- Which questions do you think don’t work so well? Why?
- Are there any questions that you think are missing, that could be important?

**Group 4**

In response to the findings from the first three focus groups, the topic guide for the final focus group was modified into a two-part group session which focused on a deeper exploration of the developing axSpA fatigue and energy specific PROM. This is a similar approach to recent PROM development studies (107,270).

The new topic guide was made up of two parts: Part A remained the same (a discussion and appraisal of the measurement framework); and Part B became an extensive and detailed discussion about the developing first-version PROM. This included a detailed review of a first version of the new questionnaire for a robust appraisal of the current items (including wording, relevance and anything potentially missing), response options, layout options and formatting.

4.5.5. **Recording and transcription**

Each focus group was digitally audio-recorded, transcribed verbatim, and a detailed summary was produced for each focus group. To ensure accuracy of interpretation, the recording was then listened to a second time and checked against the summary document to ensure no key information had been omitted. The purpose of the groups was to review and appraise the measurement framework of fatigue by incorporating the views of the two key user groups: patients and clinicians. Summaries were disseminated to the full research team for group discussion. This discussion included a detailed appraisal of the measurement framework and findings of the nominal groups, any modifications or tweaks suggested and detailed consideration of the developing item list.
4.5.6. **Data analysis**

The purpose of the analysis was to examine the content of the meetings and identify key areas of discussion, including resonance or challenges participants raised in relation to their experiences, the measurement framework or the developing items. The analysis findings would then inform development and refinement decisions to the model and items to support finalising the working measurement blueprint of axSpA fatigue and energy.

Thematic analysis has been defined as (221):

> ‘a method for identifying, analysing and reporting patterns within data’

Braun and Clarke describe a structured, six-stage process for conducting a thematic analysis (221). This process was followed for the analysis of the focus groups:

1. Familiarisation with the data – this may be done through data transcription, reading and re-reading transcripts, listening to meeting recordings and noting initial thoughts.
2. Generating initial codes – systematically coding the entire data set and drawing together data for codes.
3. Searching for themes – grouping codes into potential themes with relevant and supporting data.
4. Reviewing themes – checking the themes for each coded extract from the data set, and then across the entire data set (developing sub themes and themes; thematic map).
5. Defining and naming themes – iterative development and refinement of themes to produce clear definitions and overall thematic story.
6. Producing a report – producing a detailed report that explains the extracted themes, supported by related and relevant quotes taken from the data.

Rigour was demonstrated by following the four criteria proposed by Yardley (241) and described in Chapter 3a which included ensuring immersion in the data, maintaining an audit trail of evidence which is demonstrated by the thematic table and supporting quotes and through maintaining a
reflexive log. All data analysis and findings were discussed with the research team members.

4.5.7. **Patient research partners**
The aim of PPI in this study was to modify and review the co-produced framework underpinning PROM content, and co-produce the new questionnaire (including verbal probes for pretesting interviews) (145). A meeting with the study PRP group was held following completion of the focus groups to discuss the working axSpA fatigue and energy blueprint and developing first version questionnaire. The meeting was completed in two parts: initially, a discussion with PRP members about their own experiences and whether there was resonance or dissimilarity with the findings of the qualitative work. This was followed by a review and detailed appraisal of the working axSpA fatigue and energy blueprint and its content. Finally, the group reviewed the developing questionnaire with members discussing and commenting on the layout, formatting, item ordering and item content. Members were encouraged to make suggestions or changes with all recommendations actively discussed amongst the group before being agreed. I was responsible for actioning these changes and sharing updated documents with the group to confirm they were fulfilled as agreed.

4.6. **Results 1: reviewing the axSpA-fatigue and energy measurement framework**

4.6.1. **Participants**
41 patients were approached to participate in the focus groups. Five patients (3 females; mean age 55.6 years; range 32 to 73 years) with diagnosed axSpA (mean axSpA duration 24.25 years) took part across two focus groups lasting around 2 hours 30 minutes and 3 hours 10 minutes respectively. Each patient participant was allocated a number 1 to 5 to ensure anonymity.

30 healthcare professionals were approached to participate in the focus groups by JM using the ASSIGNw network. Additionally, six professionals were approached by JP in his specialist rheumatology hospital. Overall,
seven healthcare professionals (4 females; mean age 43.23; range 30 to 54 years) who have experience working with axSpA patients or fatigue (experience ranging from 1 to 33 years) participated across two focus groups lasting around 1 hour 30 minutes each. Again, participants were allocated a number from 6 to 12 to ensure anonymity.

All focus groups were arranged at least one month in advance. Unfortunately, attendance to three focus groups (two with patients, one with professionals) was affected by multiple, substantive weather disruptions at the time of hosting; snowfall, ice and national weather warnings. This weather could not be anticipated at the time of arranging the groups and despite best efforts by the research staff responsible for approach, and my own contact with participants, group numbers were low. A final focus group (with professionals) occurred outside the disruptive weather period and was fully attended. Other significant weather events at this time also led to the cancelling and rescheduling of two research team meetings to discuss the focus group findings.

4.6.2. Findings
The analysis is presented in two parts. Part one provides a summary of the thematic analysis of data from Part A (summary discussion and measurement framework review) and Part B (ranking task) – the detailed thematic analysis is provided in Appendix 4B. Table 4.2 reports the identified themes and subthemes. Part two lists key action points to be actioned to develop the first iteration axSpA fatigue and energy specific PROM. The action points are derived from data from Part C (review of candidate items). Table 4.3 lists the final, refined items ready to be taken into pretesting interviews.
Table 4.2: List of themes and subthemes with supporting quotes from thematic analysis

<table>
<thead>
<tr>
<th>Theme</th>
<th>Subtheme</th>
<th>Quotes from focus group meeting</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Resonance and challenges</strong></td>
<td>Financial considerations</td>
<td>HPR group 2: “if they can’t participate in work or have to take off work … there’s a financial impact from that due to reducing hours or changing work”</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HPR group 2: “perhaps money is a sacrifice they have to make, it’s not important it’s just about being able to keep going”</td>
</tr>
<tr>
<td></td>
<td>Use in practice and measurement</td>
<td>HRP group 2: “part of the problem is historically and even currently, the actual measures that we use don’t really factor in the fatigue elements … the fatigue isn’t asked about really”</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HRP group 2: “fatigue needs to be addressed … purposefully asking the questions normalises that it’s a symptom of their AS and we should take it into account … maybe there are some elements we can help with”</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HRP group 1: “with this we can tell whether the people who are not coping are actually depressed, and the ones who are managing are mentally well”</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HRP group 1: “from a clinician point of view there’s no point having a questionnaire if you’re not going to look at it, and it’s just being a [sum] score. We want to be able to see change and use domain scores”</td>
</tr>
<tr>
<td></td>
<td>Meanings ascribed to fatigue and energy</td>
<td>Patient group 1: “I would identify more with the lack of energy than the fatigue”</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HPR group 1: “I think fatigue is mental … and level of energy is the physical side of fatigue”</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HPR group 2: “it’s like energy is a finite source – they have a limited amount of it, or their perception of it is less so everything they do has to be pre-planned”</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HPR group 1: “a lot of people look back at me blankly when I say fatigue as if to say, ‘what do you mean by that?’”</td>
</tr>
<tr>
<td>Daily life and functioning</td>
<td>Making sense of sleep</td>
<td></td>
</tr>
<tr>
<td>---------------------------</td>
<td>------------------------</td>
<td></td>
</tr>
</tbody>
</table>
| **Patient group 1**: “I just don’t sleep well, it’s nothing to do with my AS” | **Patient group 1**: “My condition wakes me up, but I can’t get back to sleep because of other things”  
**HPR group 2**: “we ask about pain related sleep disturbance but not how’s your lack of sleep impacting on your day-to-day function and activity”  
**HPR group 2**: “sleep quality is something that’s really important to ask about”  
**HPR group 1**: “I never know what that actually means that, waking up feeling refreshed” |

| **Cognitive abilities** | **Patient group 1**: “I don’t know whether it’s because of my age because that’s another issue, I find now when I sit down, I doze”  
**Patient group 2**: [with the fatigue] “it’s hard to concentrate”  
**Patient group 1**: “[fatigue] it certainly affects my ability to concentrate … I find now that I can’t read any serious book for more than half an hour before either I’ve dozed off, or I’ve lost the thread… and that worries me as reading is important to me”  
**Patient group 1**: “I have memory issues, but I don’t think they’re significant – I can remember what I did yesterday” |

| **Isolation** | **Patient group 2**: “I think I just want to be left alone”  
**Patient group 1**: “I avoid, at all costs, social occasions I find them totally exhausting”  
**Patient group 1**: [describing isolation] “people say you rescue dogs, but I feel like my dog has rescued me because if I didn’t do those 2 walks a day, I don’t know what I would do” |
| Mental wellbeing | Being active | HPR group 2: “I get a lot of patients with AS fatigue who say that they couldn’t fit their exercise in because they are that drained … then it becomes an uphill battle for us as physio’s then because we are trying to get them to stretch on a regular basis but you’re fighting against the fatigue side of things … tricky balance”

**Patient group 1**: “I do feel heavy although I’m not heavy and that does sometimes make me reluctant to start doing things. I like running … but I have the utmost difficulty going for a run … I don’t know why I have this great difficulty but maybe it comes back to a lack of energy because I don’t feel energised”

**Patient group 2**: “it’s like spoon theory really isn’t it, you sort of get up and you’ve got so many spoons and you know, you’ve got to use those spoons wisely”

| Oversight in practice | HPR group 2: “the psychological aspect in clinic doesn’t really get addressed, is there any avenue there worth exploring for treatment purposes as there’s a lot of anxiety”

| Strain of being misunderstood | HPR group 1: “I think being misunderstood at work is quite common; so, I’ve had patients say… somebody will say to me ‘we’re all tired’”

**Patient group 2**: “I noticed that for me, I have a membership at the gym, but I find excuses all the time not to go but now me and my partner argue all the time”

| Deteriorating mood and loss of control | Patient group 1: “the fatigue affects my mood in the sense that I get a short fuse”

**Patient group 2**: [referring to fatigue] “I couldn’t even talk! You get so frustrated you know that you feel down and you have to do your job! … why does it have to happen to me today when I have to do my job”

**HPR group 1**: “I like the word downhearted but not depressed. I say we need to take depressed out because you can’t diagnose depression … you shouldn’t diagnose yourself or someone else with depression”

| Self-management | Re-energising | HPR group 2: “the number of energy drinks they use really surprised me” [referring to verbal summary]

**HPR group 1**: “I’ve had people eat sugar a lot, sweets if they can’t get drinks… so obviously they crash”
| Learning processes | Patient group 1: “you’ve just got to grit your teeth, you’re in for a long-term condition so you just have to keep going … don’t give in”  
Patient group 1: “I just have to keep going until 8 o’clock because when I sit down, I won’t move”  
Patient group 2: “my life changed after I got a dog … it changed the mood and everything … you feel like someone needs you for something for some reason” |
4.6.3. Part one: Thematic analysis findings

Four themes emerged from focus group analysis: (i) Resonance and challenges; (ii) Daily life and functioning; (iii) Mental wellbeing, and; (iv) Self-management. A summary of key points for each theme is provided below (with respective subthemes) and a thematic table with supporting quotes is provided in Table 4.2.

Resonance and challenges

The theme of resonance and challenges describes the extent to which participants related to or felt at odds with the domains and PRO outcomes described in the measurement framework. This theme consists of four subthemes and is briefly summarised below: (i) financial considerations; (ii) use in practice and measurement; (iii) meanings ascribed to fatigue and energy, and; (iv) making sense of sleep.

(i) Financial considerations were highlighted by professionals as a challenge to the measurement framework, however, patients in focus groups (and previous work described in Chapter 3) did not share this view that difficulties with finance were due to fatigue.

(ii) Suggestions for use in practice by professionals included ensuring the questionnaire was not too long, onerous and they preferred the prospect of domain-level scoring to assist them when using the questionnaire in practice.

(iii) Fatigue was felt to be more internal by participants, and energy related to being active. Professionals found the term fatigue potentially problematic but could not generate a term they felt was more accessible to replace it.

(iv) Sleep was considered by all participants to be problematic, with items likely capturing issues other than fatigue in axSpA, warranting further exploration in pretesting interviews and potential removal.

Daily life and functioning

The theme of daily life and functioning explores possible reasons why axSpA patients may be affected in certain aspects of their lives or functioning because of fatigue, e.g. impaired memory. A brief summary of this theme and its four subthemes is provided below: (i) changing cognitive abilities; (ii) isolation, and; (iii) being active.
Professionals and patients noted the negative impact of fatigue on cognitive function, but patients questioned whether age was important too.

Isolation was linked, by patients, to difficulties communicating and patients described withdrawing socially due to the extra effort required, and energy demands it places on them.

Being active was considered important by patients and professionals with feeling fatigued or low on energy affecting patients’ activity levels. Pain and stiffness were not considered a main reason for reduced activity and it was linked to fatigue.

Mental wellbeing

This theme captures the challenges that fatigue and low energy present both for participants and professionals in routine practice. This is briefly summarised below in three subthemes: (i) oversight in practice; (ii) strain of being misunderstood, and; (iii) deteriorating mood and loss of control.

Professionals raised the issue that mental wellbeing was often overlooked in practice with no current routine measure in use to assess fatigue impact on mental wellbeing. Professionals highlighted this was a common issue raised with them and the psychological wellbeing domain in the measurement framework was welcomed.

Feeling misunderstood was a major issue for patients and professionals recognised this as commonly discussed. Can be a source of frustration and negatively affect relationships and family life.

Deterioration in mood and loss of control was directly influenced by their fatigue. Some linked this to a feeling that they were losing control in life, others felt fatigue was more directly affecting their mood. Professionals challenged using the term ‘depression’ in a measure of fatigue (not diagnosing depression) and preferred the term ‘downhearted’ taken from the EASi-QoL (62).
Self-management

This theme captured some of the processes of self-management that participants engaged in. This was captured in two subthemes and is briefly summarised below: (i) re-energising and (ii) learning processes.

(i) Re-energising was important to improve energy levels and the approaches used in the previous qualitative study (Chapter 3) resonated with patients and professionals. Professionals however did not realise the extreme lengths patients went to re-energise.

(ii) Learning processes captured approaches taken by patients to deal with fatigue and low energy. Trial and error were their main, experiential learning process. Organising, prioritising and balancing activity were also part of their approach.

4.6.4. Part two: Action points for PROM development

Overall comments: Health professionals were very supportive and complimentary of the measurement framework suggesting that lots could be learnt from a measure based on the model. They also felt it was very thorough and that nothing was missing.

(i) Size 14 font to enable ease of reading (recognising some may experience iritis).

(ii) Maximising white space between items to aid reader and avoiding overwhelming them.

(iii) Reducing the length of any final version used in clinical practice to around 20 questions.

(iv) Domain-level scoring (if psychometrically possible).

Language and meaning: The concept of fatigue was not a term easily understood by professionals and there was concern about patient understanding. Similarly, the distinction between fatigue and energy, whilst recognised, risked being conflated together. Professionals accepted use of the term 'fatigue' but some indicated that if fatigue was to be kept, the phrasing should not be ‘my fatigue’ as it implies ownership. Professionals also challenged the use of the word depression in the measurement framework. As the measurement framework is intended to underpin a fatigue-specific PROM, the measure would not be seeking to diagnose depression.
(i) Providing a definition of fatigue at the start of the questionnaire.
(ii) Providing a definition of energy at the start of the questionnaire.
(iii) Removing ownership from fatigue items by removing ‘my’.
(iv) Changing the term ‘depression’ for ‘downhearted’.

**Response scale and options:** All groups discussed their preferred response scales (NRS or Likert scale) and three preferred the Likert scale. One patient group preferred an 11-point numeric rating scale as they felt it could be more detailed. Due to one patient group disagreeing, this issue was taken to the study PRP group to consider and determine.

(i) Discuss response scale options with PRP group members.

**Recall period:** The recall period was subject to much discussion and no consensus was found between groups. Issue was initially taken with the phrase ‘over the past week’ and patients recommended asking ‘on average’. Across groups, a range of recall periods from the past month to the past year were suggested. When conversely, professionals saw value in asking over the last week as this would allow for intervention. This was taken to the PRP group for discussion.

(i) Specify ‘on average, over the past week’.
(ii) Discuss recall period with PRP group members.

**Question suggestions or modifications:** Sleep questions were difficult for the patient participants to discern – moreover, it was recognised as a very complex topic and questions about sleep risked capturing lots of things other than fatigue. Patient groups recommended that energy questions should be neutral or positively phrased and should avoid using terms which might be leading, for example, lack of energy, or low energy. A question about support was introduced by the first patient group and following discussion in the other groups, it was agreed that this was a necessary question that captures what is important to patients and has clinical relevance to professionals, helping them to identify patients who may be isolated and therefore struggling to cope. One patient group felt strongly that questions about socialising related to energy rather than fatigue and this should be reflected in the way the question is posited. The jogging example for an activity question felt
distant for participants, and they felt swimming or walking would be more appropriate examples. To summarise:

(i) Monitor how sleep items are interpreted in pretesting interviews.
(ii) Modify language of energy items to be neutral (e.g. energy levels instead of lack of energy).
(iii) Change social item language to ask about patient energy levels instead of fatigue.
(iv) Retain a suggested support question on being left alone in the questionnaire for pretesting.
(v) Changing examples to questions to better relate to the lives of axSpA patients.

4.6.5. Patient research partner group meeting
A PRP group meeting was held with five members lasting 2.5 hours. A student clinician also asked to participate in the group which JM and I agreed to. The purpose of the meeting was to further explore the distinctiveness between the fatigue and energy concepts, review and appraise the working measurement blueprint and co-produce a first version PROM informed by the item generation and refinement work conducted in focus groups. Three names for the PROM were shared with the group members and they unanimously agreed for it to be called the Warwick Axial Spondyloarthritis faTigue and Energy questionnaire (WASTEd).

Discussing the qualitative findings
The meeting began with a second look at the findings of the qualitative work. I provided a verbal summary of the key findings from the IPA interview study (Chapter 3). Group members felt the summary resonated strongly with their own experience and discussion moved onto how distinct the concepts of fatigue and energy were for the purpose of measurement. Members felt they were distinct and JM (group facilitator) asked for further clarification. Members provided examples stating they have felt completely empty of energy and unable to do things but are wide awake. Conversely, they have also experienced having the energy to do things but feeling tired. This reaffirmed the qualitative conceptualisation of axSpA fatigue and energy and further supported the findings from focus groups with both patients and professionals. Group members also discussed whether age was an important
factor in fatigue experience for them and experiences varied between members. As this did not emerge as a leading modulator in fatigue experience in the qualitative work, I agreed to revisit the analysis to double-check age had not been overlooked.

**Discussing the working measurement blueprint**

Following the initial discussion, members undertook a detailed and robust review of the working axSpA fatigue and energy blueprint generated from focus groups. I began the meeting by asking members to consider each domain, its subdomains and the PRO components associated with them. This started out as an individual appraisal, but it slowly became a group discussion as members described resonance with the model and how it captured their experience. The model was positively received, and no modifications were suggested. However, members did discuss the sleep domain at length and views were mixed. Some felt sleep was important for fatigue the next day, however, others felt axSpA was a bigger factor influencing sleep. As the next phase would be pretesting interviews, members suggested that this domain and its items undergo further, more precise investigation. In addition, the clinician highlighted how important the psychological wellbeing domain would be for them when working in clinic as this is not readily captured, yet her experience was that fatigue was significantly affecting patients in this domain. The members agreed that this was an important part of their experience that needed to be shared. The group noted a language preference for the word "downhearted" when referring to the impact of fatigue on mood as opposed to the word "depression", agreeing with the change identified in focus groups.

The agreed working measurement blueprint of axSpA-fatigue and energy is provided in Figure 4.1. This represents the agreed version following focus groups and PRP feedback. Although sleep had been identified as important, concerns about its usefulness within a measurement context had been consistently challenged. Therefore, the interpretation of sleep items would be tested using verbal probes in pretesting interviews to determine its usefulness in measurement. If sleep items perform poorly, the items and domain would be removed.
Reviewing the questionnaire and developing pretesting interview verbal probes

Review of the first questionnaire began with me presenting two versions of the questionnaire with the same content but different formatting: version 1 was a compact, 2-page questionnaire; version 2 was more spaced out using size 14 font and white space to assist reading (as suggested by focus group participants). The group unanimously agreed on using version 2 stating a preference for the structure being clear and easier to follow. The font size was considered acceptable and spacing between the questions was sufficient, making it easier to read questions. Additional formatting such as the use of lines to divide between questions was well received. The standardisation of the questions was liked by members, and the response options were clear – including on the few questions where they change.

An item-by-item review of the measure was conducted to ensure all important domains and PRO concepts were captured, as items, within the questionnaire. Members found no important concepts were missing, items were all relevant and verbal probes were generated for five items for further exploration in pretesting interviews.

Table 4.3 contains the final item list that forms the WASTEd, including verbal probes co-produced with the PRP group to be used in pretesting interviews.
Figure 4.1: Working measurement blueprint for the WASTEd

### Construct: Fatigue and energy in axial spondyloarthitis

#### Conceptual domains

- **Symptoms**
  - Fatigue
  - Energy
- **Impact**
  - Cognitive
  - Physical
  - Social
- **Function**
  - Sleep
- **Psychological/emotional wellbeing**
  - Mood
  - Anxiety and worrying
  - Sense of self
  - Self-isolation
- **Self-management**
  - Achieving balance
  - Energy
  - Support

#### Subdomains

- Frequency, severity, duration
- Reading, talking, memory, ability to concentrate
- Self-care, eating, exercise
- Participation, usual work, job, leisure activity
- Difficulty getting to sleep, disturbed, feeling when awake
- Downhearted, low mood, feeling down, losing control
- Worry, dread, frustration, stress, unpredictability
- Guilt, embarrassment, helplessness
- Pushing others away, hiding fatigue, invisible illness
- Avoiding over/under-exertion, coping and adapting, learnt expertise, being flexible
- Difficulty starting things, re-energising, declining events to avoid wipe-out, feeling drained
- Delegating responsibilities, friend’s family and colleagues
### Table 4.3: Item list for pretesting interviews with all verbal probes

<table>
<thead>
<tr>
<th>#</th>
<th>Item</th>
<th>Verbal probes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>How often have you felt fatigued?</td>
<td>Response process</td>
</tr>
<tr>
<td></td>
<td></td>
<td>How did you reach that answer?</td>
</tr>
<tr>
<td>2</td>
<td>How severe was the fatigue?</td>
<td>-</td>
</tr>
<tr>
<td>3</td>
<td>How much energy have you had?</td>
<td>-</td>
</tr>
<tr>
<td>4</td>
<td>How often have you felt drained?</td>
<td>PRP - Interpretation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>What is it you feel drained of?</td>
</tr>
<tr>
<td>5</td>
<td>To what extent have your energy levels interfered with your ability to chat to other people? (e.g. friends, family, work colleagues)</td>
<td>Confidence</td>
</tr>
<tr>
<td></td>
<td></td>
<td>How sure are you that it is your energy levels interfering with your ability to chat to other people?</td>
</tr>
<tr>
<td>6</td>
<td>To what extent have your energy levels interfered with your ability to remember things? (e.g. being more forgetful than normal)</td>
<td>Confidence</td>
</tr>
<tr>
<td></td>
<td></td>
<td>How sure are you that it is your energy levels interfering with your memory?</td>
</tr>
<tr>
<td>7</td>
<td>To what extent have your energy levels interfered with your ability to concentrate? (e.g. reading, writing, jigsaws, crosswords)</td>
<td>-</td>
</tr>
<tr>
<td>8</td>
<td>To what extent have your energy levels interfered with your ability to take care of yourself? (e.g. cooking, eating well throughout the day)</td>
<td>Difficulty</td>
</tr>
<tr>
<td></td>
<td></td>
<td>How hard was this to answer?</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PRP – Interpretation and comprehension</td>
</tr>
<tr>
<td></td>
<td></td>
<td>What does ‘take care of yourself’ mean to you?</td>
</tr>
<tr>
<td>9</td>
<td>To what extent have your energy levels interfered with your ability to do physical things?</td>
<td>General</td>
</tr>
<tr>
<td></td>
<td></td>
<td>How did you arrive at that answer?</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PRP – Interpretation and comprehension</td>
</tr>
<tr>
<td></td>
<td></td>
<td>What does ‘physical things’ mean to you?</td>
</tr>
<tr>
<td>10</td>
<td>To what extent have your energy levels interfered with your ability to take part in more demanding exercise of moderate intensity? (e.g. going for a walk, swimming)</td>
<td>PRP – Interpretation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Would you consider the examples provided to be ‘moderate intensity’ exercises for you?</td>
</tr>
<tr>
<td>11</td>
<td>To what extent have your energy levels interfered with your ability to do your usual work? (e.g. at work or at home)</td>
<td>-</td>
</tr>
<tr>
<td>12</td>
<td>To what extent would, or have your energy levels interfered with your ability to do the things you enjoy? (e.g. hobbies, leisure activities)</td>
<td>-</td>
</tr>
<tr>
<td>13</td>
<td>Because of fatigue would you, or have you avoided making plans?</td>
<td>PRP – Confidence</td>
</tr>
<tr>
<td></td>
<td></td>
<td>How sure are you that it is because of fatigue that you avoided making plans?</td>
</tr>
<tr>
<td>14</td>
<td>Because of fatigue would you, or have you cancelled plans? (e.g. a trip somewhere, holiday)</td>
<td>-</td>
</tr>
<tr>
<td>15</td>
<td>Because of fatigue would you, or have you turned down invitations? (e.g. to meet a friend, socialise)</td>
<td>-</td>
</tr>
<tr>
<td>No.</td>
<td>Question</td>
<td>Response process</td>
</tr>
<tr>
<td>-----</td>
<td>----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>------------------</td>
</tr>
<tr>
<td>16</td>
<td>On average, what was the quality of your sleep?</td>
<td>How did you reach that answer?</td>
</tr>
<tr>
<td>17</td>
<td>Have you woken up ready to face the day?</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>Has fatigue made you feel downhearted?</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>Has fatigue made you feel like you have less control in your life?</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>Has fatigue left you feeling worried? (e.g. about being able to cope)</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>Has fatigue left you feeling frustrated?</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>Has fatigue left you feeling overwhelmed?</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>Have you preferred to be alone because of fatigue?</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>Have you felt like the fatigue is invisible to others? (e.g. people do not seem to understand)</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>I have struggled to find the balance to maintain my energy levels (e.g. doing too little or too much).</td>
<td>PRP – Difficulty</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Response process</td>
</tr>
<tr>
<td>26</td>
<td>I feel I have coped well with fatigue.</td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>Because of your energy levels, have you found it physically difficult to start, or finish doing things?</td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>Because of your energy levels, have you found it mentally difficult to start, or finish doing things?</td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>Have you run out of energy quickly and needed to take a break? (e.g. have a nap or rest)</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>Have you used stimulants to boost your energy? (e.g. coffee, sugary drinks or foods)</td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>Because of your fatigue, would you, or have you had to ask friends and/or family to do things for you?</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>Would you, or have you been able to share your experience or feelings of fatigue with someone? (e.g. friends, family, healthcare professional)</td>
<td></td>
</tr>
</tbody>
</table>
4.7. **Method 2: pretesting the WASTEd**
NHS ethical approval was granted following review by a research ethics committee (REC reference: 16/WM/0147).

4.7.1. **Aims and objectives**
To pretest the new WASTEd PROM (using two pretesting methods) to ensure item clarity, comprehensibility, precision and their acceptability to respondents.

Objectives:

1. To determine whether the concepts of fatigue and energy are interpreted as distinct concepts by respondents at the point of questionnaire completion.
2. To review items comprehensiveness, relevance, acceptability, comprehension and precision (measuring one concept).

4.7.2. **Justification for approach**
Pretesting interviews are the final moment for significant PROM revision (126) and provide an opportunity for the practical properties of the developing PROM and its items to be assessed for comprehensiveness, relevance, acceptability and comprehension. By investigating these practical properties, the goal of pretesting interviews is to improve the design and use of a developing PROM by investigating potential sources of misunderstanding for respondents, identifying errors and reduce the likelihood of missing data. Current guidance from ISOPOR recommends the use of pretesting interviews (in the form of cognitive interviews) to understand the comprehensiveness, relevance, acceptability and comprehension of the questionnaire by exploring different aspects of it, such as the instructions, item content, response options and recall period (276).

Pretesting interviewing is underpinned by a four-stage cognitive process model proposed by Tourangeau (1984) that describes how respondents understand, retrieve, judge and respond to a question (129). These processes can occur consciously or unconsciously – that is, automatically and outside of the respondent’s awareness. The four stages of the cognitive process model (129) are:

1. **Comprehension**
   a. What do respondents think the question is asking (question intent)?
b. What do certain terms or phrases within a question mean to respondents (meaning of terms)?

2. Retrieval
   a. What information needs to be recalled for the respondent to be able to answer (recallability of information)?
   b. How does the respondent retrieve response information (recall strategy)?

3. Judgement
   a. Are respondents actively attempting to answer the question (motivation)?
   b. Does the respondent want to be truthful in their answer or do they want to give a perceived socially desirable response (sensitivity and social desirability)?

4. Formulating a response
   a. Can the respondent match their internal answer to a question to an available response category (response mapping)?

Other models describing the respondent item-response process have been developed (277–279) including one model specifically designed for appraising Quality of Life (QoL) questionnaires (280). However, no model could be found that was developed specifically for self-reported measures or PROMs. Additionally, three of the alternative models are largely similar to Tourangeau’s model but provide less detailed descriptions for each of the stages (281). The QoL-specific model focuses more closely on the cognitive processes that are engaged and used to respond to an entire QoL questionnaire (280). Given the large similarity between models, and the lack of evidence to indicate any greater value of their use in PROM development, the Tourangeau model was used.

4.7.3. Techniques in pretesting interviews
There are two techniques commonly used in pretesting interviews to assess the cognitive processes of participants: think aloud and verbal probing (Table 4.4). The first technique ‘think aloud’ is a respondent-led technique whereby the participant talks aloud whilst completing items to share information about their thoughts. The purpose of think aloud is to get participants to verbalise otherwise unobservable cognitive processes.
The second technique is verbal probing – an interviewer-led technique whereby participants can be asked probe questions about their decisions or thinking following completion of a question. The purpose of probing is to target specific stages in processing system (described by Tourangeau) and investigate specific stages such as judgement, and subcomponents such as motivation (130), to check questions perform as intended. Two types of probing are described: concurrent and retrospective probing. These probing types relate to when probes should be administered within the interview process. Probing can be integrated throughout an interview (concurrent probing) following completion of an item or group of items. Alternatively, probes may be retrospectively administered upon conclusion of the pretesting interviews (retrospective probing) (130).

In practice, this means that probing can be standardised prior to interview, informed during interview (e.g. in response to comments made during ‘think aloud’), or administered in response to respondent behaviours exhibited during interview (e.g. not answering an item). This creates a flexible approach that can support targeted review at an item level as well as afford the interviewer flexibility to probe on new unexpected issues as they arise.

Table 4.4: Advantages and disadvantages of using ‘think aloud’ and ‘verbal probing’ (130)

<table>
<thead>
<tr>
<th>Think aloud</th>
<th>Verbal probing</th>
</tr>
</thead>
</table>
| **Advantages** | - Free from interviewer bias  
- Minimal interviewer training requirements  
- Open-ended format | - Interviewer maintains control  
- Investigative focus  
- Training participants is not difficult to do |
| **Disadvantages** | - Participant needs to be trained  
- Participant may lack proficiency in TA  
- Burden on participant  
- Risk of participant digressing from task or topic  
- May not generate useful TA streams | - Potential for reactivity (participants may respond differently than normal due to probing)  
- Potential for probes to bias participant responses  
- Interviewer needs to be carefully trained |
4.7.4. Cognitive interviews

To date, cognitive interviewing has been the primary approach taken to pretesting questionnaires during their development (105,107). However, there is little information on how compatible cognitive interviewing is for appraising self-reported questionnaires such as PROMs. Cognitive interviewing was initially developed to support identification of item-level problems in interviewer-administered questionnaires (130). The interview process would typically follow a question-probe format, which in the context of questionnaire appraisal, lends itself to very precise item-level review but may offer a much less detailed insight into challenges relating to, or the acceptability of the wider PROM. This is in part due to the fragmented completion of questionnaire (item-by-item) rather than as a full questionnaire which can dilute any interactions between the items, and between the respondents’ answers. For example, in my study, separate items on making and cancelling plans were developed due to their perceived distinction by patients in focus groups. However, any similarities or differences between the items may be more pronounced and observable to respondents during completion of the full questionnaire compared to a question-probe cognitive interview format. Consequently, the way in which the cognitive interview is conducted could have significant implications for the types of problems identified, and the approach taken to rectify those problems. Such a per-item cognitive interview approach can therefore provide substantial opportunity to focus on, identify and remedy item-level issues but potentially at the expense of a wider, more encompassing review of the wider questionnaire. A method has since been developed specifically for pretesting self-report questionnaires – the three-step test interview (TSTI). Similarly to cognitive interviewing, this approach incorporates think aloud and verbal probing techniques, but the procedure is much more structured and allows for a full PROM completion and review, contrasting with the typical cognitive interview pretesting approach taken in PROM development (105,107). A combination of cognitive interviews to focus on conceptual and specific item-level issues coupled with the TSTI approach could produce a much more thorough and robust pretesting of the WASTEd.

4.7.5. Three-step test interview (TSTI)

The TSTI method is an approach to questionnaire pretesting specifically designed for self-completion questionnaires (131). Its authors describe the aim of TSTI as (282):
‘to produce observational data on actual response behaviour of respondents who respond to a self-completion questionnaire’

To do this, TSTI utilises both ‘think aloud’ and ‘verbal probing’ techniques delivered in a structured format to enable pretesting of the questionnaire under more natural completion conditions. The three steps of the TSTIs (131,282) are:

1. **Respondent-led:** concurrent think aloud with interviewer making observational notes. Observational notes may consider hesitation, skipping items, altering response.

2. **Interviewer-led:** focused interview addressing gaps or points identified during stage one. Observational notes form the basis for verbal probing.

3. **Interviewer-led:** semi-structured interview to explore respondent experiences and their opinions about the questionnaire. This stage can be applied flexibly depending on the research needs and can be used for detailed discussion about the respondent’s response decisions or seek their opinion on possible improvements.

4.7.6. **Chosen pretesting approach in this study**

Cognitive interviews can be used as a precise pretesting method, allowing me to conduct detailed, targeted reviews of specific items at great depth. Within my work, this type of precision would enable me to better ascertain whether conceptual differences between fatigue and energy are relevant and meaningful during questionnaire completion; further confirming, or challenging, the relevance of the concepts for measurement. However, its limitations include the use of frequent probing throughout which could be leading to participants; moreover, there is a risk of administering too many verbal probes and biasing responses or creating problems that may not exist (130). Further, this method will provide very limited insight into more global, questionnaire-wide problems. Therefore, only the first round of pretesting interviews used cognitive interviewing.

The second round of pretesting interviews used the TSTI method. This addressed the gaps of a ‘cognitive interview only’ approach, providing a more holistic review of the developing PROM under more natural completion conditions, whilst minimising potential interference by me (e.g. leading participants with repeated probes throughout the completion process).
4.7.7. Sampling

Guidance for pretesting interview sample sizes varies with upper limits suggested for cognitive interviews of 7-10 participants per item (130). In PROM development, smaller sample sizes of five participants per interview round have been reported. For example, a recent body of work to develop a stiffness PROM in rheumatoid arthritis interviewed 11 participants to produce a final item list for psychometric evaluation (107). Another pretesting study to develop a pressure ulcer risk assessment measure used think aloud interviews in three rounds of four participants (283) alongside pretesting focus groups. This study found most issues in the first interview round, minor problems in a second round and very minimal changes were made in a third round. As per recent practice in PROM development studies, a minimum of two rounds of interviews with at least 5 participants each was sought (107,124).

4.7.8. Participants

A purposive sample of participants with axSpA were identified and approached by a member of their local rheumatology at two participating sites. Patients who were interested in participating were contacted by me. I explained the study and arranged the pretesting interviews to suit the participants’ schedules. I also approached patients who had previously expressed interest in previous interviews and focus groups but were unable to participate.

Pretesting interviews were conducted in two rounds: Round 1 used cognitive interviewing methods and its primary purpose was to confirm that fatigue and energy were interpreted as separate constructs, by respondents, at the point of PROM completion. Round 2 used the TSTI method for a more holistic pretesting approach to consider the full PROM for its practical properties, but also explore formatting, layout and presentation. Separate interview guides were developed for each round and are provided in the procedure section below.

4.7.9. Methods and data collection

Patients were given a study cover letter and information sheet by clinical staff, and if interested in the study provided permission to be contacted by me. Willing participants were contacted by me and interviews were arranged to suit the schedule of participants – typically to coincide with appointments or weekly AS groups hosting at their local rheumatology clinic. All interviews were conducted in a
private room within participants local rheumatology department. Written informed consent was taken prior to the interview beginning and all interviews were digitally recorded and transcribed verbatim.

Procedure 1 – round 1 cognitive interview

Evidence from cognitive psychology suggests that individuals can retain up to five (+ or − 2) pieces of information in their short-term memory (284). Therefore, to minimise memory burden on participants and maximise their ability to speak accurately and confidently about their thought processes, the PROM – consisting of 32 items – was divided into six blocks: five blocks of six items and one block of eight. Blocks were colour-coded to aide participants in identifying the end and start of new item blocks. During an item block, participants were asked to think aloud. Once the item block was completed, participants were then verbally probed by me.

Interviews began with me explaining the purpose of the interview and the unique elements to cognitive interviews: think aloud and verbal probing (see Appendix 4C for topic guide). Training participants in think aloud is recommended (130) but is not reported in pretesting studies. I used a brief training exercise to introduce participants to thinking aloud and to give them insight into how that information assists the study. The training question was: “Try to visualise the place where you live and think about how many windows there are in that place. As you count up the windows, tell me what you are seeing and thinking about.” (130). Participants were asked to think aloud whilst completing a block of items, which would then be followed by interviewer-administered standardised probes (co-produced by me and the PRP group) and occasionally reactive probes based on my in vivo observations. This process was repeated six times until all blocks of items were completed. Upon conclusion of the interview, participants were then asked to complete BASDAI (76), BASFI (77) and HADS (238) assessments to ensure a range of experiences were expressed.

Procedure 2 – round 2 TSTI

The TSTI process was followed (131) as described in section 4.7.5 and a topic guide is provided in Appendix 4D.
Interviews began with an introduction and think aloud training exercise as described in procedure 1, followed by the three stages of the TSTI approach. The interviewer had an annotated copy of the PROM being completed by participants, containing additional notes of important points raised in previous interviews and systematic probes. All interviewer notes and observations were made on the interviewers copy. Interviews closed with participants completing the BASDAI, BASFI and HADS questionnaires.

4.7.10. Data analysis and modifying the PROM

There is no one agreed approach to analysing pretesting interview data, with multiple approaches to analysis frequently being used: including but not limited to listening to audio recordings only, using notes only, producing reports for each participant or producing a group-level report (130, 285). The Cognitive Interview Reporting Framework (CIRF) was developed to assist with comprehensive reporting of cognitive interview studies and was used as a reporting guide in my study (286).

What participants say as part of think aloud, and in response to probes, per item, is important to identify problems and inform revisions to the PROM. Therefore, summaries were generated for each participant per item as per another study (285). For cognitive interviews, each summary consisted of two sections: a think aloud segment per item, and a verbal probing segment per item. This allowed for the thought processes of participants to be considered alongside responses to probes, minimising the risk of diluting the distinct data each technique provides into a more generic summary.

For TSTIs, summaries were crafted for each of the three stages of the interview. Drawing on guidance for conducting and ensuring transparency in cognitive interview studies (89, 126), data sheets were crafted to structure the findings from the pretesting interviews. Each data sheet listed the items, the issues raised, referenced verbatim quotes to support the points raised and potential changes to resolve the issues raised. The information sheets were abbreviated and simplified for team and PRP group review to maximise engagement and discussion during the meetings.

The reading level of the PROM was assessed using the Flesch Kincaid Reading level, which assesses the number of syllables per word and the number of words in a sentence (287).
Modifying the PROM

Currently there is no agreed guidance on how to formalise modifications to a PROM following pretesting (285). Previous studies have used different approaches to analysis such as thematic analysis (288) – which has been recommended (130) – developing their own coding framework (289,290) or using Tourangeau’s cognitive process model (129) as an analysis framework (291). Another study took a more transparent, structured approach by drawing on the Questionnaire Appraisal System (QAS-99 (292)) categories to act as a framework for data coding and to inform item-refinement decisions (285). The QAS-99 is:

‘designed to assist questionnaire designers in evaluating survey questions, and in finding and fixing problems, before the questions ‘go into the field’ (292)

The checklist consists of eight leading categories which can be used to review and appraise items within a questionnaire prior to pretesting interviews. However, researchers are increasingly using the categories as a framework for analysing pretesting interview data. For example, in a study using cognitive interviews to evaluate fatigue items developed for the PROMIS initiative, the QAS-99 categories were used to structure interview findings (285).

The current pretesting study used the QAS-99 categories to identify the issue with each item and inform changes (292). These categories were further supplemented by FDA guidance on exploring variability (89) therefore, an additional category was added to the coding framework. Categories 1 and 2 of the QAS-99 are focused on whether the interviewer can read questions uniformly to participants, and any instructions in the questionnaire. As the focus of the pretesting interviews was to focus on the conceptual distinction between fatigue and energy from the patient perspective, these categories were not used.

Using the QAS-99 categories, each item was assigned a code (Table 4.5) which indicated the challenges raised in relation to each item. This captured issues with comprehension, item interpretation, appropriateness of response options and was summarised across participants to home in on the important, underlying challenges stemming from each item. To ensure transparency of process, all reasons for modifications and any decisions made were recorded in a spreadsheet, providing an audit trail (89). Decisions to modify, remove or add new items were made on a per-item basis following review by the team, and PRP groups. Decisions were informed
by a synthesis of comments and feedback by the two groups with no one group unilaterally able to edit the measure. During team meetings, I would press the case made from patient interviews and represent the views of the PRP. Conversely, during PRP group meetings, I would represent the view formed in the team meeting and seek to find a practicable way forward that best reflected the findings from the qualitative work.

Table 4.5: Coding framework for analysis pretesting data adapted from QAS-99 and including an FDA recommendation (89,292)

<table>
<thead>
<tr>
<th>QAS-99 categories</th>
<th>Subcategories</th>
<th>Challenges</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Clarity</td>
<td>a. Wording</td>
<td>Lengthy question, poor grammar, complicated syntax, awkward to read</td>
</tr>
<tr>
<td></td>
<td>b. Technical terms</td>
<td>Complex, lacks definition or clarity</td>
</tr>
<tr>
<td></td>
<td>c. Vague</td>
<td>Multiple interpretations making response difficult</td>
</tr>
<tr>
<td></td>
<td>d. Reference periods</td>
<td>Missing, poorly specified or conflicting</td>
</tr>
<tr>
<td>2. Assumptions</td>
<td>a. Inappropriate assumptions</td>
<td>Question inappropriately assumes something of the respondent</td>
</tr>
<tr>
<td></td>
<td>b. Assumes constant behaviour</td>
<td>Fails to recognise that situations vary</td>
</tr>
<tr>
<td></td>
<td>c. Double-barrelled</td>
<td>Asks more than one question of the respondent</td>
</tr>
<tr>
<td>3. Knowledge/ Memory</td>
<td>a. Knowledge</td>
<td>Respondent may not know an answer</td>
</tr>
<tr>
<td></td>
<td>b. Recall</td>
<td>Respondent may not be able to recall the information</td>
</tr>
<tr>
<td></td>
<td>c. Computation problem</td>
<td>Difficult mental calculations affecting responses</td>
</tr>
<tr>
<td>4. Sensitivity/ Bias</td>
<td>a. Sensitive content (general)</td>
<td>Embarrassing or private question for respondents</td>
</tr>
<tr>
<td></td>
<td>b. Sensitive wording (specific)</td>
<td>Question should be worded to minimise sensitive responses</td>
</tr>
<tr>
<td>c. Socially acceptable</td>
<td>Implied response by the question</td>
<td></td>
</tr>
<tr>
<td>------------------------</td>
<td>----------------------------------</td>
<td></td>
</tr>
</tbody>
</table>

5. Response categories

<table>
<thead>
<tr>
<th>a. Open-ended</th>
<th>Difficult or inappropriate</th>
</tr>
</thead>
<tbody>
<tr>
<td>b. Mismatch</td>
<td>Do response options match the question</td>
</tr>
<tr>
<td>c. Technical terms</td>
<td>Complex, poorly defined or unclear language</td>
</tr>
<tr>
<td>d. Vague</td>
<td>More than one interpretation for a given response option</td>
</tr>
<tr>
<td>e. Overlapping</td>
<td>Response options are not distinct from one another (conflated)</td>
</tr>
<tr>
<td>f. Missing</td>
<td>Categories that should be affirmed are missing data</td>
</tr>
<tr>
<td>g. Illogical order</td>
<td>Categories should be logically ordered</td>
</tr>
</tbody>
</table>

6. Other problems

| a. Other             | Problem other than those defined |

7. Variability in responses

| a. Inactive response options | One or more response options not being endorsed by respondents. |

4.7.11. Review and refinement meetings
Analysis was iterative and findings from the first round were discussed in both team and PRP group (128,130,293), then operationalised ready for review in the next round of pretesting interviews. I generated summaries and discussed them with ET to ensure sufficient detail was captured in the files. All decisions relating to PROM modification were made following research team and PRP group review and actioned by me. These changes were then returned to the team and group by email clearly documenting the points raised in the meetings and how they had been addressed – the team and group members were asked to confirm that modifications were what had been agreed. The final long-form version of the PROM was taken to the PRP group for their agreement to proceed with the co-produced measure.

The aim of PPI in the pretesting interview study was to co-produce and refine the WASTEd, including co-producing additional items and verbal probes for the second round of pretesting interviews (145).
4.8. Results 2: pretesting the WASTEd

Ten male patients diagnosed with axSpA participated in two rounds of pretesting interviews: five in cognitive interviews (round 1) and five in TSTIs (round 2). Participants were aged between 28 and 75 years (mean age 52.8 years). Patient characteristics across key variables are presented in Table 4.6. Participants were interviewed in their local rheumatology clinic at two sites in Stoke-on-Trent and Wigan. Interviews lasted between 33 minutes and 2 hours (mean duration approximately 1 hour 20 minutes). Prior to cognitive interviews, reading ease was calculated at 75.7, grade level 4.3 which indicates the questionnaire is fairly easy to read (287).

Table 4.6: Characteristics of pretesting interview participants

<table>
<thead>
<tr>
<th>ID</th>
<th>BASDAI</th>
<th>BASFI</th>
<th>HADS depression</th>
<th>HADS anxiety</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5.7</td>
<td>7.6</td>
<td>8</td>
<td>22</td>
</tr>
<tr>
<td>2</td>
<td>5.3</td>
<td>8.4</td>
<td>8</td>
<td>16</td>
</tr>
<tr>
<td>3</td>
<td>2.1</td>
<td>1.6</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>5.9</td>
<td>5.7</td>
<td>2</td>
<td>9</td>
</tr>
<tr>
<td>5</td>
<td>2.8</td>
<td>6.8</td>
<td>7</td>
<td>15</td>
</tr>
<tr>
<td>6</td>
<td>2.1</td>
<td>0.7</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>7</td>
<td>3.1</td>
<td>2.7</td>
<td>12</td>
<td>27</td>
</tr>
<tr>
<td>8</td>
<td>7.6</td>
<td>6.1</td>
<td>4</td>
<td>13</td>
</tr>
<tr>
<td>9</td>
<td>2.4</td>
<td>1.1</td>
<td>11</td>
<td>19</td>
</tr>
<tr>
<td>10</td>
<td>5.8</td>
<td>7.9</td>
<td>9</td>
<td>13</td>
</tr>
</tbody>
</table>

Analysis was conducted iteratively following each round, therefore, each round and the findings pertaining to those interviews are outlined separately below. Research team meetings and PRP groups were held after each round of interviews. An item list is provided in Table 4.7 to assist with the reading of the results.

Table 4.7: Reference 32-item list tested in round 1 cognitive interviews

<table>
<thead>
<tr>
<th>#</th>
<th>Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>How often have you felt fatigued?</td>
</tr>
<tr>
<td>2</td>
<td>How severe was the fatigue?</td>
</tr>
<tr>
<td>3</td>
<td>How much energy have you had?</td>
</tr>
</tbody>
</table>
4. How often have you felt drained?

5. To what extent have your energy levels interfered with your ability to chat to other people? (e.g. friends, family, work colleagues)

6. To what extent have your energy levels interfered with your ability to remember things? (e.g. being more forgetful than normal)

7. To what extent have your energy levels interfered with your ability to concentrate? (e.g. reading, writing, jigsaws, crosswords)

8. To what extent have your energy levels interfered with your ability to take care of yourself? (e.g. cooking, eating well throughout the day)

9. To what extent have your energy levels interfered with your ability to do physical things?

10. To what extent have your energy levels interfered with your ability to take part in more demanding exercise of moderate intensity? (e.g. going for a walk, swimming)

11. To what extent would, or have your energy levels interfered with your ability to do your usual work? (e.g. at work or at home)

12. To what extent would, or have your energy levels interfered with your ability to do the things you enjoy? (e.g. hobbies, leisure activities)

13. Because of fatigue would you, or have you avoided making plans?

14. Because of fatigue would you, or have you cancelled plans? (e.g. a trip somewhere, holiday)

15. Because of fatigue would you, or have you turned down invitations? (e.g. to meet a friend, socialise)

16. On average, what was the quality of your sleep?

17. Have you woken up ready to face the day?

18. Has fatigue made you feel downhearted?

19. Has fatigue made you feel like you have less control in your life?

20. Has fatigue left you feeling worried? (e.g. about being able to cope)

21. Has fatigue left you feeling frustrated?

22. Has fatigue left you feeling overwhelmed?

23. Have you preferred to be alone because of fatigue?

24. Have you felt like the fatigue is invisible to others? (e.g. people do not seem to understand)

25. I have struggled to find the balance to maintain my energy levels (e.g. doing too little or too much).

26. I feel I have coped well with fatigue.

27. Because of your energy levels, have you found it physically difficult to start, or finish doing things?

28. Because of your energy levels, have you found it mentally difficult to start, or finish doing things?

29. Have you run out of energy quickly and needed to take a break? (e.g. have a nap or rest)

30. Are you using stimulants to boost your energy? (e.g. coffee, sugary drinks or foods)

31. Because of your fatigue, would you, or have you had to ask friends and/or family to do things for you?

32. Would you, or have you been able to share your experience or feelings of fatigue with someone? (e.g. friends, family, healthcare professional)

4.8.1. Round 1 cognitive interview results

32 items were evaluated in the first round of cognitive interviews: 13 items remained unchanged; eight items were identified for minor modification (adding or updating examples, tweaking language or response option anchor) and six items were
identified for major changes (changing concept of the item from fatigue or energy, changing meaning of the question). Additionally, six new items were crafted, three existing items (items 25, 27 and 28) were split, and two items (items 22 and 24) were identified for deletion. In total, a pool of 39-items was taken to research team for review. All identified issues and changes made are detailed in Table 4.8 and a summary is provided below.

(i) Six items (items 6, 7, 12, 14, 15 and 26) were identified as referring to the wrong concept and therefore six alternative items were crafted.

(ii) Two items were split into more precise items: item 27 (found it mentally difficult to start or finish doing things) and item 28 (found it physically difficult …).

(iii) One item example (item 25) was considered confusing by participants, indicating it might not be measuring its intended concept. This item was split for review in team and PRP group meetings.

(iv) One item (item 26) was changed from asking whether participants ‘cope’ to ‘manage’ their fatigue. An energy management question was also crafted.

(v) Two items (items 22 and 24) were identified for deletion due to inconsistent interpretation by participants.

Following interview-informed revisions, 39-items were taken through to research team and PRP groups for deeper consideration and further revision.

4.8.2. Research team review

The research team review lasted for three hours and informed modifications to the developing PROM and its items in two ways: (i) through discussion of changes informed by cognitive interviews, and (ii) the ideas and suggestions of the research team. ET and I ensured that changes proposed and accepted did not undermine the qualitative underpinnings of the PROM, and that the measurement framework was respected.

Following team review of the full 39 items: 10 items were left unchanged; nine items underwent minor modifications; four items were subject to major changes and nine items were identified for removal (Table 4.8). Context on how modifications decisions were reached are described below under three headers: deletions, major changes and minor changes.
**Deletions**

Previously raised concerns about the sleep domain were confirmed in cognitive interviews, indicating that the domain and its items (items 16 and 17) were capturing information unrelated to fatigue such as pain, stiffness and restless thinking. This corroborated concerns raised in focus groups, research team and PRP group meetings and therefore, it was unanimously agreed to remove the sleep domain and its two items from the PROM.

There were concerns that four items would not perform well in terms of measurement reliability and responsiveness: two items were previously identified in cognitive interviews (items 22 and 24), and two (items 30 and 32) were highlighted in the team meeting. The latter two items were also identified as potentially lacking relevance to professionals.

Two fatigue items (items 14 and 15) were removed and an energy version of them was retained. This resulted from a consensus within the team that activities fit the energy definition.

One item (item 25) was split following cognitive interviews into two items: (i) doing too much and (ii) doing too little. There was concern that 'doing too little' would attract responses relating to stiffness and pain rather than energy levels and was therefore highlighted for removal.

**Major changes**

One item (item 26) was focused on in discussions, with the team proposing that the term 'cope’ be changed to ‘manage’ to improve accessibility to patients and improve focus on self-management of fatigue. This corroborated the findings in cognitive interviews.

Two items (items 27 and 28) were split following cognitive interviews: (i) difficulty starting physical tasks; (ii) finishing physical tasks; (iii) starting mental tasks, and; (iv) finishing mental tasks. The team review felt the essence of these items was about lacking energy to fulfil mental and physical tasks. As these items are known to be included in the BRAF-MDQ (106), I proposed these items be merged and refocused into two questions: (i) have you lacked physical energy? and (ii) have you lacked mental energy?. The group accepted my proposed change.
One item (item 13) was changed from ‘fatigue’ to ‘energy’ as the wording of the item had a physical connotation, better aligning it to the energy definition (chapter 3).

**Minor changes**

Thirteen items underwent minor changes of which four items (items 6, 10-12) were changed to standardise the language, tense or syntax; three items (items 4, 5 and 19) were changed to improve precision by modifying language to be less vague, and examples were added, modified or removed for six items to assist with interpretation (items 7, 9, 18, 20, 23 and 31). Underlining the concept in each question – fatigue or energy – was recommended to assist respondents when reading questions.

4.8.3. **PRP group review**

The PRP group meeting lasted for around 3 hours and was attended by five PRP members. Almost all suggested changes from cognitive interview findings and team review were accepted by the PRP group. Substantive changes to the structure and layout of the PROM and further reduction of item pairs (i.e. identifying whether a question should refer to energy or fatigue) was recommended. The group also identified items for further scrutiny in TST1 pretesting interviews and new verbal probes were co-produced for these items.

The meeting began with brief discussion of the PROM including a reminder of the measurement framework for reference. To assist with item-by-item review, I suggested members complete the questionnaire themselves to refresh their memory of the items and assist their appraisal of the PROM. Overall, members were happy with the PROMs development and felt the layout and formatting was accessible. However, the mix of fatigue and energy questions was considered jarring for the reader and therefore, members proposed dividing the questionnaire into two sections: Section 1 – Fatigue; Section 2 – Energy with definitions at the start of each section.

One energy item (item 5) was challenged by members who felt that this was instead a fatigue issue. Difficulty in conversations was due to fatigue affecting their ability to mentally engage, not low energy. This reasoning also supported refinement of two previously split items with their energy versions being deleted (items 6 and 7) as the PRP group categorised them as being fatigue-related issues.
Three items (items 11, 31 and 19) were identified for verbal probing in TSTIs. Item 11 was identified as members were unsure whether this item was sufficiently distinct from items on light and demanding activity. Item 31 was identified due to uncertainty about the language in the question and members disagreeing between using ‘dependence’ or ‘seek help’. As no consensus was reached, it was agreed that these would be targeted by further verbal probing in interviews for discussion after round 2. Item 19 was identified for verbal probing due to its lack of precision: members felt that the item was broad and captured more than just fatigue, but felt fatigue was an important contributor to a loss of control. Further verbal probing would be conducted in TSTIs to establish more evidence about this item.

The content of two items (items 13 and 14) was considered important but some members thought a single item would be sufficient, whilst others felt the items should remain separate. In response, the group proposed a third question to attempt to merge these two items together for pretesting in TSTIs. These items would be reviewed in the round 2 PRP group meeting.

Finally, some tweaks to response options and example modifications were suggested and all suggested modifications (from interviews, the research team and PRP group) are summarised in Table 4.8. Following revisions, 29 items remained and formed the PROM for round 2 TSTIs. Reading ease was calculated at 69.7, grade level 5.9 which indicates it is fairly easy to read, albeit slightly more difficult than the previous version (287).
Table 4.8: Summary of identified changes from cognitive interviews (n=5), research team review and PRP group review

<table>
<thead>
<tr>
<th>#</th>
<th>Question</th>
<th>Changes proposed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Cognitive interviews</td>
</tr>
<tr>
<td>1</td>
<td>How often have you felt fatigued?</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>How severe was the fatigue?</td>
<td>Response options &quot;a lot&quot; changed to “very”</td>
</tr>
<tr>
<td>3</td>
<td>How much energy have you had?</td>
<td>Response options &quot;not at all&quot; changed to “none”</td>
</tr>
<tr>
<td>4</td>
<td>How often have you felt drained?</td>
<td>Needs clarity</td>
</tr>
<tr>
<td>5</td>
<td>To what extent have your energy levels interfered with your ability to chat to other people? (e.g. friends, family, work colleagues)</td>
<td>Change syntax “Have you found it difficult to chat to other people because of fatigue?”</td>
</tr>
<tr>
<td>6</td>
<td>To what extent have your energy levels interfered with your ability to remember things? (e.g. being more forgetful than normal)</td>
<td>Split item</td>
</tr>
<tr>
<td>6a</td>
<td>To what extent has fatigue interfered with your ability to remember things?</td>
<td>New version of item 6a</td>
</tr>
<tr>
<td>7</td>
<td>To what extent have your energy levels interfered with your ability to concentrate? (e.g. reading, writing, jigsaws, crosswords)</td>
<td>Split item</td>
</tr>
</tbody>
</table>


<table>
<thead>
<tr>
<th></th>
<th>Question</th>
<th>Change</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>7a</td>
<td>To what extent has fatigue interfered with your ability to concentrate? (e.g. reading, follow a film or TV program)</td>
<td>New version of item 7a</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>To what extent have your energy levels interfered with your ability to take care of yourself? (e.g. cooking, eating well throughout the day)</td>
<td>-</td>
<td>Example</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Prefer 'personal care' than showering or bathing</td>
</tr>
<tr>
<td>9</td>
<td>To what extent have your energy levels interfered with your ability to do physical things?</td>
<td>-</td>
<td>Examples added</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Vague</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Physical “things” not precise enough, “activities” a better word</td>
</tr>
<tr>
<td>10</td>
<td>To what extent have your energy levels interfered with your ability to take part in more demanding exercise of moderate intensity? (e.g. going for a walk, swimming)</td>
<td>-</td>
<td>Wording changed</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Change part to “more demanding physical activities of moderate intensity?”</td>
</tr>
<tr>
<td>11</td>
<td>To what extent have your energy levels interfered with your ability to do your usual work? (e.g. at work or at home)</td>
<td>-</td>
<td>Probe</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Is this similar to item 10?</td>
</tr>
<tr>
<td>12</td>
<td>To what extent would, or have your energy levels interfered with your ability to do the things you enjoy? (e.g. hobbies, leisure activities)</td>
<td>Split item</td>
<td>Delete this item</td>
</tr>
<tr>
<td>12a</td>
<td>To what extent would, or has fatigue interfered with your ability to do the things you enjoy? (e.g. listening to music)</td>
<td>-</td>
<td>Retain this item</td>
</tr>
<tr>
<td>13</td>
<td>Because of fatigue would you, or have you avoided making plans?</td>
<td>Split item</td>
<td>Delete this item</td>
</tr>
<tr>
<td>13a</td>
<td>Would you, or have you avoided making plans in case you might not have the energy to do them?</td>
<td>New version of item 13a</td>
<td>Retain this item</td>
</tr>
<tr>
<td></td>
<td>Question</td>
<td>Action</td>
<td>Action</td>
</tr>
<tr>
<td>---</td>
<td>--------------------------------------------------------------------------</td>
<td>--------------</td>
<td>--------------</td>
</tr>
<tr>
<td>14</td>
<td>Because of fatigue would you, or have you cancelled plans? (e.g. a trip somewhere, holiday)</td>
<td>Split item</td>
<td>Delete this item</td>
</tr>
<tr>
<td>14a</td>
<td>Would you, or have you cancelled plans because you did not have the energy to do them?</td>
<td>New version of item 14a</td>
<td>Retain this item</td>
</tr>
<tr>
<td>15</td>
<td>Because of fatigue would you, or have you turned down invitations? (e.g. to meet a friend, socialise)</td>
<td>Split item</td>
<td>Delete this item</td>
</tr>
<tr>
<td>15a</td>
<td>Would you, or have you turned down invitations because you did not have the energy to go? (e.g. to meet a friend, socialise)</td>
<td>New version of item 15a</td>
<td>Retain this item</td>
</tr>
<tr>
<td>16</td>
<td>On average, what was the quality of your sleep?</td>
<td>Performing poorly</td>
<td>Delete this item</td>
</tr>
<tr>
<td>17</td>
<td>Have you woken up ready to face the day?</td>
<td>Performing poorly</td>
<td>Delete this item</td>
</tr>
<tr>
<td>18</td>
<td>Has fatigue made you feel downhearted?</td>
<td>Reword “Has you felt downhearted because of fatigue”</td>
<td>Examples added</td>
</tr>
<tr>
<td>19</td>
<td>Has fatigue made you feel like you have less control in your life?</td>
<td>Reword “Have you felt like you have less control in your life because of fatigue?”</td>
<td>Reworked</td>
</tr>
<tr>
<td>20</td>
<td>Has fatigue left you feeling worried? (e.g. about being able to cope)</td>
<td>Reword “Have you felt worried about being able to get through the week, because of fatigue?”</td>
<td>Example removed</td>
</tr>
<tr>
<td>21</td>
<td>Has fatigue left you feeling frustrated?</td>
<td>Reword “Have you felt frustrated because of fatigue?”</td>
<td>-</td>
</tr>
<tr>
<td>22</td>
<td>Has fatigue left you feeling overwhelmed?</td>
<td>Delete this item</td>
<td>Delete this item</td>
</tr>
<tr>
<td>23</td>
<td>Have you preferred to be alone because of fatigue?</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>24</td>
<td>Have you felt like the fatigue is invisible to others? (e.g. people do not seem to understand)</td>
<td>Delete this item</td>
<td>Delete this item</td>
</tr>
<tr>
<td>Page</td>
<td>Original Text</td>
<td>Action</td>
<td>Replacement Text</td>
</tr>
<tr>
<td>------</td>
<td>-------------------------------------------------------------------------------</td>
<td>----------------------------------</td>
<td>---------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>25</td>
<td>I have struggled to find the balance to maintain my energy levels (e.g. doing too little or too much).</td>
<td>Create two replacement items and delete this item</td>
<td>-</td>
</tr>
<tr>
<td>25</td>
<td>I have struggled to maintain my energy levels because I have done too little.</td>
<td>-</td>
<td>Delete this item</td>
</tr>
<tr>
<td>25a</td>
<td>I have struggled to maintain my energy levels because I have done too much.</td>
<td>-</td>
<td>Retain this item</td>
</tr>
<tr>
<td>26</td>
<td>I feel I have coped well with fatigue.</td>
<td>Introduce new item</td>
<td>Reword to manage &quot;I have been able to manage my fatigue.&quot;</td>
</tr>
<tr>
<td>26a</td>
<td>I have been able to manage my energy levels.</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>27</td>
<td>Because of your energy levels, have you found it physically difficult to start, or finish doing things?</td>
<td>Create two replacement items and delete this item</td>
<td>Replacement item for 27, a and b Have you lacked physical energy?</td>
</tr>
<tr>
<td>27a</td>
<td>Because of your energy levels, have you found it physically difficult to start doing things?</td>
<td>-</td>
<td>Delete this item</td>
</tr>
<tr>
<td>27b</td>
<td>Because of your energy levels, have you found it physically difficult to finish doing things?</td>
<td>-</td>
<td>Delete this item</td>
</tr>
<tr>
<td>28</td>
<td>Because of your energy levels, have you found it mentally difficult to start, or finish doing things?</td>
<td>Create two replacement items and delete this item</td>
<td>Replacement item for 28, a and b Have you lacked mental energy?</td>
</tr>
<tr>
<td>28a</td>
<td>Because of your energy levels, have you found it mentally difficult to start doing things?</td>
<td>-</td>
<td>Delete this item</td>
</tr>
<tr>
<td>28b</td>
<td>Because of your energy levels, have you found it mentally difficult to finish doing things?</td>
<td>-</td>
<td>Delete this item</td>
</tr>
<tr>
<td></td>
<td>Question</td>
<td>Action</td>
<td>Note</td>
</tr>
<tr>
<td>---</td>
<td>-------------------------------------------------------------------------</td>
<td>--------------</td>
<td>----------------------------------------------------------------------</td>
</tr>
<tr>
<td>29</td>
<td>Have you run out of energy quickly and needed to take a break? (e.g. have a nap or rest)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>Are you using stimulants to boost your energy? (e.g. coffee, sugary drinks or foods)</td>
<td>Remove item</td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>Because of your fatigue, would you, or have you had to ask friends and/or family to do things for you?</td>
<td>Example added</td>
<td>Is this about dependence?</td>
</tr>
<tr>
<td></td>
<td>“e.g. run errands, do things around the house”</td>
<td></td>
<td>(language from team meeting)</td>
</tr>
<tr>
<td>32</td>
<td>Would you, or have you been able to share your experience or feelings of fatigue with someone? (e.g. friends, family, healthcare professional)</td>
<td>Delete this item</td>
<td></td>
</tr>
</tbody>
</table>
4.8.4. **Round 2 TSTI results**

29 items were evaluated in the second interview round: 24 items were unchanged; three items were identified for minor changes; one item was identified for major changes and one item was identified for removal. One new item was also proposed which focused on coping with fatigue (internal), differentiating it from an item on managing fatigue (practical steps). All identified issues and changes made are detailed in Table 4.9 and a summary is provided below.

(i) **Stage 1 (observation):** Only one item was identified for deletion (item F8). This was previously re-worded in round 1 cognitive interviews, however, verbal probing highlighted that the rewording was unsuccessful in focusing respondents on fatigue.

(ii) **Stage 2 (verbal probing):** One item (item F8) showed a range of interpretations, supporting stage 1 observations. Another item (item 23) was identified for rephrasing – participants disliked the term ‘preference’ (to be alone). Finally, in round 1, an item was changed from asking about ‘coping’ to ‘managing’ fatigue. However, in TSTIs participants noted that a question asking whether are coping with fatigue would be important, therefore highlighting the coping item for reintroduction. Coping was considered distinct from management, referring to internal processes whilst management related to practical steps taken by participants.

(iii) **Stage 3 (semi-structured review):** Three items used ‘completely’ as a response option anchor, but it was identified for change due to being perceived as ‘too absolute’. ‘Extremely’ was suggested as a suitable replacement anchor. Participants also suggested making questions bold instead of response options to better focus attention to the question. Finally, three suggestions were made for the instructions: 1) shorten length for clarity; 2) provide a justification for the recall period, and; 3) better distinguish between the fatigue and energy sections.
4.8.5. Research team review
An extensive 1.5-hour review was conducted with the research team (all members present). Recognising the suggested changes by respondents to some response options, it was agreed to standardise all response options and anchors throughout the questionnaire. For example, for the fatigue severity question the options were changed from ‘not at all’ through to ‘extremely’ to ‘not at all severe’ through to ‘extremely severe’. This was intended to add even greater clarity to responses for participants and add consistency throughout. The reintroduced item on coping was flagged by the team who asked whether this was important, however, due to it emerging in TSTIs following prior removal, I decided to retain the item for review in PRP.

4.8.6. PRP group review
A PRP group was held with 5 members lasting around 2-2.5 hours. A thorough discussion of the measure began with a detailed consideration of the front-page instructions. Members liked the text but proposed shortening it even further still, and more clearly separating the descriptions of each section by listing them. Then, the definitions of fatigue and energy were reviewed, and members suggested removing reference to Axial Spondyloarthritis to further simplify the definitions. Eight items were identified for minor tweaks such as to the language used (e.g. change ‘the fatigue’ to ‘your fatigue’) or selecting more relevant examples (e.g. changing ‘swimming’ for ‘gardening’). Other tweaks included correcting slight formatting differences between items (data not reported). Members supported the reintroduction of the ‘coping with fatigue’ item and some members proposed a new item asking about ‘coping with energy levels’, however, neither I nor some of the members were convinced this was a meaningful item. This item was relayed back to the research team for their input and they too suggested it lacked meaning. Therefore, due to limited endorsement within the PRP group, research team and not being identified in the qualitative work, I decided this item would not be introduced into the measure.

Overall comments from the group members was that they felt very positive about the questionnaire and were happy with how it has developed. They
also commented that it had been their favourite meeting to date, and they had really enjoyed seeing something from beginning to a finished product and looked forward to the next stages.
Table 4.9: Summary of identified changes from TSTIs (n=5), research team review and PRP group review

<table>
<thead>
<tr>
<th>#</th>
<th>Question</th>
<th>Changes proposed</th>
<th>TSTIs</th>
<th>Research team 2&lt;sup&gt;8&lt;/sup&gt;</th>
<th>PRP group 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F1</td>
<td>How often have you felt fatigued?</td>
<td>Response options</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Change “always” to “everyday”</td>
<td></td>
</tr>
<tr>
<td>F2</td>
<td>How severe was the fatigue?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F3</td>
<td>Has your fatigue made it difficult to remember things? (e.g. being more forgetful than normal)</td>
<td>Response options</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Change “completely” to “extremely”</td>
<td></td>
</tr>
<tr>
<td>F4</td>
<td>Has your fatigue made it difficult to concentrate on demanding tasks? (e.g. driving, puzzles, electronic games)</td>
<td>Language</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>“demanding” carries physical connotations – discuss with team and PRP</td>
<td></td>
</tr>
<tr>
<td>F5</td>
<td>Have you found it difficult to engage in conversations with other people because of your fatigue? (e.g. friends, family, work colleagues)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F6</td>
<td>Has your fatigue made it difficult to do the things you enjoy? (e.g. listening to music, watching a TV programme)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F7</td>
<td>Have you felt downhearted because of your fatigue? (e.g. feeling low or down)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<sup>8</sup> Meeting predominantly focused on response options which were standardised, with anchors added throughout.
<table>
<thead>
<tr>
<th></th>
<th>Question</th>
<th>Action</th>
<th>Notes</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>F8</td>
<td>Have you felt you have less control in your life because of your fatigue?</td>
<td>Delete this item</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F9</td>
<td>Have you felt worried because of your fatigue?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F10</td>
<td>Have you felt frustrated because of your fatigue?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F11</td>
<td>Have you preferred to be left alone because of your fatigue? (e.g. not interacting with friends or relatives)</td>
<td>Language</td>
<td>Not a “preference”</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Reworded</td>
<td>“Have you felt the need to</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>be left alone because of</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>your fatigue? (e.g. not</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>interacting with friends or</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>relatives)&quot;</td>
<td></td>
</tr>
<tr>
<td>F12</td>
<td>I feel I have been able to manage my fatigue.</td>
<td>Additional item identified</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F12a</td>
<td>I feel I have been able to cope with my fatigue.</td>
<td>New item (reintroduced)</td>
<td>Flagged Is this really different?</td>
<td>Retain this item</td>
</tr>
<tr>
<td>F13</td>
<td>Have you been more dependent on others (e.g. friends or family) because of your fatigue? (e.g. to run errands, do things around the house)</td>
<td>Example</td>
<td>Include “emotional support”</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Example</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Remove “emotional support”</td>
<td></td>
</tr>
</tbody>
</table>

**Section 2: Energy**

<table>
<thead>
<tr>
<th></th>
<th>Question</th>
<th>Action</th>
<th>Notes</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>E1</td>
<td>How much energy have you had?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E2</td>
<td>How often have you felt drained of energy?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E3</td>
<td>Have your energy levels made it difficult to take care of yourself? (e.g. cooking, personal care)</td>
<td></td>
<td>Example</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Change “cooking” to</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>“showering or “brushing</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>your teeth”</td>
<td></td>
</tr>
<tr>
<td>E4</td>
<td>Have your energy levels made it difficult to do physical activities? (e.g. light housework, making something to eat or drink)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Question</td>
<td>Answer</td>
<td>Note</td>
<td>Example</td>
</tr>
<tr>
<td>---</td>
<td>--------------------------------------------------------------------------</td>
<td>--------</td>
<td>-------------------------------------------</td>
<td>--------------------------------------------</td>
</tr>
<tr>
<td>E5</td>
<td>Have your energy levels made it difficult to take part in more demanding physical activities of moderate intensity? (e.g. going for a walk, swimming)</td>
<td>-</td>
<td></td>
<td>Change “swimming” to “gardening”</td>
</tr>
</tbody>
</table>
| E6 | Have your energy levels made it difficult to do your usual work? (e.g. at work or at home) | Repetition with E7  
Discuss with team and PRP | -      |                                           |
| E7 | Have your energy levels made it difficult to do the things you enjoy? (e.g. hobbies, leisure activities) | Repetition with E6  
Discuss with team and PRP | -      |                                           |
| E8 | Have your energy levels made it difficult to make plans?                  | -      |                                           |                                           |
| E9 | Have your energy levels cause you to cancel plans?                        | -      |                                           |                                           |
| E10| Have your energy levels made it difficult to keep to your plans?          | -      |                                           |                                           |
| E11| Have you ‘turned down’ invitations because of your energy levels? (e.g. to meet a friend, socialise) | -      |                                           |                                           |
| E12| I have struggled to maintain my energy levels because I have done too much. | -      |                                           |                                           |
| E13| I feel I have been able to manage my energy levels.                       | -      |                                           |                                           |
| E13a| I feel I have been able to cope with my energy levels                    | -      |                                           |                                           |
| E14| Have you lacked physical energy?                                          | -      |                                           |                                           |
| E15| Have you lacked mental energy?                                            | -      |                                           |                                           |
| E16| Have you run out of energy quickly and needed to take a break? (e.g. have a nap or rest) | -      |                                           |                                           |
4.8.7. **Final item list used to produce the long form WASTEd**

Following extensive pretesting utilising both cognitive interviews and TSTIs, coupled with the active involvement of the research team and PRPs, a final 30-item long-form PROM was crafted (Table 4.10). This measure is ready for psychometric evaluation in a large population of axSpA patients to support statistically informed refinement of the measure. Reading ease was calculated at 69.2, grade level 6 which indicates it is fairly easy to read, and similar to the previous version (287).

### Table 4.10: Final item list for field-test evaluation

<table>
<thead>
<tr>
<th>#</th>
<th>Question</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>Section 1: Fatigue</strong></td>
</tr>
<tr>
<td>F1</td>
<td>How often have you felt fatigued?</td>
</tr>
<tr>
<td>F2</td>
<td>How severe was your fatigue?</td>
</tr>
<tr>
<td>F3</td>
<td>Has your fatigue made it difficult to remember things? (e.g. being more forgetful than normal)</td>
</tr>
<tr>
<td>F4</td>
<td>Has your fatigue made it difficult to concentrate on demanding tasks? (e.g. driving, puzzles, electronic games)</td>
</tr>
<tr>
<td>F5</td>
<td>Have you found it difficult to engage in conversations with other people because of your fatigue? (e.g. friends, family, work colleagues)</td>
</tr>
<tr>
<td>F6</td>
<td>Has your fatigue made it difficult to do the things you enjoy? (e.g. listening to music, watching a TV programme)</td>
</tr>
<tr>
<td>F7</td>
<td>Have you felt downhearted because of your fatigue? (e.g. feeling low or down)</td>
</tr>
<tr>
<td>F8</td>
<td>Has your fatigue made it difficult to be in control of your life?</td>
</tr>
<tr>
<td>F9</td>
<td>Have you felt worried because of your fatigue?</td>
</tr>
<tr>
<td>F10</td>
<td>Have you felt frustrated because of your fatigue?</td>
</tr>
<tr>
<td>F11</td>
<td>Have you felt the need to be left alone because of your fatigue? (e.g. not interacting with friends or relatives)</td>
</tr>
<tr>
<td>F12</td>
<td>I feel I have been able to manage my fatigue.</td>
</tr>
<tr>
<td>F13</td>
<td>I feel I have been able to cope with my fatigue.</td>
</tr>
<tr>
<td>F14</td>
<td>Have you been more dependent on others (e.g. friends or family) because of your fatigue? (e.g. to run errands, do things around the house for you).</td>
</tr>
<tr>
<td></td>
<td><strong>Section 2: Energy</strong></td>
</tr>
<tr>
<td>E1</td>
<td>How much energy have you had?</td>
</tr>
<tr>
<td>E2</td>
<td>How often have you felt drained of energy?</td>
</tr>
</tbody>
</table>
4.9. Discussion

As per current best practice guidance, a combination of qualitative methods were used to derive a measurement framework of PROs – from the patient perspective – to underpin the development of an axSpA fatigue and energy specific PROM. The final measurement framework incorporated the views and input from a wide group of stake holders including patient participants, PRPs, professionals and researchers, finalising a four-domain measurement framework of symptoms, impact on self (the patient), psychological and emotional wellbeing, and self-management (see chapter 3 for a detailed description of the original framework).

Only one major change to the original measurement framework derived from qualitative interviews was made after focus group and round 1 pretesting review: the removal of the sleep domain. Whilst conceptually important to the patient experience – and therefore, justifiably should be
retained in an overall framework of axSpA-fatigue and energy – for measurement, this domain and its items presented measurement challenges that could not be overcome. These challenges relate to the lack of item and domain specificity to fatigue – risking capturing unrelated information. Otherwise, only minor revisions such as modifying language were proposed for the rest of the measurement framework. PROM mapping informed the development of item content reflective of the PROs captured in the measurement framework – producing the WASTEd PROM. Pretesting the WASTEd supported content development and item revisions, and alongside the active involvement of both the research team and the study patient research partners, a 30-item long form WASTEd was developed ready for psychometric evaluation (Chapter 5).

The following strengths of this study are described below and include the development process (‘PROM mapping’), the methodological advancement of using two approaches to pretesting, sampling strengths and how the WASTEd challenges the relevance and content validity of the PROMIS standardised fatigue measures.

PROM mapping – using PROMs that have been used to assess fatigue in axSpA, an RA-specific PROM (106) and those identified in existing reviews of fatigue measures (150–152) supported the generation of an initial item-set. This meant that the measures used to derive the content of the WASTEd had undergone a form of development and psychometric evaluation. This enhanced the likelihood that these items would be more precise to their intended target concept and could potentially reduce the number of item-level problems that could arise during pretesting, thus reducing the need for item revisions. This process of pooling items from other measures is a recommended approach (59,266). Items for the WASTEd were selected or modified to suit the language of axSpA patients (Chapter 3), better sensitising the item language to that of axSpA patients (112,126). Secondly, the process was well documented, providing transparency in how the initial item set was reached and thus adhering to current best practice recommendations (89,126).

The mixed pretesting approach I used in my study is something I have been unable to identify in any other PROM development study. My study demonstrates that cognitive interviews and TSTIs can be used in two
distinct ways, focusing pretesting toward detailed item-level evaluation (cognitive interviews) and a fuller, holistic review of the measure within more natural self-completion conditions (TSTI). Using two different approaches allowed me to apply focus to items, probing and concept elicitation in cognitive interviews, whilst focusing on non-verbal behaviour, observations, issues arising during self-completion and invite PROM-level discussions with participants when using TSTIs, producing different types of data. It is recommended non-verbal information is attended to in cognitive interviews (130) and this approach enhanced the visibility of non-verbal information and informed some of my *in vivo* verbal probes.

Review of the measurement framework and item set has further confirmed the multifaceted nature of axSpA fatigue and energy as reflected in its confirmed multidimensional measurement framework and item content. The WASTEd contrasts with the PROMIS initiative’s fatigue short-form measures with content ranging from 4 to 13 items. There is evidence that 4, 7 and 8-item short form PROMIS fatigue measures have good evidence of construct validity in RA (294) and the PROMIS fatigue item bank has been used to develop a four-domain (16-item) fibromyalgia-specific fatigue PROM (295), however, there is no evidence of such application or evaluation in axSpA. Three of the short-form PROMIS fatigue measures use item content from the FACIT-fatigue questionnaire and a fourth measure is the 13-item FACIT-fatigue questionnaire (61). The FACIT-fatigue was found to have poor quality and relevance to axSpA patients in a systematic review of fatigue PROM quality (146) indicating these measures are likely to lack relevance for use in this population and the FACIT-Fatigue is known to lack content validity (146) (Chapter 2). This content validity argument is reinforced by the lack of coping or self-management items which emerged as important concepts in the development of the WASTEd. Whilst a short unidimensional measure may have clear benefits in practice such as quick administration and summated scoring, they fail to adequately capture or reflect the multiplicity of experiences of axSpA patients important to support clinical decision making.

Limitations of this study include sampling difficulties and challenges for focus groups and pretesting interviews, and these limitations end with potential issues relating to sample representativeness.
Recruitment challenges in focus groups were predominantly linked to the severe weather conditions which led to national warnings to not drive unless absolutely necessary. However, it is recognised that larger groups within the current recommendation range of 6-8 (275) could have produced a richer diversity of views, potentially influencing the measurement framework. The bottom-up derivation of the measurement framework from a rich and detailed phenomenological framework as per recommendations (123), coupled with the PRP group, minimises the risk that anything important was omitted from the framework. Moreover, focus groups were better attended by healthcare professionals who provided a missing and important perspective on the framework, enhancing the clinical relevance of the developing PROM.

Current sampling guidance for pretesting interviews recommends interviewing between 7-10 individuals per item within the measure (130). This was achieved in this study for most items, however, new items introduced in round 2 TSTIs were only subject to 5 reviews. However, given the small number of new items introduced into the measure, and the extensive involvement of PRP members in also co-producing and reviewing the items, this is not considered to be a significant issue. Similar sample sizes have been reported in other PROM development studies (107,124) and few substantial issues – similar to the current study – were highlighted in round 2. This argument is strengthened by a recent study developing a pressure ulcer risk assessment that found a third round of pretesting interviews identified only three very minor issues – changing a box colour, tweaking the title and modifying one response option to specify no problem/occasionally (283).

Pretesting interviews were representative across disease-specific, activity and mood variables but not gender. Despite substantial time and effort from me and the recruitment teams, it was difficult to identify eligible and willing female patients to participate in the study. As females were well represented in the lived experience interviews, it suggests that this issue is a recruitment-level challenge rather than females not wanting to participate. However, the level of female participation in lived experience interviews which produced the measurement framework, coupled with the high representation of females in the PRP group, ensured that the female
perspective was incorporated into co-production and refinement decisions of the WASTEd.

4.10. Next step
Following the development of the WASTEd, the next step is to elucidate evidence of its internal structure and establish initial evidence of quality (e.g. structural validity, internal consistency). Chapter 5 describes a national field survey to evaluate the WASTEd to support statistical refinement of the questionnaire to produce a good quality short form version of the WASTEd.
5. Chapter 5: Psychometric evaluation of the WASTEd

5.1. Introduction
This chapter describes a preliminary psychometric evaluation of the WASTEd. This was to determine its structure, identify which items are contributing to measurement and perform statistically informed item reduction. Section 5.2 explains the background and purpose of a psychometric evaluation, including underlying measurement theory. The methods to perform a psychometric evaluation are described in section 5.3. Section 5.4 describes the results of the evaluation and the chapter closes with a discussion (section 5.5).

5.2. Background
Psychometrics originates in psychological research and is a methodological discipline (160) consisting of different measurement theories, such as: classic test theory (CTT) and item response theory (IRT) (118,161,166). Psychometrics is typically used in objective measurement of typically unobservable phenomenon, such as personality traits, attitudes or experience of symptoms. In PROM evaluation, psychometrics is used to evaluate the measurement properties of a given measure – for example, its validity, reliability and responsiveness. Whilst there are different psychometric methods, it is advised that both classic (89,118) and modern psychometric methods such as IRT (118) or Rasch Measurement Theory (RMT) (296) are used in PROM evaluation (59).

5.2.1. Field-testing a new PROM
This section outlines the purpose of conducting a psychometric evaluation and then describes field-testing approaches (electronic and postal) and provides some comparative evidence of response rates between approaches.

Current guidance stipulates that once content validity is established (Chapters 3 and 4), other measurement properties such as construct validity, reliability and ability to detect change should be evaluated (89). Determining the quality of a PROM is typically done across two field tests:
field test 1 is a preliminary, cross-sectional evaluation to support statistically-informed item reduction and refinement of the long form PROM to a short form version – this includes establishing evidence of the internal structure of the measure (118), and field test 2 is a comprehensive, longitudinal evaluation of the short form PROM – this includes establishing evidence of reliability (test-retest) and responsiveness to change (59). The current study describes a preliminary psychometric evaluation of the WASTEd.

A preliminary psychometric evaluation provides an opportunity to identify which items are contributing to measurement of the target construct and to elucidate the internal structure of the new PROM. It is recommended that this evaluation should use both CTT approaches and modern psychometric methods to support a more thorough and robust evaluation of the developing PROM (59). Inclusion of modern psychometric techniques as part of the evaluation goes beyond simply defining the structure of the PROM; instead exploring item targeting, ordering of item category thresholds, item fit and differential item functioning amongst other evaluative evidence sources (296). Each theoretical approach to psychometric evaluation (CTT, IRT and RMT) are described in section 5.2.2.

Field tests can be conducted electronically or by post. In a field test to psychometrically assess a new quality of life measure for axSpA – the Evaluation of Ankylosing Spondylitis Quality of Life (EASI-QoL) – a postal survey method was used and achieved a response rate of 64% (62). A similar response rate was reported in two other recent psychometric evaluation studies using postal surveys (106,293), suggesting that a postal survey would be an appropriate method for field testing in this study.

In a randomised study comparing response rates and completeness of data between electronic and paper-and-pencil questionnaires, it was found that response rates were 12.2% higher when using paper-and-pencil questionnaires (297). Moreover, in a comparative study of postal and online survey methods used with physicians, it was found that across three postal surveys (years 1995, 2000, 2005), each achieved a higher response rate compared to completion of an online version of the survey in 2011 (298). Recognising the evidence that response rates are relatively strong in postal
surveys, including evidence from an axSpA study, a postal survey method was adopted in my study. A summary of general advantages and disadvantages of each method are provided in Table 5.1.

Table 5.1: General advantages and disadvantages of postal and online survey methods (299,300)

<table>
<thead>
<tr>
<th>Method</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Postal survey</td>
<td>Can target a large cohort</td>
<td>Non-response</td>
</tr>
<tr>
<td></td>
<td>Hard copies of data available in case of technology failure</td>
<td>Data entry can be slow, and needs to be checked for errors</td>
</tr>
<tr>
<td>Electronic (online)</td>
<td>Lower costs</td>
<td>Non-response</td>
</tr>
<tr>
<td></td>
<td>Access to unique populations</td>
<td>Access difficulties</td>
</tr>
<tr>
<td></td>
<td>Can save time</td>
<td>Technology can fail</td>
</tr>
</tbody>
</table>

5.2.2. Theoretical approaches in measurement science

The main approaches used in questionnaire evaluation are CTT, IRT and RMT. These approaches and their associated statistical methods (relevant to a preliminary psychometric evaluation of a PROM) are described below.

Classic test theory (CTT)

CTT has been widely used in the development and evaluation of measures (see Chapter 2 for more detailed measurement property information). A central tenant of CTT is that observed scores are made up of a ‘true’ score plus error. That is, a patient’s response to an item is their score on the construct being measured (e.g. fatigue) plus measurement error.

Three main assumptions of CTT are: 1) items of a given measure are indicators of the target construct for measurement (representing a reflective model); 2) items are equivalent meaning it is assumed each item equally contributes to the total score, and; 3) homoscedasticity, meaning the error of measurement is the same across the scale (118). As part of a preliminary psychometric evaluation, CTT methods allow for the evaluation of the measures internal structure (structural validity), relationships
between its items (internal consistency) and determine data quality (floor and ceiling effects; missing data).

Structural validity is evaluated using an exploratory factor analysis (EFA). EFA is a data-driven statistical method that elucidates the underlying domain structure within a given data set. This will identify the number of domains within a measure and highlight which items are contributing to the measurement of each domain. If a single domain is extracted, then that indicates that the measure is unidimensional; if two or more are extracted then that suggests the measure is multidimensional.

When conducting an EFA, there are two types of rotation that can be applied to the data: an orthogonal rotation which assumes that extracted domains are uncorrelated, or an oblique rotation which assumes that extracted domains are correlated (161). It has been argued that during exploratory factor analyses, the type of rotation is not important and seeking the simplest domain structure to explain the data is the focus of the analysis, through use of an orthogonal rotation (301). The orthogonal, varimax rotation tends to produce more easily interpreted data solutions (302) with clearly defined domains (166).

Internal consistency is evaluated using Cronbach’s alpha; that is – a scale-level assessment of the interrelatedness between the items of the measure. When internal consistency is high, this indicates that the items are measuring the same concept and when low, this indicates that the items may be measuring different things. Cronbach’s alpha is sensitive to the number of items included in the analysis, with too few or too many items deflating or inflating the alpha score respectively (303).

Data quality can be evaluated through descriptive statistics generated at scale and item level. This includes checking missing data at item level and floor and ceiling effects.

*Item response theory (IRT)*

Item response theory is not a specific technique, but a framework of mathematical models developed to separately estimate parameters for test items, and respondents (118). This means that IRT-based measurement approaches seek to better evaluate measures at item level which contrasts
with the test-level focus of CTT (118). The assumptions of IRT specify that a measure must be unidimensional, and that items must be independent. The use of modern psychometric methods has enabled more robust development and evaluation of PROMs through a more thorough investigation of each item, and its attributes, ensuring meaningful measurements can be obtained (160). IRT approaches can distinguish between the characteristics of respondents and items, meaning specific aspects of the PROM such as item-fit (to the mathematical model), targeting and response option thresholding can be evaluated.

*Rasch Measurement Theory (RMT)*

RMT has been increasingly used in the development of new measures (296,304) and can assist in identifying why items within a given scale may not work as intended. This is done through evaluating the extent to which the observed item scores fit the Rasch mathematical model. There are two assumptions for RMT analysis: 1) the scale is unidimensional, and 2) items are not locally dependent. Some of the key outputs that support PROM development and refinement are described below: this includes local dependency, item-person interactions, person-separation index, individual item fit, category characteristic curves and differential item functioning.

- **Local dependency**: An assumption of RMT is that items are independent – that is, the likelihood of positively responding to an item is unrelated to the likelihood of any other person, with the same level of underlying trait, positively responding to any other item (118). Local dependency is usually assessed through investigating the correlations between items, particularly the residual correlations once the Rasch model has been applied.

- **Class intervals**: This is the number of ability groups that can be distinguished between on the underlying trait. Using RUMM2030 software (305), person locations are ranked and divided into almost equal class intervals sizes.

---

9 Dimensionality can be determined prior to conducting RMT using EFA or CFA. If the measure is multidimensional, each dimension should be separately analysed using RMT.
- **Item-person interactions**: this provides essential evidence about the fit of data to what would be expected on the underlying trait i.e. the data fit to the Rasch model. This is evaluated using a chi-squared test with a significant result indicating that there is some item misfit to the model. Misfitting items are likely assessing more, or something other than the target construct.

- **Person-separation index**: this provides additional information regarding measure reliability: that is, how consistently the Rasch model can separate between \( n \) number of groups on the underlying trait. This is determined by the amount of variance that can be explained by the model, relative to the amount of variance that is due to error.

- **Individual item fit**: this is an assessment of the extent to which items fit the Rasch model. Three sources of evidence can be drawn on to determine fit: (i) fit residuals which can indicate over or under discrimination; (ii) item-trait test of fit (chi-squared); and (iii) visual examination of item characteristic curves.

- **Category characteristic curves**: this is graphical representation, produced for each item, that shows the probability of an individual affirming a particular response option. These should be logically ordered and not be disordered (these are described in the method).

- **Differential item functioning**: this is an assessment of whether different groups of people with the same level of underlying trait, e.g. males and females, answer a given question in the same way, or whether they answer it differently.

5.2.3. **Existing psychometric evaluations of PROMs**

Use of both CTT and modern psychometric methods in PROM evaluation is increasingly advocated for due to the different types of information they provide. However, many recently developed PROMs still rely solely on CTT approaches in the development of new PROMs.

A recently developed QoL PROM for use in coeliac disease was psychometrically evaluated in a postal survey with 412 patients (124). The initial evaluation included determining the structural validity of the measure using EFA and the measures internal consistency which identified six
domains, one of which had questionable internal consistency reliability (124). However, the study does not indicate the number of response options used, nor does it evaluate whether these options are functioning well other than exploring data quality. In addition, it remains unknown whether the items are targeting respondents across the underlying trait – information essential to know how items are contributing to measurement.

A similar approach was taken in the preliminary psychometric evaluation of the BRAF-MDQ in a sample of 299 RA patients (106). Evidence of construct validity, internal consistency and the domain structure was established, but little is known about the characteristics of items and whether they are fit for purpose.

The value of including RMT during the development of a PROM has been highlighted in a Rasch analysis on an already developed and published measure of dementia HRQoL measure (DEMQOL) (304). Given the nature of the disease, a self-report and proxy version (DEMQOL-P) was developed (63), and both were subject to a Rasch analysis (304). The analysis found that item-person threshold locations for both measures could be improved with additional item thresholds at the higher ends of the continuum (304). Response options for five items in the DEMQOL, and four in the DEMQOL-P were disordered, with evidence that the response options were not being used as intended. Whilst all items fit the Rasch model, two of five positively phrased items showed a suboptimal fit to the item characteristic curve. The DEMQOL-P had evidence of differential item functioning for three items across three different patient characteristics. Four item pairs in the DEMQOL, and fourteen in the DEMQOL-P showed signs of local dependency, and neither of the measures were unidimensional (304). The findings of this evaluation demonstrate the significant value that can be gained from using a combination of both CTT and RMT approaches in the development and evaluation of PROMs. Similar applications of RMT to improve measures have been reported in other recent evaluation studies (306–308). This enables these types of problems to be identified early in the development process and remedied.

Recognising the value of using both CTT and modern methods in PROM development, and particularly recent empirical evidence demonstrating how RMT can further inform development decisions (304), I will use both CTT and RMT methods to evaluate the WASTEed. This chapter presents a study
which aims to refine and evaluate the quality and acceptability of the Warwick Axial Spondyloarthritis faTigue and Energy questionnaire (WASTEd) following completion by a large cohort of UK patients with diagnosed axSpA.

5.1.1. Aims and objectives
To refine the initial long-form WASTEd questionnaire, informing construction of the final short-form measure.

Objectives:

1. Evaluate the quality (internal consistency reliability; structural validity) and acceptability (data quality; missing data) of the new measure.
2. Evaluate the measure using both CTT and modern psychometric (RMT) methods.

5.3. Method
5.3.1. Design
A cross-sectional postal survey was conducted across England with participating sites responsible for making patient approaches and conducting mail-outs, including follow-ups. All material needed by sites was printed, packed, stamped and boxed by me prior to the site becoming active, minimising the burden on local research teams.

This study was granted ethical approval by a Research Ethics sub-committee (REC reference: 18/SC/0188) and the Health Research Authority (approval dated: May 17, 2018)

5.3.2. Sample size and eligibility criteria
Sample size calculations were informed by best current practice for psychometric evaluation of new PROMs (118,160,293,296,309). Numbers required for RMT analysis should exceed 250 (310). Evidence from two postal survey studies conducted in axSpA indicated a likely response rate of 60% (62,293). Therefore, I calculated that approximately 420 axSpA
patients needed to be approached to achieve 250 responses to the postal survey.

Eligibility criteria
Participants were included if: (i) they had a primary diagnosis of axSpA (including Ankylosing Spondylitis); (ii) were aged 18+ years, and; (iii) were able to read and write in English.
Participants were excluded if: (i) axSpA or AS was not the patients’ primary diagnosis; (ii) they lacked sufficient proficiency to read and write in English.

5.3.3. Procedure
As per the study ethics, I was not allowed to have access to the patient databases or identifiable information. Therefore, all patient approach and mail-outs were handled locally at participating rheumatology sites. To assist sites in conducting the study, I produced a study process chart which provided a structured process for staff to follow from patient identification through to follow-up and completion.

Patients were identified and purposively sampled (as per the eligible criteria) from existing clinical databases of axSpA patients by nominated staff within their local participating rheumatology department. Eligible patients were assigned a unique identifier which contained two letters (denominating the site) and four digits (denominating the participant number) and sent a study package which contained an invitation letter, participant information sheet and questionnaire booklet. Patients interested in participating were asked to complete the enclosed questionnaire booklet and return it to Nathan Pearson at Warwick Medical School, using the postage-paid envelope provided. I logged all unique ID’s of respondents into a site-log, as they arrived, and updated sites with who had participated and therefore did not require follow-up.

Consent and non-respondents
The study used a process of implied consent. The questionnaire booklet contained a front sheet that explicitly stated that, by completing and returning the documents, the participant was giving their implied consent to
participate in the study. Non-respondents to the study were handled using a three-stage process (outlined below) which began two weeks after the first booklet was sent, as per a previous PROM evaluation study (62) and good practise guidelines (311,312).

i. Stage 1 (2 weeks after baseline): a postcard was sent to participants from their local rheumatology department reminding them to return the questionnaire booklet.

ii. Stage 2 (4 weeks after baseline): a letter was sent explaining the purpose of the study, how to access a new questionnaire booklet (if lost) and to remind them to return their questionnaire booklet.

iii. Stage 3 (6 weeks after baseline): it was assumed the participant did not want to participate, and therefore would receive no further contact in relation to the study.

5.3.4. Materials: the questionnaire booklet
The questionnaire booklet consisted of a front sheet explaining the implied consent process, followed by demographic questions which asked about age, gender, years diagnosed and years experiencing symptoms of axSpA, ethnicity and occupation. The following questionnaires were then presented – these included fatigue-specific (WASTEd; MFI-20), axSpA-specific (BASDAI, BASFI, EASi-QoL), emotional wellbeing specific (HADS) and a generic health status measure (SF-36v2). The questionnaire ended with a study summary request form for respondents who wished to be provided with the results of the study.

The order in which questionnaires were presented was decided with two key considerations in mind: (i) its importance to the current study and (ii) respondent comfort (see Table 5.2). Therefore, the demographic questions, WASTEd and MFI-20 were placed at the front of the booklet and the BASDAI and BASFI at the end – two short questionnaires that respondents would be very familiar with due to their frequent use. To ensure acceptability of the questionnaire booklet to participants, and to allow me and the participating site team to test the proposed study process, three PRP members participated in a test-run. PRP members were asked to provide feedback on the acceptability of the full study package and to time completion of the questionnaires. The PRP members reported no concerns
or issues, and a mean completion time of 33 minutes (range 25-40 minutes).

Table 5.2: Questionnaire type, name and number of items in order of presentation

<table>
<thead>
<tr>
<th>Type of measure</th>
<th>PROM name</th>
<th>Number of items</th>
</tr>
</thead>
<tbody>
<tr>
<td>AxSpA-fatigue</td>
<td>Warwick Axial Spondyloarthritis measure of fatigue and Energy (WASTEd)</td>
<td>32</td>
</tr>
<tr>
<td>Generic fatigue</td>
<td>Multidimensional Fatigue Inventory – 20 (MFI-20)</td>
<td>20</td>
</tr>
<tr>
<td>AxSpA-specific</td>
<td>Evaluation of Ankylosing Spondylitis Quality of Life (EASi-QoL)</td>
<td>20</td>
</tr>
<tr>
<td>Generic</td>
<td>Short Form 36-item Health Survey version 2 (SF-36 v2)</td>
<td>36</td>
</tr>
<tr>
<td>Emotional Wellbeing specific</td>
<td>Hospital Anxiety and Depression Scale (HADS)</td>
<td>14</td>
</tr>
<tr>
<td>AxSpA-specific</td>
<td>Bath Ankylosing Spondylitis Disease Activity Index (BASDAI)</td>
<td>6</td>
</tr>
<tr>
<td>AxSpA-specific</td>
<td>Bath Ankylosing Spondylitis Functionality Index (BASFI)</td>
<td>10</td>
</tr>
</tbody>
</table>

5.3.5. Data processing and cleaning

All research data was input at secure facilities within Warwick Medical School into a formatted, password protected database. All data was anonymous at the point of receipt. An electronic log linking patient identity with their unique identifier was prepared and maintained at each site and was inaccessible to the research team. All physical copies of responses were stored in a locked cupboard, in a locked room within the medical school.

Upon conclusion of the field test, all data were imported into a spreadsheet (Microsoft Excel) for cleaning. Where demographic information was missing, key non-identifying information (age and sex) was requested from study sites. Missing data were coded as “NA” and all reverse scores were checked. Subscale and total scores for all measures were calculated in the spreadsheet. SF-36v2 subscale scores were calculated using software provided by the makers.

To assess the quality of data entry, a 10% random subset was selected (from all responses) by Dr Helen Parsons. This subset was then checked
for data entry errors and the number of errors were recorded. Acceptable error rates were pre-defined in this study, as per Warwick Clinical Trials Unit Standard Operating Procedures, at <1%, for key data (demographics, WASTEd) and <5% for the remaining measures. Finally, item-level descriptive statistics (mean, range) were computed for all measures to check for erroneous data input.

5.3.6. Data analysis

Both classic and modern psychometric methods were used to evaluate the WASTEd – these are described below in greater detail. A category system was crafted *a priori* to group items based on their statistical performance, enabling the research team and patient research partners to easily identify good or poor performing items (see Table 5.3). To maintain the content validity of the WASTEd, all refinement decisions were made in consideration of the measurement framework described in Chapter 3, thus ensuring concepts identified as important in axSpA-fatigue and energy experience are not completely removed from the measure.

All descriptive statistics and CTT analysis (internal consistency, structural validity) were conducted in SPSS 25 (313). All Rasch analysis was conducted in RUMM2030 software (305).

Table 5.3: Category scheme to identify good or poor performing items

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Category</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>Item fits as intended and clearly contributes to measurement</td>
<td>Recommend to research team and PRP group that item remains unchanged</td>
</tr>
<tr>
<td>?</td>
<td>Item fit is uncertain or its contribution to measurement is unclear (strong qualitative or PRP argument for retention)</td>
<td>Focus of discussions with research team and PRP group</td>
</tr>
<tr>
<td>-</td>
<td>Item does not fit or is not contributing to measurement (irrelevant item or measuring the same as another item within the questionnaire)</td>
<td>High likelihood of being removed</td>
</tr>
</tbody>
</table>
5.3.6.1. Sample overview and data quality

The representativeness of the sample was assessed using descriptive statistics across key variables: age, gender, occupation, ethnicity, disease duration and symptom duration. To ascertain the disease severity of the study population, the NICE guidance for biologics eligibility were applied (314). A BASDAI total score >4 indicates the participant as biologics eligible (indicating more severe disease), whilst a score ≤4 indicates they are not eligible for biologic treatment.

To assess whether the full range of response options were being used in the WASTEd, and to check the amount of missing data per item, item-level descriptive statistics were computed. Item-level missing data below 10% was considered acceptable (315). The analysis was conducted in two stages: Stage 1 – classic test theory approaches, and Stage 2 – modern psychometric approaches.

5.3.6.2. Stage 1: CTT analysis

The CTT analysis included evaluating the factor structure of the WASTEd using exploratory factor analyses, and both item (corrected ITC) and scale-level (Cronbach’s Alpha) evaluations.

Exploratory factor analysis (EFA)

To determine the factor structure and item-loadings of the WASTEd, an exploratory factor analysis (EFA) was conducted. Eigenvalues >1 were accepted as indicating a new factor. As the purpose of analysis is to elucidate the simplest factor structure of the WASTEd and develop the clearest understanding of which items contribute to the measurement of each factor, a varimax rotation was applied. This rotation supports ease of interpreting factor solutions.

Items were considered part of a factor when they had a minimum loading of 0.4 on a single factor (161). If a minimum loading of 0.4 was observed for an item on two or more factors (i.e. cross-loading) then the item was highlighted for further scrutiny. A 0.6/0.4 loading rule (across two factors) has been reported in other studies whereby the higher 0.6 item loading is considered the substantive loading and therefore, the item is considered to
belong to the substantive factor (316,317). In this study, cross-loadings with a difference of ≥+0.2 were considered to indicate a substantial loading on a single factor. Cross-loading differences below this threshold were identified for further scrutiny.

When making decisions for factor extraction, the amount of data variance explained by each factor, and the model overall, was considered. A minimum of 60% of variance explained by the model was sought (318). Communalities – that is, the ratio of unique-to-shared variance explained, was checked per item. Low communalities mean that item variance cannot be well explained by the extracted domains, meaning there are potentially other sources of variance – for example, the item may measure a different construct, or be substantively affected by error. Items with communalities <0.2 were highlighted for removal (319).

Model refinement decisions were informed by the findings of the EFA and, following removal of poor fit items, further EFA models were run until no items showed a poor model fit.

**Item and scale-level evaluation**

Internal consistency was evaluated using Cronbach’s alpha, with alpha values ≥0.7 being considered acceptable (166,320). Values >0.95 were considered evidence of excellent internal consistency, but likely indicating redundancy within the measure (i.e. multiple items may be measuring the same concepts) (161).

Item-total correlations were judged using the following criteria (166):

- Correlations <0.4 were considered poor.
- Correlations between 0.4 and 0.6 (inclusive) were considered moderate.
- Correlations >0.6 were considered well correlated.

**5.3.6.3. Stage 2: Modern psychometrics (RMT)**

The analysis began with identifying the appropriate model to use, followed by checking key assumptions of RMT and finally, a thorough RMT evaluation of the WASTEd.
Identifying the appropriate modelling approach

There are two commonly used parametrisation models used in RMT analysis for categorical item responses: Unrestricted (Partial-Credit) Model or Rating Scale Model (RSM). The approach most suitable to the study data needs to be identified prior to full analysis beginning. The Partial-Credit Model (PCM) specifies that each item within a given item set possesses its own rating scale structure (321). This means that the item characteristic curves, and threshold parameters are individual to each item. In contrast, the RSM specifies that all items within a given item set share the same rating scale structure (322).

Determining which mathematical model to use was evaluated using a likelihood-ratio test which assesses the goodness of fit between two competing statistical models: the PCM and RSM. This was run for each domain and all had excellent power except the control factor which had good power (as determined by the RUMM2030 software). A statistically significant result supports using the PCM (321), and a statistically insignificant result supports using the simpler RSM (322).

Checking RMT assumptions

There are two important assumptions of RMT that need to be determined (118,323):

1. Is the measure unidimensional? This would be determined from the findings of the EFA analysis.

2. Is there any item dependency? This would be determined as part of the RMT analysis with dependent items being highlighted for scrutiny.

There is currently no consensus on what criteria to apply when determining whether there is local dependency between items, however, it is suggested that from the model generated by RMT, the residual correlations between items can be assessed for dependency (324). There are disputes about how best to determine local dependency between items, but current thinking is that items with correlations between 0.2 and 0.3 above the average of all item residuals may be indicate dependency, and thus be
problematic (324). Therefore, item dependency was assessed as part of the Rasch analysis.

**RMT analysis**

The RMT analysis followed a structured procedure to evaluate key features of the WASTEd to support refinement and modification decisions. The analysis included determining the number and size of ability groups (classic intervals), evaluating item-person interactions, calculating person separation indexes, checking individual item fit with the Rasch Model, assessing item targeting, inspecting category probability curves and evaluating differential item functioning. It is recommended that when many statistical analyses are being conducted without pre-planned hypotheses, Bonferroni corrections should be applied (325). This recognises the increased risk of type I errors occurring (false positive results) and adjusts for the type I error rate. As the RMT analysis requires many statistical tests, this correction was applied (where necessary).

**Class intervals**

Class intervals describe the number of ability groups a given item set can distinguish between on the underlying construct. 250 respondents could be grouped into five severity groups of 50 persons. Relatively equal distributions between each class interval were sought, and these distributions were monitored throughout the analysis process.

**Item-person interaction**

This provides essential evidence about the fit of the data to what would be expected on the underlying trait and was determined by assessing the fit residual means and standard deviations at both item and person level. The RUMM2030 software transforms person and item fit statistics to approximate a Z-score (i.e. a standardised normal distribution). As such, where the data fit the model precisely, a mean of 0 and standard deviation of 1 would be expected (326).
**Model fit**

Model fit was assessed using the test-of-fit statistic, where a significant value (p<0.05) indicates misfit within the model and therefore suggests there may be items not contributing to measurement of the target construct, as intended.

**Person separation index**

Person separation index was calculated to determine the reliability of person separation on items. A person separation index >0.7 is considered acceptable (327) as this indicates that it is possible to statistically differentiate between two groups on the underlying trait (see Table 5.4 for further detail).

**Table 5.4: Person separation index – statistical strength and number of distinct groups adapted from (328)**

<table>
<thead>
<tr>
<th>Person separation index</th>
<th>Proportion of variance not error (due to error)</th>
<th>Number of groups</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0 (100)</td>
<td>1</td>
</tr>
<tr>
<td>0.5</td>
<td>50 (50)</td>
<td>1</td>
</tr>
<tr>
<td>0.7</td>
<td>70 (30)</td>
<td>2</td>
</tr>
<tr>
<td>0.8</td>
<td>80 (20)</td>
<td>3</td>
</tr>
<tr>
<td>0.9</td>
<td>90 (10)</td>
<td>4</td>
</tr>
<tr>
<td>0.94</td>
<td>94 (6)</td>
<td>5</td>
</tr>
<tr>
<td>0.96</td>
<td>96 (4)</td>
<td>7</td>
</tr>
<tr>
<td>0.97</td>
<td>97 (3)</td>
<td>8</td>
</tr>
<tr>
<td>0.98</td>
<td>98 (2)</td>
<td>9</td>
</tr>
</tbody>
</table>

**Individual item fit**

The extent to which each item fits the Rasch Model was explored using three sources of evidence: chi squared ($\chi^2$) probability (p<0.05), fit residuals (between -2.5 and 2.5) (329) and F-statistics (p<0.05). Items that did not meet the specified criteria (shown in brackets) for any of the three tests were highlighted as indicating misfit to the model and therefore could be targeted in item refinement (329).
**Targeting**

To evaluate targeting of the scale to the sample, the spread of person and item (threshold) locations were compared. This was done by generating person-item threshold distribution maps which could be visually inspected as per another RMT evaluation study (304).

**Category probability curves**

For each item, the likelihood of affirming a specific response option at a given point on the underlying trait was evaluated. Each probability curve was checked for model fit. This involved identifying disordered thresholds to ensure appropriate ordering of categories and that categories represented a distinct point on the underlying trait – that is, the category probability curve for one response option was not subsumed by another. Thresholds should also be logically ordered meaning that as a participant moves along the underlying trait (e.g. experiences greater fatigue), they move sequentially through the available response options.

**Differential item functioning (DIF)**

Two types of DIF can be observed at item level: uniform DIF and non-uniform DIF. Uniform DIF is present when one group consistently displays a greater ability to confirm an item compared to another group, for example, males always score higher than females. Non-uniform DIF is present when class interval groups are inconsistent in how they respond to the item between the DIF-evaluated groups, such as males scoring higher at low abilities and females scoring higher at high abilities. Splitting an item has been reported as a common solution to uniform DIF – that is, allowing different item parameters to be applied to each group, allowing the item to be ‘specific’ to each group (330). In contrast, non-uniform DIF is problematic and there is no accepted solution to it.

DIF was evaluated per item across a range of key characteristic variables: age group, gender, ethnicity and biologics eligibility. This was done using ANOVA, with significance values <0.05 considered evidence that DIF is present. Uniform DIF presents as a difference between the groups (clear,
consistent difference). Non-uniform DIF presents as an interaction between the groups (no clear, consistent difference).

A summary of the statistical analyses and quality criteria are provided in Table 5.5.

**Table 5.5: Summary of statistical tests and applied quality criteria**

<table>
<thead>
<tr>
<th>Measurement property</th>
<th>Statistical test</th>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Classic Test Theory</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Internal consistency</td>
<td>Cronbach’s alpha (α)</td>
<td>Cronbach’s α ≥0.7 (118) and &lt;0.95 (161) indicates adequate internal consistency Cronbach’s α ≥0.95 indicates possible item redundancy (161)</td>
</tr>
<tr>
<td></td>
<td>Item-total correlation (ITC)</td>
<td>ITC should minimally be &gt; +0.3 (118,166)</td>
</tr>
<tr>
<td>Structural evaluation</td>
<td>Exploratory factor analysis</td>
<td>Eigen value &gt;1 to be considered a factor (160) Item factor loading ≥ +0.4 indicates acceptable loading (161) If cross-loading was present, a difference of ≥ +0.2 was needed to indicate a substantive loading on a single factor Model variance explained acceptable if ≥ 60% (318) Communalities (per item) &lt;0.2 identified for removal (319)</td>
</tr>
<tr>
<td><strong>Rasch Measurement Theory</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RUMM2030 software</td>
<td>Class interval structure</td>
<td>50 or more individuals per ability group Groups should be relatively similar in size <em>(subjective judgement; group sizes should be manually modified if these criteria are not met)</em></td>
</tr>
<tr>
<td></td>
<td>Item-person interaction</td>
<td>Item and persons fitting the model should have a mean near 0 (329), and a standard deviation near 1 (326)</td>
</tr>
<tr>
<td><strong>Local dependency</strong></td>
<td>Moderate dependency evidenced if item pairs have residual correlations between 0.2 and 0.3 above the average (324) Significant dependency evidenced if item pairs have residual correlations &gt;0.3 of the average (304)</td>
<td></td>
</tr>
<tr>
<td>----------------------</td>
<td>--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td><strong>Person separation index (PSI)</strong></td>
<td>PSI &gt;0.7 minimally necessary to distinguish between two groups on the latent trait (327)</td>
<td></td>
</tr>
<tr>
<td><strong>Test of fit (model fit)</strong></td>
<td>Chi-squared analysis should not show significant misfit (p&gt;0.05)</td>
<td></td>
</tr>
<tr>
<td><strong>Individual item fit</strong></td>
<td>Chi square ($\chi^2$) is significant (p&lt;0.05) Fit residual within ± 2.5 (304,329) ANOVA is significant (p&lt;0.05)</td>
<td></td>
</tr>
<tr>
<td><strong>Targeting</strong></td>
<td>Person-item threshold distribution maps visually inspected (304)</td>
<td></td>
</tr>
<tr>
<td><strong>Category probability curves</strong></td>
<td>Thresholds should be logically ordered Thresholds should not be disordered Thresholds should cover an appropriate area of the continuum</td>
<td></td>
</tr>
<tr>
<td><strong>Differential item functioning</strong></td>
<td>Significant main effect p&lt;0.05 indicates uniform DIF Significant interaction effect p&lt;0.05 indicates non-uniform DIF</td>
<td></td>
</tr>
</tbody>
</table>

**5.3.6.4. Research team review**

Following statistical analysis, a research team meeting was held to identify potential items for removal due to having poor statistical evidence of their model fit or performance. The purpose of this meeting was to produce an item revision list for review in the PRP group.

I began the meeting by giving a detailed description of the analysis process, a full summary of results and any key decisions made. I provided a physical copy of the full analysis which was available to the group and referred to throughout discussions. Each item was assigned a category as described in Table 5.3 and this formed the basis for a per item review. The group deliberated over each item. To maintain content validity of the WASTEd I ensured all decisions considered the qualitative importance of the item, the PRP group perspective and that the measurement framework
defined in Chapter 3 would still be reflected following refinement suggestions. All suggestions were manually recorded on a per-item basis, producing an item-refinement list for consideration in the PRP group meeting. Only items identified as performing statistically poor as part of the EFA were subject to removal without PRP approval, however, these items would still be discussed with the PRP group review to ensure nothing of conceptual importance would be lost.

The item refinement list was underpinned by the statistical findings and contained the additional suggestions and recommendations of the research team.

5.3.6.5. PRP group review
The aim of PPI in this study was to co-refine the WASTEd following statistical evaluation (145).

A PRP group meeting was scheduled after the team review and was organised to begin three weeks after the research team review. To support ease of engagement and reduce the jargon and heaviness of statistical analyses, the item revision list was modified into a refinement file. The refinement file contained each item, a lay summary of any statistical concerns and proposed solutions. I also took a copy of items grouped by their conceptual domains (Chapters 3 and 4).

The meeting began the same as the team meeting, however, more time was spent explaining the principles behind key statistical tests. Extensive discussions took place with different group members putting forward their ideas, suggestions or criticisms and final refinement decisions were agreed by PRP consensus. A summary of the group was prepared and provided to both the research team and PRP group and where consensus could not be reached, an agreed suggestion or group of recommendations would be returned to the research team for their consideration.
5.4. Results

5.4.1. Participants

Seven sites across England participated in the study. 580 individuals were identified and contacted to participate, of which 372 provided responses (64.13% response rate). The mean age of respondents was 54.16 years (SD 13.4, range: 20 – 58 years). 351 (94.1%) respondents were white, seven were Asian, two were Black and five belonged to other ethnic groups. Population characteristics across key demographic variables are described in Table 5.6.

Table 5.6: Demographic details of respondents

<table>
<thead>
<tr>
<th>Variable</th>
<th>Number (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>242 (65.1)</td>
</tr>
<tr>
<td>Female</td>
<td>129 (34.7)</td>
</tr>
<tr>
<td>Age</td>
<td></td>
</tr>
<tr>
<td>20-29</td>
<td>16 (4.3)</td>
</tr>
<tr>
<td>30-39</td>
<td>41 (11.1)</td>
</tr>
<tr>
<td>40-49</td>
<td>72 (19.5)</td>
</tr>
<tr>
<td>50-59</td>
<td>103 (27.7)</td>
</tr>
<tr>
<td>60-69</td>
<td>86 (23.3)</td>
</tr>
<tr>
<td>70-79</td>
<td>48 (13.0)</td>
</tr>
<tr>
<td>80-89</td>
<td>3 (0.8)</td>
</tr>
<tr>
<td>Occupation</td>
<td></td>
</tr>
<tr>
<td>Employed or in education</td>
<td>205 (55.9)</td>
</tr>
<tr>
<td>Not employed or retired</td>
<td>163 (44.1)</td>
</tr>
<tr>
<td>Biologics eligibility</td>
<td></td>
</tr>
<tr>
<td>Eligible (BASDAI total ≥4)</td>
<td>222 (59.7)</td>
</tr>
<tr>
<td>Not eligible (BASDAI total &lt;4)</td>
<td>150 (40.3)</td>
</tr>
</tbody>
</table>

5.4.2. Data quality

Across all items of the WASTEd there were 63 missing data (unanswered items), with missing data for items ranging from one to four. This was almost identical to the EASI-QoL – another axSpA-specific questionnaire – which recorded across all items, 62 missing data (unanswered items), with missing data for items ranging from three to five. Missing data for other questionnaires, in the form of unanswered questions, was between 15
(BASDAI) and 255 (HADS). All WASTEd items had endorsements for the full range of response options.

5.4.3.  Structural evaluation
A factor solution for the WASTEd was achieved following analysis of three EFA models. Each model was constructed, analysed and modified prior to the next being conducted. The process was performed sequentially and the models, including item removal decisions, are described below.

5.4.3.1.  Model 1
An initial EFA was conducted on the all 30 candidate items. This model produced a four-factor solution to the data with 18 items clearly loading onto distinct factors, and 12 items cross-loading (Table 5.7). Six of the cross-loading items had a major factor (cross-loading difference was >0.2). Three items were highlighted to be monitored (cross-loadings between 0.1 and 0.2). Three items had narrow cross-loadings (magnitude within 0.1) meaning a major factor could not be determined, therefore, they were removed (Table 5.8). Communalities for each item were checked and all were above accepted levels. Variance explained by each factor was checked and the four-factor solution explained 72.28% of total variance.

All but two energy items loaded onto factor 1 (n=9). These items related to physical, mental and social activity, matching the proposed definition of ‘having energy’ in Chapter 3. Therefore, this factor was this was labelled “Energy”. All but two fatigue items loaded onto factor 2 (n=8). These items captured the experience of fatigue as defined in Chapter 3, thus, this factor was labelled “Fatigue”. Five items loaded onto factor 3 and captured the experience of fatigue and energy in terms of severity and frequency and was therefore labelled “Symptoms”. Factor 4 consisted of two items capturing managing and coping and was therefore labelled “Control”. The remaining six items cross-loaded between multiple factors with no clear ‘major’ factor loading and therefore their contribution to measurement were explored in subsequent EFA models. See Table 5.7 for the factor loadings of Model 1.
Cronbach’s alpha for the 30-items was calculated to be 0.97 (n=363), indicating possible item redundancy within the measure. Item-total correlations ranged from 0.609 to 0.847 suggesting good relationships between item and total scores.

Table 5.7: Extracted factors for Model 1

<table>
<thead>
<tr>
<th>Item</th>
<th>Factor 1 (energy)</th>
<th>Factor 2 (fatigue)</th>
<th>Factor 3 (symptoms)</th>
<th>Factor 4 (control)</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1</td>
<td>How often have you felt fatigued?</td>
<td></td>
<td></td>
<td>0.740</td>
</tr>
<tr>
<td>F2</td>
<td>How severe was your fatigue?</td>
<td>0.410</td>
<td></td>
<td>0.661</td>
</tr>
<tr>
<td>F3</td>
<td>Has your fatigue made it difficult to remember things?</td>
<td></td>
<td></td>
<td>0.728</td>
</tr>
<tr>
<td>F4</td>
<td>Has your fatigue made it difficult to concentrate on demanding tasks?</td>
<td></td>
<td></td>
<td>0.736</td>
</tr>
<tr>
<td>F5</td>
<td>Have you found it difficult to engage in conversations with other people because of your fatigue?</td>
<td></td>
<td></td>
<td>0.728</td>
</tr>
<tr>
<td>F6</td>
<td>Has your fatigue made it difficult to do the things you enjoy?</td>
<td></td>
<td></td>
<td>0.607</td>
</tr>
<tr>
<td>F7</td>
<td>Have you felt downhearted because of your fatigue?</td>
<td></td>
<td></td>
<td>0.661</td>
</tr>
<tr>
<td>F8</td>
<td>Has your fatigue made it difficult to be in control of your life?</td>
<td>0.504</td>
<td></td>
<td>0.563</td>
</tr>
<tr>
<td>F9</td>
<td>Have you felt worried because of your fatigue?</td>
<td></td>
<td></td>
<td>0.667</td>
</tr>
</tbody>
</table>

\(^{10}\) item examples removed from the table for brevity
<table>
<thead>
<tr>
<th></th>
<th>Question</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>F10</td>
<td>Have you felt frustrated because of your fatigue?</td>
<td>0.608</td>
</tr>
<tr>
<td>F11</td>
<td>Have you felt the need to be left alone because of your fatigue?</td>
<td>0.642</td>
</tr>
<tr>
<td>F12</td>
<td>I feel I have been able to manage my fatigue.</td>
<td>0.843</td>
</tr>
<tr>
<td>F13</td>
<td>I feel I have been able to cope with my fatigue.</td>
<td>0.830</td>
</tr>
<tr>
<td>F14</td>
<td>Have you been more dependent on others because of your fatigue?</td>
<td>0.643</td>
</tr>
<tr>
<td>E1</td>
<td>How much energy have you had?</td>
<td>0.666</td>
</tr>
<tr>
<td>E2</td>
<td>How often have you felt drained of energy?</td>
<td>0.719</td>
</tr>
<tr>
<td>E3</td>
<td>Have your energy levels made it difficult to take care of yourself?</td>
<td>0.709</td>
</tr>
<tr>
<td>E4</td>
<td>Have your energy levels made it difficult to do every day activities?</td>
<td>0.715</td>
</tr>
<tr>
<td>E5</td>
<td>Have your energy levels made it difficult to take part in more demanding physical activities?</td>
<td>0.614</td>
</tr>
<tr>
<td>E6</td>
<td>Have your energy levels made it difficult to do your usual work?</td>
<td>0.685</td>
</tr>
<tr>
<td>E7</td>
<td>Have your energy levels made it difficult to do the things you enjoy?</td>
<td>0.631</td>
</tr>
<tr>
<td></td>
<td>Question</td>
<td>Factor 1</td>
</tr>
<tr>
<td>---</td>
<td>-------------------------------------------------------------------------</td>
<td>----------</td>
</tr>
<tr>
<td>E8</td>
<td>Have your energy levels made it difficult to make plans?</td>
<td>0.689</td>
</tr>
<tr>
<td>E9</td>
<td>Have your energy levels caused you to cancel plans?</td>
<td>0.727</td>
</tr>
<tr>
<td>E10</td>
<td>Have your energy levels made it difficult to keep to your plans?</td>
<td>0.722</td>
</tr>
<tr>
<td>E11</td>
<td>Have you ‘turned down’ invitations because of your energy levels?</td>
<td>0.657</td>
</tr>
<tr>
<td>E12</td>
<td>I have difficulty maintaining my energy levels when I have done too much.</td>
<td>0.483</td>
</tr>
<tr>
<td>E13</td>
<td>I feel I have been able to manage my energy levels.</td>
<td></td>
</tr>
<tr>
<td>E14</td>
<td>Have you lacked physical energy?</td>
<td>0.444</td>
</tr>
<tr>
<td>E15</td>
<td>Have you lacked mental energy?</td>
<td></td>
</tr>
<tr>
<td>E16</td>
<td>Have you run out of energy suddenly and needed to take a break?</td>
<td>0.493</td>
</tr>
</tbody>
</table>

5.4.3.2. Model 2

A second EFA was conducted on the remaining 27 items using the same criteria as Model 1. Eight items cross-loaded: five with a major factor, two identified for further scrutiny and one removed due to cross-loading outside of accepted ranges (reported in Appendix 5A). Of the two items that were identified for further scrutiny in Model 1, one was removed due to narrow cross-loading and the other was retained due to its qualitative importance (Table 5.8).
Table 5.8: Items removed in model 1 and 2 with brief rationale

<table>
<thead>
<tr>
<th>Number</th>
<th>Item</th>
<th>Reason</th>
<th>Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F8</td>
<td>Has your fatigue made it difficult to be in control of your life?</td>
<td>Cross-loading on two factors within 0.2 – no major factor</td>
<td></td>
</tr>
<tr>
<td>E12</td>
<td>I have difficulty maintaining my energy levels when I have done too much</td>
<td>Cross-loading on two factors within 0.2 – no major factor</td>
<td></td>
</tr>
<tr>
<td>E16</td>
<td>Have you run out of energy suddenly and needed to take a break? (e.g. have a nap or rest)</td>
<td>Cross-loading on two factors within 0.2 – no major factor</td>
<td></td>
</tr>
<tr>
<td>Model 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E5</td>
<td>Have your energy levels made it difficult to take part in more demanding physical activities? (e.g. going for a walk, gardening)</td>
<td>Measurement contribution unclear: other physical activity-intensity items load as expected except this item</td>
<td></td>
</tr>
<tr>
<td>E13</td>
<td>I feel I have been able to manage my energy levels</td>
<td>Cross-loading on two factors within 0.2 – no major factor</td>
<td></td>
</tr>
</tbody>
</table>

5.4.3.3. Model 3

A final EFA was conducted on the remaining 25 items (Table 5.9). As refinement decisions to this point were made on a four-factor solution, this structure was retained. The rest of the analysis was conducted as described in Models 1 and 2. Four items (E6, 8 and 14) were identified as cross-loading: three met accepted criteria (cross-loading difference >0.2); and one (E14) had a cross-loading difference of 0.152. Item E14 was not removed as it was identified as an important concept within both the qualitative element of the study and by the study PRP group. Therefore, this item required further investigation and probable modification.

After the removal of energy management items, factor 4 (Control) was left with only two items. Current recommendations suggest a minimum of three
items are necessary to form a factor (331). Therefore, the importance of the control factor in understanding axSpA fatigue and energy would be a key topic for discussion in both research team and PRP group meetings. If this factor is considered important then there are two possible solutions which would require future statistical testing: (i) to modify the control items that did not perform statistically as intended, or (ii) develop new items.

There were no substantive changes from Model 1 for either Cronbach’s alpha or item-total correlations (analysis reported in Appendix 5B).

Table 5.9: Model 3 item factor loadings (bold numbers represent major factor loadings)

<table>
<thead>
<tr>
<th>Item</th>
<th>Factor</th>
<th>1 (energy)</th>
<th>2 (fatigue)</th>
<th>3 (symptoms)</th>
<th>4 (control)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Energy</td>
<td>Have you been more dependent on others (e.g. friends or family) because of your fatigue?</td>
<td></td>
<td>0.648</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E3</td>
<td>Have your energy levels made it difficult to take care of yourself?</td>
<td>0.714</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E4</td>
<td>Have your energy levels made it difficult to do everyday activities?</td>
<td>0.714</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E6</td>
<td>Have your energy levels made it difficult to do your usual work?</td>
<td>0.681</td>
<td>0.453</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E7</td>
<td>Have your energy levels made it difficult to do the things you enjoy?</td>
<td>0.638</td>
<td>0.486</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E8</td>
<td>Have your energy levels made it difficult to make plans?</td>
<td>0.708</td>
<td>0.436</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E9</td>
<td>Have your energy levels caused you to cancel plans?</td>
<td>0.754</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E10</td>
<td>Have your energy levels made it difficult to keep to your plans?</td>
<td>0.742</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E11</td>
<td>Have you ‘turned down’ invitations because of your energy levels?</td>
<td>0.685</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fatigue
<table>
<thead>
<tr>
<th>F3</th>
<th>Has your fatigue made it difficult to remember things?</th>
<th>0.743</th>
</tr>
</thead>
<tbody>
<tr>
<td>F4</td>
<td>Has your fatigue made it difficult to concentrate on demanding tasks?</td>
<td>0.742</td>
</tr>
<tr>
<td>F5</td>
<td>Have you found it difficult to engage in conversations with other people because of your fatigue?</td>
<td>0.732</td>
</tr>
<tr>
<td>F6</td>
<td>Has your fatigue made it difficult to do the things you enjoy?</td>
<td>0.619</td>
</tr>
<tr>
<td>F7</td>
<td>Have you felt downhearted because of your fatigue?</td>
<td>0.655</td>
</tr>
<tr>
<td>F9</td>
<td>Have you felt worried because of your fatigue?</td>
<td>0.645</td>
</tr>
<tr>
<td>F10</td>
<td>Have you felt frustrated because of your fatigue?</td>
<td>0.608</td>
</tr>
<tr>
<td>F11</td>
<td>Have you felt the need to be left alone because of your fatigue?</td>
<td>0.649</td>
</tr>
<tr>
<td>E15</td>
<td>Have you lacked mental energy?</td>
<td>0.571</td>
</tr>
</tbody>
</table>

**Symptoms**

| F1   | How often have you felt fatigued? | 0.770 |
| F2   | How severe was your fatigue? | 0.692 |
| E1   | How much energy have you had? | 0.676 |
| E2   | How often have you felt drained of energy? | 0.716 |
| E14  | Have you lacked physical energy? | 0.460 |

**Control**

| F12  | I feel I have been able to manage my fatigue. | 0.852 |
| F13  | I feel I have been able to cope with my fatigue. | 0.850 |

5.4.4. **Rasch analysis**

The four-factor solution described in EFA Model 3 was used for Rasch Measurement Theory (RMT) analysis. To conform to the Rasch assumption of unidimensionality, each EFA-defined factor was analysed separately. A likelihood-ratio test was conducted for each of the four factors to determine
whether a PCM or RSM should be used. All factors produced a statistically significant result (p<0.001) supporting use of the PCM.

A full analysis for each of the four factors is described below. Local dependency between items within a given factor was assessed as part of the RMT analysis and is reported per factor.

**Energy**

The nine-item energy factor was able to distinguish between six ability groups (class intervals). All ability groups had over 50 individuals and were relatively similar in size (range: 51–61 individuals per ability group). The item-person interaction statistic indicated no issue at person level (mean near 0, SD <1.4), but at item level there was deviation far exceeding accepted levels at 2.03, suggesting a potential issue at item level. Test-of-fit analysis was significant (p<0.001) therefore suggesting misfit to the Rasch model.

Evaluation of local dependency highlighted three item pairs with dependency outside of current recommendations. These item pairs were: E3 (self-care) with E4 (everyday activities); E9 (cancel plans) with E10 (keep plans), and; E9 with E11 (turn down invites). Another item pair – E6 (usual work) and E7 (do things you enjoy) – was 0.13 above the residual average and therefore highlighted for further scrutiny.

Person separation index (PSI) was calculated at 0.91, indicating that at least 4 ability groups can be distinguished. Cronbach’s alpha was calculated at 0.94 (n=368) showing the factor items have excellent internal consistency.

Evaluation of individual item fit using chi-squared and ANOVA tests (with post-hoc Bonferonni correction) highlighted one item (F14) as problematic, with both tests showing significant misfit (p<0.001). Further, this item had a fit residual of 4.55, suggesting that the item is under discriminating.

Evaluation of person-item threshold distributions indicates a good item spread across the continuum, indicating good targeting (Figure 5.1).
Category threshold curves highlighted five items with problematic thresholds: two items (E6 and E7) had one disordered threshold (0: not at all); two items (E9 and E10) had two disordered thresholds (0: not at all; 4: extremely) and one (E11) with a single disordered threshold (4: extremely). See Figure 5.2 for an example. Here, the first response option (option 0) is subsumed by response option 1, meaning respondents across the range of the underlying trait covered by option 0 will likely affirm option 1.

Differential item functioning was separately evaluated for all items by age group, gender, ethnicity, biologics eligibility and occupation. ANOVA analysis found that there was no evidence of uniform DIF for any item or group (p>0.05) or non-uniform DIF (p>0.05). The findings of the analysis collated and categorised in Table 5.10.
Figure 5.2: Example disordered threshold for item E6 (response option 0)

Table 5.10: Summary of energy factor items identified as problematic following RMT analysis

<table>
<thead>
<tr>
<th>Number</th>
<th>Item</th>
<th>Problem</th>
<th>Considerations</th>
<th>Category</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Energy</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>F14</td>
<td>Have you been more dependent on others (e.g. friends or family) because of your fatigue?</td>
<td>Under-discriminating – limited contribution to measurement</td>
<td>? / -</td>
</tr>
<tr>
<td></td>
<td>E3</td>
<td>Have your energy levels made it difficult to take care of yourself?</td>
<td>Dependency with E4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>E4</td>
<td>Have your energy levels made it difficult to do everyday activities?</td>
<td>See E3 – same problem</td>
<td>? / -</td>
</tr>
<tr>
<td></td>
<td>E6</td>
<td>Have your energy levels made it difficult</td>
<td>Disordered threshold (0)</td>
<td>? / +</td>
</tr>
<tr>
<td></td>
<td>to do your usual work?</td>
<td>Dependency with E7</td>
<td>Delete one, or merge with E7</td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>------------------------</td>
<td>---------------------</td>
<td>-----------------------------</td>
<td></td>
</tr>
<tr>
<td>E7</td>
<td>Have your energy levels made it difficult to do the things you enjoy?</td>
<td>See E6 – same problem</td>
<td>See E6 – same solution</td>
<td>? / +</td>
</tr>
<tr>
<td>E8</td>
<td>Have your energy levels made it difficult to make plans?</td>
<td>NA</td>
<td>NA</td>
<td>+</td>
</tr>
<tr>
<td>E9</td>
<td>Have your energy levels caused you to cancel plans?</td>
<td>Disordered thresholds (1,4) Dependency with E10 and E11</td>
<td>Reduce response options Evidence they are asking about the same thing: can two be deleted?</td>
<td>? / -</td>
</tr>
<tr>
<td>E10</td>
<td>Have your energy levels made it difficult to keep to your plans?</td>
<td>Disordered thresholds (0,4) Dependency with E9 and 11</td>
<td>Reduce response options Evidence they are asking about the same thing: can two be deleted?</td>
<td>? / -</td>
</tr>
<tr>
<td>E11</td>
<td>Have you ‘turned down’ invitations because of your energy levels? (e.g. to meet a friend, socialise)</td>
<td>Disordered threshold (4) Dependency with E9 and 10</td>
<td>Reduce response options Evidence they are asking about the same thing: can two be deleted?</td>
<td>? / -</td>
</tr>
</tbody>
</table>

**Fatigue**

The nine-item fatigue factor was able to distinguish between six class intervals. Again, all groups had over 50 individuals and although the range was slightly larger, the group sizes were still relatively close (range: 51-66). The item-person interaction statistic indicated no issue at person level, but at item level the standard deviation was just outside of accepted levels at
1.43. Test-of-fit analysis was not significant (p=0.659), suggesting a good fit to the Rasch model.

No items showed signs of significant dependency (>0.3 of the residual average), however, three item pairs showed sign of moderate dependency (between 0.2 and 0.3 of the average residual correlation) and therefore warrant further scrutiny. These item pairs were: F3 (memory) with F4 (concentration); F7 (downhearted) with F9 (worried), and; F7 with F10 (frustrated).

PSI was high at 0.91, indicating that at least 4 ability groups can be distinguished between. Cronbach’s alpha was calculated at 0.94 (n=369) showing excellent internal consistency.

Evaluation of individual item fit identified no problematic items.

Evaluation of person-item threshold distributions indicates a good item spread across both upper and low ends of the continuum (Figure 5.3).

*Figure 5.3: Person-item threshold location distributions for the fatigue factor*

Category threshold curves examination highlighted eight items with problematic thresholds for the edge categories (0 and 4): Seven items had one disordered threshold (0: not at all; items F2, F4, F7, and F10. 4: extremely; items F1, F3 and F4). One item (E15) had two disordered thresholds (0: not at all; 4: extremely). The findings of the analysis collated and categorised in Table 5.11.
<table>
<thead>
<tr>
<th>Number</th>
<th>Item</th>
<th>Problem</th>
<th>Considerations</th>
<th>Category</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>Fatigue</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F3</td>
<td>Has your fatigue made it difficult to remember things?</td>
<td>Disordered threshold (4) Dependency with F4</td>
<td>Reduce response options Are F3 and F4 asking about the same things?</td>
<td>?</td>
</tr>
<tr>
<td>F4</td>
<td>Has your fatigue made it difficult to concentrate on demanding tasks?</td>
<td>Disordered threshold (0,4) Dependency with F3</td>
<td>Reduce response options Are F3 and F4 asking about the same, or different things?</td>
<td>?</td>
</tr>
<tr>
<td>F5</td>
<td>Have you found it difficult to engage in conversations with other people because of your fatigue?</td>
<td>Disordered threshold (4)</td>
<td>Reduce response options</td>
<td>+</td>
</tr>
<tr>
<td>F6</td>
<td>Has your fatigue made it difficult to do the things you enjoy?</td>
<td>Disordered threshold (4)</td>
<td>Reduce response options</td>
<td>+</td>
</tr>
<tr>
<td>F7</td>
<td>Have you felt downhearted because of your fatigue?</td>
<td>Disordered threshold (0) Dependency with F9 and F10</td>
<td>Reduce response options Merge F9 and 10 into F7</td>
<td>?</td>
</tr>
<tr>
<td>F9</td>
<td>Have you felt worried because of your fatigue?</td>
<td>Disordered thresholds (2,4) Dependency with F7 and F10</td>
<td>- Merge into F7</td>
<td>-</td>
</tr>
<tr>
<td>F10</td>
<td>Have you felt frustrated because of your fatigue?</td>
<td>Disordered threshold (0) Dependency with F7 and F9</td>
<td>- Merge into F7</td>
<td>-</td>
</tr>
<tr>
<td>Item</td>
<td>Question</td>
<td>Symptom</td>
<td>Disordered Thresholds</td>
<td>Reduce Response Options</td>
</tr>
<tr>
<td>------</td>
<td>---------------------------------------------------------------------------</td>
<td>---------</td>
<td>-----------------------</td>
<td>--------------------------</td>
</tr>
<tr>
<td>F11</td>
<td>Have you felt the need to be left alone because of your fatigue?</td>
<td>Disordered thresholds (1,4)</td>
<td>Reduce response options</td>
<td>+</td>
</tr>
<tr>
<td>E15</td>
<td>Have you lacked mental energy?</td>
<td>Disordered thresholds (0,4)</td>
<td>Reduce response options</td>
<td>+</td>
</tr>
</tbody>
</table>

Symptom

The five-item symptom factor was able to distinguish between six class intervals, however, the sixth ability group only had 12 individuals – significantly below the minimum group size of 50. When reduced to five class intervals, all groups consisted of over 50 individuals and were similar in size (range: 66-75). Item-person interaction statistic indicated no issues at person or item level. Test-of-fit analysis was not significant (p=0.477) suggesting a good fit to the Rasch model.

Only two items showed signs of moderate dependency (between 0.2 and 0.3 of the average residual correlation): F1 (fatigue frequency) with F2 (fatigue severity). As both items are known to have utility in identifying and detecting individuals experiencing major fatigue (frequent and severe) (40) they were not highlighted for further scrutiny.

PSI was extremely high at 0.91, indicating that at least 4 ability groups can be distinguished between. Cronbach’s alpha was calculated at 0.92 (n=369) showing excellent internal consistency.

Evaluation of individual item fit identified no problematic items.

Evaluation of person-item threshold distributions indicates a good item spread across both upper and low ends of the continuum. The distribution across persons indicates that the relative difficulty of affirming items between 0 and 5 logits remains roughly the same, indicating that the items may not be discriminating as intended (Figure 5.4).
Category threshold curves identified three items (F2, E2 and E14) with two disordered thresholds (0: not at all; 4: extremely). One item (F1) had three disordered thresholds (0: not at all; 1: a little; 3: a lot). One item (E1) was substantially disordered with the middle (moderate) response option almost always likely to be endorsed regardless of underlying energy levels (Figure 5.5). The findings of the analysis are collated and categorised in Table 5.12.
Table 5.12: Summary of symptom factor items identified as problematic following RMT analysis

<table>
<thead>
<tr>
<th>Number</th>
<th>Item</th>
<th>Problem</th>
<th>Considerations</th>
<th>Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1</td>
<td>How often have you felt fatigued?</td>
<td>Disordered thresholds (0,1,3)</td>
<td>Reduce response options</td>
<td>+</td>
</tr>
<tr>
<td>F2</td>
<td>How severe was your fatigue?</td>
<td>Disordered thresholds (0,4)</td>
<td>Reduce response options</td>
<td>+</td>
</tr>
<tr>
<td>E1</td>
<td>How much energy have you had?</td>
<td>Only one response option active</td>
<td>Removal</td>
<td>-</td>
</tr>
<tr>
<td>E2</td>
<td>How often have you felt drained of energy?</td>
<td>Disordered thresholds (0,4)</td>
<td>Reduce response options</td>
<td>+</td>
</tr>
<tr>
<td>E14</td>
<td>Have you lacked physical energy?</td>
<td>Disordered thresholds (0,4)</td>
<td>Reduce response options</td>
<td>+</td>
</tr>
</tbody>
</table>

**Control**

Due to the small size of this two-item factor, any RMT analysis would produce unstable findings. Despite this, an analysis was attempted to test whether anything meaningful could be extracted, however, results across tests were extreme. Consequently, it was not possible to have confidence in the results (findings not reported). Limited information about thresholding could be extracted and is reflected in Table 5.13.

Following completion of the analysis, each summary table per domain (including EFA-excluded items) were compiled into one single table to facilitate ease of engagement for both the research team and PRP group. This table formed the basis of refinement discussions and decisions to inform the development of the short form WASTE'd.
Table 5.13: Summary of control factor items identified as problematic following RMT analysis

<table>
<thead>
<tr>
<th>Number</th>
<th>Item</th>
<th>Problem</th>
<th>Considerations</th>
<th>Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F12</td>
<td>I feel I have been able to manage my fatigue.</td>
<td>Disordered thresholds (0,4)</td>
<td>Reduce response options</td>
<td>+</td>
</tr>
<tr>
<td>F13</td>
<td>I feel I have been able to cope with my fatigue.</td>
<td>Disordered thresholds (0,4)</td>
<td>Reduce response options</td>
<td>+</td>
</tr>
</tbody>
</table>

5.4.5. Research team review

A 3-hour item-refinement meeting was held with the full research team: Drs Elizabeth Tutton, Jon Packham, Helen Parsons and Kirstie Haywood.

Items F3 (remembering) and F4 (concentrating) had evidence of dependency and were discussed, with suggestions that one should be deleted. However, both items emerged as separate concepts from the qualitative work, therefore, I suggested that merging the items may be better for content validity.

Items E6 (usual work) and E7 (things you enjoy) showed signs of dependency and due to E6 also having disordered thresholds, this item was identified for removal. I agreed to take this suggestion to the PRP group meeting.

Items E3 (take care of self) and E4 (everyday activities) showed evidence of dependency and in discussions, the team felt the items were two takes on the same idea (self-care is usually an every day activity for people). Therefore, these items were identified for merging.

Items E9 (cancel plans), E10 (keep plans) and E11 (turn down invitations) were all showing dependency indicating they capture the same concept. However, item E8 (make plans) was statistically robust across other assessments. First, the group discussed whether E8 was a separate concept to E9-11 and suggested that removing all three would be an acceptable approach. I challenged this, highlighting that these items had
been subject to significant debate in a previous item-generation and refinement PRP group meeting. As I felt PRPs would raise an important content validity argument based on previous discussions, I proposed reaching a solution with the PRP group instead.

Item F14 (dependent on others) was under-discriminating and therefore contributed little to measurement. The group agreed this item should minimally be modified but would probably be best removed due to its lack of specificity, thus likely capturing multiple concepts rather than just axSpA fatigue. Additionally, there were doubts the item had clinical relevance to healthcare professionals.

Item E1 (how much energy) was a poorly performing item with almost all participants affirming the middle response option (moderate amount of energy). Therefore, it was unanimously agreed that this item should be removed.

An extensive discussion ensued regarding whether physical and mental energy were important questions, with the idea of lacking physical energy being likened to being drained of energy. No agreement was reached; therefore, this item was taken to the PRPs to determine whether the two items are distinct.

Finally, to ensure that the control factor meets the requirements for the minimum number of items, a third item needed to be crafted. The team felt that the control factor is an important domain to understand how patients are coping or self-managing, and identify those who may be struggling; thus, having greater clinical relevance. The group consensus was to retain the four-factor solution and to generate new energy management or coping items to replace those that did not perform as expected. A new item would be co-produced with the PRPs.

5.4.6. PRP group review

A 3-hour item-refinement meeting took place, co-facilitated Dr Jane Martindale (group convener) and myself. The meeting was structured into six parts: the four domains (fatigue, energy, symptoms, control), a consideration of items removed as part of the EFA modelling to ensure no important concepts were removed, and finally a discussion about the
number of response option categories. I did not allow items with clear evidence of poor statistical performance to be readmitted into the measure as they are known not to work (e.g. cross loading items identified in EFA). However, if PRP group members felt those items were conceptually important then a modified or new item could be introduced into the measure.

**Factor 1: Fatigue**

Discussions considered the potential removal or modification of some items and a conceptual discussion about mental and physical energy, and their importance in measurement.

Regarding memory and concentration questions, the group feeling that whilst both concepts are important and should be retained, they are both part of the same process. Good concentration allows them to better remember things whereas poor concentration can lead to a poorer memory. This was the groups proposed explanation for why items were showing dependency and proposed merging them together to form a single, modified question.

*Modified question:* Has your fatigue made it difficult to concentrate or remember things? (e.g. difficulty concentrating on driving or puzzles, being more forgetful than normal).

Three items: F7 (downhearted), F9 (worried) and F10 (frustrated) showed dependency and I proposed merging the items. The group members discussed the distinctness of these concepts extensively, stating that “downhearted” may not be the right word to capture mood. Members shared their own understanding of the term “downhearted” and it was clear there were multiple interpretations. Members highlighted that feeling down, worried or frustrated felt different to them, but were quite specific and using a broader, more simple term like “mood” would work better. Consequently, the group rejected my proposal to merge the items and opted to modify item F7 (downhearted) and remove items F9 (worried) and F10 (frustrated), instead using these as examples in the modified question. Further, members shared their feelings of guilt associated with fatigue (pulling out of events, cancelling things, letting others down) and noted that although guilt
was omitted in a previous meeting, it fits well in the modified question and would be a good, suitable example. The group unanimously agreed, and the changes were adopted.

*Modified question:* Has your mood been affected by your fatigue? (e.g. feeling low, worried, frustrated, guilty).

Finally, the group discussed the concepts of physical and mental energy, stating they are distinct, important and should not be removed. The group argued that they have clear meaning and afford insight into the overall picture of a person’s energy levels. Therefore, it was agreed that these would not be changed.

**Factor 2: Energy**

Discussions focused on the issue of dependency for multiple items and reducing repetition of social activity items.

Questions E6 (doing usual work) and E7 (doing things you enjoy) were showing dependency and members accepted that the two concepts were similar. They reasoned that doing your usual activity might also include tasks and hobbies and therefore a single, better honed question would be the best solution. Members coalesced around the word “routine” following discussions, and proposed a new, replacement question for both items.

*New question:* Have your energy levels made it difficult for you to keep to your routine? (e.g. usual work at work or home, hobbies, leisure activities)

Again, dependency between two items: E3 (difficult to take care of self) and E4 (difficult to do everyday activities) was observed. The group felt that the new routine item (proposed earlier) would capture item E4 on everyday activity. The discussion then focused on how they “take care” of themselves and whether rewording E3 to “maintaining personal care” would be more easily understood by respondents. Members felt that the term “maintaining” allowed for respondents to change their answer if their energy improved or worsened; therefore, a new item was crafted.

*New question:* Have your energy levels made it difficult to maintain your personal care? (e.g. showering, brushing your teeth, eating your usual meals)
Four questions relating to social activity were discussed: one performed very well (E8: making plans) and three others – E9 (cancel plans), E10 (keep to your plans) and E11 (turn down invitations) – were highly dependent on one another and needed to be reduced. The group felt that making plans was distinct from the focus of the other three items. They felt that the essence of the dependent group of items was to tap into changing plans due to changing energy levels. Therefore, they proposed merging the three questions into a single question that focused on ‘changing plans’ due to changes in energy levels.

*New question to merge E9-11:* Have your energy levels caused you to change your plans? (e.g. cancel or reschedule)

Item F14 (dependent on others) was heavily discussed. The group felt strongly that this was an important question and should not be dropped, however, I explained that the statistical evidence for this item was incredibly poor. I also highlighted the issue raised in the research team meeting; the item may be too broad and capturing non-fatigue issues such as pain, stiffness or challenges with sleep. The group accepted my point and agreed that keeping the current item would not be acceptable, however, they proposed modifying the question to be evaluated in a future field test. Members crafted a modified item that brought fatigue to the beginning of the question and developed the examples.

*Modified question:* Do you feel that your fatigue has made you more dependent on others? (e.g. having to ask for help to do everyday tasks from family, friends, carers)

*Factor 3: Symptoms*

Discussion focused on one poor performing item – E1 (how much energy have you had) – and whether two other items: E14 (lacking physical energy) and E15 (lacking mental energy) were conceptually distinct. The group unanimously agreed with the removal of question E1 due to its poor performance and they felt it was not essential for energy measurement. The group felt that the question about being drained of energy captured both physical and mental energy and provided an energy overview. Questions specific to physical or mental energy were more precise and
conveyed ‘how’ respondents were affected by their energy levels. This
global and specific distinction were arguments members made to retain all
three items.

**Factor 4: Control**

Conversations focused on how important this factor is for measurement
and whether coping and management are different concepts in axSpA.

There was unanimous group agreement that the control factor is important
and should be retained. Members explained that the concept of coping was
about the person and how they were doing personally, whilst management
reflected the practical steps people took to live with their fatigue or low
energy. As only two fatigue-specific control items performed well
psychometrically, it was essential that at least one new energy item was
generated for the factor to meet minimum statistical requirements. The
group deliberated extensively and produced four potential questions,
however, they struggled to agree on which question to use. As further
discussion failed to produce an answer, members stated they were
satisfied with the four items and suggested returning them to the research
team to choose one. The four proposed questions were:

1. Have you been able to maintain your energy levels?
2. Have you been able to control your energy levels?
3. Do you feel you have been able to control your energy levels?
4. Have you been able to maintain your energy levels to achieve what
   you wanted to do?

Following the PRP group meeting, these items were emailed to the
research team. After discussion with everyone, I chose to adopt item 4 due
to it best meeting the purpose of the item, and its greater resonance with
both PRPs and research team members.

*Items removed as part of EFA modelling and response options*

I explained to the PRP group that, following the EFA modelling process,
five items were identified and removed due to poor statistical performance.
Only one question (F8: has your fatigue made it difficult to be in control of
your life?) was discussed at length for potential reintroduction. Members felt this resonated strongly with them and was important to the narrative of fatigue experience. Jane (co-facilitator) and I both explained that this question could be incredibly multifactorial. PRP rewrites of the question did not remedy the broadness of question. Based on the statistical evidence and concerns of poor item focus, it was decided that this item would not be reintroduced.

I explained that the issue with response option endorsements and that frequently, the edge categories were not being used as intended, and the central options (moderately, sometimes) appeared to be overused. The group members felt the middle option was an easy choice for people and a default choice for those not paying attention when completing the questionnaire. It was therefore agreed that the middle option would be removed to leave four response options.

Summary of revisions

Psychometric evaluation using classic and modern approaches, coupled with both the research team and PRP group review has supported refinement of the WASTEd. The new short-form WASTEd is a multidimensional measure of axSpA fatigue and energy made up of 18-items across four domains: energy, fatigue, symptoms and control. A version of the new, short form WASTEd is available in Appendix 5C. A final measurement blueprint derived from the statistical analysis is presented in Figure 5.6.
Figure 5.6: Final measurement blueprint of axSpA fatigue and energy

<table>
<thead>
<tr>
<th>Construct</th>
<th>Conceptual domains</th>
<th>Subdomains</th>
<th>PRO concepts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fatigue and energy in axial spondyloarthritis</td>
<td>Symptoms</td>
<td>Fatigue</td>
<td>Frequency, severity</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Energy</td>
<td>Feeling drained, lacking physical energy, lacking mental energy</td>
</tr>
<tr>
<td></td>
<td>Energy impact (activity)</td>
<td>Physical</td>
<td>Personal care, usual routine</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Social</td>
<td>Making plans, changing plans</td>
</tr>
<tr>
<td></td>
<td>Fatigue impact (self)</td>
<td>Psychological and emotional wellbeing</td>
<td>Mood, need to be alone</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cognitive</td>
<td>Concentration, enjoyment, engaging in conversations</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Support</td>
<td>Feeling dependent on others</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>Coping</td>
<td>Coping with fatigue</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Managing</td>
<td>Managing fatigue, maintaining energy levels</td>
</tr>
</tbody>
</table>
5.5. Discussion
This study used statistical evidence from both CTT and RMT analyses to refine the long-form WASTEd to a good quality, short-form version ready for future testing; as per current recommendations (59, 89). The WASTEd has strong psychometric evidence of its structural validity, internal consistency reliability and item-level performance. Its content is underpinned by extensive qualitative research and captures important aspects of axSpA fatigue and energy experience.

Application of CTT and RMT analysis methods has allowed for a detailed and robust evaluation of the WASTEd, from its overall structure down to how individual items are functioning, and whether response options are performing as intended. Using RMT, it was possible to identify mis-fitting or poorly performing items and dysfunctional response options, supporting the reduction and refinement of the WASTEd from its initial 30-item longform version to an 18-item short form version. RMT analysis has demonstrated that the WASTEd is capable of distinguishing between at least four ability groups for the energy, fatigue and symptom domains. This capacity to distinguish between people across different ability groups on the underlying trait is promising for future psychometric evaluation which would focus on longitudinal measurement properties including test-retest reliability, responsiveness and interpretability.

To date, reliance on the single-item VAS on fatigue severity has meant the multifaceted and wide-ranging impact of fatigue on axSpA patients has not been fully understood or explored. Moreover, PROMs currently used in fatigue assessment fail to capture key features of axSpA fatigue and energy (Chapter 2), frequently fail to adequately conceptualise fatigue and often have limited and poor psychometric evidence (146). This gap in measurement can now be addressed with the new short-form WASTEd: a patient co-produced, 18-item multidimensional PROM developed to specifically measure axSpA fatigue and energy. It consists of four domains which measure fatigue, energy, symptoms and control. For the first time in axSpA fatigue measurement, researchers and clinicians will be able to use a specific measure developed for this population that has demonstrable evidence of its content and psychometric quality. In addition, the ability to produce domain scores will provide healthcare professionals with detailed information about how people with axSpA are affected by their fatigue and
changing energy levels, supporting the provision of targeted and tailored care.

In statistically refining the WASTEd, my study has demonstrated that patient partners have an important role in decision-making and are integral to the process, supporting item reduction whilst ensuring content validity of the PROM is retained. This extends on other PROM development studies such as the BRAF-MDQ, whereby patient involvement during the preliminary psychometric evaluation was limited to labelling domains and ordering the items for presentation (106).

The four domains of the WASTEd contain items reflective of the measurement framework derived in Chapter 3, each which could have significant implications on how routine clinical practice and healthcare provision is tailored and delivered. The energy domain has items that capture the impact of changing energy levels on daily and leisure activity, social activity and usual work. Despite these issues often being described as related to fatigue (33–36), the EFA-derived domains support the qualitative evidence that underpins the WASTEd: energy is a distinct, related component of fatigue that relates to the person’s ability to be physically and mentally active. Better understanding the concept of energy and its relationship with exercise – a common intervention recommended for axSpA symptoms (26,332) and recently demonstrated as being effective for axSpA-fatigue (333) – could help identify patients less likely to adhere to their axSpA-exercise regime and thus be more likely to show worse health outcomes. Moreover, there is evidence from HIV research that suggests fatigue has different diurnal patterns which were associated with different reported impacts of fatigue (334). This could not be explored in this cross-sectional survey, however, the diurnal patterns of fatigue and energy in axSpA – and their potential clinical implications – should be explored in future research.

The fatigue domain contains items on memory, mood, talking with others and enjoying things. These items reflect the conceptualisation of fatigue as internal for the participant, predominantly affecting cognitive and psychological/ emotional wellbeing, as described in Chapter 3. These items provide insight into any practical implications of fatigue on cognitive function, as well as whether the participant is struggling psychologically or
emotionally. A recent meta-analysis has reported that depression in axSpA affects between 38% and 52% of patients (335). This suggests that the inclusion of the domain can address a current gap in standard measurement in axSpA, supporting both the identification and understanding of how patients are affected their fatigue and its impact on their psychological and emotional wellbeing, enabling the provision of tailored care and support.

The symptoms domain contains items that capture the frequency and severity of the fatigue and energy, two components known to be important to detect patients experiencing major fatigue (frequent and severe) (40). The addition of energy items may enable better identification of individuals suffering from both major fatigue and significantly reduced energy levels, thus most likely to struggle with activities of daily living. Thus, this can contribute to a fuller profile of the patients experience of axSpA fatigue and changing energy levels.

A key strength of this study is the utilisation of both classic and modern psychometric methods to evaluate and refine the WASTEd: CTT and RMT. This contrasts with the CTT-only approach adopted in other PROM development studies (106,124). CTT analysis identified only one problem with the WASTEd: problematic items that were cross-loading and therefore not clearly providing useful measurement information, thus undermining validity. However, RMT highlighted multiple problems including an under discriminating item, item dependency and disordered thresholds for most items which all could undermine validity and reliability.

Odd-numbered Likert scales are known to present issues with the middle category option, thus warranting careful development and testing to ensure the scale performs as intended (336). This problem was observed in the WASTEd, however, using RMT during the development stages has ensured this was identified early and can be rectified prior to a comprehensive psychometric evaluation. Existing PROMs that use Likert scales but have not used modern psychometric methods to evaluate how well they function are failing to demonstrate that their response scale works as intended. Using both CTT and RMT has ensured a robust evaluation of the WASTEd and supported better informed item reduction decisions.
The control domain can enable professionals to tease apart patients who are coping and not coping, and those who are managing and therefore have developed practical ways of living with their symptoms of fatigue. However, the current absence of an energy management item that performs well statistically needs to be addressed in future pretesting and subsequent psychometric evaluation. Moreover, for this domain to meet statistical requirements (331), a third item is necessary – its current retention in this study is driven by the qualitative argument that this is an important patient-derived outcome for measurement.

There was some evidence that the energy domain was failing to target the higher end of the continuum, however, the source of this issue is unknown. There are two potential explanations: firstly, there was no one to sample at the higher end of the continuum, or alternatively the current item set does not target the higher end of the continuum. These are important considerations for future evaluation and should be explored in a new field test with a new sample.

This study was a cross-sectional evaluation using a nationally recruited, purposive sample identified using existing clinical databases. Whilst strengthened by its representativeness, there was an almost 2:1 male/female ratio and participants tended to be older – however, there was a better balance of representation in this study compared to a HRQoL measure developed for axSpA (62). Nonetheless, future psychometric evaluations need to include a greater proportion of females to ensure the WASTEEd performs as intended in all respondents.

A second methodological limitation in this study is how disease severity was determined for the sample – using biologics eligibility as a proxy of disease status. This may not be an accurate assessment of how a patient feels about their health, or accurately reflect those with severe, or without severe illness. Moreover, groups would contain a mix of patients who are or are not taking biologics meaning the groups are not as distinct as it would seem. Nonetheless, this was a useful way to group patients to begin exploring how the measure performs but should be considered in future field tests.
5.6. **Next step (outside of the PhD)**

The current study has provided a preliminary evaluation of the WASTEd and its data quality, structural validity and internal consistency within a cross-sectional application. Further pretesting work is necessary to appraise all newly crafted or modified items for their content validity and interpretation and to establish evidence of its longitudinal properties (test-retest reliability, responsiveness). This evidence is essential to underpin its uptake and usage in clinical practice and research.
6. Chapter 6: Overall discussion of the thesis

An overview of the thesis, its findings and its contributions to knowledge are outlined in section 6.1. The strengths of the thesis, a detailed discussion of how it contributes to knowledge and my reflections on the research process – including what I would do differently – are described in section 6.2 and the limitations of the research process are presented in section 6.3. Future research initiatives, implications for clinical practice and education are described in section 6.4, and the chapter and thesis end with a conclusion presented in section 6.5.

6.1. Overview of thesis

The aim of this thesis was to explore the experience of fatigue in axSpA patients to inform the co-production of a high quality and relevant axSpA-fatigue specific PROM that is suitable for use in routine clinical settings and research. Fatigue was highlighted as an important symptom of axSpA almost two decades ago (77) and in 1994, the development of the BASDAI (76) – which included a single item on fatigue severity – became the recommended assessment method for axSpA fatigue (2). However, the research reported in this thesis has demonstrated the insufficiency of relying on a single-item VAS on fatigue severity; instead, the multiplicity of fatigue experience and wide-ranging physical, social, psychological and emotional impact has been demonstrated (chapter 3).

This thesis consisted of four phases which sought to understand and conceptualise fatigue experience in axSpA, identify what is important for measurement to underpin the development of an axSpA-fatigue specific PROM – the WASTEd. A brief summary of each phase of the thesis (Chapters 2-5) and the original contributions of this thesis, to knowledge, are described below.

Phase 1: identifying a need (evidence synthesis)

This thesis provides a systematic evidence synthesis that determined whether: (i) there was an existing fatigue measure of suitable relevance and quality; (ii) whether an existing measure was partially fit for purpose...
and could be modified for use in axSpA-fatigue; or (iii) no existing PROMs are appropriate for use in axSpA-fatigue assessment, requiring the development of a new PROM, or modification of another measure (e.g. the BRAF-MDQ (106)).

The findings identified nine PROMs; one with evidence of practical evaluation and eight with evidence of psychometric evaluation. Of these measures, evidence of their quality was limited, disparate and poor meaning no PROM could be recommended for use. In addition, no identified PROM reflected the components of the RA-fatigue conceptual model raising concerns about the content validity of these measures. This study adds to current knowledge by identifying the lack of a suitable (high quality and relevant) measure for use in the assessment of fatigue in axSpA.

In the absence of a relevant, high quality measure of fatigue for use in axSpA, it was then essential to develop an understanding of axSpA and fatigue experience. This was done through a qualitative exploration of the patients' lived experience of fatigue in axSpA. The primary outcomes of this study were to develop our understanding of axSpA and, the associated fatigue experience to produce a lived experience framework – something that was missing in the literature. Moreover, this would inform the development of a measurement framework to underpin the content of a new axSpA-fatigue specific PROM; addressing the issue of poor content validity and relevance of currently used PROMs identified in the systematic review.

**Phase 2: conceptualising fatigue in axSpA**

This thesis provides an in-depth qualitative exploration of the patients' lived experience of fatigue within the experience of axSpA. A scoping review of patient experience within qualitative research highlighted a very limited qualitative evidence-base of only four axSpA-specific studies, most of which had an implied assumption that fatigue experience could be dissociated from axSpA and thus investigated (and understood) separately. This failed to recognise the complexity of living with axSpA-fatigue, the interaction between disease and fatigue experience and meant findings about axSpA-fatigue may wrongly identify the impact of axSpA as a fatigue
issue, or vice versa. Drawing on an IPA approach, face-to-face semi-structured interviews were used to:

i. Develop a deeper understanding of the patients' lived experience of axSpA-fatigue in daily life.

ii. Inform the generation of a measurement framework of fatigue in axSpA grounded in the patients’ experiences and perspective.

Findings from the patient interviews adds to existing knowledge through the identification of the overarching concept of achieving balance expressed through three superordinate themes: having energy, engaging in everyday life and living with axSpA. An important finding was that energy was perceived as a separate and distinct component of fatigue experience. Energy was defined as an internal resource that participants draw on to be physically and mentally active, enabling them to be engaged in, and live their lives (Chapter 3) – similar to one proposed in HIV research (209). The concept of energy was investigated further in the development of a PROM specific to axSpA-fatigue through focus groups and pretesting interviews (Chapter 4).

In addition to the lived experience data, further knowledge was developed in the form of an axSpA fatigue and energy measurement framework. The lived experience data and evidence synthesis was used to generate a model with five domains: symptoms, impact on self (the patient), function, psychological/emotional wellbeing and self-management. A key finding was the contested place of sleep within the framework. Whilst an important component in modulating fatigue experience, its contribution to meaningful measurement was unclear (i.e. is it too multifaceted for measurement? Would it be reliable, responsive and would people answer it about their fatigue only?). Therefore, greater interrogation of sleep was undertaken in focus groups with professionals, patients, pretesting interviews and with the PRP group. Additionally, it was recognised that there was a lack of similarity between the RA-fatigue model (50) and the new axSpA-fatigue and energy measurement framework; for example, the theoretically distinct meaning of ‘energy’ and the inclusion of self-management as a concept for measurement. The measurement framework was used to develop a new PROM specific to axSpA-fatigue and energy.
Phase 3: developing and pretesting the WASTEd

This thesis adds to knowledge through the provision of a new 30-item, long form PROM (WASTEd) to assess axSpA-fatigue. PROM development and refinement was conducted across three phases: (i) item generation (PROM mapping); (ii) confirmation of the axSpA-fatigue and energy framework, and refinement of an item set following focus groups, and (iii) pretesting the new PROM.

i. An item-set (n=32) was generated through consideration of the qualitative work conducted as part of this thesis (Chapter 3) and as per good guidance recommendations (60,89), reviewing and extracting items from fatigue PROMs identified in the systematic review (Chapter 2) (146) and from other published reviews (150–152).

ii. The measurement framework and its content were endorsed across focus groups and only minor modifications were proposed. The sleep domain was identified as potentially problematic due to concerns it may capture more than just fatigue. However, it was considered prudent to test items in pretesting interviews to determine how sleep items were interpreted by respondents and whether they could provide meaningful measurement data. Important outcomes for both patients and professionals were identified, and revisions, informed by focus group findings, were made to the item-set for PRP group review and development/modification in anticipation of pretesting interviews.

iii. Pretesting interviews used two approaches (cognitive and TSTI to conduct both item-level and PROM-wide evaluation, utilising verbal probes co-produced with the PRP group. The comprehensiveness, relevance, acceptability and comprehension of the PROM tested, and item-revisions were made drawing on three sources of input: patients as participants, the research team and patient research partners. At this stage it was clear the sleep items and domain were not performing well even following revisions, confirming previously raised concerns about interpretation. A final 30-item long-form
PROM named ‘WASTEd’ was produced ready for psychometric evaluation.

**Phase 4: psychometric evaluation of the WASTEd**

This thesis adds to knowledge through the provision of a short form 18-item PROM (WASTEd) to measure axSpA-fatigue. A preliminary, cross-sectional psychometric evaluation of the WASTEd was conducted in a large UK sample of axSpA patients. This evaluation used both classic and modern psychometric methods to support statistically informed item refinement; producing a short form 18-item WASTEd.

CTT analysis produced a four-factor model and the factors (domains) were labelled: fatigue, energy, symptoms and control. Five items were removed following CTT analysis, seven items were removed following RMT analysis, and one item on energy management (“to achieve what you wanted to do”) was added to the measure by the PRP group following review. The findings of the analysis were discussed in both research team and patient partners group meetings. In these meetings, statistically identified changes were considered against the measurement framework and qualitative work completed in Chapter 3 to ensure content validity of the revised short form WASTEd. Following item deletions, modifications and additions from the research team and PRP meetings, I then synthesised these recommendations to create the final 18-item short-form WASTEd.

6.2. **Strengths and contributions of this PhD**

This thesis adds to knowledge at all stages; the incorporation of a PRP group as co-production partners is described first, including reflections on how I would do this differently in future. Rigour – as per current best practice and reporting guidance – is then described, followed by the contribution to knowledge made at each phase of the study.

6.2.1. **Incorporating active involvement of a patient research partner group**

This thesis adds to knowledge through the active participation of user involvement in the research process as co-producers of the WASTEd.
There is currently no agreed approach to how best to incorporate PRP involvement in PROM development despite it being advocated (132–134). Substantial disparity exists in current PROM development practice which diminishes the evidence-base from which a high-quality approach can be derived from. As described in the opening chapter, studies report a variety of approaches to PRP involvement from none or minimal involvement in inconsequential research tasks (106,124,125) through to embedded PRP at all stages (105). This thesis better aligns with and supports the embedded PRP approach taken in the development of the PsAID (105), and further demonstrates that PRP can be incorporated throughout the PROM development process to meaningfully contribute to research activity, data analysis and the co-production of a PROM. However, whilst the PsAID had high levels of PRP involvement throughout the development, there was no PRP reported in the psychometric evaluation of the measure (105). My thesis has demonstrated that PRP members have an important and valuable voice in statistical refinement, challenging decisions that could undermine content validity and co-producing new items, or modifying existing, to ensure changes retain the content validity and relevance of the measure.

Currently recommended measures for use in axSpA such as the BASDAI have failed to engage with patients as partners and as demonstrated in this thesis, the single item VAS on fatigue severity fails to adequately capture the multiplicity of fatigue experience, or measure what matters to patients about their fatigue. The BASDAI claims to have been developed with ‘major input’ from patients with AS, but this input is not described and seems to merely refer to 154 patient participants who completed the questionnaire as part of its psychometric evaluation (76). There is no indication that patients were consulted as research partners or had input in the research process, data analysis or item generation and refinement. This contrasts with the development of the EASi-QoL where over 550 patients were involved, as participants, in the development stages of the measure (81,337). Development included a national survey and semi-structured interviews which informed the item content of the questionnaire, enhancing the relevance of the measure and its content and better capturing the patient perspective. However, my thesis goes further to incorporate PRP into the development and co-production of a measure – transitioning PROM
development away from reliance on patients as participants and toward a development process that is for patients, by patients.

INVOLVE advocate for greater PPI and regarding co-production, specifying principles of sharing power, including all perspectives and skills, respecting and valuing everyone and their knowledge, reciprocity and relationships (137); principles echoed by other researchers too (338). These principles are difficult to demonstrate as they predominantly speak to the dynamics of a group and how PRPs are engaged with – however, the richness of the PRP contribution in this study and their visible role in a range of research tasks supports the notion that this thesis followed the principles of co-production. The groups involvement in the study was enabled through frequent face-to-face meetings, dissemination of post-meeting summaries (beneficial also for those who could not attend all meetings and new members), email updates and communication through Dr Jane Martindale (PRP group facilitator).

The PRP group represented a diverse range of health conditions and experiences and brought a unique experiential perspective to the study. Their involvement includes: developing and refining topic guides; supporting ethics applications (including reviewing patient information sheets, testing the acceptability of the questionnaire package (chapter 5) and assisting with testing research processes); analysing qualitative data and reviewing lived experience measurement frameworks; developing verbal probes for pretesting interviews; reviewing, modifying and co-producing items for the WASTEd and making item development and refinement decisions whilst considering psychometric evidence. This goes beyond what has been described in other recent PROM development studies (107,108,124,125,143) and ensured the ‘P’ was firmly placed in ‘PROM’ when developing the WASTEd, moving decisively beyond only including patients as participants to ‘co-production’ with patient research partners (137). However, whilst this thesis had a clear and consistent PRP presence compared to other recent PROM development studies, there is scope for greater involvement.
Reflections and recommendations for PRP in PROM development

On reflection, there are two ways in which the input of the PRP group could have been developed, further enhancing the quality of the research output: (i) including them within the item-content appraisal completed in the systematic review (Chapter 2) (146); and (ii) during the PROM mapping process to further enhance face and content validity of the item-set.

The comparative item-content appraisal used the existing RA-fatigue model (50) to judge relevance and face/ content validity of the PROMs identified in the review (146). However, on reflection, I now see three ways in which the PRP group could have been incorporated into the review process to further align the item-content appraisal with the patient perspective:

i. PRP members could have been invited to identify which domains of the RA-model they felt were reflective of, or relevant to axSpA fatigue. RA-fatigue domains could then have received PRP-endorsements to support interpretations.

ii. If domains were thought to be missing or not captured in the existing model, members could have generated their own domains for item-content to be appraised against. This would have supported a more robust item-content appraisal that incorporated the patient perspective.

iii. A face-to-face group meeting could have been facilitated to review the item content of the PROMs and assign them to RA-fatigue conceptual domains, producing a PRP-led content appraisal rather than a researcher-led appraisal.

In the development of the item-set for review in focus groups, items could have been pooled and reviewed in a PRP meeting with members performing the PROM mapping process. This could have further enhanced the face and content validity of items linked to their PRO concepts through the inclusion of the PRP perspective, and ignited PROM content conversations earlier within the WASTEd’s development process. Earlier involvement would better align with the principle of power sharing (137,338) and something I would advocate future PROM development researchers to do.
Throughout the thesis I followed current best practice and reporting guidance to ensure maximum transparency and quality in the research (Chapters 2-5). The overall structure of the development programme followed best practice guidance for PROM development \((60,89)\). The lived experience interviews ensured rigour through application of principles specified by Yardley et al \((241)\); the interview study also used reporting guidance from the COREQ checklist \((242)\), and I drew on the cognitive reporting framework for pretesting interviews \((286)\). Moreover, during the qualitative phases of the research, another researcher was involved in coding, analysis, interpretation of findings and developing a lived experience framework. This ensured rigour through reflection and discussion of the emerging findings. This combined with immersion in the data and provision of a clear audit trail ensures the findings are trustworthy \((339)\). In the quantitative research the systematic review followed current guidance from COSMIN to appraise study methodological quality \((102,103)\) and used a synthesis of current recommendations to support transparency in measurement property quality appraisal \((97,98,163,176)\). Further, the postal survey utilised both CTT and RMT to ensure a thorough and robust evaluation of the WASTEd, as recommended \((59)\). Finally, although short-form tables are not completed and provided in this thesis, all PPI elements of the study were performed using the GRIPP2-SF checklist \((145)\) as a guide to ensure purposeful engagement, alongside drawing on the principles of co-production proposed by INVOLVE and others \((137,338)\). This is evidenced by the inclusion of a clear aim of PPI in each study method, and a qualitative reporting of their involvement, important contributions, and the discussion and reflection of PPI in this current chapter.

6.2.2. Summary of knowledge contributions per phase of study

A short summary of the new contributions to knowledge, per phase of the study, is provided below. This includes both empirical and methodological contributions, however, PRP will not be covered as it has already been extensively discussed.
Phase 1: the systematic review of measurement properties addressed a gap in the literature by providing clear evidence of the quality and relevance of existing fatigue measures used in axSpA fatigue assessment. This review highlighted the poor quality and limited evidence-base from which interpretations could be drawn. The use of item-content appraisal in this review, using the RA-fatigue model as a conceptual framework to comparatively appraise item content, is something that has not yet been done in reviews of fatigue measures. This method allowed for content validity to be assessed in a clear and transparent way and despite the recent update to the COSMIN checklist (104), this checklist does not provide a method for appraising item content.

Phase 2: the lived experience interviews highlighted the importance of energy as a related but distinct component in fatigue experience, from the patient perspective. This research informed the development of two frameworks: a lived experience framework and measurement framework; both of which were absent in axSpA fatigue literature. The derivation of a measurement framework from a methodologically strong qualitative evidence base provides a content appraisal framework for future systematic reviews of fatigue PROM quality in axSpA. Moreover, this framework can be empirically assessed in other cultures to investigate its cross-cultural applicability.

Phase 3: a qualitatively refined 30-item questionnaire specific to axSpA-fatigue was generated following a process of PROM mapping, focus group review (incorporating the important clinical perspective into the measurement framework and developing PROM) and pretesting interviews. Two approaches to pretesting were used with clearly defined purposes and processes which has not yet been reported in other studies. Reporting guidance (CIRF (286)) was used to support transparency in the PROM development process – as per current development guidance (89). Inclusion of self-management and coping items and understanding their conceptual distinction enhanced the relevance and utility of the WASTEd, supporting measurement which can better inform the provision of targeted and tailored care. This contrasts with other fatigue PROMs which do not include self-management or coping items (42,61,106,177,179).
Phase 4: psychometric evaluation in a large national sample of axSpA patients used both CTT and RMT, demonstrating the added value of using modern methods during the PROM development process. The statistically refined 18-item WASTEd has good quality evidence – addressing the insufficiency and inadequacy of current measurement reliance on the single-item VAS on fatigue severity, and providing a multi-item measure of fatigue that is relevant and has good preliminary psychometric evidence supporting its quality – addressing the measurement gap identified in Chapter 2.

6.3. Limitations

Despite my best efforts and those of the recruitment teams at each site, recruiting female participants was particularly difficult. Almost half the participants in the lived experience interviews were female, however, this recruitment challenge quickly became pronounced at the point of conducting focus groups and even more so when recruiting for pretesting interviews. A female-only and intensive recruitment effort over the course of a month resulted in no new female participants for the study. Further to this, there was almost a 3:1 male to female ratio recorded in the postal survey. Similar gender ratios have been reported for interviews in one UK study interview study (36) and in the development of an axSpA HRQoL-specific PROM (62). Only two studies conducted in Norway are known to have reported a majority-female axSpA sample (33,206).

This issue may have arisen due to recruitment and interviewing taking place over a defined period, potentially limiting the number of females participating. The use of a sampling matrix in the qualitative interviews was effective in helping to offset this issue, helping me to identify who was participating and understand their demographic details, supporting purposive sampling. The implications of this are thought to be minimal due to a balanced female involvement in the interviews which informed the development of the measurement framework that underpinned the content of the WASTEd. Moreover, there was a strong female presence and voice throughout all PRP group meetings meaning a female perspective was reflected throughout the development of the WASTEd.
Unfortunately, the impact of severe weather conditions on three of the four focus groups could not be anticipated and I recognise that bigger focus groups could have produced different results. Despite this limitation, the involvement of a strong professional group in reviewing and refining the measurement framework ensured that it was clinically relevant, informing the development of items and a PROM that is relevant and acceptable to both end-users.

The RMT analysis of the WASTEd has identified an interesting question regarding item targeting for the energy domain which warrants further investigation. This domain showed that many people reported higher levels of the underlying trait, but most items were clustered further down the continuum highlighting a potential item-targeting issue. There are two possible explanations for this; neither of which can be answered based on the current cross-sectional evaluation alone. Firstly, the sample could just be more heavily affected on the domain and therefore meaning this is a sample-specific issue. An alternative explanation could be that the WASTEd possesses insufficient items at the mid-to-high end of the continuum on this domain, meaning it may not be able to discriminate well, between participants, across the continuum. Irrespective of the cause, this issue can be rectified, however, it first needs to be assessed in another psychometric evaluation to identify what the issue is before a solution can be reached.

### 6.4. Future research initiatives

There are a range of future research initiatives that could further develop on the work of the WASTEd. These include a comprehensive psychometric evaluation to finalise the WASTEd for use in clinical practice and research; cross-cultural translation studies to explore the relevance and acceptability of the WASTEd in different cultures; and integrating the WASTEd into clinical practice and research.

#### 6.4.1. Comprehensive psychometric evaluation of the WASTEd

The cross-sectional psychometric evaluation conducted as part of this thesis (Chapter 5) was an important first step in elucidating the internal
structure of the WASTEd and refining its content to produce a short-form version. Evidence is now needed to demonstrate the validity, reliability (test-retest) and responsiveness of the WASTEd to ensure it is suitable for use in clinical practice and research.

A future field test should be a comprehensive, longitudinal psychometric evaluation conducted in a large, representative sample and, to ensure suitability for use in research and clinical practice, should evaluate essential measurement properties of validity, reliability, responsiveness and interpretability (59). Both CTT and modern psychometric methods such as RMT (296) and item response theory (IRT) (118) should be used to provide a robust evaluation of the PROM (59). Additionally, changes made in the preliminary evaluation – such as removing the middle response option to encourage respondents to use the response scale – can be evaluated to ensure they had the intended effect. This evaluation will further establish evidence of the internal structure of the WASTEd whilst providing essential evidence necessary for its uptake in clinical practice and research. I currently have ethical approval to conduct the comprehensive evaluation of the WASTEd in England (until April 2020).

### 6.4.2. Cross-cultural translation studies

Cross-cultural translations have been successfully completed for a variety of developed PROMs, including (but not limited to) the BASDAI and BASFI (340), and SF-36 (341). Current guidance specifies that evidence needs to be established that demonstrates any translated version of a PROM should perform the same as the original (89). To ensure the WASTEd is suitable for use in other countries, cross-cultural translation studies are needed to ensure the PROMs relevance and acceptability in a new language or country, and confirm the measure works as intended. This would include confirming the measurement framework, identified PROs and demonstrating the quality of the measure. Consequently, such studies present new development needs and challenges which range from the practical (translating the language, terms, formatting) through to theoretical (e.g. is energy conceptually distinct and important to fatigue experience in other cultures?) and relevance of the measurement framework (e.g. does the axSpA fatigue and energy framework reflect fatigue experience in other
cultures?). To date, the WASTEd has received international interest for possible translation studies.

6.4.3. Implications for clinical practice

The current 18-item WASTEd has good evidence of initial quality and is short enough to be used routinely with an estimated completion time of 2-5 minutes; this is in keeping with the BRAF-MDQ – an RA-fatigue specific PROM (106). The four domains of the WASTEd provide clear utility within a clinical setting or research study, providing specific scores on key domains that can support the provision of tailored care. This is most applicable to the psychological and emotional wellbeing domain which healthcare professionals highlighted (in focus groups) was poorly captured and not readily measured in clinic. The WASTEd enables clinicians to better understand how their patients are experiencing, affected by and able to manage or cope with their fatigue, assisting identification of patients most affected by their fatigue or low energy, and support the delivery of targeted interventions to reduce the burden on patients.

The inclusion of a self-management domain in the initial measurement framework, and the statistically derived domain on control, the WASTEd becomes increasingly distinct from existing fatigue measures. Of the nine measures identified in the systematic review described in Chapter 2 (146), none of them contain items that ask whether respondents are coping with or managing their fatigue. Three of the PROMs identified in the review (MFI-20, FACIT-fatigue and MFSI-SF) were designed specifically for cancer-related fatigue – yet, despite qualitative research specifically exploring self-management or coping in cancer (342,343), these important concepts are not reflected in the measures. The omission of such an important component in fatigue experience exemplifies the challenges associated with relying on legacy measures; namely, their poor development processes that fail to incorporate and adequately capture the patient perspective and thus, are of questionable relevance (as demonstrated in chapter 2 (146)).

Self-management has been specifically explored in two UK axSpA studies (35,36) and was identified as important throughout the development process of the WASTEd. Readily capturing this information in a routine
assessment can help healthcare professionals to identify patients most severely affected by fatigue and contextualise their reports of how they experience and are impacted by fatigue and low energy. Further, those struggling to cope or manage could be targeted with specific education, training and support programmes to better equip them to manage their fatigue and energy levels. Many programmes exist and a recent feasibility study exploring the “Inpatient energy management education” designed for use in MS fatigue has exciting potential (344). This type of training programme could be the focus of future research to evaluate its acceptability and effectiveness for use in axSpA fatigue and energy. Moreover, research could seek to evaluate other existing interventions to identify a “best” fatigue and energy intervention for use with axSpA patients.

6.4.4. Challenges with mental health and emotional wellbeing

In each study in this thesis, almost half of the participants reported experiencing at least mild depression or anxiety, echoing anxiety and depression estimates in axSpA (254–259). However, despite this substantive impact of axSpA and fatigue on mood and emotional wellbeing, there is little existing experiential evidence about the impact of axSpA and fatigue on mental health and emotional wellbeing. Evidence indicates that depression, particularly in younger patients, is associated with worse scores across disease specific measures (BASDAI, BASFI) (256,257,259) and can negatively impact on other symptoms such as pain, or more broadly on HRQoL (256,257). The reported levels of anxiety and depression in this study cohort further highlights this serious issue and warrants further investigation. Qualitative evidence in this thesis (chapter 3) demonstrates that fatigue in axSpA can negatively affect psychological and emotional wellbeing and patients reported a link between fatigue and mood – a finding that corroborates findings of existing UK qualitative studies (35,36). By using the WASTEd, it will now be possible for healthcare professionals to detect patients experiencing low mood due to fatigue and enable them to direct patients to appropriate care and support services in a timely manner. Moreover, these items can also facilitate research that
explore the relationship between important fatigue outcomes and mood to identify potential targets for interventions.

6.4.5. Implications for research and PROM development

There are two implications of this thesis for research: the first is to lay the foundation for potential methodological guidelines or recommendations for future PROM development studies for how to best capture the patient perspective and their lived experience. The second is the possible development of PRP-specific guidance for how best to incorporate a PRP group into PROM development which typically follows the four phases of development described in Appendix 1A.

Despite there being recommendations for PRO development to draw on a phenomenologically derived framework (123), there is still a tendency for researchers to conduct large numbers of interviews, with patients, analysed using thematic analysis. This approach fails to adequately capture or reflect the lived experiences or perspectives of patients due to theme derivation often being descriptive rather than interpretative. In this thesis, the findings of Chapter 3 – using an IPA methodological underpinning – produced themes that captured processes of living and ‘being’ with axSpA and fatigue which are distinct from the more pragmatic, descriptive themes often observed in fatigue studies. This produced results that deviated away from the central domains of physical and mental fatigue, which are commonly prescribed in fatigue studies, and instead drawing out important experiences for patients, such as the concept of energy which appears to have only been conceptualised in one other disease group (HIV patients) following a detailed concept analysis (209). This thesis demonstrates the value of using a phenomenologically based methodological approach to qualitative interviews and I would encourage researchers to invest the time in developing a measurement framework that is meaningful to and captures what matters to patients about their experience.

Since the introduction of the FDA guidance in 2009, PROM development studies often follow a four-phase process (as in this thesis) to identify the need for a new PROM, conceptualise what the phenomenon is, craft an initial PROM to capture important PROs and evaluate the quality of the PROM (89). This current best practice guidance has provided a relatively
standardised approach to PROM development which may now present an
opportunity for the development of best practice guidance for PPI
involvement in PROM development studies. For example, the reflections in
this thesis of how PRP members could be included in a complex systematic
review of PROM quality, relevance and acceptability provides a foundation
for the potential development of PPI-in-PROMs methodological guidance.
Such guidance would assist in moving the PPI agenda forward in PROM
development, supplementing current guidance on how to interact with a
group with recommendations for how to best incorporate PPI from a
methodological perspective. This guidance could better enable researchers
with little or no PPI experience to meaningfully incorporate a PPI group into
their research with the medium-to-long term implication of advancing the
evidence-base to support future evaluative reviews of PPI impact.

6.4.6. Implications for intervention studies for axSpA fatigue

PROMs with demonstrable evidence of validity, reliability and grounded
within the patient perspective are essential for clinical practice and research
(112,262,345,346). Currently, there is no known treatment for fatigue in
axSpA and no known studies that address the distinct energy component of
fatigue experience. The WASTEd is a well-developed, good quality PROM
that provides a means to assess what matters to patients about their
fatigue and has demonstrable evidence of its face, content and structural
validity. The WASTEd provides a way to evaluate the effectiveness of
treatments or interventions across PROs meaningful to patients.

A recent scoping review of existing systematic reviews on fatigue
interventions for long term physical health conditions explored the
effectiveness of both pharmacological and non-pharmacological
interventions across a range of chronic conditions (including RA) (347).
However, the review also highlighted that there was a difficulty in
determining whether any observed changes were of minimal, clinically
important difference. Moreover, there were no reports about interventions
or their effectiveness for axSpA fatigue. The limitation of this review, and
the lack of evidence for axSpA fatigue interventions may relate to the lack
of good quality fatigue measures that capture the multiplicity of fatigue
experience, what matters for patients and capturing that from the patient
perspective. The WASTEd provides essential information to inform an evidence-based evaluation and selection of appropriate fatigue interventions. There is demonstrable evidence of the WASTEd’s content validity and quality evidence is good. A second comprehensive evaluation will provide evidence of the WASTEd’s the ability to detect and measure change across key fatigue-related outcomes specific to axSpA fatigue experience.

6.5. Conclusion

Current assessment recommendations for axSpA fatigue are restricted to a single-item VAS on fatigue severity (2), and both single and multi-item measures used in axSpA fatigue assessment are of limited quality and relevance (146). This thesis identifies the complex, multidimensional symptom experience of axSpA-fatigue, including its substantive impact on patients that permeates all spheres of their lives. The lived experience of axSpA centres around a difficulty with achieving balance in symptom experience, impact and efforts to self-manage. The importance of energy emerged as a distinct but related component of axSpA fatigue – a finding that has not yet materialised in other qualitative studies in axSpA.

The need for a good quality and relevant measure of fatigue that captured the nuances of axSpA-fatigue experience was demonstrated in this thesis. The development and refinement of a measurement framework to underpin the content of a new, axSpA-fatigue specific PROM – the WASTEd – provides such a measure and addresses the inadequacy of current measurement with a PROM that is demonstrably relevant to patient needs and has good initial evidence of its quality.

Development and refinement of the WASTEd has produced an 18-item multidimensional measure of axSpA-fatigue that can be used in the development and assessment of interventions, support clinical decision-making for targeted supported tailored to the needs of patients, and better inform treatment decisions in relation to fatigue.
References


9. Rudwaleit M, Khan MA, Sieper J. The challenge of diagnosis and classification in early ankylosing spondylitis: do we need new


27. Taurog JD, Chhabra A, Colbert RA. Ankylosing Spondylitis and Axial


55. Kvien TK, Heiberg T. Patient perspective in outcome assessments--


75. NHS Information Centre. National Patient Reported Outcomes Programme [Internet]. Available from: www.ic.nhs.uk/proms


Available from:


265


126. Patrick DL, Burke LB, Gwaltney CJ, Leidy NK, Martin ML, Molsen E,


134. Staniszewska S, Adebajo A, Barber R, Beresford P, Brady L-M, Brett


http://bmjopen.bmj.com/content/3/8/e003311.abstract


158. COConsensus-based Standards for the selection of health Measurement INstruments (COSMIN) website.


175. Terwee CB, Jansma EP, Riphagen II, de Vet HCW. Development of


183. Durmus D, Alayli G, Cil E, Canturk F. Effects of a home-based exercise program on quality of life, fatigue, and depression in


https://www.ncbi.nlm.nih.gov/pubmed/22985544


220. Smith J. Beyond the divide between cognition and discourse: Using


233. Chun Tie Y, Birks M, Francis K. Grounded theory research: A design framework for novice researchers. SAGE open Med [Internet]. 2019


239. QSR International Pty Ltd. NVivo qualitative data analysis software. 2014. Version 10.


243. Brophy S, Calin A. Definition of disease flare in ankylosing


275. Ritchie J, Lewis J, Nicholls CM, Ormston R. Qualitative Research


278. KROSNICK JONA, ALWIN DF. AN EVALUATION OF A COGNITIVE THEORY OF RESPONSE-ORDER EFFECTS IN SURVEY MEASUREMENT. Public Opin Q [Internet]. 1987 Jan 1;51(2):201–19. Available from: https://doi.org/10.1086/269029


290. Rapley T, Girling M, Mair FS, Murray E, Treweek S, McColl E, et al. Improving the normalization of complex interventions: part 1 -


298. Sepp A. Online versus postal questionnaires: a comparison between two data collection methods [Internet]. Actas do VIII Colóquio de


Available from: https://doi.org/10.1186/s12955-015-0337-y


Appendix 1A: Flowchart of project stages

**DEVELOPMENTAL PHASES (1 – 3)**

1. Identifying need
   - i) Semi-structured interviews with axSpA patients
   - ii) Separate focus groups with patients, or healthcare professionals and researchers

2. Developing conceptual model
   - Long-form PROM created using existing items from fatigue PROM’s or newly generated items to reflect conceptual domains

3. Item generation, reviewing and modifying the PROM
   - Pretesting interviews with patients using think-aloud and verbal probing techniques

**PSYCHOMETRIC EVALUATION (PHASE 4)**

Field Test 1 (long-form)

Long-form PROM finalised for psychometric evaluation

- Preliminary evaluation (long-form)
  - Cross-sectional postal survey with axSpA patients using classic and modern psychometric methods to evaluate the measure and support an evidence-based reduction of PROM length.
  - Evaluated properties include: Data quality (end-effects; missing data); Structural validity; Internal consistency reliability; Category Characteristic Curves (and others).

Systematic review evaluating PROM-based assessment of fatigue in axSpA

Patient Research Partner (PRP) group input
Appendix 2A: search strategy

Search 1 – PROM evaluation studies

Database: Ovid MEDLINE(R) <1946 to August Week 5 2017>, Embase Classic+Embase <1947 to 2017 September 08>, PsycINFO <1967 to September Week 1 2017>

Search Strategy:

--------------------------------------------------------------------------------
1 Spondylitis, Ankylosing/ or (Axial Spondyl* or Spondyl*, Axial or (Ankylosing Spondyl* or Spondyl*, Ankylosing) or Spondyloarthopath* or Spondyloarthritis).ti,ab. (49090)
2 (Fatigue or Asthenia* or Lassitude or Exhaust* or Inertia or Drows* or Letharg* or (Tiring or Tired* or Weary or Weariness)).mp. (544147)
3 (HR-PRO or HRPRO or HRQL or HRQoL or QL or QoL or (PRO or PROs or PROM or PROMs)).ti.ab. or (VAS or NRS).mp. or visual analogue scale*.mp. or numeric* rating scale*.mp. or quality of life.mp. or (health index* or health indices or health profile*).ti,ab. or health status.mp. or ((patient or self or proxy) adj (appraisal* or appraised or report or reported or reporting or rated or rating* or based or assessed or assessment*)).ti,ab. or ((disability or function or functional or functions or subjective or utility or utilities or wellbeing or well being) adj2 (index or indices or instrument or instruments or measure or measures or questionnaire* or profile or profiles or scale or scales or score or scores or status or survey or surveys)).ti,ab. (2092925)
4 (instrumentation or methods).sh. or (Validation Studies or Comparative Study).pt. or exp Psychometrics/ or psychometr*.ti.ab. or (clinimat* or clinometer*).tw. or exp "Outcome Assessment (Health Care)"/ or outcome assessment.ti.ab. or outcome measure*.tw. or exp Observer Variation/ or observer variation.ti.ab. or exp Health Status Indicators/ or exp "Reproducibility of Results"/ or reproducibility.ti.ab. or exp Discriminant Analysis/ or (reliable or unrelia*ble or valid* or coefficient or homogeneity or homogeneous or "internal consistency").ti.ab. or (cronbach* and (alpha or alphas)).ti,ab. or (item and (correlation* or selection* or reduction*)).ti,ab. or (agreement or precision or imprecision or "precise values" or test-retest).ti,ab. or (test and retest).ti,ab. or (reliable* and (test or retest)).ti,ab. or (stability or interrater or inter-rater or intrarater or intra-rater or intertester or inter-tester or intratester or interobserver or inter-observer or intraobserver or intertechnician or inter-technician or intratechnician or intra-technician or interexaminer or inter-examiner or intra-examiner or interassay or interassay or intraassay or intra-assay or interindividual or inter-individual or intraindividual or intra-individual or interparticipant or inter-participant or intraparticipant or intra-participant or kappa or kappa's or kappas or repeatab*).ti,ab. or ((replicable* or repeated) and (measure or measures or findings or result or results or test or tests)).ti,ab. or (generalizable* or generalisa* or concordance).ti,ab. or (intraclass and correlation*).ti,ab. or (discriminative or "known group" or factor analysis or factor analyses or dimension* or subscale*).ti,ab. or
(multitrait and scaling and (analysis or analyses)).ti,ab. or (item discriminant or interscale correlation* or error or errors or "individual variability").ti,ab. or (variability and (analysis or values)).ti,ab. or (uncertainty and (measurement or measuring)).ti,ab. or ("standard error of measurement" or sensitiv* or responsive*).ti,ab. or ((minimal or minimally or clinical or clinically) and (important or significant or detectable) and (change or difference)).ti,ab. or (small* and (real or detectable) and (change or difference)).ti,ab. or (meaningful change or "ceiling effect" or "floor effect" or "item response model" or IRT or Rasch or "Differential item functioning" or DIF or "computer adaptive testing" or "item bank" or "cross-cultural equivalence").ti,ab. (12549272)

5  (addresses or biography or case reports or comment or directory or editorial or festschrift or interview or lectures or legal cases or legislation or letter or news or newspaper article or patient education handout or popular works or congresses or consensus development conference or consensus development conference, nih or practice guideline).pt. not (*animals/ not *humans/) (5263211)

6  (1 and 2 and 3 and 4) not 5 (347)

7  limit 6 to english language (341)

8  limit 7 to human (316)

9  limit 8 to yr="1980 -Current" (316)

***************************

Search 2 – Named measures search

Database: Ovid MEDLINE(R) <1946 to August Week 5 2017>, Embase Classic+Embase <1947 to 2017 September 08>, PsycINFO <1967 to September Week 1 2017>

Search Strategy:

--------------------------------------------------------------------------------

1  Spondylitis, Ankylosing/ or (Ankylosing Spondyl* or Spondyl*, Ankylosing or Spondyloarthropath* or Spondyloarthritis).ti,ab. (49014)

2  (Fatigue or Asthenia* or Lassitude or Exhaust* or Inertia or Drows* or Letharg* or (Tiring or Tired* or Weary or Weariness)).mp. (544147)

3  (Fatigue Assessment Scale or FAS or Fatigue Impact Scale or FIS or Fatigue Scale or FS or Fatigue Symptom Inventory or FSI or Myasthenia Gravis Fatigue Scale or Multidimensional Fatigue Symptom Inventory or Multi-dimensional Fatigue Symptom Inventory or MFSI or MFSI-SF or Parkinsons Fatigue Scale or (Pearson and Byars Fatigue Feeling Checklist) or Revised Piper Fatigue Scale or Piper Fatigue Scale or R-PFS or PFS or Rot hen Fatigue Scale or (Schedule of Fatigue and Anergia) or SOFA or Schwartz Cancer Fatigue Scale or SCFS or Visual Analog Fatigue Scale or VAS-F or Checklist Individual Strength or CIS20R or CIS8R or
Multidimensional Fatigue Scale or Multi-dimensional Fatigue Scale or MFS or Fatigue Questionnaire or Multidimensional Assessment of Fatigue or Multi-Dimensional Assessment of Fatigue or MAF or Multi-dimensional Health Assessment Questionnaire or Multidimensional Health Assessment Questionnaire or MDHAQ or Profile of Fatigue or ProF or Multidimensional Fatigue Inventory or Multi-dimensional Fatigue Inventory or MFI or Bath Ankylosing Spondylitis Disease Activity Index or BASDAI or (mini adj BASDAI) or ((Bristol Rheumatoid Arthritis Fatigue Multidimensional Questionnaire adj MDQ) or NRS) or ((Bristol Rheumatoid Arthritis Fatigue Multi-dimensional Fatigue Questionnaire adj MDQ) or NRS) or ((BRAF adj MDQ) or NRS) or BRAFMDQ or BRAFNRs or Short-Form Health Survey or SF-36 or SF36 or SF36-V2 or SF36V2 or Bath Ankylosing Spondylitis Functional Index or BASFI or Evaluation of Ankylosing Spondylitis Quality of Life or EASI-QoL or EASIQoL or Brief Fatigue Inventory or BFI or Functional Assessment of Chronic Illness Therapy or (FACIT adj (F or Fatigue)) or Evaluation of Daily Activity Questionnaire or EDAQ or Worst Fatigue-Numeric Rating Scale or WF-NRS or WFNRS or Fatigue Severity Scale or FSS or Chalder Fatigue Scale or CFS or Patient Reported Outcome* Measurement Information System or PROMIS Visual Analogue Scale* or VAS or Numeric Rating Scale* or NRS).mp. (436787)
(meaningful change or "ceiling effect" or "floor effect" or "Item response model" or IRT or Rasch or "Differential item functioning" or DIF or "computer adaptive testing" or "item bank" or "cross-cultural equivalence").ti.ab. (12549272)

5  1 and 2 and 3 and 4 (319)
6  limit 5 to english language (312)
7  limit 6 to human (288)
8  limit 7 to yr="1980 -Current" (288)
9  remove duplicates from 8 (220)

***********************
## Appendix 2B: characteristics of included studies (n=23)

<table>
<thead>
<tr>
<th>Study</th>
<th>Country</th>
<th>PROM(s) evaluated</th>
<th>Sample</th>
<th>Setting</th>
<th>Study focus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aissaoui et al. (182)</td>
<td>Morocco</td>
<td>MAF 10cm VAS</td>
<td>AS patients</td>
<td>Not reported</td>
<td>Cross-sectional study evaluating the frequency of fatigue, and its relationship with other variables (disease-specific variables, psychological status and sleep disturbance) in Moroccan patients diagnosed with AS.</td>
</tr>
<tr>
<td>Bedaiwi et al. (192)</td>
<td>Canada</td>
<td>FSS</td>
<td>AS patients and Nr-AxSpa</td>
<td>Rheumatology clinic</td>
<td>Longitudinal cohort study investigating fatigue prevalence, associated factors, and the impact of tumour necrosis factor inhibitors (TNFi) on a patient subgroup. Study was conducted in Canada with</td>
</tr>
<tr>
<td>Study Authors (Year)</td>
<td>Country</td>
<td>Measure</td>
<td>Study Population</td>
<td>Study Details</td>
<td></td>
</tr>
<tr>
<td>----------------------</td>
<td>---------</td>
<td>---------</td>
<td>------------------</td>
<td>---------------</td>
<td></td>
</tr>
<tr>
<td>Bodur et al. (200)</td>
<td>Turkey</td>
<td>SF-36 vitality</td>
<td>AS patients Mean age 39.4, 76.1% male</td>
<td>Not reported Cross-sectional study evaluating quality of life in Turkish patients diagnosed with AS.</td>
<td></td>
</tr>
<tr>
<td>Da Costa et al. (188)</td>
<td>Canada</td>
<td>MFI-20</td>
<td>SpA patients Mean age 46.5, 46.4% male</td>
<td>3 university affiliated rheumatology sites and 1 satellite community clinic Cross-sectional study investigating contributors to dimensions of fatigue. Study was conducted in Canada with SpA patients.</td>
<td></td>
</tr>
<tr>
<td>Dagfinrud et al. (44)</td>
<td>Norway</td>
<td>10cm VAS SF-36 vitality</td>
<td>AS patients Mean age 47, 58% male</td>
<td>Rheumatology clinic Cross-sectional study investigating levels of fatigue, the relationship between fatigue and other factors (demographics, self-reported, clinician measures) and the performance of a generic and a disease-specific measure of fatigue. Study conducted in Norway with AS patients.</td>
<td></td>
</tr>
<tr>
<td>Authors</td>
<td>Country</td>
<td>Method</td>
<td>Participants</td>
<td>Setting</td>
<td>Description</td>
</tr>
<tr>
<td>-------------------------</td>
<td>---------</td>
<td>--------</td>
<td>--------------</td>
<td>----------------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Dernis-Labous et al. (195)</td>
<td>France</td>
<td>10cm VAS</td>
<td>AS patients</td>
<td>Rheumatology clinic</td>
<td>Review of evidence from two double-blind, placebo controlled RCT’s evaluating fatigue prevalence and the clinical relevance and effect of NSAID therapy. Study was conducted with French AS patients.</td>
</tr>
<tr>
<td>Durmus et al. (183)</td>
<td>Turkey</td>
<td>MAF SF-36 vitality</td>
<td>AS patients</td>
<td>Not reported</td>
<td>Trial study investigating the effect of a home-based exercise programme on QoL. Study used Turkish AS patients.</td>
</tr>
<tr>
<td>Fallahi et al. (196)</td>
<td>Iran</td>
<td>10cm VAS</td>
<td>AS patients</td>
<td>Rheumatology clinic</td>
<td>Cross-sectional study evaluating the reliability of the Persian version of the ASQoL questionnaire. Study conducted with Iranian AS patients.</td>
</tr>
<tr>
<td>Fernandez-Sueiro et al. (199)</td>
<td>Spain</td>
<td>10cm VAS</td>
<td>PsA, Axial PsA and AS patients</td>
<td>AS clinic</td>
<td>Longitudinal study conducted in Spain evaluating the</td>
</tr>
<tr>
<td>Study</td>
<td>Country</td>
<td>Methodology</td>
<td>Outcomes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------------</td>
<td>---------</td>
<td>-------------</td>
<td>-------------------------------------------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gunaydin et al. (189)</td>
<td>Turkey</td>
<td>Cross-sectional study</td>
<td>Evaluating fatigue frequency and its multi-dimensional structure, and its association with other variables (demographics, disease-specific). Study conducted with Turkish AS patients.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ibn Yacoub et al. (184)</td>
<td>Morocco</td>
<td>Cross-sectional study</td>
<td>Investigating the aspects of fatigue and relationships with disease-specific variables of activity and severity. Study conducted with Moroccan AS patients.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study Authors</td>
<td>Country</td>
<td>Measure/Scale</td>
<td>Participants</td>
<td>Setting</td>
<td>Study Type</td>
</tr>
<tr>
<td>------------------------</td>
<td>---------------</td>
<td>---------------------------------------------------</td>
<td>---------------------------------------</td>
<td>------------------------------</td>
<td>------------------------------------------------</td>
</tr>
<tr>
<td>Maksymowycz et al. (191)</td>
<td>Canada</td>
<td>FACIT-Fatigue</td>
<td>AS patients</td>
<td>Rheumatology clinic</td>
<td>Cross-sectional postal survey exploring contributors to PASS and validating the PASS concept. Study conducted with Canadian AS patients.</td>
</tr>
<tr>
<td>Naegeli et al. (202)</td>
<td>US</td>
<td>BFI (worst fatigue item)</td>
<td>AS patients</td>
<td>Not reported</td>
<td>Cross-sectional qualitative study exploring the development and content validity of the WF-NRS. Study conducted with US AS patients.</td>
</tr>
<tr>
<td>Park et al. (197)</td>
<td>Korea</td>
<td>10cm VAS</td>
<td>AS patients</td>
<td>Rheumatology clinic</td>
<td>Cross-sectional study evaluating the validity and reliability of a Korean translation of the BASDAI. Study conducted with Korean AS patients.</td>
</tr>
<tr>
<td>Revicki et al. (190)</td>
<td>US, Europe and Canada</td>
<td>FACIT-Fatigue, SF-36 vitality, 10cm VAS</td>
<td>AS patients</td>
<td>Rheumatology clinic</td>
<td>A review of 2 RCT studies to evaluate the measurement properties of the SF-36 and</td>
</tr>
<tr>
<td>Study</td>
<td>Country</td>
<td>Methodology</td>
<td>Fatigue Measure</td>
<td>Participant Details</td>
<td>Site</td>
</tr>
<tr>
<td>-----------------------</td>
<td>-----------</td>
<td>-------------</td>
<td>-----------------</td>
<td>---------------------</td>
<td>-----------------------</td>
</tr>
<tr>
<td>Schneeburger et al.</td>
<td>Argentina</td>
<td>Case-control</td>
<td>10cm VAS FSS</td>
<td>AS patients, Mean age Not reported, % male Not reported</td>
<td>Not reported</td>
</tr>
<tr>
<td>Stebbings et al.</td>
<td>New Zealand</td>
<td>Cross-sectional</td>
<td>10cm VAS MAF</td>
<td>AxSpa patients, Mean age 43.42, 67% male</td>
<td>Rheumatology clinic in hospital</td>
</tr>
<tr>
<td>Turan et al. (186)</td>
<td>Turkey</td>
<td>Cross-sectional</td>
<td>MAF SF-36</td>
<td>AS patients, Mean age 37.7, 82.35% male</td>
<td>Outpatients clinic</td>
</tr>
</tbody>
</table>
van Tubergen et al. (43)
The Netherlands  MFI
10cm VAS
SF-36 vitality
AS patients
Mean age Not reported, % male Not reported
Rheumatology and outpatient’s clinic
PROM evaluation study using patients recruited from three sources. Also seeking to identify factors that influence fatigue and the association of fatigue with quality of life. Patients were Dutch with a diagnosis of AS.

Wanders et al. (194)
The Netherlands  FSS
SF-36 vitality
AS patients
Mean age Not reported, % male Not reported
University hospital
Double-blind clinical trial investigating the responsiveness and discriminative ability of, and relationships between, ASAS-recommended measures for the DC-ART core set. The study used Dutch AS patients.

Wheaton et al. (181)
Canada  Modified 10cm VAS
SpA patients
Hospital Rheumatology clinic
PROM evaluation study investigating the MID in SpA.
<table>
<thead>
<tr>
<th>Study</th>
<th>Location</th>
<th>Assessment Tool</th>
<th>Participants</th>
<th>Setting</th>
<th>Study Design</th>
<th>Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yilmaz et al. (198)</td>
<td>Turkey</td>
<td>10cm VAS SF-36 vitality</td>
<td>AS patients</td>
<td>Outpatients clinic</td>
<td>Cross-sectional study evaluating HR-QoL and to assess peripheral involvement impact on HR-QoL domains. The study used Turkish AS patients.</td>
<td></td>
</tr>
<tr>
<td>Alkan et al. (187)</td>
<td>Turkey</td>
<td>MAF</td>
<td>AS patients</td>
<td>Outpatient clinic</td>
<td>Cross-sectional study evaluating fatigue and its relationship with other variables (disease-specific, spinal mobility, HR-QoL). The study used Turkish AS patients.</td>
<td></td>
</tr>
</tbody>
</table>
Appendix 3A: scoping review search strategy for qualitative studies of axSpA and fatigue

Database: Embase Classic+Embase <1947 to 2019 Week 25>, Ovid MEDLINE(R) and Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Daily and Versions(R) <1946 to June 21, 2019>, PsycINFO <1967 to June Week 3 2019>

Search Strategy:

---------------------------------------------------------------
1  Spondylitis, Ankylosing/ or (Ankylosing Spondyl* or Spondyl*, Ankylosing or Spondyloarthropath* or Spondyloarthritis).ti,ab. (55877)
2  (Fatigue or Asthenia* or Lassitude or Exhaust* or Inertia or Drows* or Letharg* or (Tiring or Tired* or Weary or Weariness)).mp. (628338)
3  (Semi-structured interview or interview or focus group).mp. [mp=ti, ab, hw, tn, ot, dm, mf, dv, kw, fx, dq, nm, kf, ox, px, rx, an, ui, sy, tc, id, tm, mh] (708297)
4  1 and 2 and 3 (50)
5  remove duplicates from 4 (41)
6  qualitative.mp. [mp=ti, ab, hw, tn, ot, dm, mf, dv, kw, fx, dq, nm, kf, ox, px, rx, an, ui, sy, tc, id, tm, mh] (673545)
7  5 and 6 (19)

***************************
Appendix 3B: lived experience interview schedule

Topics:
1. Introduction
2. Experience of fatigue or tiredness
3. Causes of fatigue or tiredness
4. Daily life and activities – consequences of fatigue or tiredness
5. Managing your fatigue – how do you deal with your fatigue or tiredness
6. Sleep
7. Clinic

Questions:
1. Introduction
What’s it like living with your AS on a daily basis?

PROMPTS: Walking? Using stairs?

2. Experience of fatigue or tiredness
Do you experience any fatigue or tiredness?
When is it worse?
What does ‘fatigue’ mean to you?
What do you see as the difference between fatigue and tiredness?
How would you describe your fatigue?
What is your fatigue/tiredness like on a daily basis?

3. Causes of Fatigue
What do you think causes your fatigue or tiredness?

PROMPTS: Pain? Mood? When your AS is more active?

What sort of symptoms do you experience with your fatigue or tiredness?

4. Daily life and activities
How does your fatigue affect your daily life?

PROMPT: Work? Relationships? Family?

How do you change your daily routine?

Is there a difference between working days and non-working days (such as the weekend)?

5. Managing your fatigue?
How do you manage your fatigue?

What ‘works best’ for you in managing your fatigue?

PROMPT: Can you give me an example?

Does it always work? What do you do if it doesn’t?

How does that affect people around you?

6. Sleep
Do you find it easy to get to sleep?

Do you often wake up during sleep?

If your AS is more active, how does that affect your sleep?

PROMPT: Compared to when your AS is less active?

Is your fatigue worse when your sleep is disturbed?

Is there anything you can do that will help you sleep?

PROMPT: Can you give me an example?
7. Clinic

Do you mention your fatigue in your clinic appointments?

*PROMPT: With your doctor, physiotherapist, nurse?*

What would stop you from bringing it up?

What are the things that you would like a clinician/physiotherapist to ask you about your fatigue or tiredness?

If we could measure this fatigue or tiredness on a ruler (or scale), what would the words at the beginning and end of the ruler (or scale) say?

Have you been diagnosed with fibromyalgia or chronic pain syndrome?

We have reached the end of our interview. *Before we finish, is there anything else that you would like to say or add?*

Thank you for your time and for participating and contributing to this important research.
### Appendix 3C: example extract of exploratory comments and initial emergent themes from IPA analysis for participant 4

<table>
<thead>
<tr>
<th>Emerging themes</th>
<th>Original transcript</th>
<th>Exploratory comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temporal changes in AS experience</td>
<td>I: So to start, I just wanted to ask you a little bit about what’s it like with AS on a daily basis?</td>
<td>Experience can vary between time points.</td>
</tr>
<tr>
<td></td>
<td>P: Is that how I’m feeling now or is it just how I felt whilst I’ve had the..</td>
<td></td>
</tr>
<tr>
<td>Medication (positive)</td>
<td>I: Just on a day to day basis, what is it like with your AS?</td>
<td>Changed medication modulates AS - positive.</td>
</tr>
<tr>
<td></td>
<td>P: As of now it’s not as bad as I’ve changed the medication I’m on a humira injections which since I’ve come onto them I’ve found it quite..getting back to normal as I can. But before, it was trying hard to get out of bed, always feeling stiff. But now with this new injection, quite free in the movement. Like I’ve noticed in the past couple of times that I’ve come to Haywood and I’ve been able to if they’ve done rotation movements and measurements and stuff like that. But yes its going quite well.</td>
<td>When medication works, back to normal – AS makes him a different person?</td>
</tr>
<tr>
<td>Old vs new me</td>
<td></td>
<td>Medication able to give a new lease of function. Old him able to do most thing, AS-him not. Things are improving.</td>
</tr>
<tr>
<td>Improvement in condition</td>
<td>I: And does it affect you walking or..?</td>
<td>Other factors impairing his function not AS related.</td>
</tr>
<tr>
<td>Other illnesses</td>
<td>P: It does as I’ve had my hip replaced and I’ve... my right leg is actually an inch shorter than my left so I have to have a heel raise and when I’ve not got any shoes on obviously I limp about quite a bit. It’s not that bad at the moment like, I know my spine’s a little bit curved but apart from that it’s not affecting me at the moment. I: And how is it different with AS now compared to before you had AS or before you were diagnosed with it?</td>
<td></td>
</tr>
<tr>
<td>Reducing activity due to impairment</td>
<td>P: I’m a lot more wary of what activities that I do because I know I’m not as mobile and flexible as I once was. So it...and its given me a [bad 2:06??] outlook on life I just get out and do stuff like live your life whilst you can before it sort of takes over. I: And do you experience any fatigue? P: Not such now. I’ll have the odd day where I’ll feel tired and just want to stay in bed but before the injection it was quite...I did have a lot of fatigue were it would literally be get up and go to work and then just sit about cos I didn’t have the energy to really go out and do stuff. Yeh.. I always felt quite tired. I: Can you just tell me a little bit about what the fatigue felt like? P: It felt sort of like I’d been up for countless hours and I just sort of wanted to go to bed and sleep. Like I’d been awake for say 48 ...</td>
<td>Cautious about what he does due to AS-related impairments: reduced mobility, flexibility etc. Negative outlook on life – a race against time with the illness before it progresses and takes over, taking away his ability to do things he can now. Fatigue reduced since humira treatment – still occasional though. Before medication, frequent fatigue that negatively impacted work. Just sit there because he had no energy. Feels like an unrested state – deprived of sleep. The feeling of being drained, empty of energy or motivation.</td>
</tr>
<tr>
<td>Battling progressive, degenerative impact of AS on ability</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medication (positive, fatigue)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lack of energy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Drained/ empty</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Appendix 3D: example table of superordinate and subordinate themes for participant 5

<table>
<thead>
<tr>
<th>Theme</th>
<th>Page/ line</th>
<th>Key words/ phrases</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Experiencing fatigue</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Like a disability</td>
<td>1.43</td>
<td>It's like a disability</td>
</tr>
<tr>
<td>Unpredictable</td>
<td>19.543</td>
<td>If it comes it comes</td>
</tr>
<tr>
<td><strong>Symptoms of fatigue</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sudden onset</td>
<td>8.214</td>
<td>Hits me quite quickly</td>
</tr>
<tr>
<td>Constantly draining</td>
<td>5.112</td>
<td>It's constantly draining. Constantly draining.</td>
</tr>
<tr>
<td>- Physical fatigue</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Drained</td>
<td>5.112</td>
<td>I feel drained now</td>
</tr>
<tr>
<td>Slowing down</td>
<td>14.405</td>
<td>Going to take me a long longer...</td>
</tr>
<tr>
<td>- Mental fatigue</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not thinking straight</td>
<td>16.457</td>
<td>So mentally it’s draining you don’t think straight</td>
</tr>
<tr>
<td>Constant concern</td>
<td>2.43</td>
<td>Always in the back of my mind</td>
</tr>
<tr>
<td><strong>Impact of fatigue</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Powerlessness</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Feeling incapable</td>
<td>6.155</td>
<td>I should be able to do it but you can’t</td>
</tr>
<tr>
<td>- Feeling managed by fatigue</td>
<td>3.79</td>
<td>I think it manages me</td>
</tr>
<tr>
<td>Avoiding responsibilities</td>
<td>7.179</td>
<td>I ring up and say like “I’m still stuck on this job”</td>
</tr>
<tr>
<td>Personal life</td>
<td>15.413</td>
<td>Obviously when you do get home you’re annoyed</td>
</tr>
<tr>
<td>Work life</td>
<td>3.54</td>
<td>I’ll fly you a text saying “I’m stuck on a job”</td>
</tr>
<tr>
<td>Others</td>
<td>9.253</td>
<td>She just tells me to go out and see one of my mates</td>
</tr>
<tr>
<td><strong>Coping with fatigue</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive mental attitude</td>
<td>11.310</td>
<td>Being positive helps</td>
</tr>
<tr>
<td>Letting tiredness take its</td>
<td>4.99</td>
<td>Let nature take its course</td>
</tr>
<tr>
<td>course</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Humour as a coping strategy</td>
<td>19.533</td>
<td>I fell asleep when she did the acupuncture once, so no</td>
</tr>
<tr>
<td>Knowledge is power</td>
<td>7.193</td>
<td>The more people... knowledge is power, you know</td>
</tr>
<tr>
<td>--------------------</td>
<td>-------</td>
<td>---------------------------------------------</td>
</tr>
<tr>
<td>Accepting fatigue</td>
<td>19.543</td>
<td>You’ve just got to deal with it</td>
</tr>
<tr>
<td><strong>Adapting to fatigue</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Changing approach to tasks</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Task avoidance</td>
<td>8.221</td>
<td>You look for a way of avoiding it</td>
</tr>
<tr>
<td>Balancing life</td>
<td>4.86</td>
<td>If I’m really shot at, I can be in bed by 6.30-7pm</td>
</tr>
<tr>
<td>*Changing diet</td>
<td>4.107</td>
<td>I’ve cut out pop</td>
</tr>
<tr>
<td><strong>Psychological impact of fatigue</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative mood</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Mood swings</td>
<td>15.418</td>
<td>Dust yourself off and go “right come on dickhead”</td>
</tr>
<tr>
<td>- Losing control of mood</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative emotions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Guilt</td>
<td>9.243</td>
<td>It’s not fair on her to be honest</td>
</tr>
<tr>
<td>- Anxiety</td>
<td>15.438</td>
<td>It causes anxiety</td>
</tr>
<tr>
<td>- Annoyance/ angry/ temper</td>
<td>15.411</td>
<td>I get mentally annoyed</td>
</tr>
<tr>
<td>- Feeling powerless</td>
<td>3.65</td>
<td>I let it manage me, which is wrong really</td>
</tr>
<tr>
<td>Internalising frustration</td>
<td>15.411</td>
<td>I get mentally annoyed because I should have been done</td>
</tr>
<tr>
<td><strong>Understanding</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feeling misunderstood</td>
<td>9.231</td>
<td>People come up with stupid questions when you’re tired</td>
</tr>
<tr>
<td>Lack of understanding</td>
<td>10.266</td>
<td>There’s huge lack of understanding</td>
</tr>
<tr>
<td>- Friends and family</td>
<td>7.194</td>
<td>People have got to get it right in their head</td>
</tr>
<tr>
<td>- Medical professionals</td>
<td>13.379</td>
<td>They don’t understand it they don’t know anything</td>
</tr>
<tr>
<td>Understanding the cause of fatigue</td>
<td>7.195</td>
<td>Once you understand it you can probably live with it</td>
</tr>
<tr>
<td>Lost and confused</td>
<td>16.464</td>
<td>In the back of your mind just “why, why?”</td>
</tr>
<tr>
<td><strong>Rumination</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Then vs now</td>
<td>1.7</td>
<td>Over this past 12 months especially</td>
</tr>
<tr>
<td>What to do</td>
<td>16.453</td>
<td>I’ll be thinking “if I move that and then move that”</td>
</tr>
<tr>
<td>-----------------------------------------------</td>
<td>--------</td>
<td>-----------------------------------------------------</td>
</tr>
<tr>
<td>Ruminating on behaviour</td>
<td>7.186</td>
<td>You examine what you’ve just said</td>
</tr>
<tr>
<td>Internal battle/ conflicts</td>
<td>8.209</td>
<td>“shall I move, can I move it. I’ve got to move it”</td>
</tr>
<tr>
<td><strong>Support</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quality of support</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Inappropriate/ Poor professional advice</td>
<td>7.198</td>
<td>“yeah, go and have a lie down” … that’s not an answer.</td>
</tr>
<tr>
<td>Awareness of support</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Family support</td>
<td>8.228</td>
<td></td>
</tr>
<tr>
<td>- Utilising support</td>
<td>15.427</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>The wife probably catches the brunt of it</td>
</tr>
<tr>
<td></td>
<td></td>
<td>I know they’ll pick the slack up so I’m not worried</td>
</tr>
</tbody>
</table>
## Appendix 3E: completed COREQ checklist

<table>
<thead>
<tr>
<th>Item</th>
<th>Short description</th>
<th>Location (section/ page)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Domain 1: Research team and reflexivity</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Personal Characteristics</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Interviewer/facilitator</td>
<td>Nathan Pearson</td>
<td>3.3.4. / 104</td>
</tr>
<tr>
<td>2. Credentials</td>
<td>BSc (Hons), MSc</td>
<td>3.3.4. / 104</td>
</tr>
<tr>
<td>3. Occupation</td>
<td>Doctoral research student</td>
<td>3.3.4. / 104</td>
</tr>
<tr>
<td>4. Gender</td>
<td>Male</td>
<td>3.3.4. / 104</td>
</tr>
<tr>
<td>5. Experience and training</td>
<td>Undergraduate degree in psychology, introductory training course in qualitative research</td>
<td>3.3.4. / 104</td>
</tr>
<tr>
<td><strong>Relationship with participants</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Relationship established</td>
<td>Yes</td>
<td>-</td>
</tr>
<tr>
<td>7. Participant knowledge of the interviewer</td>
<td>The study purpose was explained verbally and in the study information sheet prior to giving informed consent to participate in interview</td>
<td>3.4.1. / 101</td>
</tr>
<tr>
<td>8. Interviewer characteristics</td>
<td>NP is a psychologist with an interest in health outcomes. NP also has a chronic health condition. No other potential biases identified</td>
<td>3.3.4. / 104</td>
</tr>
<tr>
<td><strong>Domain 2: study design</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Theoretical framework</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Methodological orientation and Theory</td>
<td>Interpretative phenomenological analysis (critical realist epistemology)</td>
<td>3.3. / 96 - 99</td>
</tr>
<tr>
<td><strong>Participant selection</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. Sampling</td>
<td>Purposive sampling</td>
<td>3.4.1. / 101</td>
</tr>
<tr>
<td>11. Method of approach</td>
<td>Face-to-face by members of their local rheumatology team</td>
<td>3.4.1. / 101</td>
</tr>
<tr>
<td>12. Sample size</td>
<td>Seventeen</td>
<td>3.5.1. / 105</td>
</tr>
<tr>
<td>13. Non-participation</td>
<td>21 people approached, 17 participated. Reasons described in text</td>
<td>3.5.1. / 105</td>
</tr>
</tbody>
</table>
### Setting

<table>
<thead>
<tr>
<th>14. Setting of data collection</th>
<th>Private room at the local rheumatology department of participants</th>
<th>3.4.3. / 103</th>
</tr>
</thead>
<tbody>
<tr>
<td>15. Presence of non-participants</td>
<td>In one interview (participant 2) a carer was present only 'commented' during interview toward the end</td>
<td>-</td>
</tr>
</tbody>
</table>
| 16. Description of sample | Age range: 22 – 72 years
9 males; 8 females
Disease duration range: 1 – 42 years | Table 3.2. / 105 - 106 |

### Data collection

<table>
<thead>
<tr>
<th>17. Interview guide</th>
<th>Semi-structured interviews were conducted using a patient co-produced topic guide. Follow-up questions and verbal prompts were used</th>
<th>Appendix 3B</th>
</tr>
</thead>
<tbody>
<tr>
<td>18. Repeat interviews</td>
<td>No</td>
<td>-</td>
</tr>
<tr>
<td>19. Audio/visual recording</td>
<td>Digital audio recordings using an encrypted device</td>
<td>3.4.1. / 101</td>
</tr>
<tr>
<td>20. Field notes</td>
<td>Reflexive notes were taken for the researchers' personal use</td>
<td>-</td>
</tr>
</tbody>
</table>
| 21. Duration | Mean duration: 47.4 minutes
Range: 22.1 – 85.49 minutes | 3.5.1. / 105 |
| 22. Data saturation | Not applicable to methodology | - |
| 23. Transcripts returned | No | - |

### Domain 3: analysis and findings

#### Data analysis

<table>
<thead>
<tr>
<th>24. Number of data coders</th>
<th>Three (NP, ET, KH) coded early interviews. Two coders (NP, ET) completed one third. One coder (NP) completed all transcripts.</th>
<th>3.4.4. / 104</th>
</tr>
</thead>
<tbody>
<tr>
<td>25. Description of the coding tree</td>
<td>Coding described in text and in master table.</td>
<td>Table 3.3.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td><strong>26. Derivation of themes</strong></td>
<td>Derived from data</td>
<td>-</td>
</tr>
<tr>
<td><strong>27. Software</strong></td>
<td>NVivo10</td>
<td>3.4.4. / 104</td>
</tr>
<tr>
<td><strong>28. Participant checking</strong></td>
<td>Patient research partners took part in data analysis</td>
<td>3.5.4. / 123</td>
</tr>
</tbody>
</table>

**Reporting**

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>29. Quotations presented</strong></td>
<td>Yes</td>
<td>3.5. (results)</td>
</tr>
<tr>
<td><strong>30. Data and findings consistent</strong></td>
<td>Yes</td>
<td>3.5. (results)</td>
</tr>
<tr>
<td><strong>31. Clarity of major themes</strong></td>
<td>Yes</td>
<td>3.5. (results)</td>
</tr>
<tr>
<td><strong>32. Clarity of minor themes</strong></td>
<td>Yes</td>
<td>3.5. (results)</td>
</tr>
</tbody>
</table>
## Appendix 4A: Results of PROM mapping

<table>
<thead>
<tr>
<th>Conceptual domains</th>
<th>Subdomains</th>
<th>Components</th>
<th>PROMs used</th>
<th>Identified items</th>
</tr>
</thead>
</table>
| Symptoms           | Fatigue    | Frequency; Severity; Duration | BRAF-MDQ (items 1-3)  
BASDAI (item 1)  
MAF (items 2 and 15)  
FACIT-F (item 3)  
SF-36v2 VT (items 9g and 9i) | Frequency  
BRAF-MDQ: “How many days have you experienced fatigue?”  
MAF: “Over the past week, how often have you been fatigued?”  
SF-36v2 VT: “Did you feel worn out?”  
SF-36v2 VT: “Did you feel tired?”  
Severity  
BRAF-MDQ: “How severe is the fatigue that you have experienced?”  
BASDAI: “How would you describe the overall level of fatigue/tiredness you have experienced?”  
MAF: “How severe is the fatigue which you have been experiencing?”  
FACIT-F: “I feel listless (‘washed out’)”  
Duration  
BRAF-MDQ: “How long, on average, does each spell of fatigue last?” |
<table>
<thead>
<tr>
<th>Impact on self</th>
<th>Low energy</th>
<th>Frequency; Severity; Duration</th>
<th><strong>BRAF-MDQ</strong> (items 1-3) SF-36v2 VT (item 9e) FACIT-F (item 7)</th>
<th><strong>BRAF-MDQ</strong>: Modification of items 1 to 3 to reflect energy Level of energy SF-36v2 VT: “Did you have a lot of energy?” FACIT-F: “I have energy”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cognitive</td>
<td><strong>Talking</strong></td>
<td><strong>BRAF-MDQ</strong> (item 15) CFS (items 8 and 9) MFI-20 (item 11) MFSI-SF (item 1)</td>
<td><strong>Talking</strong> CFS: “do you make slips of the tongue when speaking?” <strong>Concentration</strong> BRAF-MDQ: “Has fatigue made it difficult to concentrate?” CFS: “do you have difficulties concentrating?” <strong>Memory</strong> CFS: “How is your memory?” MFI-20: “I can concentrate well.” MFSI-SF: “I have trouble remembering things”</td>
<td></td>
</tr>
<tr>
<td>Physical</td>
<td><strong>Self-care</strong></td>
<td><strong>MAF</strong> (items 5 and 14) FACIT-F (item an14) FSS (item 2)</td>
<td><strong>Self-care</strong> MAF: “In the past week, to what degree has fatigue interfered with your ability to cook?” FACIT-F: “I am too tired to eat” <strong>Exercise</strong> MAF: “In the past week, to what degree has fatigue interfered with your ability to exercise, other than walking?” FSS: “Exercise brings on my fatigue”</td>
<td></td>
</tr>
<tr>
<td>Function</td>
<td>Sleep</td>
<td>Social participation – usual work; job; leisure activities (family or friends);</td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>-------</td>
<td>--------------------------------------------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>BRAF-MDQ (items 8, 10)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>MFI-20 (item 15)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>FSS (item 9)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>BFI (item 4f)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>MAF (item 11)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>FACIT-F (item An7)</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Interference (generic)**

**FSS:** “Fatigue interferes with my work, family, or social life”

**BFI:** “Circle the one number that describes how, during the past 24 hours, fatigue has interfered with your relations with other people”

**Usual activity**

**FACIT-F:** “I am able to do my usual activities”

**Leisure activities**

**MAF:** “In the past week, to what degree has fatigue interfered with your ability to engage in leisure and recreational activity?”

**Social participation**

**BRAF-MDQ:** “Have you cancelled plans because of fatigue? e.g. plans to go out, or do jobs around the home or garden”

**BRAF-MDQ:** “Have you avoided making plans because of fatigue? e.g. plans to go out, or do jobs around the home or garden”

**MFI-20:** “I have a lot of plans.”

**No items identified**
| Psychological/ emotional wellbeing | Mood | Depression; Low mood; feeling down; Losing control; helplessness | **BRAF-MDQ** (items 17 and 20)  
**BFI** (item 4b)  
**MFSI-SF** (items 3 and 21) |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Control</strong></td>
<td><strong>BRAF-MDQ:</strong> Have you felt you have less control in areas of your life because of fatigue?</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| **Depression** | **BRAF-MDQ:** “Have you felt down or depressed because of fatigue?”  
**BFI:** “Circle the one number that describes how, during the past 24 hours, fatigue has interfered with your mood.”  
**MFSI-SF:** “I feel upset”  
**MFSI-SF:** “I feel depressed” |
| Anxiety/ worrying | Worry; Dread; Frustration; Stress; Unpredictable | **MFI-20** (item 9)  
**FACIT-F** (item An15) |
| **Dread** | **MFI-20:** “I dread having to do things.”  
**Frustration**  
**FACIT-F:** “I am frustrated by being too tired to do the things I want to do” |
<p>| Sense of self | Guilt – letting others down; Embarrassment – falling asleep in social situations; helplessness | <strong>BRAF-MDQ</strong> (item 18) |
| <strong>BRAF-MDQ:</strong> “Have you felt embarrassed because of fatigue?” |</p>
<table>
<thead>
<tr>
<th>Self-management</th>
<th>Self-isolation</th>
<th>Pushing others away; hiding the fatigue; invisible illness</th>
<th>No items identified</th>
</tr>
</thead>
<tbody>
<tr>
<td>Achieving balance</td>
<td>Avoiding overexertion; avoiding underactivity; dealing with new needs (coping and adapting); learnt expertise; taking a flexible approach; lost – don’t know what to do.</td>
<td>MFI-20 (items 2 and 8)</td>
<td>Exertion and activity</td>
</tr>
<tr>
<td>Energy</td>
<td>Difficulty starting things; Re-energising (coffee/sugar, napping); declining events to avoid wipe-out; feeling drained</td>
<td>CFS (item 4) FACIT-F (item An3)</td>
<td>Difficulty starting things</td>
</tr>
<tr>
<td>Support</td>
<td>Delegating responsibilities (family, friends, colleagues)</td>
<td>FACIT-F (item An14)</td>
<td>FACIT-F: “I need help doing my usual activities”</td>
</tr>
</tbody>
</table>
Appendix 4B: Thematic analysis findings from focus groups (Chapter 4)

Four themes emerged from focus group analysis: (i) Resonance and challenges; (ii) Daily life and functioning; (iii) Mental wellbeing, and; (iv) self-management. A detailed analysis of each theme is provided along with respective subthemes. Table 4.2 provides supporting quotes for each identified theme.

Resonance and challenges

The theme of resonance and challenges describes the extent to which participants related to or felt at odds with the domains and PRO outcomes described in the measurement framework. This theme consists of four subthemes: (i) financial considerations; (ii) use in practice and measurement; (iii) meanings ascribed to fatigue and energy, and; (iv) making sense of sleep.

Both patients and professionals described substantive resonance with the measurement framework and interview findings – particularly around fatigue and energy and the negative impact it has on patients across multiple areas of their lives (social, leisure, mentally etc). Finance arose as an important issue for healthcare professionals and they were inquisitive about why it was not in the measurement framework. However, patients never referred to finance – similarly, this did not arise in the qualitative interviews. Moreover, the PRP group members did not identify finance as an issue either. Only one professional noted that they understood financial challenges to be related to broader axSpA symptoms and not fatigue.

Use in practice was important and specific to healthcare professionals, along with scoring processes. Professionals discussed how a new questionnaire, underpinned by the model, could be utilised within a clinical context, including how its scores may be interpreted to support their decision making. They expressed a clear preference for domain scoring rather than summated scoring. They considered domain scoring to be more informative, supporting a better understanding the circumstances of the patient and enabling them to tailor support accordingly. Additionally, professionals discussed the utility and relevance of specific items such as...
asking about management which could help distinguish between patients who are significantly affected by fatigue but coping compared to those who are not. However, healthcare professionals challenged the usefulness of the sleep domain in terms of its contribution to fatigue measurement. Professionals were cautious about further development of this domain due to its complexity and the perceived risk that it may distract from fatigue which is the focus of the intended measure. This further supported criticism that the sleep domain is multifactorial, and this may prove problematic for a fatigue-specific questionnaire.

Meanings ascribed to fatigue was a particular issue for healthcare professionals. Some felt that the term fatigue was one that they struggled to articulate and understand and were concerned this would be true for patients too. One professional drew a parallel between the term fatigue and stress, asking “what does that mean?” Professionals accepted that fatigue was different to normal or everyday tiredness but felt that tiredness was more easily understood; attempted to generate a replacement term to avoid conflating fatigue and tiredness but were unsuccessful. To assist understanding the meaning of fatigue, professionals suggested providing a definition. Meanings ascribed to energy were also extensively discussed, including whether energy was a distinct component of fatigue or was a synonym of it. There was consensus amongst healthcare professionals that fatigue is more of a mental issue. One healthcare professional described energy as a finite resource that required careful use. One participant stated that the term energy reflected the feeling of being disinclined to do activity.

Professionals felt that phrases such as ‘being drained of energy’ resonated well with patients, and elicited responses from them. In patient groups, they confirmed this stating that they liked the term energy and indicated it related to activity more so for them, reinforcing the qualitative conceptualisation that fatigue and energy are not the same.

Making sense of sleep was an important theme and captured the wide range of views relating to sleep from both patients and professionals. For patients, axSpA symptoms (pain, stiffness) were much more problematic for their sleeping behaviour than fatigue or low energy. The impact of poor sleep varied between participants, with some suggesting it worsened feelings of tiredness, and others indicating it worsened their fatigue.
Daily life and functioning

The theme of daily life and functioning explores possible reasons why axSpA patients may be affected in certain aspects of their lives or functioning because of fatigue, e.g. impaired memory. This theme consists of four subthemes: (i) changing cognitive abilities; (ii) isolation; (iii) the relationship between pain and fatigue, and; (iv) being active.

Healthcare professionals felt that fatigue was a mental challenge and noted the negative impact fatigue could have on patients’ cognitive abilities (e.g. poorer concentration). However, patients questioned whether age was an important factor in how fatigue impacts on them. They described better social circumstances for older axSpA patients due to different life circumstances, such as being retired, making life easier to live with and deal with fatigue. The patients then contrasted that with younger axSpA patients and considered their circumstances, such as how they “have to” work. Whilst no one agreed that age was fully responsible for changes in cognitive ability, it was clear patients were unsure how much change they could attribute to it.

Communicating with others felt much harder with fatigue for patients and this concept resonated strongly with them. Patients described difficulty with concentrating during conversations and felt this was a key contributor to the challenges they experience in conversation. Here – worsening fatigue acted as an amplifier to these difficulties, and during low fatigue periods they witnessed improvement in their function.

Isolation emerged as an important issue with some patients experiencing withdrawing socially; socialising can be demanding, requiring extra effort, therefore withdrawal can help to preserve energy levels. The mental aspect of fatigue was again contrasted against the physical aspects of energy with participants arguing that fatigue and energy are separate concepts that should asked about separately.

The potential for pain to drive fatigue experience was discussed amongst patient participants with them disputing its link and feeling unsure about its role in fatigue experience. However, participants did agree that they experienced fatigue irrespective of whether they were experiencing pain.
Being active emerged as an important theme for both patients and professionals. Some healthcare professionals highlighted the difficulty participants had with exercise due to fatigue. This was attributed to a reduced tolerance to formal exercise (running, swimming, cycling) due to fatigue rather than axSpA symptoms (pain or stiffness). Conversely, some participants felt heaviness affected their ability to do things and be active.

**Mental wellbeing**

This theme captures the challenges that fatigue and low energy present both for participants and professionals in routine practice. This is described in three subthemes: (i) oversight in practice; (ii) strain of being misunderstood, and; (iii) deteriorating mood and loss of control.

Oversight in practice was an issue raised by healthcare professionals. Professionals stated that currently, there is no measure in use to assess the impact of fatigue on mental wellbeing meaning this information is not readily captured in practice. They felt that this was hugely problematic given that their patients often describe a significant impact of fatigue on them. Additionally, professionals discussed that mental health often had a negative tone to it that gave it a negative connotation. Recognising this as a potential problem, they suggested that psychological wellbeing items within the developing axSpA-fatigue and energy measure could challenge the negative connotations by adopting the use of positively valenced language. For example, instead of asking about feeling down, patients could be asked about how happy they feel.

The strain of being misunderstood was difficult for patients to deal with and was recognised by healthcare professionals, with professionals noting patients frequently report this as a problem to them. Specifically, they noted patients felt they were being dismissed or treated as though they were exaggerating, demonstrated by comments from others such as “we all get tired”. This was corroborated by patient accounts too where they described being perceived as lazy when they were experiencing fatigue. This was a source of frustration that negatively affected family life and relationships with others.
Deteriorating mood and a loss of control was extensively discussed by both patients and professionals. For patients, they acknowledged that feeling depressed was common with the condition and felt that fatigue was a significant influencing factor in that. For some patients, they felt this was due to a change in control over their lives. Healthcare professionals recognised low mood but challenged using the term ‘depression’ in the measurement framework as a fatigue-specific measure would not seek to diagnose depression. The term downhearted is used in the EASi-QoL (62) and was suggested as a replacement term for depression.

**Self-management**

This theme captured some of the processes of self-management that participants engaged in. This was captured in two subthemes: (i) re-energising and (ii) learning processes.

Re-energising was highlighted by patients and professionals during discussion as important to improving energy levels. Patients described efforts to re-energise such as drinking energy drinks or consuming sugary foods or drinks – confirming what professionals were hearing in clinic from their patients. However, the extreme lengths participants would go (e.g. drinking 15 energy drinks a day) to re-energise was something that professionals were not aware of, nor did they realise how extreme participants would go in their efforts to re-energise.

Learning processes captured the approaches participants took to learn to deal with their fatigue and low energy, and some of the strategies that they adopted. Trial and error was the main method participants used and they reported an experiential learning process. Part of this process was learning to maintain energy levels by balancing activity with their energy levels. Patient participants also described focusing on responsibilities and prioritising important tasks to minimise the impact of fatigue or low energy on daily life. Some participants acquired new responsibilities that necessitate their involvement, such as getting a pet dog, to help them push through feelings of fatigue.
Appendix 4C: interview schedule for cognitive interviews

Outline of procedure

1. Introduction to study
2. Pre-interview training: “Think aloud”
3. Interview using mixed approach: think aloud and concurrent probing
4. Administration of demographic questions (close)

1. Introduction to study

Thank you for agreeing to come along today and agreeing to help us better understand and improve this new fatigue questionnaire for AxSpA/AS.

This questionnaire – called the Warwick Axial Spondyloarthropathy Tiredness and Energy Scale – has been developed to specifically assess fatigue in patients with Axial Spondyloarthropathy and Ankylosing Spondylitis. We developed the questionnaire after conducting interviews with patients to better understand ‘what’ fatigue is: how it makes them feel, how it affects them and how they try to deal with it. We also learned that patients are affected by fatigue, and sometimes they may have ‘lost’ their energy which stops them being able to do the things they want to do. From this we developed a ‘picture’ of fatigue and energy in AxSpA/AS and explored this in some separate group discussions with patients and healthcare professionals. This helped us identify ‘what’ is important to measure about fatigue and energy. Now we need to find out how best to measure it when you attend the clinic or participate in research, so we co-developed this questionnaire with our patient research partners.

In this interview today, we are asking you to work with us to help us double-check the questionnaire and your thoughts and opinions about it, and the questions. The interview will involve you completing the questionnaire. Whilst you complete it I want you to say everything that you are thinking (I’ll explain this in a moment and we can do a short practice task) – I may make some notes, too. After some questions I may ask you to pause a moment so that I can ask you some follow-up questions. There are no right or wrong answers so please feel free to be honest about things you do and do not like or understand.
I will be audio-recording our session so that I capture all of your thoughts and feedback which will help me to make any modifications to the questionnaire, accurately. The results of these interviews will help us refine the questionnaire about AxSpA or AS-related fatigue or loss of energy.

Do you have any questions, or is there anything you would like to ask?

2. **Pre-interview training: “Think aloud”**

Try to visualise the place where you live and think about how many windows there are in that place. As you count up the windows, tell me what you are seeing and thinking about.

3. **Interview using mixed approach – block design**

In the training we discussed “thinking aloud” and how talking about what you are thinking will allow me to better understand how you are making sense of things and reaching an answer. Now, I would like you to complete this questionnaire [provide questionnaire] and as you go through the instructions and each question, please talk about what you are thinking. Whilst you do this I will simply listen, and I might make some notes about things I would like to ask you about later. The questions are colour coded. After you complete all the ‘red’ questions, for example, you can then pause, and I will ask you some questions. Please remember there are no right or wrong answers and your help will be invaluable in supporting improvements to the questionnaire and its content.

<table>
<thead>
<tr>
<th>Block</th>
<th>Colour code</th>
<th>Num. in block</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Red</td>
<td>6</td>
</tr>
<tr>
<td>2</td>
<td>Blue</td>
<td>6</td>
</tr>
<tr>
<td>3</td>
<td>Green</td>
<td>6</td>
</tr>
<tr>
<td>4</td>
<td>Purple</td>
<td>6</td>
</tr>
<tr>
<td>5</td>
<td>Orange</td>
<td>8</td>
</tr>
</tbody>
</table>
4. Administration of questionnaires (close)

Thank you so much for taking the time to volunteer and participate in this session. Your help is greatly appreciated and will be invaluable in supporting the refinement and improvement of the questionnaire. If I could ask you to provide some details before you think that would be great [provide demographics form and BASDAI/ BASFI/ HADS]. Thank you.
Appendix 4D: interview schedule for TSTI

Outline of procedure

1. Introduction to study
2. Pre-interview training: “Think aloud”
3. TSTI Step 1: Observation of patient questionnaire completion with concurrent think aloud (+note taking)
4. TSTI Step 2: Follow-up based on observation and field notes
5. TSTI Step 3: Semi-structured interview and probing
6. Administration of demographic questions (close)

1. Introduction to study

Thank you for agreeing to come along today and agreeing to help us better understand and improve this new fatigue questionnaire for AxSpA/AS.

This questionnaire – called the Warwick Axial Spondyloarthropathy Tiredness and Energy Scale – has been developed to specifically assess fatigue in patients with Axial Spondyloarthropathy and Ankylosing Spondylitis. We developed the questionnaire after conducting interviews with patients to better understand ‘what’ fatigue is: how it makes them feel, how it affects them and how they try to deal with it. We also learned that patients are affected by fatigue, and sometimes they may have ‘lost’ their energy which stops them being able to do the things they want to do. From this we developed a ‘picture’ of fatigue and energy in AxSpA/AS and explored this in some separate group discussions with patients and healthcare professionals. This helped us identify ‘what’ is important to measure about fatigue and energy. Now we need to find out how best to measure it when you attend the clinic or participate in research, so we co-developed this questionnaire with our patient research partners.

In this interview today, we are asking you to work with us to help us double-check the questionnaire and your thoughts and opinions about it, and the questions. The interview will be conducted in 3 steps: 1) you will be asked to complete the questionnaire, but whilst you do it I want you to say everything that you are thinking (I’ll explain this in a moment and we can do a short practice task) – I may make some notes, too; 2) I will ask you some follow-up questions; and 3) an interview about your experience of fatigue and energy, the questionnaire and what your thoughts and opinions are
about it. There are no right or wrong answers only good or bad questions, so please feel free to be honest about things you do and do not like or understand.

I will be audio-recording our session so that I capture all of your thoughts and feedback which will help me to make any modifications to the questionnaire, accurately. The results of these interviews will help us refine the questionnaire about AxSpA or AS-related fatigue or loss of energy.

Do you have any questions, or is there anything you would like to ask?

2. **Pre-interview training: “Think aloud”**
   
   Try to visualise the place where you live and think about how many windows there are in that place. As you count up the windows, tell me what you are seeing and thinking about.

3. **Step 1: Observation of completion**
   
   In the training we discussed “thinking aloud” and how talking about what you are thinking will allow me to better understand how you are making sense of things and reaching an answer. Now, I would like you to complete this questionnaire [provide questionnaire] and as you go through the instructions and each question, please talk about what you are thinking. Whilst you do this I will simply listen, and I might make some notes about things I would like to ask you about later. Please remember there are no right or wrong answers and your help will be invaluable in supporting improvements to the questionnaire and its content.

4. **Step 2: Follow-up based on observation and field notes**
   
   [written notes during think aloud to reflect key issues: non-verbal's (hesitation on questions?), possible probes based on responses or thought process]

5. **Step 3: Semi-structured interview guide**
   
   *Instructions*
   
   - What did the instructions ask you to do?
Definitions
- Can you tell me what fatigue means, in your own words?
- Can you tell me what energy means, in your own words?

Recall period
- What period of time were you thinking about when you answered the questions?
- Over the past 7 days, what has your experience of fatigue/energy levels been?

Conceptual model (comparator)
- In your opinion, what do you think these questions are asking about?
- Are there any questions missing from this section that you think are important?
- Are there any questions included in this section that you think should be removed?

Per-item probes
- What do you think this question is asking? [Comprehension]
  o Question F4 - are these examples relevant?
  o Question F8 – is this more than just fatigue?
  o Question F13 – is this about help?
  o Question E6 – is this question asking the same thing as E4 and E5?
  o Question E10 – are E8, E9 and E10 different questions for you?
- What does [word] mean to you? [Interpretation]
- Were there any questions that you found difficult for any reason?
  o If so, how would you improve it?

Response options
- How did you come to choose that answer? [Response process and judgement]
- Can you tell me what each response option means to you? [Interpretation]
- Are you able to answer using those response options?
  o What do you think about them?

Structure and format
- What do you think about the questionnaire layout?
- Was it clear to read and complete?

**Missing items**
- Is there anything that is important to you when thinking about fatigue and energy, and how that might affect you, that you think is missing?

**General reflection**
- What are your thoughts of the questionnaire overall?
- Is there anything that you would change?

6. **Administration of demographic questions (close)**

Thank you so much for taking the time to volunteer and participate in this session. Your help is greatly appreciated and will be invaluable in supporting the refinement and improvement of the questionnaire. **If I could ask you to provide some details before you think that would be great** [provide demographics form]. Thank you.
## Appendix 5A: factor structure for model 2

<table>
<thead>
<tr>
<th>Item</th>
<th>Factor Description</th>
<th>Factor</th>
<th>1 (energy)</th>
<th>2 (fatigue)</th>
<th>3 (symptoms)</th>
<th>4 (control)</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1</td>
<td>How often have you felt fatigued?</td>
<td>0.740</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F2</td>
<td>How severe was your fatigue?</td>
<td>0.410</td>
<td>0.661</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F3</td>
<td>Has your fatigue made it difficult to remember things?</td>
<td>0.728</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F4</td>
<td>Has your fatigue made it difficult to concentrate on demanding tasks?</td>
<td>0.736</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F5</td>
<td>Have you found it difficult to engage in conversations with other people because of your fatigue?</td>
<td>0.728</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F6</td>
<td>Has your fatigue made it difficult to do the things you enjoy?</td>
<td>0.607</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F7</td>
<td>Have you felt downhearted because of your fatigue?</td>
<td>0.661</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F9</td>
<td>Have you felt worried because of your fatigue?</td>
<td>0.667</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F10</td>
<td>Have you felt frustrated because of your fatigue?</td>
<td>0.608</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F11</td>
<td>Have you felt the need to be left alone because of your fatigue?</td>
<td>0.642</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F12</td>
<td>I feel I have been able to manage my fatigue.</td>
<td>0.843</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F13</td>
<td>I feel I have been able to cope with my fatigue.</td>
<td>0.830</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Question</td>
<td>Score 1</td>
<td>Score 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>--------------------------------------------------------------------------</td>
<td>---------</td>
<td>---------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F14</td>
<td>Have you been more dependent on others because of your fatigue?</td>
<td>0.643</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E1</td>
<td>How much energy have you had?</td>
<td></td>
<td>0.666</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E2</td>
<td>How often have you felt drained of energy?</td>
<td></td>
<td>0.719</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E3</td>
<td>Have your energy levels made it difficult to take care of yourself?</td>
<td></td>
<td>0.709</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E4</td>
<td>Have your energy levels made it difficult to do everyday activities?</td>
<td></td>
<td>0.715</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E5</td>
<td>Have your energy levels made it difficult to take part in more demanding physical activities?</td>
<td>0.614</td>
<td>0.525</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E6</td>
<td>Have your energy levels made it difficult to do your usual work?</td>
<td>0.685</td>
<td>0.458</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E7</td>
<td>Have your energy levels made it difficult to do the things you enjoy?</td>
<td>0.631</td>
<td>0.488</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E8</td>
<td>Have your energy levels made it difficult to make plans?</td>
<td>0.689</td>
<td>0.437</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E9</td>
<td>Have your energy levels caused you to cancel plans?</td>
<td>0.727</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E10</td>
<td>Have your energy levels made it difficult to keep to your plans?</td>
<td>0.722</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E11</td>
<td>Have you 'turned down' invitations because of your energy levels?</td>
<td>0.657</td>
<td>0.409</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Question</td>
<td>Score 1</td>
<td>Score 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>-------------------------------------------------------------------------</td>
<td>---------</td>
<td>---------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E13</td>
<td>I feel I have been able to manage my energy levels.</td>
<td>0.485</td>
<td>0.556</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E14</td>
<td>Have you lacked physical energy?</td>
<td>0.444</td>
<td>0.669</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E15</td>
<td>Have you lacked mental energy?</td>
<td>0.583</td>
<td>0.404</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix 5B: item-total correlations and Cronbach’s α per model

<table>
<thead>
<tr>
<th>Item</th>
<th>Model 1 (30 items; α = 0.971)</th>
<th>Model 3 (25 items; α = 0.972)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Item</td>
<td>ITC</td>
<td>ITC</td>
</tr>
<tr>
<td>f1: How often have you felt fatigued?</td>
<td>0.731</td>
<td>0.731</td>
</tr>
<tr>
<td>f2: How severe was your fatigue?</td>
<td>0.766</td>
<td>0.766</td>
</tr>
<tr>
<td>f3: Has your fatigue made it difficult to remember things? (e.g. being more forgetful than normal)</td>
<td>0.703</td>
<td>0.705</td>
</tr>
<tr>
<td>f4: Has your fatigue made it difficult to concentrate on demanding tasks? (e.g. driving, puzzles, electronic games)</td>
<td>0.756</td>
<td>0.764</td>
</tr>
<tr>
<td>f5: Have you found it difficult to engage in conversations with other people because of your fatigue? (e.g. friends, family, work colleagues)</td>
<td>0.725</td>
<td>0.730</td>
</tr>
<tr>
<td>f6: Has your fatigue made it difficult to do the things you enjoy? (e.g. listening to music, watching a TV programme)</td>
<td>0.759</td>
<td>0.763</td>
</tr>
<tr>
<td>f7: Have you felt downhearted because of your fatigue? (e.g. feeling low or down)</td>
<td>0.807</td>
<td>0.811</td>
</tr>
<tr>
<td>f8: Has your fatigue made it difficult to be in control of your life?</td>
<td>0.847</td>
<td></td>
</tr>
<tr>
<td>f9: Have you felt worried because of your fatigue?</td>
<td>0.746</td>
<td>0.746</td>
</tr>
<tr>
<td>f10: Have you felt frustrated because of your fatigue?</td>
<td>0.813</td>
<td>0.814</td>
</tr>
<tr>
<td>f11: Have you felt the need to be left alone because of your fatigue? (e.g. not interacting with friends or relatives)</td>
<td>0.763</td>
<td>0.762</td>
</tr>
<tr>
<td>f12: I feel I have been able to manage my fatigue.</td>
<td>0.609</td>
<td>0.609</td>
</tr>
<tr>
<td>f13: I feel I have been able to cope with my fatigue.</td>
<td>0.62</td>
<td>0.621</td>
</tr>
<tr>
<td>f14: Have you been more dependent on others (e.g. friends or family) because of your fatigue? (e.g. to run errands, do things around the house for you)</td>
<td>0.7</td>
<td>0.690</td>
</tr>
<tr>
<td>e1: How much energy have you had?</td>
<td>0.731</td>
<td>0.721</td>
</tr>
<tr>
<td>e2: How often have you felt drained of energy?</td>
<td>0.784</td>
<td>0.773</td>
</tr>
<tr>
<td>e3: Have your energy levels made it difficult to take care of yourself? (e.g. personal care, showering, brushing your teeth)</td>
<td>0.695</td>
<td>0.697</td>
</tr>
<tr>
<td>Question</td>
<td>Score 1</td>
<td>Score 2</td>
</tr>
<tr>
<td>-------------------------------------------------------------------------</td>
<td>---------</td>
<td>---------</td>
</tr>
<tr>
<td>e4: Have your energy levels made it difficult to do every day activities? (e.g. light housework, making something to eat or drink)</td>
<td>0.776</td>
<td>0.773</td>
</tr>
<tr>
<td>e5: Have your energy levels made it difficult to take part in more demanding physical activities? (e.g. going for a walk, gardening)</td>
<td>0.769</td>
<td></td>
</tr>
<tr>
<td>e6: Have your energy levels made it difficult to do your usual work? (e.g. at work or at home)</td>
<td>0.813</td>
<td>0.800</td>
</tr>
<tr>
<td>e7: Have your energy levels made it difficult to do the things you enjoy? (e.g. hobbies, leisure activities)</td>
<td>0.819</td>
<td>0.812</td>
</tr>
<tr>
<td>e8: Have your energy levels made it difficult to make plans?</td>
<td>0.825</td>
<td>0.825</td>
</tr>
<tr>
<td>e9: Have your energy levels caused you to cancel plans?</td>
<td>0.791</td>
<td>0.794</td>
</tr>
<tr>
<td>e10: Have your energy levels made it difficult to keep to your plans?</td>
<td>0.812</td>
<td>0.810</td>
</tr>
<tr>
<td>e11: Have you ‘turned down’ invitations because of your energy levels? (e.g. to meet a friend, socialise)</td>
<td>0.772</td>
<td>0.772</td>
</tr>
<tr>
<td>e12: I have difficulty maintaining my energy levels when I have done too much.</td>
<td>0.764</td>
<td></td>
</tr>
<tr>
<td>e13: I feel I have been able to manage my energy levels.</td>
<td>0.708</td>
<td>0.780</td>
</tr>
<tr>
<td>e14: Have you lacked physical energy?</td>
<td>0.791</td>
<td>0.731</td>
</tr>
<tr>
<td>e15: Have you lacked mental energy?</td>
<td>0.731</td>
<td></td>
</tr>
<tr>
<td>e16: Have you run out of energy suddenly and needed to take a break? (e.g. have a nap or rest)</td>
<td>0.775</td>
<td></td>
</tr>
</tbody>
</table>
Appendix 5C: the final short form WASTEd following psychometric evaluation

Warwick Axial Spondyloarthritis Fatigue and Energy questionnaire (WASTEd)

Instructions

We would like to know how your fatigue and energy levels associated with your Axial Spondyloarthritis (axSpA) have affected you, **on average**, over the past 7-days.

We understand that your fatigue and energy levels may have changed day-to-day, but we would like you to answer the questions about how you have been feeling **on average** over the past 7-days. This will let us understand how you feel and are affected by your axSpA-fatigue and help us decide how best we can support you.

The questionnaire is separated into two sections:

**Section 1 – Fatigue:** here, we ask 10 questions about your experience of fatigue associated with your axSpA and how it has affected you, **on average, over the past 7-days**.

**Section 2 – Energy:** here, we ask 8 questions about your energy levels associated with your axSpA and how they have affected you, **on average, over the past 7-days**.

Please read each question carefully and answer each question with a single cross.
Section 1: Fatigue

Please read the following statement before completing this section.

Fatigue: Everyone gets tired or even worn-out at times, but after a few good night’s rest they usually feel refreshed. It is known that people with this condition experience fatigue which is not like normal tiredness. Fatigue can last for weeks at a time and no amount of sleep or rest will relieve it.

On AVERAGE over the PAST 7-DAYS

1. How often have you felt fatigued?

☐ ☐ ☐ ☐ ☐
Not at all Rarely Often All the time

2. How severe was your fatigue?

☐ ☐ ☐ ☐ ☐
Not at all severe A little Very Extremely severe

3. Has your fatigue made it difficult to concentrate or remember things? (e.g. concentrating on driving or puzzles, more forgetful than normal)

☐ ☐ ☐ ☐ ☐
Not at all difficult A little Very Extremely difficult

4. Have you found it difficult to engage in conversations with other people because of your fatigue? (e.g. friends, family, work colleagues)

☐ ☐ ☐ ☐ ☐
Not at all difficult A little Very Extremely difficult
On **AVERAGE** over the **PAST 7-DAYS**

5. Has your **fatigue** made it difficult to do the things you enjoy? (e.g. listening to music, watching a TV programme)

   □  □  □  □  □
   Not at all difficult  A little  Very  Extremely difficult

6. Has your mood been affected by your **fatigue**? (e.g. feeling low, worried, frustrated, guilty)

   □  □  □  □  □
   Not at all affected  A little  Very  Completely affected

7. Have you felt the need to be left alone because of your **fatigue**? (e.g. not interacting with friends or relatives)

   □  □  □  □  □
   Not at all  A little  A lot  All the time

8. I feel I have been able to manage my **fatigue**.

   □  □  □  □  □
   Not at all  A little  A lot  All the time

9. I feel I have been able to cope with my **fatigue**.

   □  □  □  □  □
   Not at all  A little  A lot  All the time
On **AVERAGE** over the **PAST 7-DAYS**

10. Do you feel that your fatigue has made you more dependent on others? (e.g. having to ask for help to do everyday tasks from family, friends or carers)

☐ ☐ ☐ ☐

Not at all  A little  A lot  All the time
Section 2: Energy levels

Please read the following statement before completing this section.

**Energy:** Everyone usually has the energy levels to do things in their day, but with this condition it could be a real struggle to find that 'get up and go' to do the things you want or need to do. You may feel 'drained' and need to stop for a quick rest which might help you generate some energy.

On **AVERAGE** over the **PAST 7-DAYS**

1. **How often have you felt drained of energy?**
   - Not at all
   - Rarely
   - Often
   - All the time

2. **Have your energy levels made it difficult for you to maintain your personal care?** (e.g. showering, brushing your teeth, eating your usual meals)
   - Not at all difficult
   - A little
   - Very
   - Completely difficult

3. **Have your energy levels made it difficult for you to keep to your routine?** (e.g. usual work, hobbies, leisure activities)
   - Not at all difficult
   - A little
   - Very
   - Extremely difficult

4. **Have your energy levels made it difficult to make plans?**
   - Not at all difficult
   - A little
   - Very
   - Extremely difficult
On **AVERAGE** over the **PAST 7-DAYS**

<table>
<thead>
<tr>
<th>5. Have your energy levels caused you to change your plans? (e.g. cancel or reschedule)</th>
</tr>
</thead>
<tbody>
<tr>
<td>☐ ☐ ☐ ☐ ☐</td>
</tr>
<tr>
<td>Not at all</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>6. Have you been able to maintain your energy levels to achieve what you wanted to do?</th>
</tr>
</thead>
<tbody>
<tr>
<td>☐ ☐ ☐ ☐ ☐</td>
</tr>
<tr>
<td>Not at all</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>7. Have you lacked physical energy?</th>
</tr>
</thead>
<tbody>
<tr>
<td>☐ ☐ ☐ ☐ ☐</td>
</tr>
<tr>
<td>Not at all</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>8. Have you lacked mental energy?</th>
</tr>
</thead>
<tbody>
<tr>
<td>☐ ☐ ☐ ☐ ☐</td>
</tr>
<tr>
<td>Not at all</td>
</tr>
</tbody>
</table>