Measuring outcome in clinical trials for adults with an ankle fracture

Rebecca Louise McKeown

BSc(Hons) MCSP PGCert

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

Warwick Clinical Trials Unit, Warwick Medical School

University of Warwick

July 2020

Table of Contents

List of Tables

List of Figures

Acknowledgements

Thank you to the participants of the AIR trial, particularly those who kindly invited me into their homes and talked so openly to me about the impact of their fractures on their lives.

This thesis would not be possible without the guidance and support of my supervisors, Rebecca Kearney and David Ellard. I would like to sincerely thank them both for their time, advice, knowledge and patience throughout this PhD. Thank you also to Martin Underwood for his guidance as a temporary supervisor in the early stages of this project.

Thank you to the many colleagues who have provided help and support throughout this project; Helen Parsons, Jaclyn Brown, Siobhan Kefford, Sunayna Bora, Nicola Cashin, Sophie Rees, Samantha Johnson, Hannah James, Vivien Nichols, Julie Bruce, Nick Parsons and the AIR trial management group. Thank you also to Abdul-Rasheed Rabiu, Emma Sutton, Jonathan Young, Eleni Karasouli, Zi Heng Liew, Richard Grant and Karen Keates for their help with the work included in this thesis.

Thank you to all my wonderful friends. Particular thanks to Harriet Rehmert, Emily Ogalbe, Sarah Loftus and Louise Clarkson for always being on hand to listen and provide advice.

Thanks to all my lovely family, with an extra special thank you to Joanne and Jeremy. Thanks for everything, I'm incredibly lucky to have you as both parents and friends.

Finally, a huge thank you to my partner, Rakesh, for putting up with me throughout all this. Thank you for your unwavering patience, understanding and belief.

Declaration

This thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy. It has been composed by myself and has not been submitted in any previous application for any degree.

The work presented (including data generated and data analysis) was carried out by the author except in the cases outlined below:

Chapter 2: Abdul-Rasheed Rabiu (ARR), specialist registrar in trauma and orthopaedics, acted as a second reviewer for the articles for eligibility assessment. Emma Sutton (ES), clinical lecturer, Abdul-Rasheed-Rabiu and Jonathan Young (JY), consultant trauma and orthopaedic surgeon, provided independent assessment and comments on the outcome measure categorisation framework in order to display results of the review.

Chapter 3: Abdul-Rasheed Rabiu acted as second reviewer for the article review for eligibility. Eleni Karasouli (EK), senior research fellow, provided independent data extraction and risk of bias assessment.

Chapters 4 and 5: Richard Grant (RG) and Karen Keates (KK), patient and public involvement representatives, reviewed and provided recommendations for the patient facing materials used in this study. Zi Heng Liew (ZHL), research fellow and physiotherapist, performed the duplicate coding for the analysis of the qualitative interview data.

Chapter 6: Helen Parsons (HP), associate professor in medical statistics, retrieved the data used in this project from the AIR trial online application and provided advice on statistical analysis. The AIR CRFs used to collect data for this project were originally developed for the purposes of the AIR trial and were developed in collaboration between the AIR chief investigator (CI) Rebecca Kearney (RSK), the AIR trial management group (TMG) and myself.

All chapters: This programme of research was embedded within an existing multi-centre randomised controlled trial (Ankle Injury Rehabilitation; AIR), for which I was the trial manager and member of the TMG. I had an active role in all trial management processes to ensure successful set up of the study, including collaborative development and design of the CRFs, ethics applications, contracts, trial risk assessment, monitoring plan, programming requirements, site initiation visits and ongoing support with sites once open to recruitment. The programme of

research for this thesis was embedded within the main AIR trial and was included as either a substantial amendment to the main AIR protocol or as an appendix to the main AIR protocol. All of these were prepared and submitted by myself, with approvals from the CI, sponsor and other colleagues as appropriate.

Funding

This work was funded by a National Institute for Health Research (NIHR) Career Development Fellowship for this research project (reference CDF-2016-09-009).

Parts of this thesis have been published by the author.

Publications as a result of work presented in this thesis

McKeown, R., Rabiu, A., Ellard, D. R. and Kearney R. S. (2019) Primary outcome measures used in interventional trials for ankle fractures: a systematic review. *BMC Musculoskeletal Disorders*. 20:388

McKeown, R., Ellard, D. and Rabiu, A., Karasouli, E. and Kearney, R. K. (2019) A systematic review of the measurement properties of patient reported outcome measures used for adults with ankle fracture. *Journal of Patient Reported Outcomes.* 3:70.

McKeown, R., Kearney, R. S., Liew, Z. H. and Ellard, D. R. (2020) What are the patient experiences of ankle fracture and the factors of most importance to them in their recovery? A qualitative interview study. *BMJ Open.* 10: e033539

Submitted publications as a result of work presented in this thesis

McKeown, R., Parsons, H., Ellard, D. R and Kearney, R. S. (2020) A study of the measurement properties of the Olerud Molander Ankle Score in adults with an ankle fracture. *Bone and Joint Research* (submitted for publication)

Published abstracts as a result of work presented in this thesis

McKeown, R., Kearney R. S., Liew, Z. H. and Ellard, D. R. (2020) Proceedings of BSRM and SRR: The patient experiences of ankle fracture: a qualitative study. BSRM and SRR conference proceedings. *Clinical Rehabilitation.* 34 (4). 551-567

Publications during PhD registration period related to wider AIR project

Kearney, R. S., McKeown, R. and Stevens, S. et al. (2018) Cast versus brace in the rehabilitation of patients treated for an ankle fracture: a protocol for the UK study of

ankle injury rehabilitation (AIR) multicentre randomised controlled trial. *BMJ Open.* 8:e027242

Kearney, R. S., McKeown, R. and Gallacher, D. et al. (2019) Ankle injury rehabilitation (AIR): a feasibility randomised controlled trial comparing functional bracing to plaster cast in the treatment of adult ankle fractures. *Pilot and Feasibility Studies*. 5:55

Submitted publications during registration period related to wider AIR project

Kearney, R. S., McKeown, R. and Nicholas Parsons et al. (2020) Ankle Injury Rehabilitation (AIR): a multicentre randomised controlled trial and economic evaluation. *The Lancet* (submitted for publication)

Podium presentations

- Warwick Medical School Postgraduate Research Symposium 2019 –*Measuring outcome in clinical trials for adults with ankle fracture* – University of Warwick, 21/05/2019
- Podium Presentation at The British Orthopaedic Association Annual Congress: *Primary outcome measures used in interventional trials for ankle fracture: a systematic review* – Liverpool, 12/09/2019
- Podium Presentation at The British Orthopaedic Association Annual Congress: *A systematic review of the measurement properties of ankle specific patient reported outcome measures used in clinical trials for adults with ankle fracture* – Liverpool, 12/09/2019
- Podium Presentation at The British Society of Rehabilitation Medicine and Society for Research in Rehabilitation Joint Scientific Meeting: *The patient experience of ankle fracture: a qualitative study* - Coventry, 14/10/2019
- British Orthopaedic Association Annual Congress. *A study of the validity, reliability and interpretability of the Olerud Molander Ankle Score in adults with ankle fracture* - September 2020 (accepted to present at BOA virtual conference)

Poster presentations

- Poster presentation at University of Warwick Postgraduate Research Showcase - *A feasibility study comparing plaster cast to functional bracing in the management of adults with ankle fracture.* University of Warwick, 07/06/2017 and winner of medicine category

- Poster presentation at University Hospitals Coventry and Warwickshire (UHCW) Research and Development Summit Event: *A feasibility study comparing plaster cast to functional bracing in the management of adults with ankle fracture.* Coventry, 14/07/2017
- Poster presentation at The Midlands and East Multidisciplinary Research Conference: *A systematic review of the measurement properties of ankle specific patient reported outcome measures (PROMs) used in clinical trials of interventions for adults with ankle fracture***.** Birmingham, 01/05/2019 and winner of the poster prize
- Poster presentation UHCW Research and Development Summit Event: *What are the patient experiences of ankle fracture recovery?* Coventry, 05/07/2019 and winner of 2nd place in the Non-Medic and Allied Health Professions category
- Poster presentation at the European Federation of National Associations of Orthopaedics and Traumatology (EFORT) congress: *A qualitative study of the relevance, comprehensiveness and comprehensibility of the Olerud Molander Ankle Score in adults with an ankle fracture* – Vienna, June 2020 (accepted but not presented due to the COVID-19 pandemic)
- Poster presentation at the European Federation of National Associations of Orthopaedics and Traumatology (EFORT) congress: *A study of the construct validity, internal consistency and interpretability of the Olerud Molander Ankle Score in adults with ankle fracture* - Vienna, June 2020 (accepted but not presented due to the COVID-19 pandemic)

Abstract

Aim and objectives: The aim of this thesis is to explore how outcomes are measured for adults with ankle fracture, and to assess whether they are appropriate, within the context of randomised controlled trials. To achieve this aim the following objectives were met using a variety of research methods. These were systematic reviews in chapters 2 and 3; qualitative methods to explore the construct of interest in the population and the content validity of the Olerud Molander Ankle Score in chapters 4-5 and finally secondary analysis of clinical trial data to determine the validity, reliability, responsiveness and interpretability of the Olerud Molander Ankle Score in chapter 6.

Key findings: Key findings from this programme of research demonstrated that there is a wide range of outcome measures collected in RCTs for adults with ankle fracture. Patient reported outcome measures were the most commonly used, but a further systematic review highlighted that these measures had insufficient measurement properties. The qualitative work showed that individuals experience a broad range of issues following ankle fracture, including reduced mobility, issues adhering to weight-bearing restrictions, loss of independence, sleep problems and impacts on psychological wellbeing. Some participants didn't feel the Olerud Molander Ankle Score was a relevant and comprehensive reflection of their recovery and some had issues with the response options. Results of quantitative analysis showed an underlying structure of two subscales within the score, sufficient convergent validity, responsiveness, internal consistency and no edge effects.

Conclusions: The Olerud Molander Ankle Score is likely measuring the construct of patient reported ankle function, rather than a holistic representation of the multifactorial concept of recovery from ankle fracture. It may be that this is appropriate for some research questions, but not all, dependent upon the construct of interest for the specific research question.

Abbreviations

AAOS	American Academy of Orthopaedic Surgeons Foot and Ankle Outcome Questionnaire
AAOS-CS	American Academy of Orthopaedic Surgeons Core Score
AAOS-SCS	American Academy of Orthopaedic Surgeons Shoe Comfort Scale
ADL	Activities of daily living
A-FORM	Ankle Fracture Outcome of Rehabilitation Measure
AIR	Ankle Injury Rehabilitation (trial)
ARR	Abdul-Rasheed Rabiu
ANOVA	Analysis of variation
AO	Arbeitsgemeinschaft für Osteosynthesefragen
AOFAS	American Orthopaedic Foot and Ankle Society (Ankle-Hindfoot Scale)
BOA	British Orthopaedic Association
CCC	Close contact cast
CFA	Confirmatory factor analysis
CFI	Comparative fit index
CI	Chief investigator
COS	Core outcome set
COSMIN	COnsensus-based Standards for the selection of health Measurement INStruments
CTT	Classical test theory
DRI	Disability Rating Index
EFORT	European Federation of National Associations of Orthopaedics and Traumatology
EK	Eleni Karasouli
ES	Emma Sutton

EQ-5D	EuroQol EQ-5D score
EQ-5D-M	EQ-5D-Mobility domain
EQ-5D-SC	EQ-5D-Self Care domain
EQ-5D-UA	EQ-5D-Usual Activities domain
EQ-5D-P&D	EQ-5D-Pain&Discomfort domain
EQ-5D-A&D	EQ-5D-Anxiety&Depression domain
FAAM	Foot and Ankle Ability Measure
FAAM–ADL	Foot and Ankle Ability Measure-Activities of daily living subscale
FAAM–S	Foot and Ankle Ability Measure-Sports subscale
FAOS	Foot and Ankle Outcome Survey
FAOS-P	Foot and Ankle Outcome Survey –pain subscale
FAOS-S	Foot and Ankle Outcome Survey – symptoms subscale
FAOS-ADL	Foot and Ankle Outcome Survey – activities of daily living subscale
FAOS-Sport	Foot and Ankle Outcome Survey – sport subscale
FFI	Foot Function Index
FFI-P	Foot Function Index-pain domain
FFI-AL	Foot Function Index-activity limitation domain
FFI-D	Foot Function Index-disability domain
FWB	Full weight-bearing
GCP	Good Clinical Practice
GIC	Global impression of change (score)
GRADE	Grading of Recommendations Assessment, Development and Evaluation
GSRF	Global self-rated function
HP	Helen Parsons
HRA	Health Research Authority

HRQoL	Health related quality of life
HSD	Honest significant difference
ICC	Intra-class Correlation Coefficient
IRAS	Integrated Research Application System
IRT	Item response theory
ISRCTN	International Standard Randomised Controlled Trial Number
ITS	Information Technology Services
JSSF	Japanese Society for Surgery of the Foot Score
JY	Jonathan Young
KK	Karen Keates
KOOS	Knee Injury and Osteoarthritis Outcome Score
KS	Karlsson Score
LCRN	Local Clinical Research Network
LEFS	Lower Extremity Functional Scale
LoS	Length of stay
MCAR	Missing completely at random
MGFA	Multiple group factor analysis
MIC	Minimally important change
MOXFQ	Manchester-Oxford Foot Questionnaire
NICE	National Institute for Health and Care Excellence
NIHR	National Institute for Health Research
NHS	National Health Service
ORIF	Open Reduction and Internal Fixation
OHS	Oxford Hip Score
OMAS	Olerud Molander Ankle Score
PCA	Principal component analysis

PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROM	Patient Reported Outcome Measure
PROMIS	Patient Reported Outcome Measurement Information System
QQ	Quantile-Quantile (plot)
REC	Research Ethics Committee
ROM	Range of Movement
RCT	Randomised Controlled Trial
RG	Richard Grant
RSK	Rebecca Samantha Kearney
RMSEA	Root mean square error of approximation
SEFAS	Self-reported Foot and Ankle Score
SEM	Standard Error of Measurement
SDC	Smallest Detectable Change
SF-12	12-Item Short Form Survey
SF-12-MCS	12-Item Short Form Survey – Mental Component Summary
SF-12-PCS	12-Item Short Form Survey – Physical Component Summary
SF-36	36-Item Short Form Survey
SF-36-MCS	36-item Short Form Survey – Mental Component Summary
SF-36-PCS	36-item Short Form Survey – Physical Component Summary
SFMA	Short Form Musculoskeletal Assessment
SOPs	Standard operating procedures
SRQR	Standards for Reporting Qualitative Research
TLI	Tucker Lewis Index
TMG	Trial Management Group
UHCW	University Hospitals Coventry and Warwickshire
VAS	Visual Analogue Scale

WCTU	Warwick Clinical Trials Unit
ZHL	Zi Heng Liew

Researcher training and development

Course	Organiser	Date
Good clinical practice (GCP) research training	Warwick Clinical Trials Unit (WCTU)	06/02/2017
GCP refresher training	WCTU	27/02/2019
Dealing with sensitive phone calls	WCTU	30/06/2017
Warwick Introduction to Management	University of Warwick	23/02/2018 and 20/04/2018
Chief Investigators Course	WCTU	31/01/2018
Fundamentals of the general data protection regulations	WCTU	23/04/2018
HRA training for non-commercial studies	Health Research Authority (HRA)	13/01/2017
Randomised controlled trials: a guide to design, analysis, interpretation and reporting	Centre for Statistics in Medicine, Oxford	25/09/2017 – 29/09/2017
Site file management and delegation of duties	Local Clinical Research Network (LCRN)	28/02/2017
Cost attribution in research and development	LCRN	17/03/2017
Making Integrated Research Application System (IRAS) work for your research amendments	LCRN	25/04/2017
Performing quality control checks for accurate data collection	LCRN	17/10/2017
Clinical Academic Research Development course	Coventry University and University Hospital Coventry and Warwickshire (UHCW)	03/05/2018 – 04/05/2018
Cochrane interactive learning course for completing and	Cochrane online training	March – October 2018

conducting an intervention review (Modules 1-9)		
EQUATOR Publication School	EQUATOR Network, University of Oxford	13/06/2019 – 14/06/2019
EndNote desktop	University of Warwick Information Technology Services (ITS)	22/02/2018
Microsoft Access	University of Warwick ITS	09/02/02/2017
Introduction to NVivo	University of Warwick ITS	01/10/2018
NVivo – working with data	University of Warwick ITS	19/02/2019
Microsoft Word - theses and long documents	University of Warwick ITS	15/10/2018
SPSS - introduction for beginners	University of Warwick ITS	31/05/2019
SPSS - statistical methods	University of Warwick ITS	03/06/2019
SPSS - data operations	University of Warwick ITS	12/06/2019
Literature searching	University of Warwick Research Skills Programme	27/04/2017
Academic poster design and creation	University of Warwick Research Skills Programme	10/05/2018
Academic writing: thesis structure	University of Warwick Research Skills Programme	10/05/2017
Academic writing: how not to write	University of Warwick Research Skills Programme	24/05/2017

Team working in a research environment	University of Warwick Research Skills Programme	08/07/2017-11/07/2017
How to be an effective researcher	University of Warwick Research Skills Programme	20/02/2018
Writing for publication – taking intentional steps to publication	Warwick Medical School	09/07/2018
Preparing to teach in higher education	Warwick Medical School	01/05/2018
Academic scientific writing: writing extended scientific articles and reports	University of Warwick Research Skills Programme	October 2018-December 2018 (2 hours a week)
Influencing and negotiating	University of Warwick Research Skills Programme	25/10/2018
Understanding research and critical appraisal online distance learning module	Warwick Medical School	April-July 2017
Qualitative research methods module	Warwick Medical School	3/12/2018 - 6/12/2018
Mixed methods research module	Warwick Medical School	25 and 26/02/2019 – 13 and 14/05/2019
Case Based Learning Facilitator Training for MBChB medical students	Warwick Medical School	18/12/2019
CBL facilitator – year 1, block 4 module, locomotion	Warwick Medical School	1 day per week from 20/02/2020 – 12/03/2020
Presenting on the virtual stage – how to design and deliver online presentations and present with impact when doing so virtually	National Institute for Health Research (NIHR) Academy	03/06/2020
Member of Qualitative and Mixed Methods Interest	Warwick Medical School	Various dates

Group – includes seminars and discussion groups		
Peer reviewer of articles for the Canadian Medical Association Journal and BMC Musculoskeletal Disorders	N/A	December 2019 and March 2020

Conference attendance

- 5th NIHR Musculoskeletal Trauma Trials Conference – Coventry, 11/01/2017
- British Orthopaedic Foot and Ankle Society – Sheffield, 1/11/2017 - 3/11/2017
- Warwick Medical School Research Symposium - Warwick University, May 2017, 2018 and 2019
- 6th NIHR Musculoskeletal Trauma Trials Conference – Bristol, 10/01/2018
- British Orthopaedic Association Annual Congress – Liverpool, 11/09/2019-13/09/2019
- The British Society of Rehabilitation Medicine and Society for Research in Rehabilitation Joint Scientific Meeting – Coventry, 14/09/2019 - 15/10/2019

1. Introduction and Aims

1.1 Ankle anatomy and physiology

1.1.1 Bones and joints of the lower limb and ankle

The lower limb begins superiorly at the joint between the sacrum and the pelvis known as the sacroiliac joint [1]. The pelvis and femur join in a ball and socket articulation of the hip joint [1]. The femur meets the tibia to form the knee joint and the sesamoid bone known as the patella sits anteriorly [1]. The lower leg is the portion of the limb below the knee joint and is formed of the tibia and fibula, which meet inferiorly to form the tibiofibular joint, with the tibia and the talus forming the talocrural or ankle joint [2]. The ankle forms the articulation between the lower leg and the foot, which is the body's contact to the ground [3]. The function of the lower limb is to support the body's weight in an upright standing position and for the purposes of locomotion [1].

The tibia bears weight transferred from the talus in the foot and extends inferiorly to the medial malleolus found on the medial aspect of the ankle [4]. The fibula is a comparatively thinner bone and is lateral to the tibia. The distal end of the fibula extends to a bony landmark known as the lateral malleolus, found on the lateral aspect of the ankle joint. The posterior malleolus is formed from the lower posterior ridge of the distal tibia [5]. The tibia and fibula join inferiorly to form a mortise, which allows articulation with the body of the talus. The distal section of the tibia, which forms the mortise articulation is known as the tibial pilon or plafond [6]. The mortise articulation of these three bones is the ankle or talocrural joint. The subtalar or talocalcaneal joint, is the articulation between the talus and the calcaneus is inferior to the ankle joint [7].

1.1.2 Muscles and ligaments of the ankle

The ankle is a hinge joint capable of dorsiflexion and plantarflexion [2]. Small amounts of inversion, eversion, abduction, adduction and rotation do, however, also occur at this joint [8]. Dorsiflexion is primarily produced by the tibialis anterior muscle, which is found on the anterior aspect of the tibia. Other muscles involved in dorsiflexion of the ankle are the extensor hallucis longus, extensor digitorum longus and the fibularis tertius [7]. The gastrocnemius and soleus muscles are primarily responsible for producing plantarflexion at the ankle joint [2]. The plantaris, tibialis posterior, flexor hallucis longus and flexor digitorum longus muscles assist with plantarflexion of the

ankle [7]. Inversion and eversion of the foot occurs at the subtalar joint [2]. The peroneal muscles evert the foot, made up of the peroneus longus, peroneus brevis and peroneus tertius muscles [2]. The tibialis anterior and tibialis posterior muscles invert the foot at the subtalar joint [7].

The tibiofibular complex is passively supported by several ligaments; lateral support is provided by the anterior talofibular ligament, the posterior talofibular ligament and the calcaneofibular ligament [2]. Medial support is provided by the three portions of the triangular deltoid ligament [4]. The syndesmosis joint is comprised of four ligaments to prevent the tibia and fibula from separating apart; the anterior tibiofibular ligament, the posterior tibiofibular ligament, the transverse tibiofibular ligament and the interosseous ligament [4 9].

1.1.3 Ankle biomechanics and role in the gait cycle

The joints of the foot and ankle play an important role in the facilitation of normal walking and locomotion. The gait cycle is the term used to describe one complete step during walking; beginning when the heel strike occurs of the first foot and ending when the heel strike occurs on the same foot again, in preparation for the subsequent step forward [8]. The gait cycle can be divided into two main phases; the stance phase and the swing phase [8]. The stance phase is the weight-bearing phase, which starts with the heel strike to the floor and ends with the toe off the ground and the swing phase consists of the foot being non-weight bearing as it moves forward for the start of the next step [8]. These phases are reliant on many complex interactions between the joints and structures of the lower limb which help to reduce fluctuations of the body's centre of gravity and achieve efficient locomotion [1].

The ankle joint withstands forces of approximately five times the body weight during walking and up to 13 times the body weight when running [3]. Studies of joint biomechanics throughout the gait cycle have shown that 83% of the total load is transmitted through the tibiotalar complex and 17% through the fibula [3]. Injuries of the ankle can have a profound effect on the stability of the joint and its ability to bear weight for purposes of locomotion [3].

1.2 Bone fractures

A fracture is a break or discontinuity of the cortex of a bone which occurs as a result of energy applied to it [10]. There are many different fracture patterns which can occur in bones within the human body. The pattern of the fracture will depend upon the direction upon which the excess energy is applied to the bone [10]. For example, a

spiral fracture is the result of a twisting energy applied to a bone [10]. Tension results in a transverse fracture and axial compression results in an oblique fracture pattern [10]. An avulsion fracture occurs when a tendon or ligament is pulled off from it's attachment along with a portion of the bone of which it is attached to [10].

Fractures can also be described as open or closed; open fractures describe those in which the fractured bone comes in to contact with an epithelialized surface, for example the skin [10]. In contrast, closed fractures are those which do not come in to contact with an epithelialized layer. Open fractures are associated with a higher risk of complication due to soft tissue involvement, such as infection [10]. A study by Donaldson, et al. [11] estimated that the yearly fracture incidence is 3.6 fractures per 100 individuals in the UK population.

Some types of fractures can occur as a result of other conditions. A pathological fracture occurs following application of a force which should not be sufficient to cause a fracture in healthy bone, for example in osteoporosis or types of cancer [10]. Osteoporosis is a systemic condition that affects the bones, leading to a decrease in the bone mineral mass along with changes in the microstructure of bones [12]. It mainly occurs due to the increased turnover of bone cells which is triggered by the hormonal changes, principally the reduction in oestrogen, during the menopause in females [12]. The decrease in bone mineral density leads to a reduction in bone strength and thus more susceptible to fracture [13].

1.2.1 Fractures of the lower limb

The symptoms of lower limb fracture are pain, swelling and an inability to bear weight on the affected limb [14]. Fractures of the lower limb cause a significant reduction in an individuals' mobility, ability to complete activities of daily living (ADL) and their health related quality of life (HRQoL) [15]. Fractures are broadly described as low energy fractures or high energy fractures. Low energy fractures are defined as fractures occurring as a result of falling from a standing height or less [16]. In contrast, high energy fractures are characterised by a fall from a height greater than standing or other high trauma situations such as crush injuries and road traffic accidents [16].

Fractures of the lower limb carry a social burden, as well as a financial one. A study by Pasco, et al. [15] assessed the consequences of fractures in a sample of females aged 35 years or over for 12 months following their injury. Participants reported a reduction in ability to perform activities of daily living (ADL), such as personal care,

household chores, work and leisure activities, as well as a general loss of independence. Those who sustained a lower limb fracture reported reduced confidence mobilising and a significant decrease in their pre-fracture mobility status [15]. Fractures of the proximal femur are the most common lower limb fracture, followed by ankle fractures which are the second most common fracture of the lower limb [17].

1.3 Ankle fractures

A malleolar ankle fracture is one which involves the distal part of the tibia and fibula, without any involvement of the central part of the ankle joint [18]. There are various patterns of malleolar ankle fracture and the resulting pattern of injury is dependent upon the amount of energy and the manner in which the energy is applied to the bone [10]. This is discussed in more detail in section 1.3.3.

A pilon fracture is a subgroup of ankle fractures and describes a more severe injury than a malleolar ankle fracture. It involves an intra-articular fracture of the distal tibia and the central articulating surfaces of the ankle joint [10]. Pilon fractures are less prevalent than malleolar fractures and are usually the result of a high energy force applied to the lower leg. They are associated with more complex soft tissue injuries than malleolar fractures, as well as higher risks of complications [19]. This thesis is primarily concerned with the measurement of outcome and progress in individuals with a malleolar ankle fracture, therefore pilon fractures are outside the scope of this thesis. The term ankle fracture will be used to describe specifically malleolar fracture throughout.

1.3.1 Epidemiology of ankle fractures

A report by the National Health Service (NHS) hospital episode statistics showed there were 59, 266 adult individuals (≥18 years) admitted to hospitals in England with a fracture of the lower leg or ankle from 2017-2018 [20]. Ankle fractures were found to be the fifth most common type of fractures treated in a single trauma centre in Scotland [17] and it was estimated by the same author in a separate study that they contribute to approximately 9% of a trauma surgeons workload [21]. The injuries occur in both males and females, displaying a bimodal distribution occurring most frequently in younger males and older females [21]. Whilst ankle fractures are not described as osteoporotic fractures as a group, bimalleolar and trimalleolar ankle fractures are considered to be osteoporotic in nature [17].

There have been several epidemiological studies assessing the incidence of ankle fracture and projecting future trends. A study completed by Kannus, et al. [22] assessed the incidence of ankle fractures in Finland in individuals aged 60 years and over from 1970-2000. The authors found that the incidence of ankle fractures was 369 per 100,000 per annum in 1979 which increased to 1545 in 2000 [22]. The researchers predicted this increasing incidence to continue to the year 2030.

A study by Daly, et al. [23] assessed the epidemiology of ankle fractures in Minnesota from 1979-1981. The authors found an overall incidence of 184 fractures per 100,000 per year in the local population. Authors also noted that a high proportion of the injuries occurred in a young male population. Bengnér, et al. [24] studied the incidence of ankle fractures in Malmo, Sweden in 1950-1952 and compared the figures with those found in 1980-1982. The total number of fractures was found to be 383 in the period of 1950-1952 compared with 739 in the period of 1980-1982. The incidence was high in both younger males and older females in this study.

Thur, et al. [25] studied 91,410 inpatients in Sweden between 1987 and 2004 to assess the epidemiology of ankle fractures. Authors found that, in the study period, 91, 410 patients were treated as an inpatient for an ankle fracture. Of this, 57% were female with a mean age of 58 years, compared to the mean age of males in the sample being 45 years. Females were most likely to present with a bimalleolar or trimalleolar fracture, whereas unimalleolar fractures of the lateral malleolus were more common in males. In the 18 years study period, ankle fractures increased by 0.2% overall, by 0.3% in females and 0.1% in males.

Court-Brown, et al. [26] studied 1500 ankle fractures which presented to a single hospital in Scotland over three years. The highest incidence of injuries were found to be in males between the ages of 15 and 24 years and females between the ages of 75 and 84 years. Whilst authors found the incidence of ankle fractures was higher in males than females, their findings supported those of the above study in that the incidence of ankle fractures in females were growing at the fastest rate.

1.3.2 Ankle fracture presentation and diagnosis

The most common mechanism of injury for ankle fractures is a low energy fall or twisting inversion motion at the ankle [9]. For this reason, the majority of ankle fractures are isolated injuries, with approximately 5% of individuals presenting with concurrent injuries [9 27].

Individuals with an ankle fracture usually present with pain, swelling and inability to weight bear on the affected limb, describing an event of trauma to the ankle [2 4]. In the more severe cases or instances where the joint has dislocated, a bony deformity at the ankle is evident on observation [9]. If the injury has occurred from high energy trauma patients should be examined according to advanced trauma life support principles due to the possibility of other injuries [9].

There are many ligaments and soft tissues around the ankle complex, as described previously, which are damaged during trauma to the ankle, therefore a proportion of individuals presenting to the hospital with a history of ankle trauma will have sustained a ligament sprain rather than a fracture. The Ottawa ankle rules can be used to limit unnecessary radiographs [28]. If one of the four Ottawa ankle criteria are a met by the patient's symptoms, then a fracture cannot be ruled out and diagnostic imaging is required [29]. The use of the Ottawa ankle rules are recommended by the National Institute for Health and Care Excellence (NICE) in the immediate assessment of ankle injuries [30].

If further investigations are warranted following the Ottawa assessment, radiographs (X-Rays) are used to diagnose an ankle fracture. The X-Ray views most commonly used in this instance are anteroposterior and lateral views [9]. Examples of an anteroposterior and lateral X-Ray of an ankle fracture is shown in figure 1. Computed tomography scans can also be used in the cases of severe fractures to assess intra-articular involvement and assist with pre-operative planning [9].

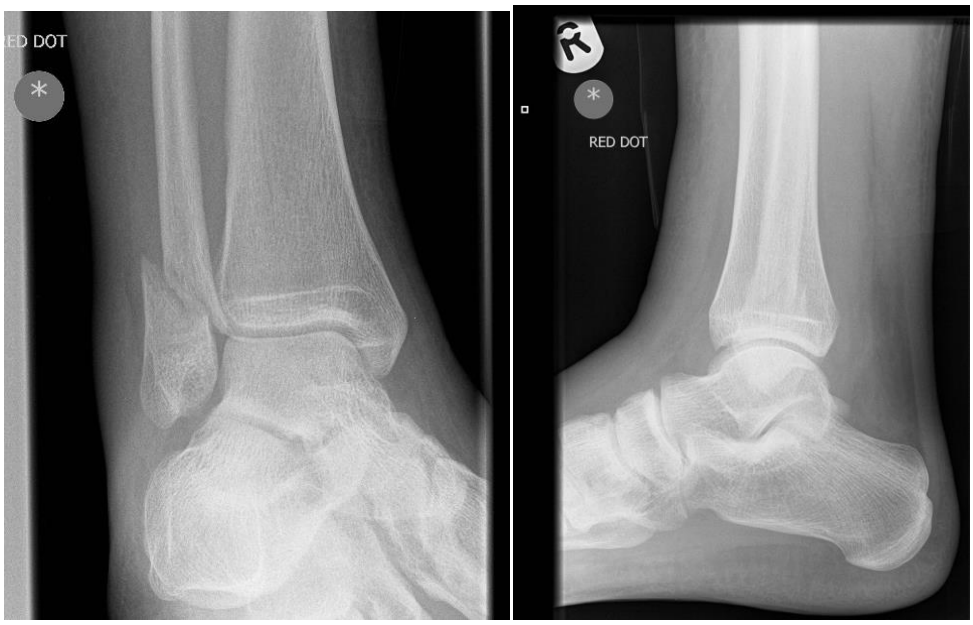


Figure 1 - Plain X-Ray radiograph, anteroposterior and lateral views of a unimalleolar Weber B ankle fracture. (Case courtesy of Dr Jeremy Jones, Radiopaedia.org, rID: 8787) [31]

1.3.3 Classification of ankle fractures

Classifying fractures into groups of similar injuries is useful for standardising treatment and understanding the prognosis of different fracture types [32]. Griend, et al. [4] explain that any classification system which incorporates all types of ankle fractures will be too long and complex to remember or be clinically useful because of the wide range of combinations of bony and ligamentous injuries which occur at the ankle. One method of describing ankle fractures is using the number of malleoli involved in the injury [9]. One, two or three of the malleoli can be involved in the fracture, known as unimalleolar, bimalleolar and trimalleolar ankle fractures respectively.

The ankle joint can be considered as a circle shape, which is made up of bones (tibia, fibula and talus) and ligaments (lateral collateral ligaments, deltoid ligaments and the syndesmosis ligaments) to remain stable [33]. Depending upon the nature of the disruption to this circle, caused by bony fracture/s and ligament damage, ankle fractures can be described as either stable or unstable. Whilst previously it was believed that the lateral structures were most important in whether an ankle fracture is stable, it's now known that the integrity of the medial deltoid ligaments is of key importance in determining the stability of an ankle fracture [33]. Bimalleolar and trimalleolar fractures are usually unstable, as the ankle circle previously described

has broken in two places [33]. Ligamentous damage is more difficult to assess because this is not visible on radiographs.

The Weber ankle fracture classification categorises injuries based upon the level of the fibular fracture in relation to the syndesmosis [4]. The Weber classification system is shown in figure 2. Weber A ankle fractures are those of the lateral malleolus which are distal to the syndesmosis complex, thus this structure remains intact. Weber B fractures involve a fibular fracture at the same level of the syndesmosis. The syndesmosis can be intact, partially ruptured or completely ruptured which can cause variation in the stability of the joint in the case of Weber B fractures [34]. Weber C fractures describe a fracture of the fibula proximal to the syndesmosis, which is completely ruptured in most cases [34]. Whilst the Weber classification is simple to use, it's main disadvantages are that it does not describe the insult of the medial portion of the ankle joint, nor can it determine stability of the joint [33].

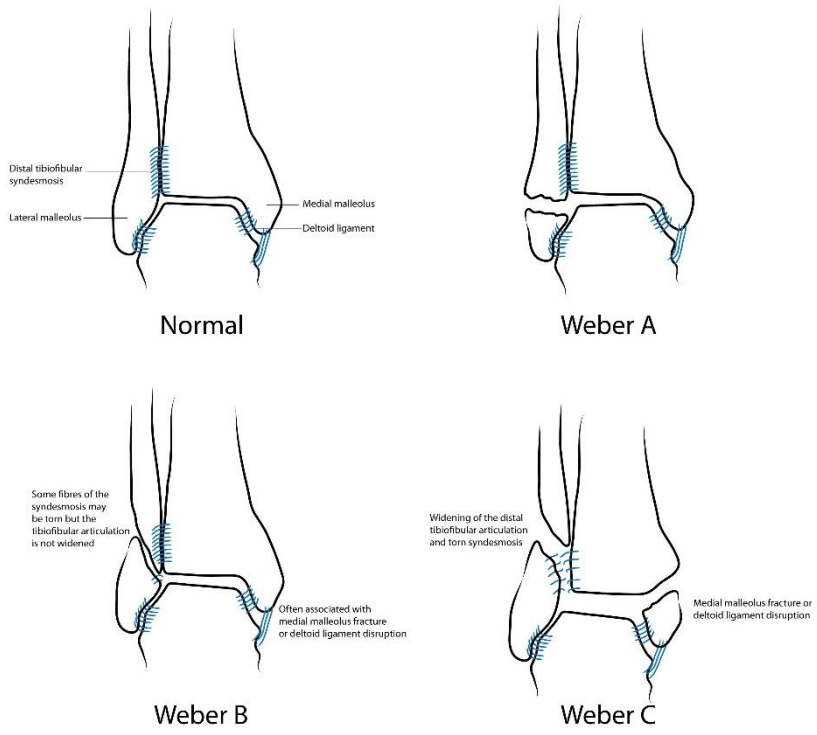


Figure 2 - Weber ankle fracture classification system. (Case courtesy of Assoc Prof Frank Gaillard, Radiopaedia.org, rID: 9642) [35]

The Lauge-Hansen classification categorises fractures based upon the position of the foot and the direction of force applied to the joint at the time of the injury [9]. The categories are pronation-abduction, pronation-external rotation, supination-adduction and supination-external rotation [34]. The benefits of this system are that it

is able to consider all components of the injury, inclusive of posterior and medial structures [33].

The most recently developed classification system is the Arbeitsgemeinschaft für Osteosynthesefragen (AO) classification, which divides the three Weber fracture types into three main subgroups (A1, A2, A3, B1, B2 etc.). There are further subdivisions of each of these subgroups to form a total of 27 different fracture patterns [34]. The advantages of this system is that it is very comprehensive due to the large variety of fracture patterns it encompasses. This advantage limits its usefulness in clinical practice, however, because of the number of options available to describe the many different injuries [33].

1.3.4 Fracture healing and repair

Secondary fracture healing occurs when the fracture is not fixed rigidly, such as in the case of external fixation or a plaster cast [32]. Secondary bone healing can be divided into five stages; haematoma, inflammation, soft callous formation, hard callous formation and remodelling [32]. The haematoma stage involves localised tissue damage and bleeding at the fracture site. The inflammatory process begins which lasts up to seven days following fracture, and involves the release of cytokines and growth factors to stimulate production of fibroblasts and osteoblasts [32 36]. The soft callous formation begins approximately two to three weeks following fracture and soft callous is built around the fracture site, comprised of collagen initially, which eventually forms cartilage [37]. Hard callous formation will commence after this and involves the formation of an immature woven bone [37]. Finally, bone remodelling occurs, whereby lamellar bone replaces the woven bone, and this can last from a period of several months to several years [32]. This process allows the bone to return to full strength by adapting to mechanical stresses placed upon it, a principle known as Wolff's law [36].

Primary bone healing, in contrast to secondary bone healing, occurs when the ends of the bone involved in the fracture are completely stable and not able to move, usually after they've been fixed with plates and screws [10]. In order for bone to heal primarily, there must be no micro-movement between the bones and no gap between the bone ends, creating a condition known as absolute stability [10]. This type of bone healing occurs in the same way as secondary bone healing, but without the stages of callous formation [10]. There are several factors which can delay fracture healing. These include nutrition, hormonal influences, diabetes, non-steroidal anti-

inflammatory drugs, age, smoking, blood supply to the fracture, severity of fracture and the adequacy of the fracture fixation [10 36 37].

1.4 Management of individuals with an ankle fracture

The principle of general fracture management can be summarised in three phases; reduce, hold and rehabilitate [10]. Reducing a fracture is when the bones are realigned to their correct position [14 38]. The key for this stage is to regain stability and alignment of the ankle [33]. All fracture dislocations should be reduced as an emergency to ensure vascular supply to the foot and ease skin tension created by the injury [9]. Holding a fracture involves keeping the fracture stable to ensure the bone heals in a suitable position, and this is usually either performed with an open reduction and internal fixation (ORIF), a plaster cast or close contact or moulded cast [10]. Rehabilitation aims to restore and individual's function with the goal of returning to pre-fracture ability to perform usual activities of daily living [39].

1.4.1 Operative management of individuals with ankle fracture

As described previously, ORIF is usually required for unstable ankle fractures [29]. NICE guidelines advise that, for individuals requiring ankle fracture fixation, the operation should occur either on the day of injury or the following day [30]. If the swelling is too severe, operative fixation may need to be delayed [40]. The decision on whether to operate on an ankle fracture or treat conservatively depends on the stability of the fracture [33]. In some cases, operative fixation may be avoided due to patient factors such as age, skin condition or presence of diabetes and in these instances, a close contact or moulded cast may be preferable (see section 1.4.2).

When surgical fixation is required it is usually accomplished by reducing the fibula to its natural length and alignment and placing a plate and screws laterally [40]. The syndesmosis stability is assessed intraoperatively and fixed if necessary using syndesmosis screws or an alternative fixation device [40]. If a medial malleolar fragment is present, this is usually fixed with screws into the tibia [40]. There is debate surrounding the fixation of the posterior malleolus and when to perform this fixation, but if this is fixed, it is usually done so with screws [9].

1.4.2 Close contact casting and moulded cast

A particular type of first line treatment is called a close contact cast (CCC) or moulded cast and is usually applied under anaesthetic in a theatre setting following manipulation and reduction of the fracture [41 42]. The CCC is applied using foam pads

under the cast at particular points about the ankle to maintain the reduction and stability of the fracture [41]. This intervention is often used for older individuals, who may not be suitable for ORIF due to skin conditions or for diabetic patients who have particular issues with soft tissue healing [43]. A large multi-centre RCT published in 2016 studied ORIF against CCC for individuals aged 60 or above with an unstable ankle fracture and found that CCC provided in similar functional outcomes for this patient population in comparison to ORIF [41]. A similar study is currently being undertaken to compare outcomes of these interventions in a younger patient population [44].

1.4.3 Immobilisation methods for individuals with ankle fracture

Regardless of whether a fracture is operatively or conservatively managed, a period of immobilisation is usually required to protect the fracture site and allow bone healing to take place with minimal disruption [10]. The traditional method of ankle immobilisation is using a plaster cast, which is applied to the lower leg to keep the bones stable during healing [38 45].

A fixed angle removable orthotic or a functional brace is a newer alternative to the more traditional immobilisation method of a plaster cast [46 47]. The functional brace is able to be removed to complete ROM exercises and for personal care. Currently the optimal method of immobilisation for ankle fractures is not known, which was highlighted in a 2012 Cochrane review on rehabilitation for adults with ankle fractures [48].

1.4.4 Rehabilitation for individuals following ankle fracture

Rehabilitation is a process which allows optimisation of an individual's full recovery potential following an episode of disease, injury, surgery or illness [49]. A Cochrane review regarding the rehabilitation of adults with an ankle fracture has previously been completed. This concluded that further high quality RCTs were required to draw conclusions on rehabilitation strategies for this patient population. This includes immobilisation methods, early exercises, stretching, manual therapy and exercise interventions in the later stages of rehabilitation [48].

For stable ankle fractures which are conservatively managed, rehabilitation can commence quickly and aims of the initial phase are to reduce swelling, increase weight-bearing as guided by weight bearing restrictions and restore full range of movement (ROM) at the joint [27]. Strengthening exercises aim to increase the strength of the muscles around the ankle, paying particular attention to the muscle which evert

the foot due to their role in stabilising the ankle joint [27]. Stretching exercises to maintain mobility in the muscles around the ankle are often indicated, particularly following a period of prolonged immobilisation [50]. Exercises to improve the ankle proprioception, such as single leg stands and wobble board exercises are examples of late stage rehabilitation, which are beneficial in progression to higher level sports activities [51 52].

For operatively managed ankle fractures, the immediate post-operative phase of rehabilitation should focus on mobilisation with weight-bearing restrictions as guided by the operating surgeon, the use of walking aids and basic early ROM exercises [52]. With regards to weight-bearing status, there is uncertainty surrounding the optimal weight-bearing protocols for individuals following an ankle ORIF and research shows the clinical practice is variable in the UK [53]. An ongoing RCT is assessing the optimal weight-bearing protocols following ankle fracture fixation [54] but other studies suggest early weight-bearing might be beneficial as opposed to non-weight bearing [55]. Once weight-bearing restrictions are relaxed and immobilisation removed, increasing the ROM at the joint, as well as functional strength is the main priority. Exercises can be progressed to higher level strengthening and proprioceptive work as able.

1.4.5 Complications associated with ankle fractures

Wound breakdown and infection following operative fixation is a potential complication and is more prevalent in patients with diabetes [40 56]. Other complications of operative interventions include nerve injury, tendon injury, excessive bleeding, failure of fixation and deep infection [43]. Deep vein thrombosis (DVT) is also a complication of both operative intervention and prolonged immobilisation of the ankle [57 58]. Plaster sores are a possible complication of plaster cast treatment [59].

A fracture non-union is defined as a fracture which has not healed within nine months [10] and most commonly affects the medial malleolus in ankle fractures [4]. Factors which can contribute to non-union are the same as outlined previously regarding factors which delay bone healing. Management is dependent upon the cause of non-union and usually requires either further intervention to reduce and re-fix the bones or bone grafting [10].

A mal-union is a fracture that has healed in a poor anatomical position and can limit function of the affected joint [10]. Incongruent joint surfaces can alter joint

biomechanics leading to loss of normal joint function and degenerative changes [4]. Mal-union can be corrected using osteotomy procedures but results are often poor if degenerative changes have already occurred as a result of the mal-union [4]. Post-traumatic ankle arthritis is a potential long-term complication of ankle fracture which is often associated with mal-union and there are very few effective treatment options [40 60]. In severe cases, ankle arthrodesis (fusion) may be necessary [40].

Complex regional pain syndrome (CRPS) is a possible complication following an ankle fracture and it is a poorly understood syndrome characterised by swelling, tenderness and trophic changes around the foot and ankle [40 61]. CRPS typically involves a combination of pain which is out of proportion in both intensity and duration, changes in temperature and skin colour, swelling, reduced strength and movement, sensory changes and reduced function [61]. The management of CRPS is complex and multidisciplinary; intervention should be commenced as early as possible to optimise chances of recovery [61].

1.5 Evidence-based healthcare

From previous sections of this chapter, it is clear that the treatment pathway for ankle fractures can be varied and many questions remain regarding the optimal management protocols for the many different types of fracture in this patient population. To enable clinicians to be adequately informed on the best ways of treating conditions like ankle fractures, they rely on a collective body of research, known as evidence-based healthcare. Evidence-based healthcare is defined as "the use of mathematical estimates of risk of benefit and harm, derived from high-quality research on population samples, to inform clinical decision-making in the diagnosis, investigation or management of individual patients" [62 (p.1)]. Appropriately designed and high-quality research in healthcare is imperative for scientific progress in the diagnosis and management of disease and injury [63].

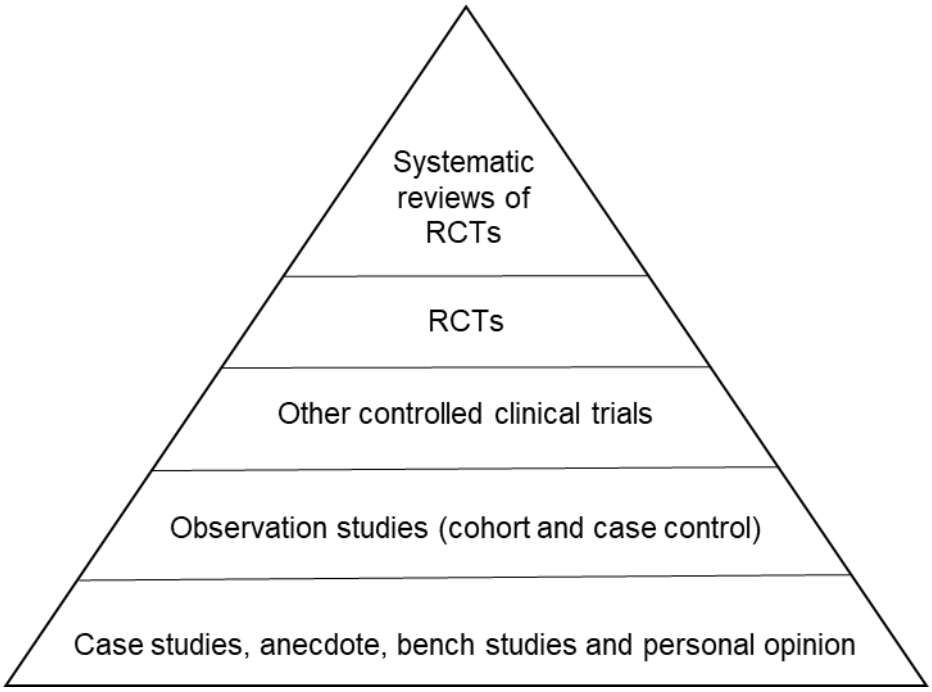


Figure 3 - Hierarchy of evidence [64]

Figure 3 shows the evidence-based medicine hierarchy of evidence, taken from Greenhalgh [64]. It depicts the levels of types of study and to that effect, the level to which each type of study can be trusted in relation to its design [64]. The further down the triangle a type of study is, the higher the risk of bias in the study due to its methodological design [65].

From figure 3, we can see that the very top of the hierarchy of evidence are systematic reviews of randomised controlled trials (RCT), which describe a synthesis of evidence done by pooling the results of similar RCTs in order to answer a particular research question. RCTs are considered the gold standard of clinical effectiveness research, and are explained in further detail in the subsequent section. Other controlled clinical trials refer to trials which compare the effects of an intervention against another in a non-randomised sample of a population [66]. Observational studies such as cohort or case controlled studies involve the monitoring of a sample of a population based upon specific characteristics or exposure and followed up over time to see which of them develop a specific outcome [66]. Generally, cohort studies are interested in the incidence or prevalence of certain diseases or outcomes [66]. Finally, the bottom of the hierarchy includes case studies and expert opinion pieces, which discuss a single patient case or accounts written on the basis of experience [66].

1.5.1 Randomised controlled trials

An RCT is considered the gold standard methodology for assessing the effectiveness of an intervention in a population of interest [67 68]. The defining feature of a clinical trial, as described by Altman [69], is the comparison of two (or more) groups who only differ by the intervention they have received with the aim of ensuring that nothing other than the intervention being studied is causing a change in an individual's health status [67]. Bias is introduced if the two groups differ in any way which is not in relation to the intervention given and bias should be reduced as much as possible [69].

In clinical trials, there are two different error types which can affect results. A type 1 error occurs when a difference is found between interventions when one doesn't actually exist, also known as a false positive result. This means the difference found between the two groups has happened for another reason (e.g. other patient characteristics) other than the intervention provided and can be a result of sampling bias. Type 2 error is when the results of an RCT fails to find a difference between interventions when one actually exists, also known as a false negative result. Type 2 error occurs from an inadequate sample size and lack of statistical power [70].

When designing and reporting clinical trials, a checklist is useful tool ensuring that key aspects of bias reducing activity has taken place to demonstrate a robust methodology, and one should not assume that they took place if they are not specifically reported in the manuscript, [69]. Perhaps the most well-known and frequently used tool for reporting RCTs is the Consolidated Standards of Reporting Trials (CONSORT) statement [71], developed to improve transparency in reporting of clinical trials and for the adequate assessment of bias which could be present in RCTs.

1.5.2 The Ankle Injury Rehabilitation trial

The Ankle Injury Rehabilitation (AIR) trial is a multi-centre RCT comparing a plaster cast to a functional brace in the management of adults with an ankle fracture (ISRCTN 15537280) [72]. The programme of research undertaken in this thesis was embedded within this trial and my role in this study is described in the declaration section of this thesis. Recruitment took place from 20 NHS hospital sites across the UK. Both operatively and non-operatively managed individuals with an ankle fracture for which the treating clinician feels plaster cast would be a reasonable management option were eligible for inclusion. The eligibility criteria can be found

in table 1 and this thesis is centred on the population described: individuals aged 18 years or over with closed malleolar ankle fractures.

Table 1 - Eligibility criteria for the AIR trial

Inclusion Criteria	Exclusion Criteria
Provision of written informed consent.	Ankle fracture secondary to known metastatic disease.
Aged 18 years or over.	Complex intra-articular fracture (e.g. Pilon Fracture).
A closed ankle fracture for which the treating clinician would consider plaster cast a reasonable management option.	In the opinion of the surgeon the patient would require manipulation and close contact casting.
Within 3 weeks of operative management or injury if non-operative.	In the opinion of the surgeon the patient would require manipulation and a moulded cast.
	Previous ankle fracture randomised in the present trial.
	Evidence that the patient would be unable to adhere to trial procedures or complete postal questionnaires.
	Known pre-existing neuropathic joint disease contraindicating functional brace intervention.

In the AIR trial, the primary outcome was the Olerud Molander Ankle Score (OMAS) at 16 weeks post randomisation. Other outcomes collected were the Disability Rating Index (DRI), Manchester Oxford Foot and Ankle Questionnaire (MOXFQ) and the EuroQol EQ-5D-5L, along with an analysis of resource use for

health economic purposes. Complications were also collected to assess the differences in adverse events between the intervention groups.

1.6 Health and health related quality of life

Health is defined by the World Health Organisation as "a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity" [73]. There are several definitions of HRQoL in the literature. Landgraf [74] defines it as "the subjective evaluation of one's ability to perform usual tasks and their impact on one's everyday physical, emotional and social well-being." (p. 346).

There are two main philosophies of health and human functioning; the biomedical model and the biopsychosocial model [75]. The biomedical model relies on the empirical scientific underpinning of disease to explain symptoms and functioning in humans, identifying dysfunctions in cells, organs and organ systems to define disease or injury [75]. The biopsychosocial model of health has become more prominent in the understanding of health and disease. This model is an expansion on the biomedical model, encompassing the psychological and social factors which can affect health and disease states [75].

This is has been further explored by Wilson and Cleary [76], who outlined a conceptual model linking clinical variables to HRQoL factors. Figure 4 shows the conceptual model, which the authors use to describe a continuum of increasing complexity in terms of biological, social and psychological factors.

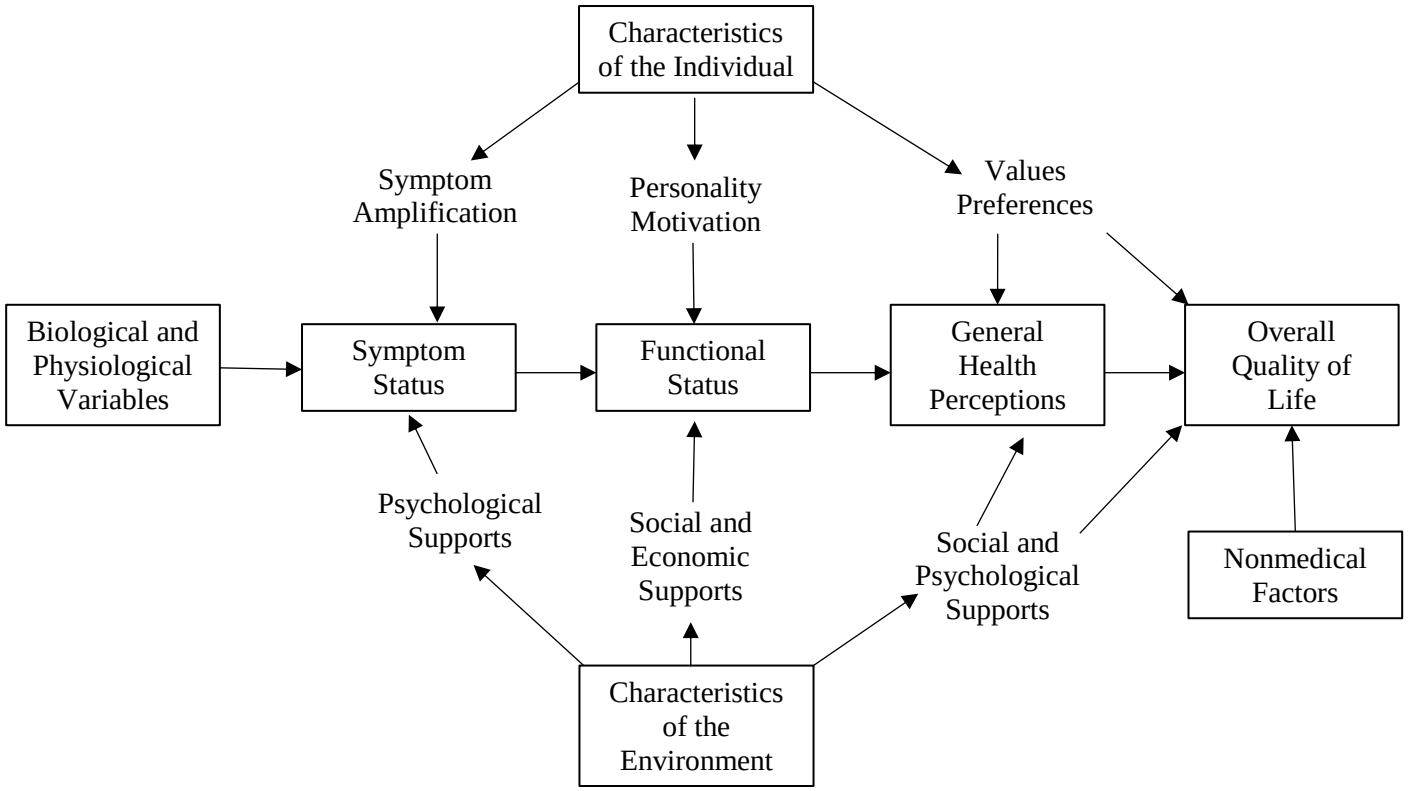


Figure 4 - Wilson and Cleary's health related quality of life conceptual model. Taken from Wilson and Cleary [76]

Biological factors include the basic and fundamental determinants of disease and injury, which are changes or disruptions to cells, organs and organ systems. Symptoms are the effects of these biological changes on the person as a whole, which can be physical or psychological in nature [76]. Despite this causal relationship between biological changes and symptoms, this relationship is not always linear or correlated with one another.

Functioning is defined as an individual's ability to complete specific tasks and activities [76]. Symptoms may influence functioning, such as a painful joint affecting walking distance, however, other factors outside of symptoms also affect function, such as inherent traits, social support and motivation [76]. General health perceptions are subjective amalgamation of the concepts of biological, symptom and functioning as well as mental well-being [76].

Overall quality of life is the final concept in the model, which is a general sense of how happy and satisfied an individual is with life. The authors explain how the psychological and emotional factors' relationship with each of the levels of the conceptual model is complex and therefore not included in the model figure, although interacts at all levels [76]. As you move from left to right in the model, the concepts become increasingly more complex in their definition and ability to be measured accurately.

1.6.1 Measuring health

Outcome measures used as an assessment of change, either over time or following an intervention, are vital in the completion of clinical research, audit and healthcare practice [77]. As explained by Altman [69], the sample size of an RCT should be calculated on a single variable or outcome measure, known as the primary outcome measure. Additional outcomes collected in an RCT are known as secondary outcome measures. As previously discussed, the CONSORT statement is a tool to ensure transparency in the reporting of RCTs and an element of this is the choice and use of an appropriate outcome measure, of which should be clearly pre-specified as primary and secondary outcome measures [71]. Furthermore, the statement outlines the requirement to demonstrate the sample size calculation, based upon the minimally important change (MIC) of an outcome measure usually, to ensure the RCT is adequately powered to detect treatment effects and thus minimising type 2 error rate [78].

1.6.2 Clinical and objective measures in trauma and orthopaedics

When considering the conceptual model of clinical variables and HRQoL described by Wilson and Cleary [76], clinical or objective measures usually assess the first two concepts of the model; biological and symptomatic. In trauma and orthopaedics specifically, clinicians often use X-Ray images to diagnose and assess the bone union at a fracture site, or assess joints for signs of joint disease [10]. Measures of the function of joints, such as ROM and strength tests allow for clinicians to assess impairments of the musculoskeletal system [79]. Some also assess functioning, such as the timed get up and go test, specific functional tests such as hopping ability or heel raises [80].

1.6.3 Patient reported outcome measures

For a long time, clinical research has focussed on measuring the avoidance of undesirable outcomes (i.e. death and complications) or clinician derived objective measures as opposed to the attainment of positive outcomes (e.g. quality of life) [77] [81]. In recent decades, however, there has been an increasing trend towards the use of patient reported outcome measures (PROMs) [82]. A PROM is defined as a "measurement of any aspect of a patient's health status that comes directly from the patient, without interpretation of the patient's responses by a physician or anyone else" [83 (p.8)]. The use of PROMs has become more frequent in clinical research studies and many researchers and policymakers are moving beyond the standard endpoints of mortality rates and clinical objective measures to measures of health status such as HRQoL [84].

In 2009 the Department of Health prioritised the routine use of PROMs in the NHS to assess effectiveness of care following health care interventions, outlining plans to collect PROMs for hip and knee replacement, hernia repair and varicose vein surgery [81 85]. PROMs serve an important role in the demonstration of value-based care and to demonstrate productivity from a health service [81 86]. As a publicly funded health service, the NHS has limited resource and it's of paramount importance that those resources are provided in the right way and the right time to produce the maximum benefit for patients and ultimately taxpayers.

In trauma and orthopaedics, several PROMs are used in clinical trials to assess outcome. These can be condition specific measures, such as the Western Ontario and McMaster Universities Osteoarthritis Score [87] or the Achilles tendon Total Rupture Score [88]. There are a range of region specific measures, for example the disabilities of the arm, shoulder and hand questionnaire [89]. Generic HRQoL measures are also

widely used in the field of trauma and orthopaedics, such as the 36 item Short-Form survey (SF-36) [90]. Some generic PROMs also enable measures of cost utility, such as the EQ-5D-5L [91] which enable analysis of the cost effectiveness of a intervention in terms of the value added in years in full health, a concept known as a quality adjusted life year [92].

1.6.4 Measurement properties of PROMs

Measurement properties (or psychometrics) of a PROM are concerned with the ability of an instrument to measure a particular construct accurately, with the ability to detect changes in an individuals' status and with resulting scores that are clinically interpretable [93]. These measurement properties are not describing the PROM alone, but the PROM used in a particular patient population in a specific context of use [77] [83]. There are four main measurement properties which I will outline below.

Validity is the ability of an outcome to measure the construct that it intends to measure [77] [94]. Reliability is concerned with an outcome measures ability to assess an individual's status in a way which is free from measurement error [77] [95]. Responsiveness is concerned with a measures ability to detect changes in the construct of interest when one actually occurs [77] [96]. Interpretability is related to the clinical meaning of scores, or changes in scores, of an outcome measure [77] [97] and there are several concepts which fall under the definition of interpretability. These measurement properties will be described in further detail in chapter 6.

1.7 Aims of thesis and research questions

The primary aim of this thesis is to evaluate the way in which outcome is assessed in a population of adults with an ankle fracture and to determine whether it's appropriate, within the context of RCTs.

There are three key objectives to achieve this aim:

1. To review the outcome measure usage in RCTs of the population of adults with ankle fractures and assess the evidence for the quality of the outcome measures.

This will answer the following research questions:

- What types of outcome measures are used in RCTs of non-pharmacological interventions for adults with an ankle fracture?
- What are the temporal trends in the use of different types of outcome measure in these studies?

- What is the evidence for the measurement properties of the PROMs collected in this study type within the patient population?
- What is the evidence for the interpretability and feasibility of these PROMs within the patient population?
- What is the best ankle specific PROM to use in this patient population?

2. To explore the construct of interest for adults recovering from an ankle fracture.

This will answer the following research questions:

- What are the patient experiences of ankle fracture recovery?
- What are the factors of most importance in individual's recovery form an ankle fracture?
- What are the patient experiences of ankle fracture in people of different ages and genders?

3. To evaluate the measurement properties of an appropriate PROM used in this patient population within the context of a multi-centre RCT.

The score will be chosen based upon the work presented in the initial two packages of work, which are systematic reviews to explore outcome measures in this area, as detailed above.

This will answer the following research questions:

- What is the relevance and comprehensiveness (content validity) of the score in a sample of adults with an ankle fracture?
- What is the validity of the score in a sample of adults with ankle fracture?
- What is the internal consistency of the score in a sample of adults with ankle fracture?
- What is the responsiveness of the score in a sample of adults with ankle fracture?
- What is the interpretability of the score in a sample of adults with ankle fracture?

1.8 Research philosophy

The differences in contrasting research approaches are most obvious when comparing quantitative and qualitative research. Positivism views scientific enquiry as objective and separate from the social world. It involves testing hypothesis through measurement of specific variables and therefore is the approach most

commonly used in quantitative research methods [98], such as in randomised controlled trials which are discussed earlier in this chapter. In contrast, interpretivism views reality as a purely socially constructed concept and attempts to generate meaning through research, usually through the use of qualitative research methods [99].

The aims of this thesis are wide ranging, and include both exploratory work, to understand the patient experience and factors of most importance to individuals with an ankle fracture as well as work to determine the measurement properties of a commonly used outcome measure. In order to explore the patient experience of ankle fracture recovery and the relevance and comprehensiveness of the outcome measure, I considered the epistemology of interpretivism. This would suit the nature of this research question which attempts to generate an understanding of patient experience and the construct of interest in relation to the measurement of outcome in this population [100]. For this reason, qualitative research methods would be most appropriate to answer this particular research question.

Interpretivism, however, would not be suitable for the subsequent project which aims to determine the measurement properties of an outcome measure. This would be more suited to a positivist approach. This is because the question is concerned with the validity and reliability of a measure on a population level and would be most appropriately answered through the use of quantitative research methods [101].

The epistemology of critical realism acknowledges the existence of a reality or truth which is considered to be separate to social narratives and experiences [100 102]. Critical realism was the most appropriate approach for this multi-methods project. This is because, whilst the initial work package is concerned with the patient experiences factors they consider important throughout their recovery, that their injury can be described and understood through methods which are outside of the individuals own narrative, for example in the scores of an outcome measure [100].

Pragmatism is described as a research paradigm which attempts to combine the understandings provided by both qualitative and quantitative methods, with an appreciation of the contextual factors in which the research is taking place [99 103]. Pragmatists will often utilise multiple research methods in order to gain understanding and answers to a particular problem [104]. A pragmatic approach allows for an appreciation that a "real" world exists alongside each individuals' personal interpretation and experience of that world and that both of these can be useful in generating knowledge and understanding [105]. Therefore throughout this thesis, I have

used pragmatism on the foundation of the epistemological position of critical realism.

This philosophical underpinning allows for the exploration of subjective and experiential understanding of the construct of interest of ankle fracture recovery as well as the objective assessment of measurement properties of a widely used outcome measure. The results of these work packages will inform conclusions on the most appropriate outcome measurement methods within the patient population of ankle fracture, creating a holistic and complete picture for these conclusions, more so than if only one of these approaches was used in isolation.

I do not describe this work as being mixed methods, because there is no integration of the data in the process of the research, which is a defining feature of this methodology [106] [107]. Rather I have chosen to employ multiple types of research methodologies to evaluate the ways in which outcome is measured in this context and patient population in the aim of these approaches complementing each other to answer the research questions.

2. A systematic review of outcome measures used in randomised controlled trials of interventions for adults with an ankle fracture

Aim: The aim of this systematic review is to identify what outcome measures are used in RCTs of non-pharmacological interventions for adults with ankle fractures. The temporal trends in the use of outcome measures in this type of study will also be explored.

Methods: This was a systematic review which followed Cochrane methodology. Five databases were searched using a comprehensive search strategy. Two trial registries were searched for registered but unpublished studies. Included articles were randomised or quasi randomised controlled trials assessing two or more non-pharmacological interventions in adults with malleolar ankle fracture. Information on the outcome measures collected as primary and secondary measures were recorded and a coding framework was devised and used to display the results concisely.

Results: Many different measures were found to be collected in this type of study and a variety of different measurement methods. PROMs were the most widely used outcome measure type in this study, followed by clinician assessments. Of these, OMAS was the most commonly used PROM and radiographic assessment was the most frequently used type of clinician assessment. Clinician assessments have been decreasing in popularity since records began and PROMs have been increasing in popularity over the same period.

Conclusion: There are a wide range of outcomes collected in this study type. There was also wide variation in the measurement methods of clinician assessments, leading to variation in practice across studies. With the growing importance and use of PROMs in this area, it's important that the quality of these are investigated further.

This work has been published in a peer reviewed journal

McKeown, R., Rabiu, A.R., Ellard, D.R. and Kearney, R.S. (2019) Primary outcome measures used in interventional trials for ankle fractures: a systematic review. *BMC Musculoskeletal Disorders*. 20: 388

This work has been presented at a national conference

Title: Primary outcome measures used in interventional trials for ankle fracture: a systematic review. **Presenter**: Rebecca McKeown **Event**: British Orthopaedic Association (BOA) Annual Congress, 12/09/2019, Liverpool.

2.1 Introduction

As discussed in the previous chapter, RCTs are vital to ensure that healthcare provision is evidence based and cost effective, particularly in resource limited healthcare systems such as the NHS. In order for trials to provide meaningful results, the outcome measure used is of paramount importance [108]. The use of an outcome measure which is not valid, reliable or responsive in RCTs is unethical and a waste of resource [109]. The success of RCTs is dependent upon the goodwill of participants to consent to randomisation to an allocated intervention, some of which could carry considerable risk profiles. If the outcome measures are not accurately and effectively measuring the construct of interest, then this randomisation and risk exposure would be unnecessary. Furthermore, if outcome measures are not sufficient enough to provide meaningful results, this constitutes a waste of resources spent in conducting the trial. RCTs are burdensome and costly projects to deliver, estimated to cost between £0.5 and £2.4million in the UK [110]. The UK's largest funder of healthcare research is the National Institute for Health Research (NIHR) and is primarily funded by the UK Department of Health and Social Care [111] and therefore it's of paramount importance that these publicly financed projects deliver cost-effective outcomes.

As described in the previous chapter, the CONSORT statement is a set of reporting standards to ensure that RCTs are designed and reported with consistency and transparency [71]. Outcome measurement is identified as a significant consideration when reporting results of a clinical trial [71]. Item 6a of the reporting checklist is concerned with the specification and definition of the primary and secondary outcome measures [71]. As discussed by Altman [69], there should be one outcome measure which is regarded as the main focus of the RCT, known as the primary outcome measure. The use of many different outcome measures in RCTs of a specific patient population is considered as research waste, because it is difficult to pool results using different measures in systematic reviews and meta-analyses, which is the highest level of evidence to inform clinical practice [112 113].

I was the trial manager for AIR and a member of the TMG. There remains uncertainty about outcome measure selection, as I observed during the AIR TMG and oversight committee meetings. Considerations for choosing an outcome measure were those which were used frequently (enabling pooling of results in future meta-analyses to be completed more easily) and those which were used less frequently, but did have some evidence for validity in other patient populations. Based on these

considerations, OMAS was chosen because it was used by other studies in this patient population, allowing for a homogeneity of outcomes between studies. This process highlighted the lack of clarity and uncertainty surrounding the best outcome measures to use in this context and this patient population and why.

In order to commence an evaluation of outcome measures used in clinical trials for adults with ankle fracture, it's first necessary to review the literature and identify outcome measurement methods used in similar studies. Results of this review will inform future work packages of this thesis, to ensure that work focusses on the most frequently used outcome measures in this research area. This type of review has been undertaken in other patient populations including patients undergoing hip and knee replacement [114 115], ankle arthroplasty [116], patients with shoulder instability [117], low back pain [118] and foot and ankle disorders [119]. These articles have demonstrated the current and previous research practice in outcome measurement in these clinical populations, showing the trends in this particular aspect of research practice. This is useful when considering the wider picture of evidence-based medicine, when results from RCTs can be pooled to show evidence in systematic reviews. If too many different outcomes are being collected in clinical trials, this hampers the ability to perform these systematic reviews effectively and is considered to be research waste [113].

2.1.1 Aim

The aim of this systematic review is to identify the outcome measures utilised and explore temporal trends in the use of outcome measures in RCTs of non-pharmacological interventions for adults with ankle fractures.

Non-pharmacological trials were chosen because of the nature of the AIR trial as discussed in the previous chapter, which compares immobilisation methods for adults with ankle fracture and therefore trials which assess pharmacological interventions are out of the scope of this thesis.

2.1.2 Research questions

1. What outcomes are collected in clinical trials for non-pharmacological interventions for ankle fracture in adults?

2. How has the type of outcome measure collected in this type of study changed over time?

2.1.1 Objectives

1. To perform descriptive analysis of the outcome measures collected in RCTs for this patient population.

2. Explore the temporal trends in use of outcome measures in RCTs for this patient population.

3. To formulate conclusions and inform further work in this research area.

2.2 Methods

2.2.1 Protocol and registration

The protocol for this systematic review was not registered prospectively with the PROSPERO register as this type of review which report the use of outcome measures in a particular patient population are ineligible. This systematic review follows the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [120] and follows the Cochrane methodology for systematic reviews [121].

2.2.2 Eligibility criteria

For inclusion, articles must be randomised or quasi-randomised controlled trials of two or more non-pharmacological interventions for adults with malleolar ankle fracture. As discussed in the previous chapter, in the hierarchy of evidence-based medicine, RCTs are the methodological gold-standard for assessing interventions for a particular condition or injury. Because AIR is an RCT and this thesis is primarily concerned with how outcome is measured in RCTs, this was the study type eligible for inclusion in this review.

Adult was defined as skeletally mature individuals but articles could stipulate a minimum age of ≥16 years. Articles were excluded if they are not randomised or quasi-randomised controlled trials evaluating interventions for adults with ankle fractures. Quasi-randomisation methods include randomising by operating surgeon or day of admission. More robust randomisation methods, such as online computer generated randomisation or telephone randomisation is the gold standard method because they limit bias in the randomisation process [69]. Despite this, I made the decision to include articles that used quasi-randomisation due to the nature of the review. Considering that this review was concerned with outcome measure use, the

methodological quality of the articles was of lesser importance than in a review evaluating treatment effects of interventions.

Articles which studied individuals with pilon fractures were excluded as these are more complex fractures with a higher rate of associated complications, and therefore out of the scope of this thesis. Articles studying ankle fractures in paediatric populations were excluded because of the variances in fracture healing, repair and possible methods of outcome measurement in this population of patients [122]. We also excluded articles which used a sample of individuals with different injuries or samples which were not solely comprised of adults with ankle fractures.

Longer term follow up studies with no new outcome measures collected and multiple publications of the same dataset were excluded; the initial paper was included and the later publications excluded. In the instances where a prospective registry record had since been published, either the protocol or the results of the RCT, the published paper was the included article and the registry record was excluded to avoid duplicate reporting. Similarly, published protocol papers of articles already included in the review were excluded for the same reasons. Animal studies, cadaveric studies or studies using synthetic sawbones were excluded as this review is concerned with outcome measurement in human population based studies. Articles not written in the English language were excluded from this review as there was no resource for translation. Any records which did not provide sufficient levels of detail on eligibility criteria and outcome measures used were excluded. The eligibility criteria is found in table 2.

Table 2 - Eligibility criteria for inclusion in review

Inclusion criteria	Exclusion criteria
Randomised or quasi-randomised controlled trials	Articles studying one or more pharmacological interventions
Trials comparing two or more non-pharmacological interventions in the management of adults with malleolar ankle fractures	Conference proceedings or abstracts only which do not provide sufficient information about outcome measures collected
	Follow-up studies with no new outcomes measures used
	Studies assessing pilon fractures or a mixed group of different fractures or injuries
	Articles not written in the English language
	Animal, cadaveric or synthetic saw bone studies
	Protocol papers or registry records of articles already included in the review
	Studies assessing paediatric patients <16years

2.2.3 Sources, search and study selection process

A comprehensive search of online databases was completed on 12/03/2018 to search for published and registered trials. Databases searched were Medline, Embase, CINAHL, AMED, Cochrane CENTRAL Trials database, ISRCTN and ClinicalTrials.gov registries. No date limits were applied to the search. The search strategy used can be found in appendix 1.

Results from the database searches were downloaded and reviewed in EndNote X5 (Thompson Reuters). Duplicates were initially removed using the EndNote duplicate removal function. The eligibility criteria were applied firstly by reviewing the title and abstract within EndNote. If it was clear from the title and abstract that the article did not meet the inclusion criteria then it was excluded. If eligibility was not clear from the title and abstract, the full text was retrieved. Results from the ISRCTN and

ClinicalTrials.gov registries were reviewed through their respective webpages. Eligibility criteria were applied first by screening the study title of the registry records. If it was clear that the record was ineligible on this initial review, then records were excluded. If eligibility was unclear, the full record was reviewed for purposes of completeness.

A second reviewer, Abdul-Rasheed Rabiu (ARR), reviewed all articles for which the full text or record was retrieved and applied the eligibility criteria to reduce the risk of error. Results of the independent review processes were discussed between myself and ARR; any which we disagreed on after discussion were referred to a third reviewer, Rebecca Samantha Kearney (RSK), for guidance and final decision. Care was taken to ensure that each record was only included once. For example, if a registry record had been subsequently published, we ensured that the most recent record (i.e. the published article) was included and the registry record excluded. If a published protocol and published results article existed for the same study, the published results article was included in the review and the published protocol article was excluded.

2.2.4 Data items

I recorded the included study details in a data extraction sheet shown in appendix 2. This detailed information from the studies; the authors, title, year of publication, primary outcome measure and secondary outcome measures. Information on whether the authors specified their primary outcome measure was also recorded, along with sample size (or planned sample size if registry records or protocol) and the types of interventions being studied. The primary outcome was regarded as the outcome measure which the authors specified as being the primary outcome measure. If the primary measure was not specifically stated, the outcome measure for which the sample size calculation was based upon was regarded as the primary outcome measure. In cases where this also was not specified, the first outcome measure mentioned in the abstract of the article was regarded as the primary outcome measure. In instances where none of the aforementioned situations applied, the first outcome measure mentioned in the full text of the article was considered as the primary outcome measure. This method has been used in a review of a similar nature [118].

2.2.5 Analysis plan

Descriptive statistics were used to show the outcome measures collected and analysed in order to determine outcome in this type of trial. The outcome measures were categorised using a coding framework to report and analyse this information. A coding framework was used to enable concise reporting of the data which would result from this study. This was to allow for a large amount of data to be presented concisely and accessibly.

The framework was developed in consultation with three other clinicians through two rounds of consultation and feedback. The three clinicians comprised of a clinical academic physiotherapist, Emma Sutton (ES) and two orthopaedic surgeons, ARR and Jonathan Young (JY). I presented the outcome measures to the group individually and suggested categories for grouping these together. The group were invited to review and make suggestions for changes to the framework. Once changes were reviewed and implemented as appropriate, a second version was redistributed for review. After further comments, a third version was distributed and all clinicians approved this version which was used for the analysis of this review. The coding framework is outlined in Table 3.

Whilst I recognise that this is not a validated framework for reporting outcome measures, it was developed for the purposes of presenting data and showing the types of outcomes collected in these trials only. I defined the categories based upon a broad definition of who assessed the outcome measures. For example, the first category items are fully patient assessed, i.e. the patient gives the scores. The third category was fully clinician assessed. Category two was for combined scores, where both patient and clinician give an assessment. The fourth category is types of outcome which cannot be categorised into the other three, involving hospital data such as length of stay and also health economic data.

Table 3 – Coding framework for data analysis

Category 1-4	Descriptor a-h	Included outcome measures
1. Patient assessed	a. Generic PROM	Health-related quality of life PROMs and disability scores, e.g. LEFS, SF-12, SF-36, linear scoring scales (e.g. VAS), time to return to work, usual activities or sports
	b. Ankle specific PROM	Ankle specific PROMs e.g. OMAS, FAAM, MOXFQ, A-FORM
2. Combined clinician and patient assessed	c. Combined patient and clinician outcomes	Combined outcomes e.g. AOFAS, JSSF, any score which involves a clinician and patient assessment
3. Clinician assessed	d. Clinical investigations	Radiographic measures (including X-Ray, DEXA, CT, MRI), blood biochemistry analysis, intraoperative assessment, other or undisclosed clinical assessment
	e. Functional measures	Measures of function of ankle e.g. range of movement, strength, swelling, muscle wastage
4. Other	f. Health economic data	Cost utility e.g. resource use analysis, EQ-5D, cost analysis of interventions
	g. Hospital episode data	Length of stay, analgesia use, operation duration, destination of discharge
	h. Complication data	Adverse events, complications, DVT, further procedures or reoperation rates

In order to be classified as a PROM, the outcome must be fully patient reported with no clinician reported elements. The related references to outcome measures collected in the articles were retrieved and reviewed to determine the nature of the outcome measure assessed. Those which did not give sufficient evidence or reference to the

method of assessment were classified as “other or undisclosed clinician assessment” (Category 3d).

Analysis will determine the proportion of studies which use patient assessed outcome measures compared with the proportion of studies who have used clinician assessed outcomes to assess recovery in this patient population. An analysis of the high level categories will be completed to demonstrate the proportion of studies which use each of these categories of outcome assessment. The outcome measures used as the primary measures will be identified and displayed. Furthermore, a breakdown of the individual categories (a-h) will be completed to ascertain the most commonly used clinical measures, functional tests and other specific groups of outcome measurement within each higher level category. This will be explored overall and at a primary outcome measure level. Details of the outcome measurement categories collected as for secondary outcome measures will be displayed. Secondary analysis will include exploring the temporal trends of outcome measure use of the outcome measure categories in this type of trial.

2.3 Results

2.3.1 Study selection and descriptors

In total, 2,817 articles were returned from the search of the databases, 57 records from the ISCRTN registry and 84 from the ClinicalTrials.gov registry. Initially, 659 duplicates were removed and titles and abstracts for the remaining 2,299 records were reviewed. Based upon the review of the title and abstracts alone 2,100 articles were excluded as ineligible. 199 full texts and records were retrieved and reviewed to assess eligibility for inclusion.

Of the 199 articles, six articles were excluded as they were long term follow up reports of articles already included in the review and 17 records were excluded as they were published protocol papers or registry records of trials already included in the review (and their subsequent results paper had already been published and included). A further 16 articles were not randomised or quasi-randomised controlled trials and therefore excluded. A further seven duplicates were identified and excluded and nine articles were excluded as they were abstracts only which did not give sufficient information on outcomes collected. Ten articles used a mixed population of lower limb trauma patients so were excluded on this basis. Seven studies included a population of paediatric participants, therefore were excluded. Nineteen articles studied one or more pharmacological interventions as part of their

trial so were excluded for this reason. A total of 91 papers were therefore excluded, leaving 108 records included in the review, comprised of 7 ISRCTN records, 19 ClinicalTrials.gov records and 82 published articles. Details of this review can be found in figure 5.

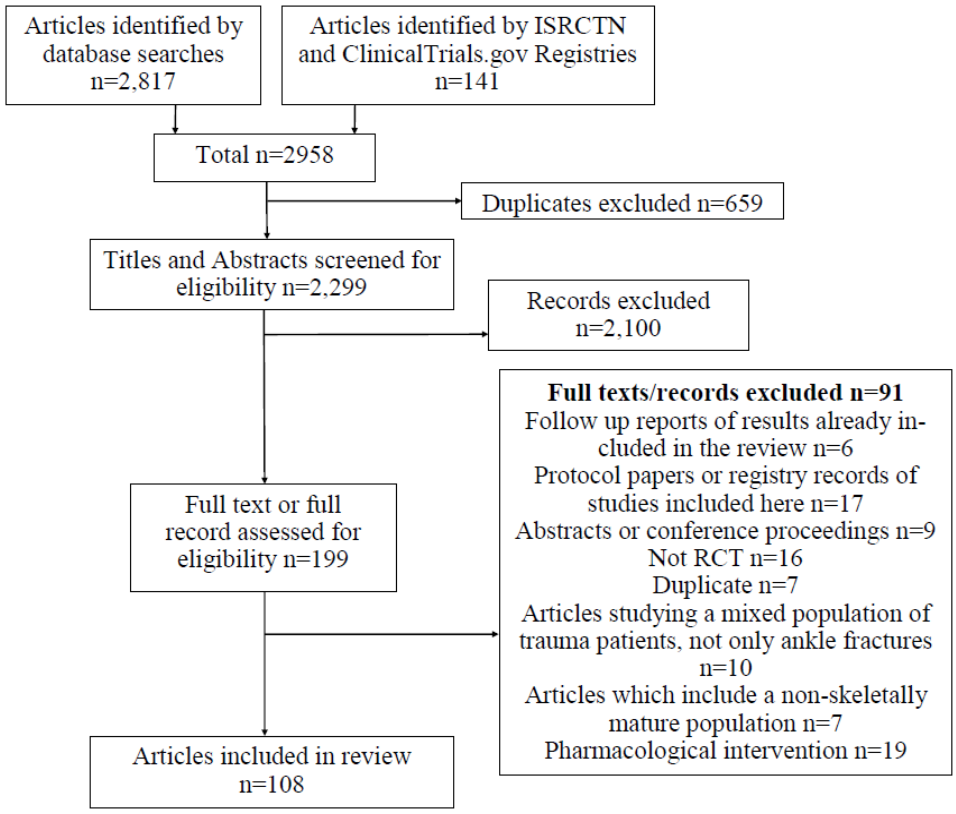


Figure 5 - PRISMA diagram

The 108 papers/records included here included a total patient population of 10,091 participants (including planned populations of protocols and registry records). Of the 108 records, 57 (53%) of them specified their primary outcome measure or specified the outcome measure on which the sample size was calculated. Fifty-one (47%) records did not specify their primary outcome measure, either explicitly or through justification or calculation of a sample size.

Table 4 - Number of studies published by decade

Decade	1980's	1990's	2000's	2010's	Registered
Number of articles published	8	16	22	36	26

Table 4 shows the number of articles published each decade. The number of RCTs in this patient population has been increasing with time, with 8 published in the 1980's, 16 in the 1990's, 22 in the 2000's and 36 in the 2010's. There are currently 26 registered trials of this type in the registries searched.

Table 5 shows the different types of interventions studied in included articles, which were grouped into six types of trial. The most commonly studied interventions were comparisons of operative techniques or implants for ankle fractures, accounting for 48 (44.4%) of the articles collected here. Thirty five articles (32%) compared post-operative rehabilitation or management for patients who have undergone surgical fixation. Nine articles (8.3%) compared operative and non-operative intervention and a further eight (7.4%) compared conservative rehabilitation strategies for patients not undergoing operative fixation. Four (3.7%) combined both operatively and non-operatively managed individuals and assessed rehabilitation strategies in these individuals. The remaining four articles (3.7%) compared different interventions for individuals awaiting operative fixation of their fracture.

Table 5 - Types of interventions assessed in included studies

Type of intervention assessed	Number of studies
Comparing operative techniques or implants	48
Comparing post-operative management or rehabilitation of operatively managed patients	35
Comparing operative to non-operative management	9
Comparing conservative management protocols for non-operatively managed patients	8
Comparing rehabilitation or treatment protocols for a mixed population of operatively and non-operatively managed patients	4
Comparing pre-operative management protocols for patients awaiting surgical fracture fixation	4
Total	108

2.3.2 All outcome measures

Table 6 shows all outcome measures used across the 108 articles returned in the search. The most commonly used category was patient assessed outcome measures (category 1), which were collected 182 times. Clinician assessed items (category 3)

were collected 175 times, followed by category 4, comprising of complications, length of stay, hospital episode data and health economic analysis, which were collected 90 times in total. Combined clinician and patient assessed outcome scores/scales were the least frequently collected at 34 times.

Table 6 - All outcome measures used in studies by category

Outcome category	Frequency of collection in articles overall	Total outcomes in each high level category 1-4
1a. Generic PROM	111	1. Patient assessed=182
1b. Ankle specific PROM	71	
2c. Combined patient and clinician score	34	2. Combined patient and clinician assessed = 34
3d. Clinical investigations	90	3. Clinician assessed = 175
3e. Functional measures	85	
4f. Health economic data	16	4. Other = 90
4g. Hospital episode data	23	
4h. Complication data	51	

As table 6 shows, when the outcome measures are displayed using detailed categories (a-h), the most commonly collected item is generic PROM data which was collected 111 times, with clinical investigations following after being collected 90 times overall. Functional measures were collected 85 times throughout the articles and ankle specific PROM data were collected 71 times. Complication data were collected 51 times throughout the articles and combined patient and clinician assessed scores 34 times. Hospital episode data were collected 23 times. Health economic data were the least widely collected type of outcome, collected 16 times.

Table 7 - Generic PROMs used in included studies

Category 1a items – generic PROMs	Frequency of collection in articles overall
Visual Analogue Scale (VAS) and linear scoring scales for pain, function, satisfaction with outcome	54
Return to work/usual activities/sports and leisure	26
36-Item Short Form survey (SF-36)	14
12-Item Short Form survey (SF-12)	6
Lower Extremity Functional Scale (LEFS)	3
Short Form Musculoskeletal Assessment (SFMA)	3
Tegner Activity Scale	2
Patient Reported Outcome Measurement Information System (PROMIS)	2
Disability Rating Index (DRI)	1
Total	111

Table 7 shows the generic PROMs collected; the most popular of this category is the Visual Analogue Scale (VAS) which was collected 54 times across the studies. Return to work or usual activities was collected 26 times across all outcomes in the included studies. The SF-36 and the 12-item Short Form survey (SF-12) were collected 14 and six times respectively. The Lower Extremity Functional Scale (LEFS) and Short Form Musculoskeletal Assessment (SFMA) were collected three times. The Tegner activity scale and the Patient Reported Outcome Measurement System (PROMIS) were collected twice and the DRI was collected once.

Figure 6 - Pie chart of outcomes collected in category 1a; ankle specific PROMs collected in included studies

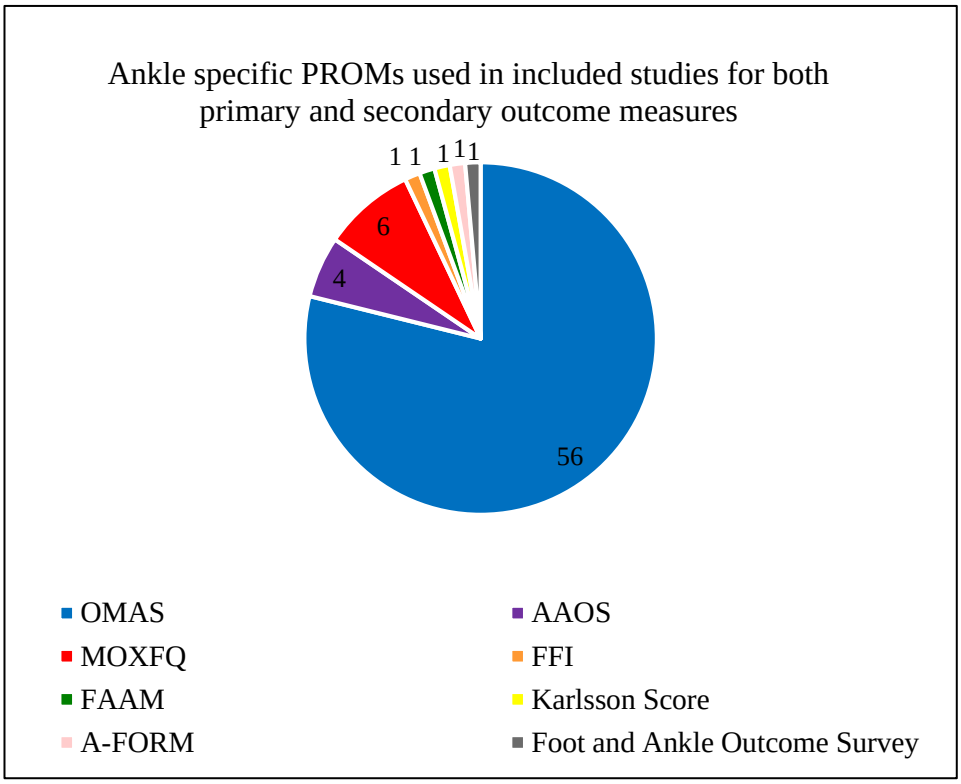


Figure 6 shows the ankle specific patient reported outcomes (Category 1b) collected here, of which the most frequently used was the OMAS [123], collected 56 times in total. The MOXFQ [124] was used six times and the American Association of Orthopaedic Surgery Foot and Ankle Outcome Questionnaire (AAOS) [125] was used four times. The Ankle Fracture Outcome of Rehabilitation Measure (A-FORM) questionnaire [126] the Karlsson Score (KS) [127], the Foot and Ankle Ability Measure (FAAM) [128], the Foot and Ankle Outcome Survey (FAOS) [129] and the Foot Function Index (FFI) [129] were all used once.

Table 8 - Category 2c items; combined patient and clinician assessed outcome measures collected

Category 2c items – combined patient and clinician scores	Frequency of collection in articles overall
American Association of Foot and Ankle Surgery ankle hindfoot score (AOFAS)	21
Rating scale described by Mazur *et al.* (1979)	2
Rating scale described by Kaikkonen *et al.* (1994)	2
Score described by Phillip *et al.* (1985)	2
Unreferenced scoring system combining patient and clinician assessments	3
Japanese Society of Surgery of the Foot Score (JSSF)	1
Weber scale described by Hughes *et al.* (1978)	1
Score described by Linde *et al*. (1984)	1
Score described by Baird and Jackson (1987)	1
Total	34

As table 8 demonstrates, the most commonly used outcome measure in category 2c is the American Orthopaedic Foot and Ankle Society Score (AOFAS) [130], which was utilised in 21 different articles. A scoring scale described by Mazur, et al. [131]and a score described by Kaikkonen, et al. [132] were used twice. The Japanese Society for Surgery of the Foot Score (JSSF) was utilised once [133], as was the Weber Scale [134]. There were three instances of an unreferenced scoring system combining patient and clinician assessments of outcome being used. The remainder of scores collected within this category were a rage of different scales devised by or described by authors [135] [136] [137].

Category 3 comprised of clinician assessed outcomes of which a total of 175 were collected. This comprised of 90 clinical investigations (3d) and 85 functional measures (3e). A breakdown of these two categories can be found is tables 9 and 10. The majority of the clinical investigations collected in these articles were radiographic analysis, comprising of X-Ray analysis, CT scan analysis and DEXA scans. These measures were collected 67 times. Blood biochemistry analysis was collected five times and there were two occasions where an intra-operative assessment was collected as an outcome measure. On 16 occasions an undisclosed

or other clinician performed assessment was completed and usually these were described as an “orthopaedic assessment” or “clinical assessment”.

Table 9 - Category 3d; clinical investigations collected

Category 3d items – clinical investigations	Frequency of collection in articles overall
Radiographic assessments	67
Clinician assessments (other or undisclosed)	16
Blood biochemistry analysis	5
Intraoperative assessment	2
Total	90

Table 10 - Category 3e; functional measures collected

Category 3e items – functional measures	Frequency of collection in articles overall
Range of movement (ROM)	55
Assessments of swelling	17
Gait analysis	8
Strength tests	5
Total	85

As table 10 demonstrates, ROM assessments were the most commonly collected functional measure in this review, being collected 55 times. Assessments of swelling and limb volume were the next most commonly collected outcome in this category, being collected 17 times. Gait analysis was collected in eight studies and strength assessments were collected in five.

Table 11 - Category 4f; health economic analysis collected

Category 4f items – health economic data	Frequency of collection in articles overall
EQ-5D	9
Resource use/cost of interventions	7
Total	16

Table 11 shows the outcomes collected in category 4f, with EQ-5D being collected in nine cases. The remaining seven cases compared the overall costs of the interventions in a resource use health economic analysis.

Table 12 - Category 4g; hospital episode data in included studies

Category 3d items – clinical investigations	Frequency of collection in articles overall
Length of stay	9
Tourniquet time/duration of operation	3
Analgesia usage	4
Number of physiotherapy sessions	2
Time to surgery	2
Time spent non-weight bearing	2
Destination of discharge	1
Total	23

A breakdown of category 4g is shown in table 12; the most commonly collected item here was length of stay, collected nine times. Analgesia usage was collected four times and operation time or tourniquet time was collected three times. The remaining items in this category were number of physiotherapy sessions and time from injury to surgery, which were both collected twice. Time spent non-weight bearing was collected twice and destination of discharge was collected once.

Table 13 - Items collected as part of category 4h; complication data in included studies

Category 4h items – complication data	Frequency of collection in articles overall
Complications (all or unspecified)	32
Wound specific complications	10
Further procedures/reoperations	7
DVT rates	2
Total	51

Category 4h captured complication data and adverse events, of which 51 were collected in total, shown in table 13. An overall assessment of complications were collected on 32 occasions and wound specific complications were captured 10 times. Rates of DVT specifically were collected in two articles and further procedures or operation related to the ankle fracture were collected on seven occasions.

2.3.3 Primary outcome measures

Table 14 shows the primary outcome measures used in each outcome category. The primary outcome measure used was patient assessed in 45 studies and combined patient and clinician assessed in 13 studies. A clinician assessed outcome was used as the primary outcome measure in 42 studies and outcomes in category 4 were the primary outcome measure used in nine studies.

Table 14 – Outcomes used as the primary outcome measure by category in included studies

Outcome category	Frequency of collection as primary outcome measure	Total outcomes in each high level category 1-4 as primary outcome
1a. Generic PROM	10	1. Patient assessed = 45
1b. Ankle specific PROM	35	
2c. Combined patient and clinician score	13	2. Combined patient and clinician assessed = 13
3d. Clinical investigations	24	3. Clinician assessed = 41
3e. Functional measures	17	
4f. Health economic data	0	4. Other = 9
4g. Hospital episode data	4	
4h. Complication data	5	

Category a made up 10 of the 45 items in category 1 used as the primary outcome measure. A variety of generic patient reported outcome measures were collected here, with four articles collecting a VAS for pain, function or satisfaction and three articles collecting LEFS. Two articles collected SF-36 as the primary outcome and one collected return to work/activities as the primary outcome. This is shown in table 15.

Table 15 - Items collected in category 1a; generic PROMs as primary outcome

Category 1a items – generic PROMs	Frequency of collection in articles as primary outcome
VAS	4
LEFS	3
SF-36	2
Return to work/usual activities/sports	1
Total	10

Table 16 shows results for category 1b, there were three ankle specific PROMs used as the primary outcome in the included studies; the OMAS, the AAOS and the MOXFQ. The OMAS was the most commonly used, collected in 31 cases of 35. The AAOS and MOXFQ were collected as the primary outcome measure in two studies each.

Table 16 - Items collected in category 1b; ankle specific PROMs as primary outcome in included studies

Category 1b items – ankle specific PROMs	Frequency of collection in articles as primary outcome measure
OMAS	31
AAOS	2
MOXFQ	2
Total	35

There were 13 articles which collected combined patient and clinician rated outcomes in category 2c as their primary outcome measure. The most commonly used of these was the AOFAS Ankle-Hindfoot Score which was collected eight times as the primary outcome measure. The rest were scores and scales devised and reported by authors in previous studies, described by [131 132 134-136] and these were each used once as the primary outcome measure. These are shown in table 17.

Table 17 - Items collected in category 2c; combined patient and clinician scores as primary outcome

Category 2c items – combined patient and clinician scores	Frequency of collection in articles as primary outcome measure
AOFAS	8
Rating scale described by Mazur *et al.* (1979)	1
Rating scale described by Kaikkonen *et al.* (1994)	1
Score described by Phillip *et al.* (1985)	1
Weber scale described by Hughes *et al.* (1978)	1
Score described by Baird and Jackson (1987)	1
Total	13

The most commonly used clinician assessed outcome measure was radiographic assessment, making up 19 of the 24 of the primary outcome measures recorded in category 3d. Other or undisclosed clinical assessments were collected as the primary outcome measure in two studies and blood biochemistry assessment in another two studies. An intraoperative assessment was the primary outcome measure in one article. These results for category 3d are shown in figure 7.

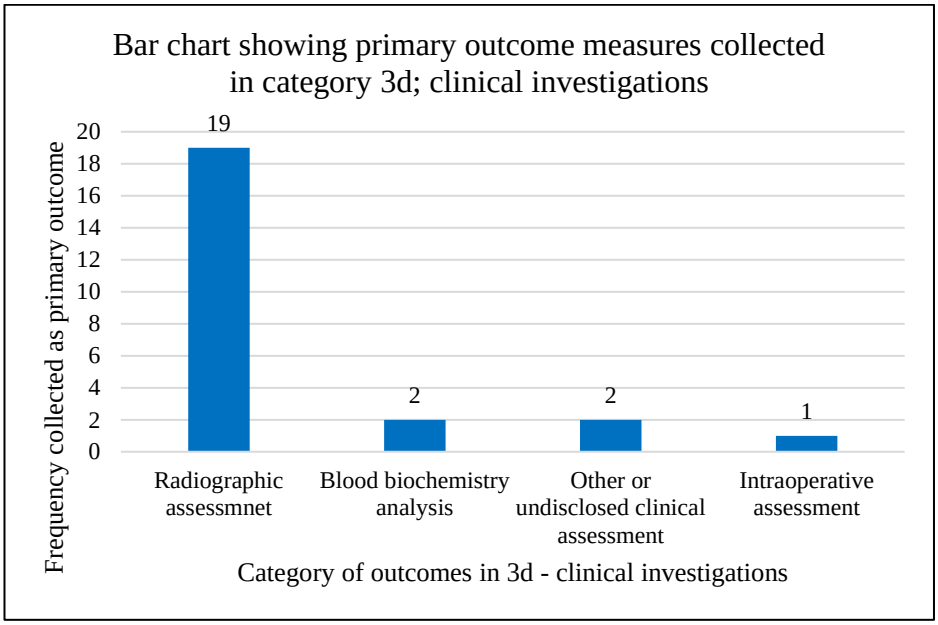


Figure 7 - Bar chart showing primary outcome measures collected in category 3d; clinical investigations

Figure 8 shows the outcome collected in category 3e. There were four different measures used; ROM was collected eight times, swelling six times, strength

assessments twice and gait assessment was collected once as the primary outcome measure.

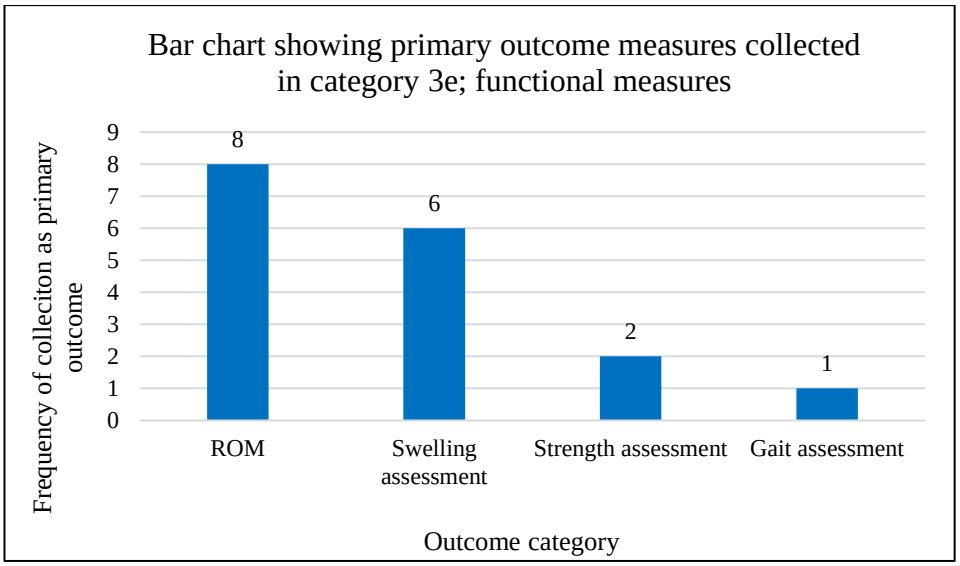


Figure 8 - Bar chart showing primary outcome measures collected in category 3e; functional measures

Table 18 - Items collected as primary outcome measure in category 4g; hospital episode data

Category 4g items – hospital episode data	**Frequency of collection in articles as primary outcome measure**
Length of stay	2
Operation/tourniquet time	1
Time from injury to surgery	1
Total	4

Table 19 - Items collected as primary outcome measure in category 4h; complication data

Category 4h items – complication data	**Frequency of collection in articles as primary outcome measure**
Wound specific complications	2
Complications	1

DVT rate	1
Further procedures/reoperations	1
Total	5

Category 4 was the least commonly used category as a primary outcome measure, being collected a total of nine times out of the 108 articles. There were no articles which collected health economic analysis outcomes (4f) as their primary outcome. Table 18 shows the breakdown of the primary outcome measures collected as part of category 4g. In category 4g, length of stay was collected twice, operation time once and time from injury to surgery was collected once. In category 4h, generic complications were collected once, specific wound complications twice, DVT rate once and further procedures and operations once. This is shown in table 19.

2.3.4 Secondary outcome measures

Table 20 - Outcome measures used a secondary outcome measures by category

Outcome category	Frequency of collection as primary outcome measure	Total outcomes in each high level category 1-4 as primary outcome
1a. Generic PROM	101	1. Patient assessed = 137
1b. Ankle specific PROM	36	
2c. Combined patient and clinician score	21	2. Combined patient and clinician assessed = 21
3d. Clinical investigations	66	3. Clinician assessed = 134
3e. Functional measures	68	
4f. Health economic data	16	4. Other = 81
4g. Hospital episode data	19	
4h. Complication data	46	

Table 20 shows the outcome measures collected as secondary measures by category 1-4 and a-h. The most commonly collected outcome as a secondary measure were

generic patient reported outcome measures (1a), which were collected 101 times. This was followed by category 3e, functional measures collected 68 times and clinical investigations (3d) collected 66 times. Category 4h, complication data, was collected 46 times and 1b, ankle specific PROMs were collected 36 times as the secondary outcome measure. Category 2c, combined patient and clinician scores were collected 21 times as the secondary outcome measure, as were category 4f (health economic data). Figure 9 shows a bar chart of outcome measures used as the primary and the secondary measure.

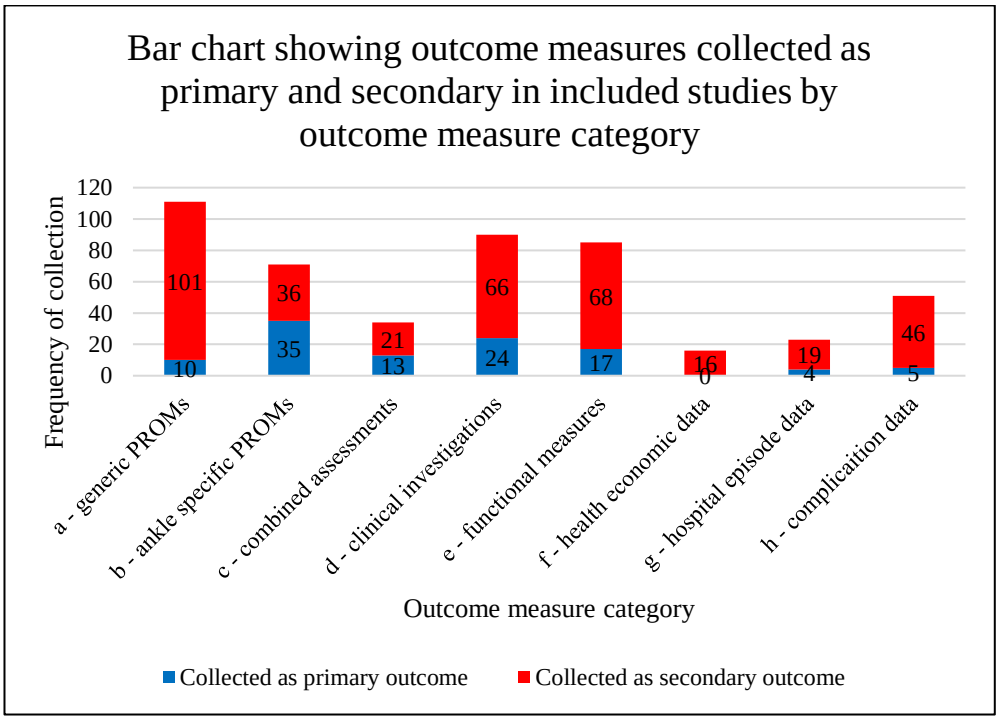


Figure 9 - Bar Chart showing overall usage of outcome measure categories split by primary and secondary measures

2.3.5 Temporal trends in outcome measure use

Table 21 - Categories of primary outcome measures collected in published studies by decade

Decade of publication	Number of studies published overall	Category 1 – patient assessed	Category 2 – combined	Category 3 – clinician assessed	Category 4 – other
1980's	8	0	6	2	0

1990’s	16	5	10	0	1
2000’s	22	8	12	2	0
2010’s	36	17	11	4	4

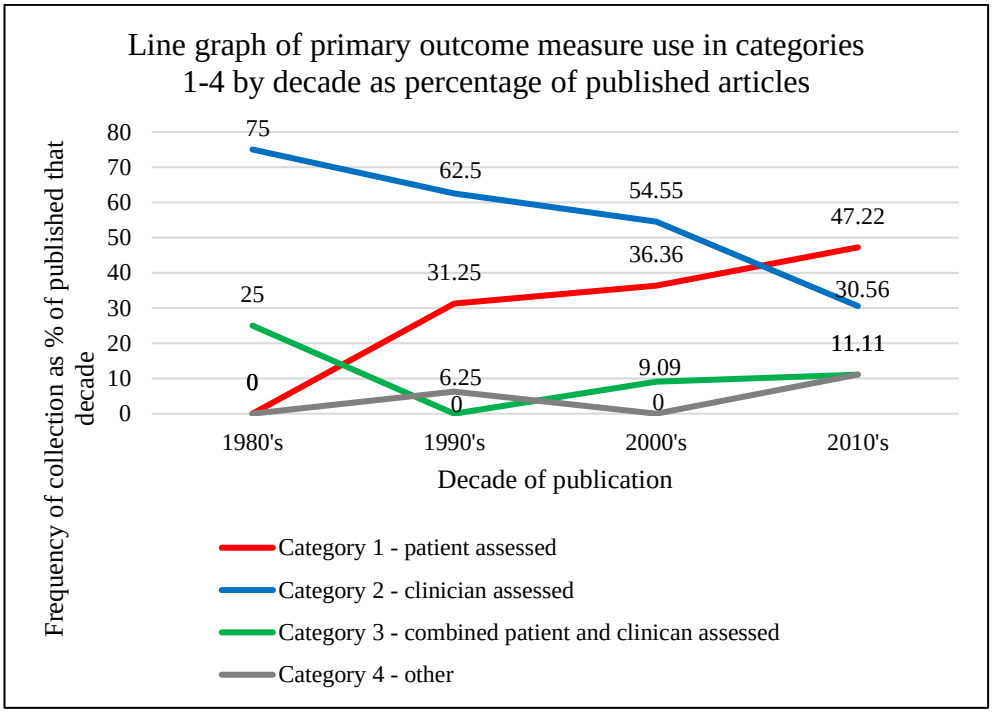


Figure 10 - Line graph showing primary outcome measure use in categories 1-4 by decade

Table 21 and figure 10 show the temporal trends in type of primary outcome measure use over time, showing the high level categories and frequencies of use as a percentage of trials published in each decade. There was a total of 82 published articles included here. Clinician assessed outcomes were the most commonly used assessments in the 1980’s, used 75% of the studies published that decade, with combined patient and clinician scores made up the 25%. The proportion of studies using clinician assessed measures has decreased over the years to 62.5% in the 1990’s, 54.55% in the 2000’s and 30.56% in the 2010’s. Patient assessed measures weren’t used at all in the 1980’s, but increased to being used in 31.25% of the studies published in the 1990’s. The use of this type of outcome increased again to 36.36% of studies in the 2000’s and again in the 2010’s to 47.22% to be the most commonly collected type of outcome collected as the primary outcome in the 2010’s so far.

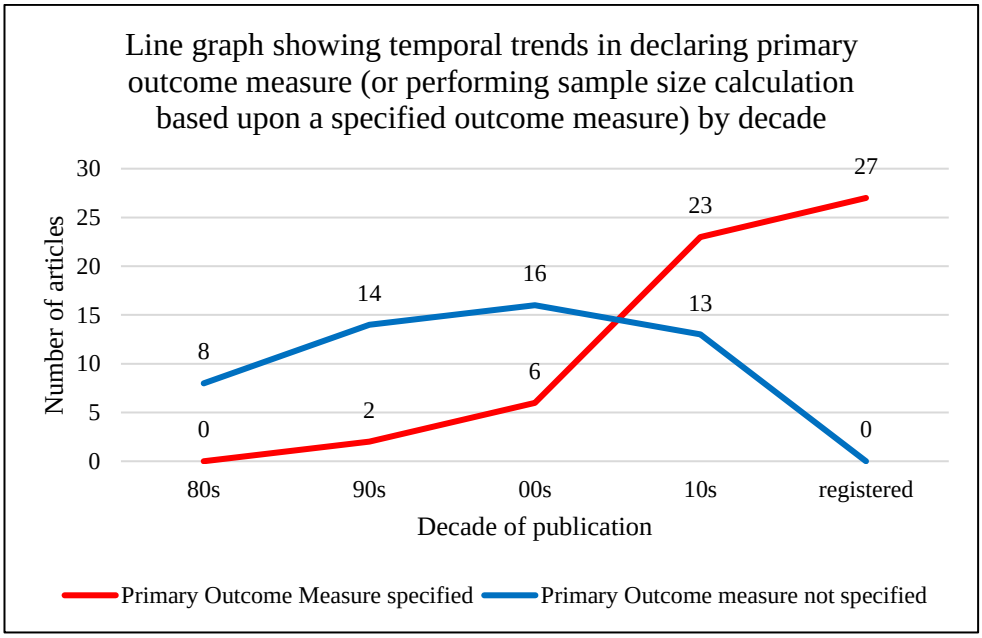


Figure 11 - Line graph showing declaration of primary outcome measure by decade of publication

Figure 11 shows the temporal trends in declaration of the primary outcome measure in this trial type. Of the eight articles published in the 1980's, none declared their primary outcome measure specifically. In the 1990's, two of 16 published articles declared their primary outcome and this was six out of sixteen in the 2000's. In the 2010's there were more articles who declared their primary measure than did not, at 23 to 13. All of the registered articles specified a primary outcome measure.

2.4 Discussion

This systematic review demonstrates the wide range of outcome measures collected in RCTs for this patient population which vary from PROMs to clinician devised scores and clinical investigations such as radiographic assessment. The number of this type of study has increased over time. Operative interventions and post-operative management protocols were the most commonly studied areas in this review, accounting for 44% and 32% of all interventions studied respectively. Comparatively, there is a lack of RCT research into conservative rehabilitation for ankle fractures, accounting for only 7.4% of the trials here. This highlights a lack of high quality RCT evidence for conservative management protocol, which was also noted by the authors of a Cochrane review in to the rehabilitation of adults with ankle fracture, who recommended further research in this areas [48].

The number of studies declaring their primary outcome measure increased from 2000-2010. The CONSORT statement, first published in 1996 [138] and updated in 2001 [139], likely explains why researchers began following this guidance in their reporting more frequently in this time period.

The most frequently measured outcome type both overall and for the primary outcome measure was category 1, patient assessed outcomes, followed by clinician assessed outcomes in category 3. Of primary outcome measures collected, ankle specific outcomes were the most popular category, made up of multi-item questionnaires, such as OMAS, A-FORM and MOXFQ. The most commonly used of these, comprising of 56 of 71 of the total in this category, was OMAS. Despite this being the most commonly used PROM in this type of study, some researchers have raised concerns with the quality of this outcome measure [140 141].

Radiographic assessment was the most frequently collected clinician assessed measure, comprising of 67% of items in the clinical investigations category. There was a large range of variety in methods described in the analysis of this data, with some authors explaining their measurement methods in detail and others not at all. Some authors stated they assessed X-Ray for fracture union or failure of fixation. For example, Bucholz, et al. [142] assessed the union of the fracture defined as "complete obliteration of the fracture line" and Lehtonen, et al. [46] used a blinded independent assessor to measure the joint alignment and fracture healing. Other authors detailed the specific joint angles and distances which were measured to assess joint alignment, for example Park, et al. [143] measured tibiofibular clear space, tibiofibular overlap and medial clear space.

Whilst X-Ray is commonly utilised in clinical practice, the application of this assessment measure in research trials is less clear and measurement methods are inconsistent between trials. Complications, such as non-union, implant loosening and signs of post-traumatic osteoarthritis can be found on X-ray [144], but the association of a well aligned ankle joint on X-Ray and satisfactory outcome for an individual has not yet been demonstrated. This has been discussed by Costa [145] in the context of distal radius fractures, discussing the lack of evidence of correlation of X-Ray evidence with patient reported outcome. With no standardised methodology for assessing radiographic imaging in ankle fractures, the value of collecting this data within RCTs for adults with ankle fractures is unclear.

Of the articles included in this review, there was a lack of health economic analysis in this area of research, with this data only being collected in 16 of the 108 studies. Brauer, et al. [146] identified health economic analysis as being less common in the field of orthopaedic trauma, when compared joint replacement surgery and suggested this type of analysis was often underrepresented in trauma and orthopaedic research compared with other clinical areas. When considering the ageing population and the subsequent increased healthcare resource use that this change in demographics will cause, it is imperative that future research includes high quality health economic analyses to ensure that interventions are cost effective.

The results of this review show similarities with studies of this nature completed in different patient populations. For example, Riddle, et al. [114] demonstrated that amongst clinical trials in hip and knee replacement populations, the outcome measures collected varied extensively and a lack of consistency in outcome measures used across trials caused difficulties in aggregating results in systematic reviews. Hunt and Hurwit [119] completed a systematic review of PROMs collected in clinical research studies of foot and ankle disorders. The authors concluded that several different PROMs were collected. The authors completed a subgroup analysis of ankle fracture studies and found the AOFAS was the most commonly collected outcome in this patient population, followed by the SF-36 and then OMAS. Here I have not classified AOFAS as a PROM because parts of the measure are clinician reported [82]. Further differences in these results might be explained by the fact that the authors included all experimental designs, not just RCTs as I have done here.

Another study reviewed the outcome measures used in foot an ankle research [147] and also found an increasing trend in the use of PROMs in foot an ankle research. Authors of this study noted the frequent use of AOFAS, despite calls from several researchers for this outcome's use to be discontinued due to the insufficient measurement properties [148 149]. Both of these reviews called for greater consistency in the use of outcome measures between studies, similar to conclusions drawn here.

This project has highlighted the wide range of outcome measures used in RCTs of interventions for ankle fracture, not only clinical investigations and functional tests but also PROMs. There is a growing emphasis on the importance of patient reported outcomes in health policy [82 81 86] which is mirrored in these results that demonstrate an increase in the use of PROMs over time. Although these results demonstrate that the use of PROMs is increasing, we are not able to draw any conclusions on the

quality of these measures from this work. Therefore, the next chapter of this thesis will identify and appraise the evidence for the measurement properties of the ankle specific PROMs we found to be used in this type of trial.

In this chapter, I have explored the use of different types of outcome measure in clinical trials of non-pharmacological interventions for adults with an ankle fracture. The next work package outlines a systematic review which identifies and critically appraises the evidence for the measurement properties of the ankle specific PROMs which were identified here. This will aim to recommend a PROM for use in the context of RCTs for this patient population, as well as identify gaps in the evidence for further evaluation.

3. A systematic review of the measurement properties of ankle specific patient reported outcome measures in adults with an ankle fracture

Aim: The aim of this systematic review was to identify and critically appraise the evidence for the measurement properties of ankle specific patient reported outcome measures used in RCTs for adults with ankle fracture.

Methods: A comprehensive search strategy was used to search three databases. Articles which detailed the development or assessed the measurement properties of one or more PROMs on the pre-specified list in a population of adults with ankle fracture were included. Guidance developed by the COSMIN group was followed, including the risk of bias assessment, assessing evidence against the criteria of good measurement properties and modified GRADE assessment. Recommendations were made regarding the most appropriate outcome measure in this patient population.

Results: Seven articles were included, detailing four different PROMs. The A-FORM had the highest quality development process, but no evidence for measurement properties of the final version of the score. OMAS demonstrated moderate quality evidence for sufficient internal consistency, reliability and construct validity in the Turkish and Norwegian versions. AAOS demonstrated low quality evidence for construct validity in the Mexican-Spanish version and FFI demonstrated very low quality evidence for sufficient construct validity in the Taiwan-Chinese version.

Conclusion: There is a lack of evidence for the measurement properties of the PROMs used in ankle fracture clinical trials. None had sufficient evidence to recommend them for use in this patient population and more evidence is required in this area.

This work has been published in a peer reviewed journal:

Mckeown, R., Ellard, D.R., Rabiu, A.R., Karasouli, E. and Kearney, R.S. (2019) A systematic review of the measurement properties of patient reported outcome measures for adults with an ankle fracture. *Journal of Patient Reported Outcomes.* 3(70).

This work has been presented at a national conference:

Title: Primary outcome measures used in interventional trials for ankle fracture: a systematic review. **Presenter**: Rebecca McKeown **Event:** BOA Annual Congress, 12/09/2019, Liverpool

3.1 Introduction

In chapter two of this thesis, I explored the current and previous use of outcome measures in clinical trials of non-pharmacological interventions for adults with ankle fracture. The key finding from this review was there are a wide range of different outcomes are used to assess outcome in this type of trial. These included a range of objective clinical measures, such as radiographic assessment, range of movement and swelling assessments. There were also a significant number of trials collecting patient reported scores, both generic scores such as the SF-12 and also ankle specific scores, such as OMAS and the AAOS. Combined patient and clinician reported scores were also a type of outcome measures, of which the most common was the AOFAS ankle hindfoot score, but this also included a range of clinician derived scores with limited information on the criteria used to calculate scores in some cases. Finally, a range of other measures, such as health economic analysis such as EQ-5D, and hospital episode data such as length of stay, were collected as outcome measures. An exploration of the temporal trends in outcome measure use showed an increasing use of patient reported measures since the 1980's and a subsequent decline in the use of clinician reported objective measures in the same time period.

The results of this initial review is in agreement with other literature which points to a growing trend towards the use of PROMs to measure outcome in clinical trials, in order to understand facets of health related quality of life from the patient's perspective [82]. For these reasons, the commonly collected ankle specific PROMs found to be used in this particular type of trial will be the focus of this chapter. Several researchers have raised concerns with the suitability of the PROMs contained within this category for assessing patients with various conditions of the foot and ankle [129 150 151].

Determining the quality of PROMs collected in clinical trials is of paramount importance, particularly for an injury of growing incidence such as ankle fracture. As previously discussed, the use of PROMs which are not valid or are unresponsive in a clinical trial is unethical and a waste of resources [108]. Clinical trials rely on the goodwill of participants agreeing to be randomised to an intervention, which could possibly be associated with differing levels of risk depending on the interventions being assessed. To subsequently assess the treatment effects of this intervention with a PROM which demonstrates insufficient measurement properties and is unable to

measure the construct of interest or detect changes in this construct is potentially exposing the patient to unnecessary risk.

One could argue it is unethical to not allocate research funding to burdensome injuries, such as ankle fractures, just because the PROMs do not have any evidence to support their use. The measures could perform suitably when utilised in the context of a clinical trial but without evidence to support this validity. A review of PROM usage in orthopaedics by Gagnier [152] highlights that there is evidence that many of the PROMs used in orthopaedic research either have evidence of insufficient measurement properties, or a lack of evidence regarding their measurement properties within their context of intended use. Understanding the current evidence base for ankle specific PROMs will guide the development of further validation research into these questionnaires, enabling an understanding of which measures we should be using and where the evidence gaps exist.

In the initial review, I found there were eight region specific PROMs used in clinical trials for ankle fracture interventions in adults; OMAS [123], AAOS [125], MOXFQ [124], FFI [153], FAAM [128], KS [127 154], A-FORM [126] and FAOS [155]. These PROMs will be the focus of this review and the search strategy will search for all evidence for the measurement properties of these PROMs in the population of adults with ankle fracture. The context of these PROMs is evaluative as the basis of this research is looking at assessment in clinical trials. The PROMs are based on a reflective model, which means that the items collected in the PROMs indicate the level of the latent variable [83 156]. The specific construct of interest is patient reported outcome following ankle fracture.

An article has previously been published to assess the psychometric properties of PROMs for ankle fractures [140]. The authors of this study concluded that the A-FORM was the most appropriate measure to use in this patient population. An update of this review is warranted as the searches for this review were completed in December 2016, further articles may have been published in this research area in the time since. Furthermore, understanding of the current evidence base of the measurement properties of these PROMs which are currently being used in clinical trials is beneficial to understand whether these should be continue to be used. Additionally, since the previous review was completed, there has been an update to the COSMIN methodology, therefore this review will use the most current version [109]. In this

review I will identify and appraise the evidence of measurement properties for the ankle specific PROMs found to be collected in trials of ankle fracture, results of which are reported in chapter two. The results of this review informed the development of further packages of work presented in this thesis.

The aim of this systematic review is to identify and critically appraise the quality of evidence for the measurement properties of ankle specific PROMs used in the population of adults with ankle fracture. Prior to outlining the methods, I will first describe the PROMs which are under review to define the outcomes under evaluation in this study.

3.1.1 Description of PROMs under review

Table 22 shows the characteristics of each article included in this review. The AAOS is an instrument consisting of 25 questions, in a combination of multiple choice, single response questions and Likert scales developed for individuals with a range of foot and ankle conditions [125]. Originally written in the English language, the questionnaire asks users to recall their foot and ankle status from the past week, or since injury if less than one week. The score consists of 2 subscales; a 20-item core score (AAOS-CS) and a five-item shoe comfort scale (AAOS-SCS). The scoring system is 0-100 with higher scores indicating better outcomes.

The A-FORM is a PROM developed specifically for individuals with ankle fracture and consists of 15 items in a combination of multiple choice, single answer questions and Likert scale questions [126]. The recall period is not specified. It was originally written in the English language and the raw score is 14-70 which users then convert into a summary score from 0-100. Lower scores indicate more favourable outcomes.

The FAAM is a questionnaire to assess self-reported outcome in patients with ankle and foot musculoskeletal disorders in the English language [128]. The questions are Likert scale questions with 29 items in total, with a recall period of one week. It consists of two subscales of ADL and sports comprising of 21 and eight items respectively. The total scores from each subscale is calculated and users then convert these raw scores to a percentage 0-100 score, with higher scores indicating a more favourable outcome for the individual.

The FAOS is an adaptation of the Knee Injury and Osteoarthritis Outcome Score (KOOS) [155] [157]. The FAOS measures self-reported functional and symptomatic

outcome in patients with a wide range of foot and ankle disorders and consists of a total of 42 items in a Likert scale question format. The score consists of five subscales; pain, other symptoms, ADL, sports and recreation and foot and ankle related quality of life. Users then convert the total score to a summary score of 0-100, with higher scores indicating a more favourable outcome.

The FFI is a questionnaire to assess patients with a range of foot and ankle conditions [153]. It is written in the English language and consists of 23 items in a Likert scale format which comprise of three subscales; pain, disability and activity limitation. Recall period for the questionnaire is current status, defined as the past week. The score is 0-100 with lower scores indicating a more favourable outcome.

The KS is a questionnaire devised by researchers in Sweden to assess functional outcome in patients following lateral ankle ligament injury [127]. It comprises of eight items in a multiple choice, single answer format with a weighted score. The score is written in the English language and there is no recall period specified. The scores are 0-100 and higher scores indicate a more favourable outcomes.

The MOXFQ [124] is a patient reported outcome measured developed initially for use in patients following hallux valgus surgery, but use has been widened to include a range of foot and ankle conditions. There are two variations of the score, the foot score and the foot and ankle score. The questionnaire was originally in the English language and consists of 16 items and 3 subscales; walking/standing, pain and social interaction. It has a recall period of four weeks, or since injury or operation if less than four weeks. Users convert the raw score to a summary score of 1-100, with lower scores indicating a more favourable outcome.

The OMAS was developed to assess outcome in individuals undergoing surgical fixation of an ankle fracture [123]. It consists of nine multiple choice, single answer questions with a weighted score and was developed by researchers in Sweden although the article is written in the English language. There is no recall period specified. The score is 0-100 with higher scores indicating a more favourable outcome.

Table 22 - Descriptions of PROMs under review

PROM	Construct	Target population	Mode of administration	Recall period	Items and subscales	Response options	Scoring	Original language
AAOS	Patient reported outcome	Patients with foot and ankle conditions	Paper based questionnaire	Past week (or since injury)	25 items, 2 subscales; core score and shoe comfort module	Combination of single answer, multiple choice and Likert scale questions	0-100; higher scores indicates better outcomes	English
A-FORM	Patient reported outcome	Patients with ankle fracture	Paper based questionnaire	Not specified	15 items	Combination of single response, multiple choice answers and Likert scales	0-100; lower score indicates better outcomes	English
FAAM	Patient reported outcome	Patients with foot and ankle musculoskeletal disorders	Paper based questionnaire	Past week	29 items, 2 subscales; ADL and sports	Likert scale questions	0-100; higher scores indicates better outcomes	English
FAOS	Patient reported outcome	Patients with a variety of foot and ankle related problems	Paper based questionnaire	Past week	42 items, 5 subscales; pain, other symptoms, ADL, sport, foot	Likert scale questions	0-100; higher scores indicates better outcomes	English

					and ankle related quality of life			
FFI	Patient reported outcome	Patients with foot and ankle disability	Paper based questionnaire	Current status, defined as the past week	23 items, 3 subscales; pain, disability and activity limitation	Likert scales questions	0-100; lower scores indicate better outcomes	English
KS	Patient reported outcome	Patients with lateral ligament injuries of the ankle	Paper based questionnaire	Not specified	8 items	Single response, multiple choice answers	0-100; higher scores indicates better outcomes	English
MOXFQ	Patient reported outcome	Patients with a range of foot and ankle conditions	Paper based questionnaire	During the past 4 weeks (or since injury or operation if less than 4 weeks)	16 items, 3 subscales; walking/standing, pain and social interaction	Single response, multiple choice answers	0-100; lower scores indicates better outcomes	English
OMAS	Patient reported outcome	Patients with ankle fracture	Paper based questionnaire	Not specified	9 items	Single response, multiple choice answers	0-100; higher scores indicates better outcomes	English

Key: ADL = activities of daily living

3.1.2 Aim

The aim of this review is to critically appraise the available evidence for the measurement properties for ankle specific self-report questionnaires to measure outcome in adults with ankle fracture and make recommendations regarding the most appropriate PROM for this patient population and context of use.

3.1.3 Research questions

1. What is the evidence for the measurement properties of ankle specific PROMs used in clinical trials for adults with ankle fracture?

2. Which is the most appropriate PROM to use in this patient population?

3.1.4 Objectives

1. To critically appraise the evidence for the development, measurement properties, interpretability and feasibility of ankle specific PROMs used in clinical trials for adults with an ankle fracture.

3. To formulate recommendations regarding the optimal outcome measure to use in this patient population and further areas for research.

3.2 Methods

3.2.1 Protocol, registration and methodology

This systematic review is reported against the PRISMA statement [120] and the protocol was prospectively registered with PROSPERO International Prospective Register of Systematic Reviews (Reference CRD42018103112) [158] prior to commencing. This review will be completed using methodology developed by Consensus-based Standards for the Selection of Health Measurement Instruments (COSMIN) group [109 159]. This methodology is frequently used in studies to evaluate the quality of available PROMs in a clinical area [160-162] and is regarded as the most appropriate methodology to use for this type of systematic [152].

3.2.2 Eligibility criteria

Table 23 shows the eligibility criteria for inclusion in this review. This review was concerned with the construct of patient reported outcome following ankle fracture, therefore to be eligible for inclusion, articles must evaluate the measurement properties of one or more of the PROMs on the pre-specified list. These PROMs were those to be found in the previous review completed as part of this thesis and

are outlined and described in section 3.1.1. This could also include the development of the PROM or a study into the interpretability of the PROM, such as change in scores or minimal differences in scores.

The measurement properties must be assessed in a majority sample of adults with an ankle fracture. In alignment with COSMIN recommendations, majority is defined as equal to or greater than 50% of the sample [109]. In cases where the whole sample did not meet the criteria for majority but a subgroup analysis was performed of only the ankle fracture participants, the analysis which included these individuals only were included and the whole group analyses or comparison of groups analyses were excluded. All other analyses of the whole sample or a different population of patients were excluded.

Articles which use the PROM/s as an outcome measure in a clinical study, trial or evaluation were excluded as there is no assessment of the measurement properties of the PROM in these types of study. Studies will be excluded if they use the PROM to validate another instrument (not on the pre-specified list of PROMs under review in this study). Abstracts and conference proceedings will be excluded as these do not provide sufficient information on methods used for assessment of risk of bias. Furthermore, articles will be excluded if they have adapted the measure in any way without formal justification for the changes as part of a comprehensive translation and/or cultural adaptation process.

Table 23 - Eligibility criteria for inclusion in review

Inclusion criteria	Exclusion criteria
The construct of interest is outcome following ankle fracture	Studies which only use the PROM/s as an outcome measurement instruments, for instance, studies in which the PROM is used to measure outcome in or in studies which the PROM is used in a validation study of another instrument
The study sample or majority (≥50%) should represent the population of adults with ankle fractures	Articles that do not appear as full texts, such as abstracts and conference proceedings

The studies included should concern one or more of the previously specified ankle specific patient reported outcomes found in the previous systematic review (OMAS, M-OXFQ, AAOS, FFI, FAAM, KS, A-FORM, FAOS)	Articles which have adapted the PROM in any way without formal justification for the changes as part of a comprehensive translation and/or cultural adaptation process.
The aim of the studies should be the evaluation of one or more measurement properties, the development of a PROM or the evaluation of the interpretability of the PROMs of interest (e.g. evaluating the distribution of scores in the study population, percentage of missing items, floor and ceiling effects, the availability of scores and change scores for relevant (sub)groups, and the minimal important change or minimal important difference)	

3.2.3 Sources, search and study selection

Comprehensive searches of the EMBASE and MEDLINE databases using the Ovid search engine and the CINAHL database using EBSCOhost were carried out on 08/08/2018. No date limits were applied. The search strategies used in this project can be found in appendix 3. The database search filters were adapted from those developed by the COSMIN group, described by Terwee, et al. [163]. The reference lists of all included studies were also searched for any other potentially eligible papers for this review.

3.2.4 Study selection

The results of the searches were downloaded to a Microsoft Excel spreadsheet (Microsoft version 11) detailing the authors, date, journal, title and abstract. I screened this list of references for eligibility by title and abstract. An independent reviewer (ARR) duplicated this exercise in order to ensure accuracy and reduction of bias in the selection process. We both independently selected references which were potentially eligible based upon the title and abstract using the Excel spreadsheet. Once the independent review of titles and abstracts was completed we discussed the results of the independent review and formulated a list of full texts to

retrieve. If we were unsure on eligibility at the stage of title and abstract review, the full text was retrieved and reviewed for purposes of completeness. The full text for all studies which at least one reviewer felt could potentially be eligible were retrieved and reviewed to reduce error and bias.

A second independent review process took place for the full text of articles deemed potentially eligible in the first review. The full texts were reviewed independently by each reviewer and a decision was made based upon the eligibility criteria. The reviewers then discussed each article and explained their reasoning for including or excluding it from the review. If the reviewers disagreed, the article was discussed in further detail to formulate a decision through consensus. Any which the reviewers disagreed on following this discussion were referred to a third reviewer (RSK) for guidance and final decision. The reference lists were also screened by each reviewer independently and full text retrieved and reviewed by both reviewers independently if at least one reviewer identified it as being potentially eligible for inclusion. The form used to assess eligibility in this process is found in appendix 4.

3.2.5 Risk of bias assessment, quality review and data extraction

The risk of bias of the included studies was assessed using the COSMIN risk of bias checklist [164]. The results of the measurement properties was compared against the criteria of good measurement properties. An overall summary of findings was generated, including the modified Grading of Recommendations Assessment, Development and Evaluation (GRADE) quality of evidence method. This process of risk of bias assessment, data extraction and overall quality assessment was completed by myself for all studies and a second reviewer, Eleni Karasouli (EK), duplicated this process for a sample of five of the seven included studies. This was to reduce bias and error in the process. Ideally this would have been completed for all seven of the included articles but resource limitation meant that this was only able to be completed for a sample of the studies. A sample of more than 50% of the studies were reviewed by a second reviewer independently so that the majority of the included studies were reviewed twice for purposes of bias reduction and consistency.

We independently assessed the risk of bias of the included studies, compared the results of the articles against the COSMIN criteria of good quality measurement properties and summarised the overall results using the modified GRADE assessment. The forms used to extract the data can be found in the summary risk of bias table in section 3.3.4 and the full version in appendix 5. The forms for the criteria

of good measurement properties is found in section 3.3.5 and the summary of findings tables in sections 3.3.6. We then discussed our findings for the sample of papers which were included in the duplicate review process with the aim of reaching agreement. In the instances of disagreement, we re-reviewed the article together and discussed the issues further with the aim of reaching consensus. In cases where we were unable to reach consensus, we referred to a third reviewer (RSK) for final decision. For the two articles which were not reviewed by both reviewers, the 2nd reviewer and consensus boxes on these tables were left blank and only the 1st reviewer boxes are completed for this reason. Finally, an assessment of the feasibility and interpretability of the outcome measures included was completed in order to formulate recommendations on the most appropriate outcome measure for use in the patient population.

3.3 Results

3.3.1 Search results

Figure 12 shows the application of the eligibility criteria to the returned articles as a result of the initial independent review process. We screened a total of 300 articles by title and abstract. A further eight items in the reference lists were also screened. We excluded 284 of these based upon the title and abstract review. 24 articles were reviewed in full text format, of which 17 were excluded; one did not evaluate one or more measurement property of the PROM in question, one used the PROM in question to validate another outcome measure, one was an abstract only and one adapted the PROM without any formal justification for the change or as part of a translation and/or cross-cultural adaptation process. A further 13 articles did not use a majority sample of adults with ankle fracture and therefore were excluded. This left seven articles evaluating four different PROMs included in the review.

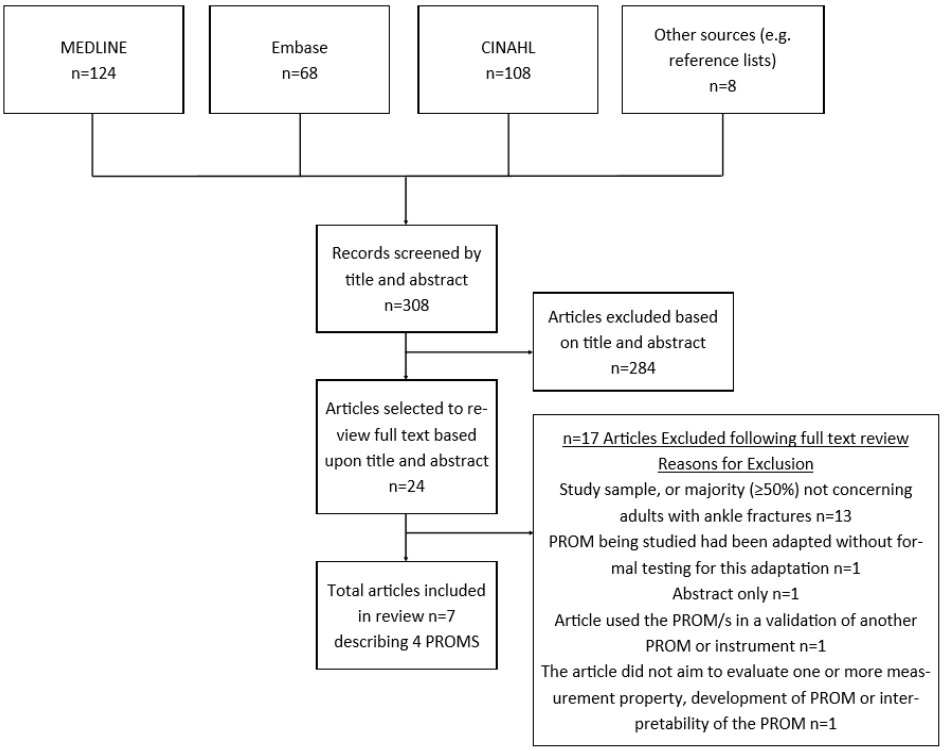


Figure 12 - PRISMA diagram

These seven articles included evaluations of AAOS, A-FORM, FFI and OMAS. The FAOS, FAAM, KS and MOXFQ outcomes had no evidence for their measurement properties in the patient population and therefore will not be included in the results tables. These PROMs, however, will be included wherever possible. For example, the PROM descriptor and feasibility of PROMs sections for purposes of completeness. Of the seven articles, one was authored by researchers in the USA, one in Sweden, one in Norway, one in Australia, one in Taiwan and two by researchers based in Turkey. Of all the articles, two detailed the development of the PROM in question and five detailed the evaluation of an existing PROM.

One article included by Wu, et al. [165] studied a population of plantar fasciitis patients (n=50) as well as "ankle-foot fracture" (n=29) patients. There was confusion as to the exact nature of the fractures (termed ankle-foot and ankle/foot fractures interchangeably) the review team discussed and felt that this terminology may have been lost in translation, as the researchers go on to use the term ankle fracture in the table headings. The introduction states that two distinct groups were analysed, therefore we deducted from this that there were two distinct groups of patients rather than a mixture of foot and ankle fractures. I contacted the corresponding author for more information of the exact nature of this sample of participant's injuries (emailed

on 21/09/2018) but received no response. I made the decision to include this, with the assumption that the population described by the authors was solely individuals with ankle fractures. In this article, the researchers split the analysis into the two samples, therefore we were able to extract data on the ankle fracture sub-sample alone. All the data discussed and evaluated here were those of the ankle fracture sample alone. The data for the plantar fasciitis patients or analysis of the whole combined group of the two clinical groups was not included in the results and analysis of this review, as per the eligibility criteria.

3.3.2 Hypotheses testing for construct validity

As per the COSMIN methodology, a judgement should be made prior to evaluating the measurement properties whether there is an available gold standard test for measuring the construct of interest in the population of interest [166]. Here, there is no accepted gold standard available for measuring outcome in adults with ankle fracture, meaning that all evidence for validity presented in the included studies will be considered as evidence for construct validity or the construct approach for responsiveness. Criterion validity will therefore not be assessed throughout this review due to a lack of gold standard test for measuring the construct in this population.

In order to make decisions on whether sufficient construct validity of a measure is demonstrated within a study, a set of hypotheses should be formulated to set parameters for expected magnitudes of correlations with scores of other instruments [109]. The following boundaries were set for these hypotheses, taken from an article by Abma, et al. [167] for purposes of hypotheses setting:

1. A weak correlation is defined as <0.30
2. A weak to moderate correlation is defined as >0.20 - <0.40
3. A moderate correlation is defined as >0.30 - <0.70
4. A moderate to high correlation is defined as >0.60 - <0.80
5. A high correlation is defined as >0.70

Table 24 shows the hypotheses formulated for this particular review.

Table 24 - Hypotheses formulated for assessment of construct validity

1.	Scores of PROMs assessed here with scores of instruments/domains measuring a similar construct (i.e. outcome in individuals with ankle fractures or other conditions affecting the foot and ankle complex) will correlate highly or moderate to highly.
2.	Scores of the PROMs included here with scores of instruments/domains measuring related but not the same constructs, for example generic disability scores or health related quality of life measures, such as the Lower Extremity Functional Scale (LEFS), SF-12 or SF-36 will correlate moderately or moderate to highly.
3.	Scores of the PROMs included here with scores of instruments/domains measuring combined health related quality of life and cost utility (i.e. EQ-5D) will correlate moderately or weak to moderately.

I hypothesised that scores of the included PROMs in this review would correlate highly or moderately to highly with scores of other region specific PROMs as these instruments should be measuring a similar construct and therefore should demonstrate at least a moderate correlation of >0.6. I hypothesised that scores of generic PROMs, such as those measuring health related quality of life of lower limb disability would correlate with the scores of the PROMs included here moderately or moderate to highly (0.3 – 0.8). I chose this because the two categories of PROMs should be measuring a related construct of interest but not the same construct, therefore they should be less associated than those which measure the same construct. Finally, PROMs which measure a health related quality of life and health utility, such as EQ-5D should correlate moderately or weak to moderately (>0.2 and <0.7) with scores of PROMs included in this review.

As recommended by Abma, et al. [167], construct validity of a PROM is considered sufficient if at least 75% of the hypotheses tested are met.

3.3.3 Study characteristics

Table 25 shows the characteristics of the studies included in this review.

Buker, et al. [168] assessed the OMAS in a population of 91 patients in Turkey, using a translated Turkish version of the questionnaire. Individuals included were 20-60

years of age with a mean age of 41.54 years (±13.28), 30.8% were female. All participants had ankle fractures which were operatively managed and the mean follow up duration was 27.92 months (±8.94), with questionnaires collected in clinic or over the phone.

Garratt, et al. [169] translated the OMAS and two other PROMs into the Norwegian language for their study based in Norway, which involved a total of 959 patients. The mean age was 57.5 years with a range of 22.2-91.2 and 56.8% of the sample were female. All of the participants included received operative management for their ankle fracture and questionnaires were completed on paper and returned to researchers via post.

McPhail, et al. [126] detailed the development of the A-FORM questionnaire in their article, which included a sample of eight individuals for the Delphi study and a further 41 participants for the cohort study. Their study was completed in Australia in the English language and the mean age of participants was 36.8 years with a range of 26.1-53.8 years. 27% of the participants were female and it included a mixture of operatively managed (46.3%) and non-operatively managed (53.7%) patients. Follow ups were completed at 6-8 weeks and 12-16 weeks post injury and were completed on paper either in clinic or via post.

Turhan, et al. [170] completed an assessment of the OMAS in a translated Turkish version in their study based in Turkey, which included 100 participants. Of the sample, 57% were managed operatively and 43% were managed non-operatively and 49% were female. The mean age was 42.3 years (±17.7) with a range of 16-81 years. Participants were follow up for 4.3 years following injury, although authors did not state how this follow up took place (i.e. in clinic or using postal questionnaires).

Olerud and Molander [123] detailed the development of OMAS in their articles, which studied 90 operatively managed participants. The authors did not described the population in any further detail, so there is no data for age or gender. The location of the study was Sweden but the language of data collection was not specified, so this is assumed to be in the Sedish language. Authors did not specify the method of data collection of the questionnaires, but due to the age of the article, this can be assumed to be paper questionnaires.

Wu, et al. [165] assessed FFI in their Taiwan based study, using a version translated into the Taiwan-Chinese language. Their sample included some non-ankle fracture patients, but a subgroup analysis was completed on the ankle fractures patient only,

so the study has been included for this purpose. The study assessed a sample of 29 ankle fracture participants. The mean age of the whole sample was 37.2 years (±14.8) and 38% were female. There was no information on whether the included injuries were operatively or non-operatively managed and the questionnaires were completed on paper in a clinic setting.

Zelle, et al. [171] used a translated version of AAOS in the Mexican-Spanish version in their USA based study. The study included 100 participants with a mean age of 42.98 years and a range of 18-88 years. The sample was a mixed population of ankle and foot fractures, with 58 ankle fractures, five talus fractures, one achilles tendon rupture, 11 calcaneal fractures, six mid-foot fractures. 73 of the total sample was operatively managed and 27 were non-operatively managed, although there is no breakdown of the management of the ankle fracture injuries in isolation. 41% of the participants were female and follow up took place over a mean of 3.97 months (±4.71) either in clinic or via post.

Table 25 - Study characteristics

Article (Reference)	PROM Evaluated	Population					Instrument Administration		
		Number of participants	Age mean (SD, range) (years)	Gender (% female)	Injury information	Follow up duration mean (SD, range)	Country	Language	Method of Collection
Buker, et al. [168]	OMAS	n=91	41.54 (±13.28, 20-60)	30.8%	Operatively managed ankle fractures	27.92 months (±8.94, range not specified)	Turkey	Turkish	Initial in clinic, then follow up in clinic or telephone interview
Garratt, et al. [169]	OMAS	Cohort n=959, test-retest questionnaire n=299	57.5 (± not specified, 22.2-91.2)	56.8%	Operatively managed ankle fractures	Not specified, stated recruited over a 3 year period	Norway	Norwegian	At home via post
McPhail, et al. [126]	A-FORM	Delphi panel n=8 Cohort n=41	36.8 (± not specified, 26.1-53.8)	27%	Operatively managed (46.3%) and non-operatively	6-8 week post injury and at	Australia	English	Either in clinic or at home via post

					managed (53.7%) ankle fractures	12-16 weeks post injury			
Olerud and Molander [123]	OMAS	n=90	Not specified	Not specified	Operatively managed ankle fractures	Not specified	Not specified (assumed Sweden)	Not specified	Not specified
Turhan, et al. [170]	OMAS	n=100	42.3 (±17.7, 16-81)	49%	Operatively (57%) and non-operatively managed (43%) ankle fractures	Mean of 4.3 years	Turkey	Turkish	Not specified
Wu, et al. [165]	FFI	29 ankle fractures (50 plantar fasciitis patients)	37.2 (±14.8, range not specified)	38%	Not specified	Not specified	Taiwan	Taiwan Chinese	Paper based in clinic at follow up
Zelle, et al. [171]	AAOS	n=100 (58 ankle #s, 5 talus #s, 1 Achilles tendon rupture, 11 calcaneus #s, 6 midfoot #s)	42.98 (± not specified, 18-88)	41%	73 operatively managed and 27 non-operative management (of whole sample)	3.97 months (±4.71 range not specified)	USA	Mexican-Spanish	Initial was at clinic or home via post, follow up was at home via post.

3.3.4 Risk of bias assessment

Table 26 shows the COMSIN risk of bias summary. Full details of the individual scores for each article can be found in the full risk of bias tables in appendix 5.

Table 26 - COSMIN risk of bias summary table

Article → **Measurement property ↓**	***Buker et al. (2017) OMAS Turkish ****	***Garratt et al. (2018) OMAS Norwegian***	***McPhail et al. (2014) A-FORM English ****	***Turhan et al. (2017) OMAS Turkish ****	***Olerud and Molander (1984) OMAS***	***Wu et al. (2008) FFI Taiwan Chinese***	**Zelle et al. (2017) *AAOS Mexican-Spanish ****
Content validity							
1. PROM development			Doubtful		Inadequate		
2. Content validity							
Internal structure							
3. Structural validity		Doubtful					
4. Internal consistency	Doubtful	Doubtful		Doubtful			

5. Cross cultural validity -measurement invariance							
Remaining measurement properties							
6. Reliability	Inadequate	Doubtful		Doubtful			Doubtful
7. Measurement error		Doubtful		Doubtful			
8. Criterion validity	N/A	N/A	N/A	N/A	N/A	N/A	N/A
9. Hypotheses testing for construct validity	Doubtful (convergent validity)	Adequate (convergent and known groups validity)		Doubtful (convergent validity)		Doubtful (convergent validity)	Doubtful (convergent validity)
10. Responsiveness							

The article by Buker, et al. [168] showed a doubtful rating for internal consistency and this was because, whilst the authors reported the Cronbach's alpha of the score in their results, there is unclear whether the score is unidimensional. Both reviewers felt there were no other important flaws for the assessment of this measurement property. The reliability assessment in the article scored inadequate on the risk of bias checklist, this was because it was unclear if the test conditions for the test re-test analysis were the same each time. The article scored doubtful for the stability of the patients being tested and the time interval being appropriate as the authors did not check that the patient's had not changed status using a global rating of change score. Finally, the article scored doubtful for convergent validity because, whilst the comparator instruments were clearly defined, it was doubtful because the comparator instrument (the FAOS) has not got any validation evidence in the target population, therefore this is unclear. However, the statistical methods were appropriate and the reviewers found no other significant flaws in the methods.

Garratt, et al. [169] scored doubtful for risk of bias regarding their assessment of structural validity. The authors used confirmatory factor analysis and the sample size was adequate, however the reviewers felt that there were flaws in the methods, as no clear hypotheses/understanding for the structure of the OMAS is present in the literature. The article scored doubtful for internal consistency assessment as it is unclear whether the score is unidimensional. Reliability scored doubtful, due to the period of time which was left between the first and second test, which was six weeks and the patients are unlikely to have remained stable during this period. The test conditions were the same for each test and an ICC was presented. This article scored doubtful for measurement error, again for the reasons statement above regarding the stability of the construct in question over the time period of the test re-testing. Finally, for construct validity, the article scored adequate for convergent validity because, whilst it was clear which comparator instruments were used, it was less clear that the comparators display sufficient measurement properties in the population. There is some evidence for the measurement properties of LEFS however, not for SEFAS. For known groups validity, the article scored adequate because the groups were adequately described.

The article by McPhail, et al. [126] described the development of the A-FORM. Here, they performed a Delphi study and item-reduction exercises and therefore we did not class this as content validity as questions were not asked regarding the final version of the questionnaire. The articles scored very good on risk of bias for all factors,

except for the section regarding the moderators. In this case, there was no information on the experience levels of the moderators in the article, therefore this criteria brought the overall score to doubtful.

The article by Turhan, et al. [170] assessed the internal consistency of the OMAS Turkish version and this was scored as doubtful on the risk of bias assessment for the measurement property because there is no information on whether the score is unidimensional based upon the evidence. For reliability, this article was measured as inadequate because the methods do not provide sufficient information on how the test re-test questionnaires were completed. Furthermore, the authors did not ask patients if they had changed or remained stable using a rating of change score. This article scored doubtful for measurement error, due to the same reasons as reliability. The authors also assessed convergent validity of the OMAS. They were clear on the comparator instruments which were used, however there is no evidence that the comparators (SF-12 and FAAM) have been validated in the target population, therefore this scored doubtful on this measurement property.

Olerud and Molander Olerud and Molander [123] described the development of the OMAS questionnaire in their article. This score inadequate form PROM development against COSMIN criteria because the construct of interest was not clearly described and the origin of the construct was not clear. Because the population was not clearly described in this article, we scored this section as doubtful for the sample being representative of the population and there was no qualitative data collection completed, which scored inadequate on this section, giving an overall score for this article as inadequate.

Wu, et al. [165] scored doubtful on the COSMIN risk of bias for their assessment of construct validity in the sample of ankle fracture patients. The authors were clear on the comparator instrument (SF-36) which were used, however there is limited evidence for the measurement properties of these instruments in the specific patient population, therefore the study scored doubtful for this measurement property.

Zelle, et al. [171] scored doubtful on the risk of bias for reliability, because there was no measure on whether the participants had remained stable between the two tests using a rating of change score. Furthermore, the questionnaires were not completed in the same setting. The study also scored doubtful for convergent validity because, whilst the comparator measures were clearly stated (SF-36), there is no evidence in

the literature of sufficient measurement properties of this score for the specific patient population.

3.3.5 Criteria of good measurement properties

The articles included in the review used several comparator PROMs and domains to complete assessments for construct validity. For ease of interpretation of the table, I have listed the comparator PROMs and subscales of these PROMs in table 27.

Table 27 - PROMs and subscale of PROMs included in the construct validity assessments

PROM	Subscales	Abbreviation
36-item Short Form Survey (SF-36)	SF-36-Physical component summary score	SF-36-PCS
	SF-36-Mental component summary score	SF-36-MCS
12-item Short Form Survey (SF-12)	SF-12-Physical component summary score	SF-12-PCS
	SF-12-Mental component summary score	SF-12-MCS
American Association of Orthopaedic Surgery Foot and Ankle Questionnaire (AAOS)	AAOS-Core score	AAOS-CS
	AAOS-Shoe comfort score	AAOS-SCS
EuroQol EQ-5D	EQ-5D-Mobility subscale	EQ-5D-M
	EQ-5D-Self care subscale	EQ-5D-SC
	EQ-5D-Usual activities subscale	EQ-5D-UA
	EQ-5D-Pain and discomfort subscale	EQ-5D-P&D
	EQ-5D-Anxiety and depression subscale	EQ-5D-A&D
Foot Function Index (FFI)	FFI-Activity limitation subscale	FFI-AL
	FFI-Disability subscale	FFI-D

	FFI-Pain subscale	FFI-P
Foot and Ankle Ability Measure (FAAM)	FAAM-Activities of daily living subscale	FAAM-ADL
	FAAM-Sports subscale	FAAM-S
Foot and Ankle Outcome Survey (FAOS)	FAOS-Pain subscale	FAOS-P
	FAOS-Symptoms subscale	FAOS-S
	FAOS-Activities of daily living subscale	FAOS-ADL
	FAOS-Sports subscale	FAOS-Sports
	FAOS-Quality of life subscale	FAOS-QoL
Global self-rated function (GSRF)	None	GSRF
Self-reported Foot and Ankle Score (SEFAS)	None	SEFAS
Lower Extremity Functional Scale (LEFS)	None	LEFS
Olerud Molander Ankle Score (OMAS)	None	OMAS

Tables 28 to 34 show the results of the articles included in this review. Where the second reviewer (EK) has completed the assessment, these are marked with an asterisk (*) as per the table key and all results presented for these articles are a consensus agreement between the first and second reviewer. Those without an asterisk are the scores only from the first reviewer (RM).

Table 28 - Measurement properties 2-6 for AAOS

*Key: Meth Qual=Methodological Quality; n=Number of participants. Score: V= very good; A = adequate; D = doubtful; I = inadequate; N/A= not applicable; ? = indeterminate; + = sufficient; - = insufficient; r = correlation results; * = reviewed by both reviewers and results are a consensus.*

PROM and article	Country/language in which PROM evaluated	Structural validity			Internal consistency			Cross-cultural validity/measurement invariance			Reliability		
		n	Meth Qual	Result (rating)	n	Meth Qual	Result (rating)	n	Meth Qual	Result (rating)	n	Meth Qual	Result (rating)
AAOS, Zelle et al. (2017) *	USA/Mexican-Spanish			?			?			?	63	D	ICC or weighted Kappa not reported (?)
Summary result (overall rating):											63		?

Table 29 - Measurement properties 7-10 for AAOS

*Key: Meth Qual=Methodological Quality; n=Number of participants. Score: V= very good; A = adequate; D = doubtful; I = inadequate; N/A= not applicable; ? = indeterminate; + = sufficient; - = insufficient; r = correlation results; * = reviewed by both reviewers and results are a consensus.*

PROM and article	Country/language in which PROM evaluated	Measurement Error			Hypotheses testing for construct validity			Responsiveness		
		n	Meth Qual	Result (rating)	n	Meth Qual	Result (rating)	N	Meth Qual	Result (rating)
AAOS, Zelle et al. (2017) **	USA/Mexican-Spanish			?	83	D	AAOS-CS and SF-36-PCS r=0.667 AAOS-CS and SF-36-MCS r=0.506 AAOS-SCS and SF-36-PCS r=0.358 AAOS-SCS and SF-36-MCS r=0.356 (+4)			?
Summary result (overall rating):				?	83		+4			?

Table 30 - Measurement properties 2-6 for A-FORM

*Key: Meth Qual=Methodological Quality; n=Number of participants. Score: V= very good; A = adequate; D = doubtful; I = inadequate; N/A= not applicable; ? = indeterminate; + = sufficient; - = insufficient; r = correlation results; * = reviewed by both reviewers and results are a consensus.*

PROM and article	Country/language in which PROM evaluated	Structural validity			Internal consistency			Cross-cultural validity/measurement invariance			Reliability		
		n	Meth Qual	Result (rating)	n	Meth Qual	Result (rating)	n	Meth Qual	Result (rating)	n	Meth Qual	Result (rating)
A-FORM, McPhail et al. (2014) *	Australia/English			?			?			?			?
Summary result (overall rating):				?			?			?			?

Table 31 - Measurement properties 2-6 for FFI

*Key: Meth Qual=Methodological Quality; n=Number of participants. Score: V= very good; A = adequate; D = doubtful; I = inadequate; N/A= not applicable; ? = indeterminate; + = sufficient; - = insufficient; r = correlation results; * = reviewed by both reviewers and results are a consensus.*

PROM and article	Country/language in which PROM evaluated	Structural validity			Internal consistency			Cross-cultural validity/measurement invariance			Reliability		
		n	Meth Qual	Result (rating)	n	Meth Qual	Result (rating)	n	Meth Qual	Result (rating)	n	Meth Qual	Result (rating)
FFI, Wu et al. (2008)	Taiwan Chinese			?			?			?			?
Summary result (overall rating):				?			?			?			?

Table 32 - Measurement properties 7-10 for FFI

*Key: Meth Qual=Methodological Quality; n=Number of participants. Score: V= very good; A = adequate; D = doubtful; I = inadequate; N/A= not applicable; ? = indeterminate; + = sufficient; - = insufficient; r = correlation results; * = reviewed by both reviewers and results are a consensus.*

PROM and article	Country/language in which PROM evaluated	Measurement error			Hypotheses testing for construct validity			Responsiveness		
		n	Meth Qual	Result (rating)	n	Meth Qual	Result (rating)	n	Meth Qual	Result (rating)
FFI, Wu et al. (2008)	Taiwan Chinese			?	29	D	FFI and SF-36-PCS r= -0.61 + FFI and SF-36-MCS r= -0.34 + FFI-AL and SF36-PCS r= -0.58 + FFI-AL and SF36-MCS r= -0.38 + FFI-D and SF36-PCS r= -0.54 + FFI-D and SF36-MCS r= -0.27 - FFI-P and SF36-PCS r= -0.44 + FFI-P and SF36-MCS r= -0.22 - 6 of 8 hypotheses met			?
Summary result (overall rating):				?	29		+6			?

Table 33 - Measurement properties 2-6 for OMAS

*Key: Meth Qual=Methodological Quality; n=Number of participants. Score: V= very good; A = adequate; D = doubtful; I = inadequate; N/A= not applicable; ? = indeterminate; + = sufficient; - = insufficient; r = correlation results; * = reviewed by both reviewers and results are a consensus.*

PROM and article	Country/language in which PROM evaluated	Structural validity			Internal consistency			Cross-cultural validity/measurement invariance			Reliability		
		n	Meth Qual	Result (rating)	n	Meth Qual	Result (rating)	n	Meth Qual	Result (rating)	n	Meth Qual	Result (rating)
OMAS, Buker et al. (2017) *	Turkey/Turkish			?	91	D	Cronbach's α= 0.76 +			?	91	I	ICC 0.98 +
OMAS, Garratt et al. (2018)	Norway/Norwegian	567	D	CFI 0.99 and TLI 0.98 + RMSEA 0.087 -	567	D	Cronbach's α= 0.82 +			?	182 (For test-retest)	D	ICC 0.92 +
OMAS, Turhan et al. (2017) *	Turkey/Turkish			?	100	D	Cronbach's α= 0.84 +			?	100	D	ICC 0.98 +

Summary result (overall rating):	567		+	667		Cronbach's α= 0.76-0.84 +			?			ICC 0.92-0.98 +

Table 34 - Measurement properties 7-10 for OMAS

*Key: Meth Qual=Methodological Quality; n=Number of participants. Score: V= very good; A = adequate; D = doubtful; I = inadequate; N/A= not applicable; ? = indeterminate; + = sufficient; - = insufficient; r = correlation results; * = reviewed by both reviewers and results are a consensus.*

PROM and article	Country/language in which PROM evaluated	Measurement Error			Hypotheses testing for Construct Validity			Responsiveness		
		n	Meth Qual	Result (rating)	n	Meth Qual	Result (rating)	n	Meth Qual	Result (rating)
OMAS, Buker et al. (2017) *	Turkey/Turkish			?	91	I	OMAS and FAOS-P r= 0.788 + OMAS and FAOS-S r= 0.753 + OMAS and FAOS-ADL r= 0.798 + OMAS and FAOS-Sport r= 0.809 + OMAS and FAOS-QoL r= 0.772 + OMAS and GSRF r= -0.794 + 6 of 6 hypotheses met +			?

OMAS, Garratt et al. (2018)	Norway/Norwegian	567	D	MIC not defined ?	567	A	OMAS and LEFS r= 0.86 + OMAS and SEFAS r= -0.88 + OMAS and SF-36-PCS r= 0.77 + OMAS and EQ-5D r= 0.79 + OMAS and EQ-5D-M r= -0.67 + OMAS and EQ-5D-SC r= -0.33 + OMAS and EQ-5D-UA r= -0.60 + OMAS and EQ-5D-P&D r= -0.73 + OMAS and EQ-5D-A&D r= -0.31 + 9 of 9 hypotheses met +			?
OMAS, Turhan et al. (2017) *	Turkey/Turkish				100	D	OMAS and FAAM-ADL 0.86 + OMAS and FAAM-S 0.83 + OMAS and SF-12-PCS 0.72 + OMAS and SF-12-MCS 0.60 + 4 of 4 hypotheses met +			
Summary result (overall rating):		567		?	678		+19			?

3.3.6 Summary of findings tables

Table 35 - Summary of findings table - structural validity

Structural validity	Summary or pooled result	Overall rating	Quality of evidence
AAOS	No evidence available	Indeterminate (?)	No evidence available
A-FORM	No evidence available	Indeterminate (?)	No evidence available
FFI	No evidence available	Indeterminate (?)	No evidence available
OMAS	CFI 0.99 and TLI 0.98 and RMSEA 0.087	Sufficient (+)	Low: one study of doubtful quality available (Norwegian version)

Table 36 - Summary of findings table - internal consistency

Internal consistency	Summary or pooled result	Overall rating	Quality of evidence
AAOS	No evidence available	Indeterminate (?)	No evidence available
A-FORM	No evidence available	Indeterminate (?)	No evidence available
FFI	No evidence available	Indeterminate (?)	No evidence available
OMAS	Cronbach's α= 0.76 - 0.84; total sample 667	Sufficient (+)	Moderate: multiple studies of doubtful quality (Turkish and Norwegian versions)

Table 37 - Summary of findings table - cross-cultural validity

Reliability	Summary or pooled result	Overall rating	Quality of evidence
AAOS	No evidence available	Indeterminate (?)	No evidence available
A-FORM	No evidence available	Indeterminate (?)	No evidence available
FFI	No evidence available	Indeterminate (?)	No evidence available
OMAS	No evidence available	Indeterminate (?)	No evidence available

Table 38 -Summary of findings table - reliability

Reliability	Summary or pooled result	Overall rating	Quality of evidence
AAOS	ICC or weighted Kappa not reported; sample 63	Indeterminate (?)	Very low: one study of inadequate quality and total sample <100 (Mexican-Spanish version)
A-FORM	No evidence available	Indeterminate (?)	No evidence available
FFI	No evidence available	Indeterminate (?)	No evidence available
OMAS	ICC 0.92-0.98; total sample 373	Sufficient (+)	Moderate: multiple studies of doubtful quality

Table 39 - Summary of findings table - measurement error

Measurement Error	Summary or pooled result	Overall rating	Quality of evidence
AAOS	No evidence available	Indeterminate (?)	No evidence available
A-FORM	No evidence available	Indeterminate (?)	No evidence available
FFI	No evidence available	Indeterminate (?)	No evidence available
OMAS	MIC not defined	Indeterminate (?)	No evidence available

Table 40 - Summary of findings table - hypotheses testing for construct validity

Hypotheses Testing	Summary or Pooled Result	Overall rating	Quality of Evidence
AAOS	4 out of 4 hypotheses met for convergent validity; sample 63	Sufficient (+)	Low: one study of doubtful quality (Spanish Version)
A-FORM	No evidence available	Indeterminate (?)	No evidence available
FFI	6 of 8 hypotheses met	Sufficient (+)	Very low: one study of doubtful quality and total sample <100 (Taiwan-Chinese version)
OMAS	19 correlations met hypotheses (out of 19); sample 678	Sufficient (+)	Moderate: one study of adequate quality available (Turkish and Norwegian versions)

Table 41 - Summary table - quality of evidence for measurement properties of included PROMs

PROM	AAOS		A-FORM		FFI		OMAS	
	Overall Rating	Quality of Evidence	Overall Rating	Quality of Evidence	Overall Rating	Quality of Evidence	Overall Rating	Quality of Evidence
	+ / - / ?	High, moderate, low, very low	+ / - / ?	High, moderate, low, very low	+ / - / ?	High, moderate, low, very low	+ / - / ?	High, moderate, low, very low
Content validity	?	N/A	?	N/A	?	N/A	?	N/A
Relevance	?	N/A	?	N/A	?	N/A	?	N/A
Comprehensiveness	?	N/A	?	N/A	?	N/A	?	N/A
Comprehensibility	?	N/A	?	N/A	?	N/A	?	N/A
Structural validity	?	N/A	?	N/A	?	N/A	+	Low
Internal consistency	?	N/A	?	N/A	?	N/A	+	Moderate
Cross-cultural validity	?	N/A	?	N/A	?	N/A	?	N/A

Measurement invariance	?	N/A	?	N/A	?	N/A	?	N/A
Reliability	?	Very Low	?	N/A	?	N/A	+	Low
Measurement Error	?	N/A	?	N/A	?	N/A	?	N/A
Criterion validity	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Construct validity	+	Low	?	N/A	+	Very low	+	Moderate
Responsiveness	?	N/A	?	N/A	?	N/A	?	N/A

Table 35 to 40 shows the summary of findings tables for each measurement property and table 41 shows a summary of the overall results.

The AAOS shows some evidence for sufficient construct validity in the patient population completed in one study in the Mexican-Spanish version. The PROM met four hypotheses with regard to correlations with other instruments. Here the authors correlated the sub-scales of the AAOS (CS and SCS) to the sub-scales of the SF-36 (PCS and MCS). This evidence scored very low quality on the GRADE assessment due to only having one study of doubtful quality available and also a total sample size of less than 100. There was also evidence for reliability but this was indeterminate as the appropriate outcomes were not presented by the authors of the paper.

The A-FORM only had evidence relating to the development of the PROM and therefore the boxes for other measurement properties were not completed.

The FFI showed some evidence for sufficient construct validity and this was achieved through correlations of the FFI subscales (AL, D and P) with subscales of the SF-36 (PCS and MCS) in one article in the Taiwan-Chinese version of the questionnaire. Six of the eight correlations performed met he hypotheses set a priori, however this evidence was scored very low quality on GRADE assessment because there was only one study of doubtful quality and the sample size was below 100 in total.

The OMAS had the most evidence relating to its measurement properties, with four studies included in this review assessing this PROM. One article outlined the development of the PROM only. The OMAS showed evidence for sufficient structural validity in a study completed assessing the Norwegian version of the questionnaire. The authors of this study calculated the Tucker-Lewis index (TLI) and comparative fit index (CFI) of OMAS, as well as the root mean square error of approximation (RMSEA). The scores performed sufficiently for the TLI and CFI (0.98 and 0.99 respectively) but not on the RMSEA (0.087). COSMIN guidance advises that either TLI or CFA or RMSEA could be within the acceptable parameters to score sufficiently for structural validity, therefore this has gained a score of sufficient for this measurement property. Overall the evidence for structural validity is low as there is only one study of doubtful quality available.

OMAS also showed evidence for sufficient internal consistency, with three articles assessing this in both Turkish and Norwegian versions of the questionnaire. The

evidence for internal consistency of OMAS was moderate because there were multiple studies of doubtful quality available. OMAS demonstrated sufficient levels of reliability in the articles, with ICC reported of 0.92-0.98. The evidence overall scored low on quality in the GRADE assessment as there are multiple studies of doubtful quality. Finally, several articles assessed the construct validity of OMAS and this showed sufficient construct validity against the hypotheses set a priori. There were 19 hypotheses tested all of which were met, all through convergent validity by assessing the correlations of OMAS against other PROMs and subscales of PROMs. The evidence for this measurement property presented here was moderate in quality as there was one study of adequate risk of bias and several of doubtful.

3.3.7 Interpretability and feasibility

Interpretability of PROMs refers to the ability to attribute qualitative significance to a PROMs scores or changes in scores [97]. Data extraction for interpretability items of the included PROMs are found in table 42. There was little information documented in the included articles regarding the interpretability of the PROMs. Most articles did provide the distribution of scores using means and standard deviations. The study by McPhail, et al. [126] was testing the non-final version of the PROM, so the scores are inflated here as 53 items were being tested to refine the final 15-item questionnaire.

Turhan, et al. [170] evaluated the floor and ceiling effects for the Turkish version of the OMAS. They concluded there were no floor effects in the measure, however found that 27-29% of responses were on the maximum amount. The authors of this study defined ceiling effects to be present at ≥30% and therefore this was not deemed to be a ceiling effect by the authors. Wu, et al. [165] also considered the floor and ceiling effects of the items in the FFI, concluding that 17 out of the 23 items had ≥15% of respondents giving the lowest possible score and two items which ≥15% of respondents giving the highest possible score. The floor and ceiling effects for the A-FORM and AAOS were not evaluated in any articles in this patient population.

The levels of missing data were reported in a few instances. Most articles reported missing total data (i.e. whole questionnaires missing), with one article reported missing individual items. Garratt, et al. [169] showed in their study that 17.3% of people missed at least one item on the OMAS. The most commonly missed item in this score was jumping (6.2% missing) followed by pain (4.9% missing) and then

squatting (4.6% missing). There was no information given for any of the PROMs here on the minimally clinically important differences or information on response shift.

Table 42 - Interpretability of included PROMS

Article and PROM	Distribution of scores in study group	Percentage of missing items and percentage of missing total scores	Floor and Ceiling Effects	Minimal Important Change (MIC) or Minimal Important Difference (MID)	Information on Response Shift
Zelle et al. (2017) AAOS	Normal distribution following Shapiro-Wilks test. No mean or SD provided.	Missing total scores 83 of 100 enrolled in first test and 63 of 100 in re-test. No data on items missing.	Not reported	Not reported	Not reported
McPhail et al. (2014) A-FORM	199.1 (±71.9) – (Not completed using final summary score system but in extended 53-item score)	Not specified – items with more than 10% missing were excluded from final version of PROM (item reduction)	Not reported	Not reported	Not reported
Wu et al. (2008) FFI	37.7±22.2	Missing items not specified	Provided by item. Many more	Not reported	Not reported

			floor scores than ceiling.		
Olerud and Molander (1984) OMAS	Not reported	Not reported	Not reported	Not reported	Not reported
Garratt et al. (2018) OMAS	75.62 (±24.07)	17.3% of respondents missed at least one item and 1.6% had missing total scores. “Jumping” most commonly missed item (6.2%).	Mentioned but data not reported	Not reported	Not reported
Buker et al. (2017) OMAS	72.58 (±23.27)	Not reported	Not reported	Not reported	Not reported
Turhan et al. (2017) OMAS	74.1 (±23.7)	Not reported	Floor - 0% Ceiling - 27-29%	Not reported	Not reported

Table 43 - Feasibility of PROMs - AAOS, A-FORM, FAAM and FAOS

Feasibility Aspects	AAOS	A-FORM	FAAM	FAOS
Patient's comprehensibility	Not reported	Final version refined in Delphi panel comprising of patients and clinicians and researchers stated that all respondents to the questionnaire in the cohort phase were able to self-complete without help	Field testing completed with patients (not solely ankle fracture patients) and clinicians to assess user friendly-ness in these groups. Included assessment of usability and how clear instructions were [128]	Not reported
Clinician's comprehensibility	Not reported	Final version refined in Delphi panel comprising of patients and clinicians	Field testing completed with patients and clinicians to assess user friendly-ness in these groups. Included assessment of usability and how clear instructions were [128]	Not reported

Type and ease of administration	Paper based score in clinic or via post – easy to administer	Paper based score in clinic or via post – easy to administer	Paper based score in clinic or via post – easy to administer	Paper based score in clinic or via post – easy to administer
Length of instrument	25 questions	15 questions	29 questions	42 questions
Completion time	Approximately 10 minutes	Approximately 7 minutes	Approximately 10 minutes	Approximately 15 minutes
Patient's required mental and physical ability level	Must be able to read and comprehend questions Physical N/A	Must be able to read and comprehend questions Physical N/A	Must be able to read and comprehend questions Physical N/A	Must be able to read and comprehend questions Physical N/A
Copyright	Available for use without copyright restrictions	Available for use without copyright restrictions	Available for use without copyright restrictions	Available for use without copyright restrictions
Cost of an instrument	No cost	No cost	No cost	No cost
Required equipment	No equipment required except pen and paper	No equipment required except pen and paper	No equipment required except pen and paper	No equipment required except pen and paper

Availability in different settings	Clinical setting or remotely via post	Clinical setting or remotely via post	Clinical setting or remotely via post	Clinical setting or remotely via post
Regulatory agency's requirement for approval	No approval required	Permission required for use by developers but easily obtained when stating nature of use	No approval required	No approval required

Table 44 - Feasibility of PROMS - FFI, KS, MOXFQ and OMAS

Feasibility Aspects	FFI	KS	MOXFQ	OMAS
Patient's comprehensibility	Not reported	Not reported	Not reported	Not reported
Clinician's comprehensibility	Not reported	Not reported	Not reported	Not reported
Type and ease of administration	Paper based score in clinic or via post – easy to administer	Paper based score in clinic or via post – easy to administer	Paper based score in clinic or via post – easy to administer	Paper based score in clinic or via post – easy to administer
Length of instrument	23 questions	8 questions	16 questions	9 questions
Completion time	Approximately 10 minutes	Approximately 5 minutes	Approximately 10 minutes	Approximately 5 minutes
Patient's required mental and physical ability level	Must be able to read and comprehend questions Physical N/A	Must be able to read and comprehend questions Physical N/A	Must be able to read and comprehend questions Physical N/A	Must be able to read and comprehend questions Physical N/A

Copyright	Available for use without copyright restrictions	Available for use without copyright restrictions	Available for use without copyright restrictions	Available for use without copyright restrictions
Cost of an instrument	No cost	No cost	Cost possibly incurred depending on whether academic or commercial study, whether translation required or how many times it will be used in a study.	No cost
Required equipment	No equipment required except pen and paper	No equipment required except pen and paper	No equipment required except pen and paper	No equipment required except pen and paper
Availability in different settings	Clinical setting or remotely via post	Clinical setting or remotely via post	Clinical setting or remotely via post	Clinical setting or remotely via post
Regulatory agency's requirement for approval	No approval required	No approval required	Approval required via a licence application online	No approval required

COSMIN methodology recommends performing an assessment of feasibility of included PROMs, which is concerned with the ease of use of a measure in the intended context given resource limitations such as money or time [93]. Data extraction on the feasibility items for the included PROMs can be found in table 43 and 44. Comprehensibility was not formally tested and assessed for the majority of the instruments. The A-FORM was assessed using a Delphi panel comprising of both patients with ankle fracture and clinicians in its initial development phase, explaining how panel members commented on the wording and suitability of items, with the option of suggesting modifications if necessary.

On face evaluation of the PROMs, they are all relatively easy to administer, either in person in a clinical setting, or remotely via post or online. Furthermore, they require no specialist equipment to administer, which is advantageous for use in clinical trials comparing interventions in this patient population. The only downside would be that the questionnaires all require a level of reading ability and capacity to recall and respond to specific questions within the questionnaires. This, however, is a common feature of all PROMs, by the fact they should be reported by the patient, which can be difficult or impossible for certain patient groups, such as those who lack capacity. Additionally, there is no evidence of development of methods or validation of these instruments to be used by proxy i.e. a relative or carer providing answers on behalf of someone else. Furthermore, most of the PROMs (with exception of the MOXFQ when used for commercial research), are free to use without copyright, meaning this can be administered in situations which may not have the resources to pay for instrument licences.

3.3.8 Recommendations

The COSMIN methodology advises the use of three categories, which are shown in table 45, to formulate recommendations following the systematic review of evidence for measurement properties of PROMs.

Table 45 - Categories for PROM recommendation

Category	Description
A	PROMs with evidence for sufficient content validity (any level) and at least low quality evidence for sufficient internal consistency
B	PROMs not categorised in A or C
C	PROMs with high quality evidence for an insufficient measurement property

Below are categorisations of the PROMs as a result of this review, along with justifications for the categorisations.

Table 46 - Categorisations of PROMS in this review

PROM	Category	Reason
AAOS	B	No evidence available for content validity of this PROM and no evidence for internal consistency. No evidence for insufficient measurement properties.
A-FORM	B	Low evidence for content validity and no evidence for internal consistency. No evidence for insufficient measurement properties.
FFI	B	No evidence for content validity and no evidence for internal consistency in this patient population. No evidence for insufficient measurement properties.
OMAS	B	No evidence available for content validity of this PROM. Moderate evidence for sufficient internal consistency. No evidence for insufficient measurement properties.
MOXFQ	B	No evidence available for quality of measurement properties of this PROM in this patient population.
FAAM	B	No evidence available for quality of measurement properties of this PROM in this patient population.
KS	B	No evidence available for quality of measurement properties of this PROM in this patient population.
FAOS	B	No evidence available for quality of measurement properties of this PROM in this patient population.

Table 46 shows the categorisations of the PROMs in relation to the levels of evidence identified in this review. All PROMs included in this review were categorised as B, meaning they do not have satisfactory levels of appropriate quality evidence for sufficient measurement properties of content validity and internal consistency to for recommendation in measuring the construct of outcome in adults with ankle fractures. Conversely, there is no evidence that any of the PROMs display insufficient measurement properties in this patient population and therefore no recommendations can be formulated from these results.

3.4 Discussion

This review demonstrates that none of the ankle specific PROMs included in this review can be recommended for use in adults recovering from an ankle fracture. There is, however, no evidence the included PROMs demonstrate insufficient measurement properties in this patient population, therefore further validation work is recommended to fully understand their performance in this particular context of use.

The OMAS demonstrates low quality evidence for sufficient structural validity and reliability and moderate quality evidence for sufficient construct validity and internal consistency. Without any understanding of the scores content validity, this measure cannot be recommended for use in this patient population. The OMAS scored poorly for the PROM development, likely because of the age of this article which was published many years prior to specific guidance on the appropriate development of PROMs was published. Considering the relative frequency of use of this PROM and the lack of patient involvement in its development, the appropriateness of the questionnaire layout and content requires further investigation. The priority areas of research in the measurement properties of these PROMs explained above is a priority in the OMAS due to its popularity of use in trials for this patient population.

The authors of one study of the OMAS indicated several high percentages of missing items in OMAS, reporting that 17.3% of respondents missed at least one item [169]. This indicates that the relevance and comprehension of the content of this questionnaire would benefit from further exploration, as these maybe reasons why people are missing so many questions. This further supports the exploration of the relevance and appropriateness of the OMAS questionnaire for patients with an ankle fracture.

The previous chapter within this thesis demonstrated that OMAS continues to be the most widely utilised PROM in this specific research area, however this review demonstrates the distinct gaps in evidence for its measurement properties, particularly with regard to its content validity. Furthermore, the majority of this validation work has been completed in translated versions of the OMAS questionnaire, and there is no evidence of validation of this measure in an English version, a version used in many UK-based clinical trials for ankle fracture [41 72 172]. Understanding the relevance and comprehensiveness of the content of the PROM in patients with an ankle fracture would help formulate recommendations on whether this is an appropriate outcome measure to be using in this population.

A-FORM scored more favourably than OMAS on the PROM development process, although still scored doubtful due to the lack of information on the background of the group moderators. The authors who developed the A-FORM also utilised a conceptual framework from qualitative interviews reported and described in a separate article [173]. The experience of rating articles against the risk of bias checklist shows that it's very difficult for articles to perform well against this rather strict criteria. Furthermore, the "lowest score counts" approach means that the overall bottom line result of the checklist assessment sometimes feels unrepresentative of the overall quality of the article.

There was limited evidence for the remaining measurement properties of the A-FORM, FFI and AAOS, with the FFI demonstrating very low quality evidence for sufficient construct validity. There was no evidence in the literature of the A-FORM being tested in its final version, therefore no remaining measurement properties were assessed for this measure. The AAOS demonstrated low quality evidence for sufficient construct validity and very low quality evidence for sufficient reliability of the measure in the patient population.

The PROMs included here had no information reported on the minimally important change or response shift. Floor and ceiling effects were assessed for the OMAS, which reported 0% of respondents achieving the worst score but 27-29% of participants achieving the best score, indicating possible ceiling effects. As discussed previously, the authors did not regard this as a ceiling effect as their acceptable parameter was set at 30% [170]. COSMIN recommend the cut off criteria for floor and ceiling effects as ≥15%, which is also the criteria used by authors of similar studies

[109 174-177]. This indicates that ceiling effects may well be present in scores of OMAS and this warrants further evaluation.

When comparing the results of this review to those found in a similar article by Ng, et al. [140], there are some differences to note. The main difference is that this research team recommended that A-FORM was the most appropriate PROM to use in this patient population. Whilst the results of this review agrees with that of Ng, et al. [140] that A-FORM has the most favourable development process, the PROM did not meet criteria for recommendation in this review because the measurement properties, such as internal consistency, were not assessed in the final 15-item version of the questionnaire. Future research should address the content validity and the internal consistency of the final version of the questionnaire as a priority in order to make recommendations on the suitability of this PROM for use.

Sierevelt, et al. [129] completed a similar review appraising the evidence for PROMs used in foot and ankle conditions, subsequently reporting the measurement properties of the three most common which were FFI, FAOS and FAAM. Here, the construct of interest and patient population was broader therefore they included and analysed a higher volume of studies. Authors found that 21 different foot and ankle specific outcome measures were used in the foot and ankle literature. The studies included in the review were found to be low in methodological quality, however recommend the use of the FAOS and FAAM in the conclusion, although stated that researchers should acknowledge the limitations of these outcome measures when using. The authors of this review also noted the lack of cross-cultural validation in all of the studies which translated PROMs, which was also found in the studies included here. They also noted a significant lack of content validity assessment in the outcome measures, which is similar to results found here.

Another similar and broad review completed by Jia, et al. [150] assessed the measurement properties of all PROMs used for patient with foot and ankle disorders. The authors of this review stated that whilst the majority of PROMs used in this patient population do not have sufficient evidence for their measurement properties, they concluded that the MOXFQ scored the most favourably in this broad context. The authors also noted, however, that none of the PROMs performed favourably in all measurement properties and there were gaps in the validation evidence base for all included measures, which is similar to results presented here.

A strength of this review is the use a structured approach to identify, review and appraise evidence using well developed and recognised methodology [109]. The use of a second reviewer in both the application of the eligibility criteria and also the risk of bias review and data extraction for the studies meant that bias was reduced as a far as possible and strengthens the methods of this research. The use of a pre-specified list of PROMs which are representative of the actual practice in clinical trials for this patient population enables these results to be of high relevance for clinicians and academics preparing protocols for trials of this nature.

A limitation of this project as previously discussed is the nature of the construct. The construct of outcome following ankle fracture recovery is difficult to wholly define it's possible that this concept varies between different groups of individuals, for example in relation to a person's age, operative management and pre-injury function levels. As with generic health related quality of life measures, the construct is multi-dimensional and difficult to define[83]. McPhail, et al. [173] has contributed to this research area through the content foundation work for the A-FORM outcome measure, through qualitative interviews with ten patients with ankle fracture and six clinicians who are involved in the treatment of ankle fracture.

Another limitation of this review is that articles included used combinations of differently managed patients in terms of operative intervention. For example, three of the articles included in this review assessed a cohort of only operatively managed ankle fractures and three assessed a cohort of a mixed sample of operatively and non-operatively managed patients (with one article not specifying). Whilst there is no clear evidence that the construct of outcome in ankle fracture should vary considerably depending on operative management, this cannot be confirmed and further research to assess the exact nature of the construct of interest in patients treated operatively and non-operatively is warranted.

Several research teams translated the PROMs here into different languages, a positive contribution to the research base to enable the use of PROMs in different contexts and countries. The performance of a PROM can differ depending on the social and cultural contexts in which it's completed, which have not been formally assessed through measurement invariance testing in these articles. I have pooled the results here for the purposes of this review as there were so few in each language version, however the possible cultural differences and their effects on the

measurement properties should be considered when interpreting the results of this review.

A further limitation when assessing the articles was the difference in date of publication, a clear outlier being the OMAS development paper [123]. During the years of publication of this paper the knowledge base and understanding of PROMs and their development was not yet in existence, making it difficult to assess the development of this PROM against the criteria outlined by specialist groups like COSMIN. The developers of the A-FORM outcome measure had the use of the frameworks set out by the COSMIN groups and also a growing evidence base for other PROM development studies, which were not available when the OMAS was developed. However, this does not make up for the lack of content foundation of this questionnaire. Further research is warranted to ascertain the content validity of this PROM and further explore the measurement properties of this score. Furthermore, an exploration of the comprehension, acceptability, and relevance of this score to patients would be beneficial, especially considering the high rates of missing items found [169]. Whist the A-FORM exhibits the most appropriate development process, there is more research required into the final 15-item version of this questionnaire to determine the sufficiency of the measurement properties.

Further research should focus on the content validity of the measures, of which there is no evidence for any of the PROMs included here. The internal consistency and structural validity of the PROMs are also areas of priority, to make initial assessments based upon the COSMIN criteria of recommendations outlined above. The responsiveness of the PROMs in clinical trials is of high importance to ensure that any difference in treatment effects between the interventions are able to be effectively captured by the PROM being used. Furthermore, and understanding of the interpretability of the scores, such as minimally important changes in the target population would be beneficial to ensure that samples size calculations are accurate.

In addition to more work to evaluate the measurement properties of PROMs, further research is warranted to understand exactly what constitutes a satisfactory outcome for patients recovering from this injury. It may be that variables such as age and fracture management have an impact on the construct being measured. For example, as described in the introduction chapter of this thesis, ankle fractures display a bimodal distribution, affecting younger males and older females [17]. The construct of interest for high energy fracture occurring in a young male maybe significantly

different to the construct of interest for an elderly woman who sustained a fracture following a low energy fall. Different PROMs or versions of these PROMs may be necessary for different subgroups of patients with ankle fracture for this reason. Many RCTs in this area have treated these two groups as the same injury [72 79] but some have focussed on one subgroup depending on age [41 178-180]. This indicates that fractures in a younger population compared to an older population may be separate injuries and therefore interventions for these injuries should be evaluated as such.

For the reasons outlined above, further exploratory work into the patient experience of ankle fracture and the factors of most importance to them is needed to gain better understanding of what constitutes a satisfactory outcome in this population. This will enable a better understanding of the construct of interest. As previously discussed, there is a distinct gap in the knowledge regarding the content validity of OMAS and no evidence to suggest that patients were involved in the development of the score. The next chapters of this thesis will use qualitative research methods to explore the patient experience of ankle fracture. The relevance, comprehensiveness and overall acceptability of OMAS will also be evaluated.

In this chapter I used systematic review methodology to identify and critically appraise the evidence for the measurement properties of the ankle specific PROMs used in clinical trials for adults with an ankle fracture. The outcome of this review was that none of the PROMs had sufficient evidence for their validity and reliability to be recommended for use under COSMIN criteria. This review has identified a gap in the evidence surrounding the specific construct of interest that PROMs in this patient population measure. Furthermore, it has identified that there is a significant evidence gap in the understanding of measurement properties of PROMs which are used in this patient population, particularly an issue in those which are very commonly used such as the OMAS. The next two chapters will outline the qualitative work completed to explore the patient experience of ankle fracture, the factors of most importance in individuals recovery and also the relevance and comprehensiveness of the items in the OMAS.

4. The patient experiences of ankle fracture recovery and the acceptability of the Olerud Molander Ankle Score: a two part qualitative study.

Aim: This chapter will outline the methods for the two part qualitative chapter, as well as the results for the initial part. The initial aim of this qualitative project is to explore the patient experience of ankle fracture recovery and the factors of most importance to them in their recovery.

Methods: Semi-structured interviews were completed with a sample of individuals taking part in the AIR trial. Purposive sampling was used to ensure a diverse range of participants were recruited in terms of age, gender, fracture management and treatment allocation. Interviews were audio-recorded using a secure encrypted audio-recorder and transcribed verbatim. The transcripts were anonymised and analysed using reflexive thematic analysis and a sample of the transcripts were coded independently by a second researcher for purposes of consistency. NVivo software was used to analyse the data.

Results: Eight themes were derived from the data for the initial part of the interviews regarding patient experience. These were mobility, loss of independence, sleep issues, family and social life, ankle symptoms, psychological effects, usual activities and experience of healthcare. Older individuals found mobility with weight-bearing restrictions and loss of independence more difficult than younger participants included in the study. Factors of most importance were regaining independence, sleep improvements, ability to fully weight bear or walk normally and return to driving.

Conclusion: Ankle fractures have a wide ranging impact on individuals' lives and some issues were more difficult for older participants of the study compared to younger individuals. These impacts extend beyond the clinical symptoms of pain and swelling to include psychological effects, loss of sense of independence, sleep disturbance and impacts on family and social roles.

This work has been published in a peer reviewed journal:

McKeown, R., Kearney, R.S., Liew, Z.H. and Ellard, D.R. (2020) Patient experiences of an ankle fracture and the most important factors in their recovery: a qualitative interview study. *BMJ Open*. 10: 2.

This work has been presented at a national conference:

Podium presentation

Title: The patient experience of ankle fracture: a qualitative study **Presenter:** Rebecca McKeown **Event:** British Society of Rehabilitation Medicine and Society for Research in Rehabilitation Joint Annual Scientific Meeting, 14/10/2019, Coventry, UK.

4.1 Introduction

The initial work package completed in this thesis examined the current and previous trends of outcome measurement usage in clinical trials for ankle fracture intervention. After discovering that PROMs were the most commonly used type of outcome measure, I completed a further review and critical appraisal of the evidence for the ankle specific PROMs used in these trials. The results of this second systematic review demonstrated that the PROMs used in these types of trials do not have sufficient evidence to recommend their use. The review also highlighted an evidence gap of the experiences of individuals who sustain a fracture to their ankle to enable an understanding of the specific construct of interest which these PROMs are aiming to measure.

It is possible that some patient variables may have an effect on the ways in which people experience their recovery. A study assessing the gender differences of individuals who are in recovery from a hip fracture showed differences in the recovery of males and females in the sample, with functional recovery better in males. The female group reported higher levels of pain and a higher incidence of symptoms of low mood than those in the male group [181]. A meta-analysis of predictors of satisfaction with healthcare showed that socioeconomic factors of age and lower education were associated with differing levels of patient satisfaction with care [182]. Baumhauer, et al. [183] completed a study to determine the differences in importance of several factors for individuals recovering from surgery of the foot and ankle. The authors found that there were significant differences in the items of key importance between people of different genders and ages. Furthermore, there is evidence to suggest that there are age and gender differences in the biomechanical functioning of the ankle, such as range of movement [3]. This demonstrates that more than one construct may exist within the broader population of adults with an ankle fracture, especially when considering the bimodal pattern of injury that this fracture demonstrates [184]

There may be differences in the way in which individuals experience their injury depending on whether they have operative intervention or not, especially considering the risks associated with surgery and the subsequent soft tissue and wound care that this intervention creates. Furthermore, the type of immobilisation method given to the individual may have an influence on their experience of recovery. For example, a study which assessed patient satisfaction and preference in children with forearm fractures

showed differences in patient satisfaction between different types of immobilisation methods [185]. For these reasons, it may be that the treatment of the fracture could also influence patients experience and reality of their injury recovery.

As outlined in chapter two, OMAS is the most frequently collected measure in this trial type. Throughout chapter three, I identified a lack of evidence that patients were involved during the development of OMAS. This chapter also showed, however, that there was some evidence of sufficient validity and internal consistency of OMAS [168-170].

This demonstrates that further work is required to understand whether the content of the questionnaire is relevant and comprehensive to patients with an ankle fracture to inform its suitability for continued use [186 187]. In order for a questionnaire to be content valid it must be relevant and comprehensive for individuals completing it with regards to the construct of interest [166]. Relevance is concerned with the items within the questionnaire being important to the specific population of individuals being assessed. Comprehensiveness is concerned with the coverage of the questionnaire items, thus whether the questionnaire omits any aspects which are of importance to the target population [166].

This research project will address two separate aims using a single method of data collection. The methods for the two projects and results of the first part of this qualitative project will be reported in this chapter (chapter 4) and the results of the second part will be outlined in chapter 5, followed by a discussion of the results.

The aims, research questions and objectives for parts one and two of this project are outlined below:

4.1.1 Aim of part one

The aim of the initial phase of this qualitative research is to explore the patient experiences of ankle fracture recovery and the factors which are most important to them in their recovery.

4.1.2 Research questions – part one

1. What are the patient experiences of adults recovering from an ankle fracture?
2. What factors are of most importance to individuals' during their recovery from an ankle fracture?

4.1.3 Objectives of part one

1. To explore the patient experiences of individuals recovering from ankle fracture.
2. To understand the factors of most importance to individuals with an ankle fracture in their recovery.

4.1.4 Aim of part two

The aim of the second part of this research study is to explore the relevance, comprehensiveness and acceptability of the OMAS in a sample of individuals recovering from an ankle fracture.

4.1.5 Research questions – part two

1. What is the relevance of the items contained within the OMAS questionnaire for adults with ankle fracture?
2. What is the comprehensiveness of the OMAS questionnaire for adults with ankle fracture?
3. What is the overall acceptability of the OMAS questionnaire for adults with an ankle fracture?

4.1.6 Objectives of part two

1. To explore the relevance of the items included in the OMAS questionnaire for individuals recovering from an ankle fracture.
2. To explore the comprehensiveness of the OMAS questionnaire for individuals recovering from an ankle fracture.
3. To explore the acceptability of the OMAS questionnaire for individuals recovering from an ankle fracture.

Due to the exploratory nature of these research aims and questions, concerned with individuals experiences, beliefs and social perspectives of participants, I have chosen to employ qualitative research methods to answer these [188-192]. I used the Standards for Reporting Qualitative Research Checklist (SRQR) [193] to report these research findings.

4.2 Methods

4.2.1 Ethical approval

I sought and gained ethical approval for this study from West Midlands Edgbaston NHS Research Ethics Committee through substantial amendment to the AIR Trial Protocol on 07/09/2018 (REC Reference 17/WM/0239) [72].

4.2.3 Participants and eligibility criteria

This project used a stratified random sample of participants taking part in the AIR trial. I used a purposive sampling technique to ensure a representative samples with regards to four identified variables. These variables were age, gender, fracture management and allocated intervention in the trial. The possible fracture managements were either operative or non-operative management and the interventions allocated within the trial were functional brace or plaster cast. These variables were chosen due to the evidence highlighted in the introduction (section 4.1) which outlined the potentially differing experiences of people who received different interventions or those from different genders and age.

The data collection in this study was cross-sectional [194] due to constraints of completing the interviews within an ongoing clinical trial. The AIR TMG agreed that interviews prior to the primary end point may influence how participants complete the questionnaire, therefore a decision was made to interview individuals after the primary outcome time point of 16 weeks. This was to ensure that the process of completing the interview did not affect the answers they provided to the questionnaire at the primary outcome time point and therefore affect the outcome of the trial in any way. Therefore, in order to be eligible for this interview study, participants of AIR had to have passed the 16 week time point, even if they did not return their questionnaire at this particular time point. To minimise recall bias I decided to set a cut off for the maximum time since injury at 24 weeks, to allow for time for individuals return their 16 week questionnaire.

The eligibility criteria for the AIR trial can be found in section 1.7 of this thesis. Eligibility criteria for this embedded qualitative study are outlined in table 47.

Table 47 - Eligibility criteria for qualitative study

Inclusion criteria	Exclusion criteria
Participant of the AIR trial and had not been randomised in error.	Has withdrawn from the study prior to interview data collection.
Consented to being contacted about further research related to ankle fractures.	
Beyond the 16 week time point (primary outcome) of the clinical trial but not more than 24 weeks post randomisation.	

4.2.4 Recruitment and consent

I screened the AIR online database for participants who had indicated on their initial consent form that they were willing to be contacted about further research into ankle fractures. This involved finding participants on the database who were within the appropriate time frame for the study. Initially I invited a range of participants who fulfilled the criteria, but as the recruitment progressed, I selected participants based upon the key variable attributes of age, gender and fracture management, to ensure that the purposive sampling framework was fulfilled. I checked the online database in the appropriate sections whether they had consented to being contacted about further research, which ensured that I had permission to contact them regarding this additional piece of research. I also accessed the participant's paper case report forms locked in the study office to ensure that the individual patient had not withdrawn from the study, therefore withdrawn consent to be contacted about further research.

In the instance where an individual was eligible for the study, fulfilled the sampling criteria and had not withdrawn their consent for the study, I sent out an invite letter and patient information sheet in the post, which outlined the purpose of the research study. As detailed in the letter, I called the participant one week later to further explain the purpose of the research and answer any questions they had. If the individual agreed to participate, we arranged a mutually agreeable date and time to complete the interview. This telephone call was the first contact between the researcher and participant and no relationship had previously been established prior to this. This is important because, if a

social or professional relationship has already been established between a researcher and participant, this may cause the participant to feel additional pressure to consent to the study, rather than utilise their full autonomy throughout the consent process [192]. The patient invite letter and information sheet used in this study can be found in appendix 6.

At the scheduled interview, prior to proceeding to the written consent process, there was time for further questions which were answered. Providing the individuals was still in agreement to participate, the written consent process took place. A copy of the consent form used can be found in appendix 6. Following the data collection process, I sent a copy of the consent form to the participant in retrospect for their records. Original copies were stored in a locked cabinet, for which only authorised trial staff had access to, compliance with WCTU Standard Operating Procedures (SOPs) for data management and security.

4.2.5 Data collection

As described in section 4.1.6, qualitative research methodology was the most appropriate to use in this instance. Interviews were used to collect the data for this research project. Interviews were more appropriate than focus groups in this instance because the concept of an individuals' experiences of an injury, the factors of most importance to them and their opinions on the relevance of items in a questionnaire are likely to be highly personalised and may differ substantially between individuals. It can be difficult to explore complex topics in depth within focus groups, due to the group dynamic aspect of this data collection method [190]. Furthermore, it's possible that the group setting may mean that some opinions are not heard as much as others, potentially creating a distorted view of the true experience of the range of individuals in the group [195]. Some individuals may feel uncomfortable expressing certain opinions which might be seen to be culturally undesirable or too personal within a group setting, or may wish to express culturally desirable or expected views [195]. These issues can be mitigated through the use of one-on-one interviews.

Interviews can be described as either structured, semi-structured or unstructured, depending on how the conversation is directed by the interviewer [190]. A topic guide is usually used, which prompts the interviewer for topics to cover and questions to ask. Structured interviews follow a very detailed topic guide which use closed-ended

questions and interviewers do not readily encourage the conversation to stray from this, much like a survey in verbal format [196]. In contrast, unstructured interviews follow a less detailed topic guide, or may not use one at all. This allows the participant to direct the conversation as much as possible [197]. Semi-structured interviews lie in the middle of these two interview types; interviewers will use a topic guide with questions and prompts but these are a mixture of both questions and general topics. A semi-structured interview is performed so that information that might be outside of those listed in the topic guide can be explored as and when they naturally arise in the conversation, creating a balance between a structured and unstructured format [196].

Semi-structured interviews are the most appropriate type of interview to use here for several reasons. Structured interviews would firstly be too limited to explore a topic such as patient experience, which is often broad, diverse and complex [198-200]. Furthermore, it is unlikely that structured interviews, which tend to ask more closed ended questions, would result in the rich and detailed data required for the research question, aims and data analysis method [197]. Whilst there is limited research into the patient experiences of this injury, we do have some idea of how patients experience a fracture of the lower limb through clinical presentation of these patients, as well as some qualitative work completed in the area [173]. Semi-structured interviews allow for coverage of all relevant topics, whether included in the topic guide or whether discussed naturally in the interview by the participant. This approach will allow for the aims of the research to be achieved with flexibility for participants to discuss issues and experiences which are outside of the topic guide but are important and relevant to the individual.

I completed all interviews for this project following specific training for completing this research, which are outlined in the researcher training and development section at the beginning of this thesis. I did not inform the interviewees of my background as a physiotherapist and instead introduced myself as a researcher from the University of Warwick. This was to avoid any assumptions the participants may develop based upon my professional and clinical background, which has been shown to influence interview interactions [192 201]. The aim of this was to ensure that participants would feel comfortable discussing all aspects of their experiences of ankle fracture recovery, not just the physical or functional aspects of their recovery, which they might perceive me to be more interested in than the other aspects [190]. The participants may not want to

discuss particular issues which they encountered for fear of insulting me as a clinician. For example if they had issues with their rehabilitation, they might feel less able to discuss any negative feelings about the rehabilitation process if I introduced myself as an individual with a background in physiotherapy [201].

Interviews were completed either at participant's own homes or places of work. I chose these locations over the option to use rooms in local hospitals or GP surgeries to avoid the social context of these healthcare delivery locations and to maximise the chances of individuals feeling able to discuss the many possible aspects of their ankle fracture recovery [202]. Furthermore, the location of the participant's own homes was often more convenient for the participants, therefore lessening the burden of research participation for them. No one else was present at the interviews apart from myself and the participant.

A topic guide was used to provide direction to the interview discussion in a semi-structured way and this can be found in appendix 6. This was formulated using relevant literature in the area [186 196]. All patient facing material including the invite letter, patient information sheet, consent form and the topic guide were reviewed independently by two patient and public involvement representatives involved in the AIR trial committees, Richard Grant and Karen Keates (RG and KK). These documents are also found in appendix 6.

No repeat interviews were able to be completed due to resource limitations. I did not make notes during the interviews to ensure that I remained aware of the direction of conversation and non-verbal communication. I kept a reflexive diary which I completed after each interview and throughout analysis. This was completed to enable the practice of reflexivity throughout the data collection and analysis, ensuring that I continually examined my assumptions and beliefs [203-205]. Interviews lasted for approximately one hour to one hour and fifteen minutes. I did not return the transcripts to the participants for comment or correction due to resource limitations. I decided that data collection would be terminated once no new themes were being discussed by the participants in the data for two consecutive interviews. Therefore, I completed data analysis concurrently with the ongoing data collection. This meant that interviews could be terminated when no new themes were being established from the data and information redundancy was reached [206].

4.2.6 Data analysis

For this project I used reflexive thematic analysis to analyse the data underpinned by critical realism, as previously described in the introduction of this thesis. Reflexive thematic analysis is a systematic approach to organising information and gaining understanding of patterns of meaning within the data, by grouping them into themes with underlying codes, as described by Braun, et al. [207]. I have chosen this approach to data analysis because of its theoretical flexibility and accessibility [208]. It is described as flexible because one is able to fit the most appropriate theoretical approach for the research question being answered, rather than using a framework which is more rigid in theoretical approach. Critical realism is the theoretical approach which will underpin this method of data analysis, which is described in further detail in section 1.8.

I also chose this technique because I had no prior experience of conducting qualitative research and therefore it was important to choose an accessible method such as this [208]. As described by Braun, et al. [207] the success of TA is dependent upon continual researcher reflexivity to evaluate the choices and assumptions made when reading and analysing the data. Reflexivity is defined as "the continual evaluation of subjective responses, intersubjective dynamics and the research process itself" [205] (p. 532). I continually acknowledged and questioned my assumptions and beliefs throughout the data collection and analysis phases through the use of a reflexive diary [191].

An inductive approach was used whereby theory is generated from the data in a bottom-up approach, rather than theory guiding data collection and analysis. I chose this analytical approach because there is a lack of theory of the patient experiences of ankle fracture and view of the OMAS in the current literature [100 203].

I used the six step process of completing thematic analysis, developed by Braun, et al. [207], details of which can be found in table 48. Whilst this six-step process is described as linear, the process was more iterative which involved revisiting phases as and when necessary [209]. I completed the transcription of all interviews and doubled checked the recordings against the transcriptions for purposes of accuracy. Data analysis for this project was completed using NVivo software (version 12, QSR International).

Table 48 - Six phase approach to thematic analysis as described by Braun and Clarke [207]

Phase	Short description of phase
1. Familiarisation with the data	Immersing oneself in the data by reading transcripts and listening to audio-recordings, intimate familiarisation with the data, starting to critically review data and identify meaning from them.
2. Coding	Generating codes, providing a label for data which is possibly relevant to the enquiry, describing and summarising the content of data.
3. Generating initial themes	Grouping codes into themes by assembling the codes into patterns or commonalities, identify overlaps and differences between codes, exploring the relatedness between themes.
4. Reviewing themes	Constructed themes are reviewed in relation to entire data set, quality testing of analysis, exploring whether there is enough rich data for something to be a theme, identifying boundaries to themes, ensuring coherence within themes.
5. Defining and naming themes	Giving themes a clear but concise name to describe the concept, identifying subthemes.
6. Writing up	Produce a story around the data and themes constructed throughout the project, making critical arguments which answers the research question.

Data collection was terminated when information redundancy was reached and no further themes were able to be generated from the data [189 206]. I did not specify codes and themes a priori, instead they were derived from the data during the analysis process in an inductive approach to analysis [210]. A second researcher, Zi Heng Liew (ZHL) independently coded four of the ten transcripts to maximise consistency and agreement

in the coding process. This second researcher was selected because of his experience in conducting qualitative interviews with orthopaedic trauma patients who had sustained shoulder dislocations therefore he was best suited to act as a second independent researcher in this project. This also coincided with a period of time availability for this researcher to complete this additional workload and therefore was logistically possible for the timing of the analysis.

The same analysis method was used to answer the separate research questions of this project (outlined in section 4.1.2 - 4.1.6) , however the results will be presented in two separate sections. For purposes of demonstrating which questions were deemed relevant and not relevant, a table will be produced to show the frequencies of the results obtained in this section. As this is a qualitative sample of participants, this will only be used to visually display the results of this section in relation to each item of the OMAS questionnaire and no formal quantitative analysis of these results will be completed.

4.2.7 Data protection and risk management

I did not identify any possible harms or benefits to individuals participating in this study and the study was deemed low risk. All participants gave written informed consent prior to the commencement of any research activities. I complied with the requirements of the Data Protection Act 2018 and any other relevant legislation with regards to the processing of personal information. The WCTU SOPs regarding data management and security were followed at all times.

The data collected throughout this research project was done using encrypted files on a passcode protected audio recorder and transferred onto University of Warwick secure servers on password protected files accessible only to myself. The data collection of this project involved a researcher lone working in participant's homes or places of work and therefore I undertook training on lone working and completed a lone working risk assessment prior to commencing data collection. WCTU SOPs including those regarding risk assessment, lone working and managing difficult conversations were referred to and followed throughout this project.

4.3 Results

4.3.1 Participants

After eight interviews, no new themes were able to be constructed from the data. I continued data collection and after a further two interviews, no new themes were discussed by participants and therefore information redundancy was reached [206]. Therefore data collection was finished after ten interviews. I was confident that there was sufficient rich data which had been created from these interviews and there were several themes generated from this data.

During the recruitment phase, three participants declined to participate during the initial phone call. One declines because they were too busy with work commitments and two declines because they did not want to be interviewed. A further six individuals were sent invite letters and information sheets but were subsequently unable to be contacted by phone after three attempts. There were no incidents of participant withdrawal throughout the study. Figure 13 shows the recruitment flow chart, taken from the publication resulting from this project [211].

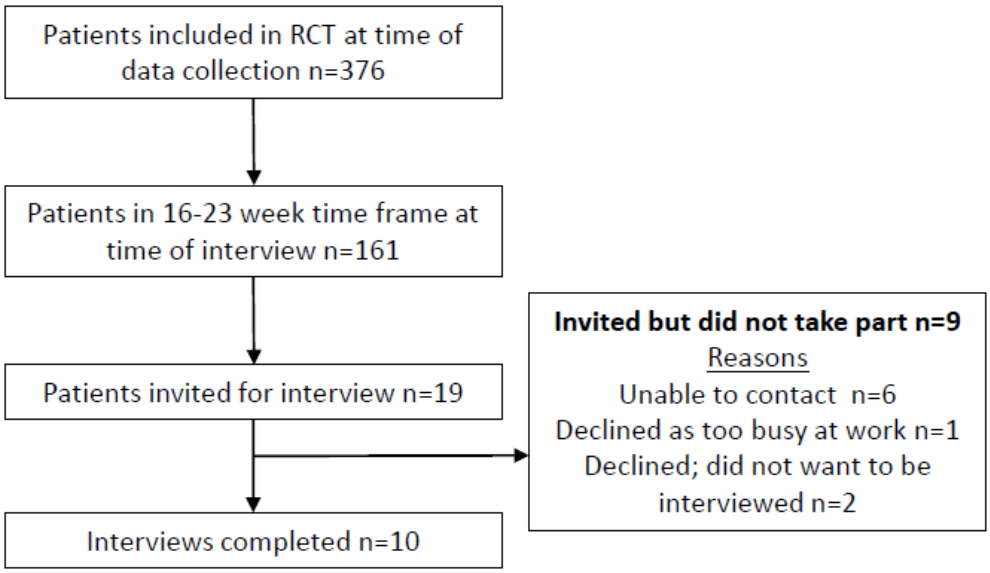


Figure 13 - recruitment flowchart

Table 49 shows the demographics of included participants, with each participant given a pseudonym to protect identity during the reporting of these results. All interviews were completed in participant's own homes, apart from two which were conducted at their places of work or study due to these locations being more convenient for these individuals.

Table 49 - Participant demographics of individuals included in interview study

Date of interview	Participant age	Age group	Gender	Fracture management	Allocated intervention	Other injuries	Mechanism of injury	Occupation	Pseudonym
12/10/2018	49	≤49	F	Operative	Functional Brace	None	Low energy fall	Teacher	Jenny
19/10/2018	59	≥50	M	Operative	Plaster Cast	None	Low energy fall	Maintenance Worker	Peter
29/10/2018	75	≥50	F	Non-operative	Plaster Cast	None	Low energy fall	Retired but volunteers for charity	Joan
31/10/2018	38	≤49	M	Non-operative	Plaster Cast	None	Low energy fall	University lecturer	Richard
01/11/2018	29	≤49	F	Operative	Plaster Cast	Foot fracture and contralateral ankle sprain	High energy fall at gymnastics	Nurse	Katie

07/11/2018	61	≥50	M	Non-operative	Functional Brace	None	High energy impact injury whilst wakeboarding	Retired	Mike
19/11/2018	73	≥50	F	Operative	Plaster Cast	Ipsilateral calcaneal fracture	High energy fall from height	Part time job in shop	Mary
28/01/2019	45	≤49	M	Non-operative	Functional Brace	Contralateral ankle sprain	On a push bike which collided with lorry	Commercial team for car manufacturer	Simon
01/03/2019	69	≥50	F	Operative	Functional Brace	Wrist Fracture	Slipped over/low energy fall	Retired	Sue
03/04/2019	21	≤49	M	Operative	Functional Brace	None	Fell off the back of a truck	Student	Adam

4.3.2 Results: patient experience of ankle fracture recovery and factors of most importance in their recovery

Eight key themes were constructed from the interview data regarding patient experience and factors of most importance in recovery. These were: mobility, loss of independence, ankle symptoms, sleep disturbance and fatigue, activities of daily living, psychological effects, family and social life and healthcare.

Each of these themes are discussed below giving examples from the data collected in the interviews. Some quotations have also been used in the publication which resulted from this chapter [211].

Mobility

Participants discussed the difficulty they encountered walking, recounting a general reduced mobility due to their injury, usually described as frustrating or inconvenient. Some related this concept to their residential environment which they now felt "confined to", with several people expressing the feeling of being a "prisoner in your own home". Participants spoke of their lack of mobility in relation to the weight bearing restrictions they were required to follow in the initial stages of their recovery, with many describing these restrictions as difficult or hard work. This was especially evident in older participants. Some individuals interviewed here referenced mobility and not being under any weight bearing restrictions as being some of the most important aspects of their recovery to them, several people cited the ultimate aim of being able to walk "normally again" or "to be back on two feet."

Sue: *"Just stuck here…I felt like a prisoner in my own home I think for the first four weeks…"*

Jenny: *"Erm not being…not being…being non-weight bearing was very, very difficult."*

Mary*: "I used to dream of walking [the dog] down the park. Yeah…I would be…you know…dream about it. Erm [pause] just…just to be on two feet that was my sort of you know [pleading] 'please let me get onto two feet'"*

Often the weight bearing restrictions and subsequent need to use walking aids were inextricable concepts and participants spoke about them as the same source of frustration. Many of the participants spoke about the impact that the use of walking aids bearing had on other parts of their body, such as the additional pressure on their

arms and hands. Crutches were the most commonly used walking aid. The older participants were generally more absolute in their dislike for walking aids in comparison to the younger participants. Older individuals used terms such as "I can't use these", "hated them" or described the difficulty they had in hopping on one leg. One individual struggled with the crutches so much that she swapped them for a walking frame, something she described as equally as difficult as crutches because it still involved hopping on one leg. In contrast to this, younger participants more often described the difficulties of using walking aids as being hard work or slowing them down with daily activities.

People discussed their preference to walking on their knees rather than using their crutches and some would often adapt their ways of performing day to day activities by performing them on their knees rather than standing. People also spoke of the inability to carry things whilst walking as they needed their hands for crutches. People mainly spoke of this in the context of making meals and hot drinks, which they had to consume standing in the kitchen if no one else was there to carry it for them. Some people described the ability to weight bear again or get rid of the walking aids as being an important milestone in their recovery and one of the defining factors of progress for them.

Joan: "*And quite honestly [pause] I…I wouldn't say I could use that [the frame] really because [pause] if you can't put your foot to the ground you've got to hop. So it was the same really as having crutches. You know so I found that very [pause] hard to take in yeah…yeah."*

Peter: *"But I preferred to walk round on my knees than the sticks it's easier for me to get around."*

Richard: *"I was moving round with crutches everywhere. It was really difficult to do…I wasn't used to the fact that I couldn't carry anything for instance erm. So I couldn't even make dinner and go and sit at the table because I couldn't carry a plate. Erm [pause] so yeah that was…that was a…it was a bit frustrating."*

Katie: *"I could put a little bit more weight through the foot which meant it was easier on my hands…it was easier on my arms. I've had previous shoulder injuries so it was easier on everything when I could just put my foot down that little bit."*

Adam: "*You couldn't go more than a hundred…two hundred metres without stopping because it just puts so much pressure on your hands."*

Some participants spoke of not complying with weight bearing restrictions. Younger participants who reported not complying with restrictions justified this because it was inconvenient, explaining it was only short distances which was unlikely to cause any damage. Older participants reported non-compliance with weight-bearing restrictions described the reasons for this as being a severe difficulty or inability to hop on one leg and use the walking aids provided. As a result, there was more dislike of weight bearing restrictions and walking aids in the older participants of this study. Some described subsequent feelings of guilt for not following weight-bearing restrictions, but others did not, justifying it by saying that no damage or consequence came from this. In contrast to this, some individuals described a reluctance to weight bear through their ankle even though the clinician had advised them to start doing so. People attributed this to pain, fears of further damage and not wanting to delay bone healing.

Adam: *"Erm but I think by the end of it I gave up I was walking on the cast to be honest...so I kind of just wasn't using the crutches and just kind of hopping or hobbling…it was only a little kind of couple of metre journeys around the house and stuff like that…"*

Joan: *"I just had to give them [crutches] up 'cos I couldn't use them at all. So they gave me something to use in the house [frame] which I certainly found difficult as well and I'm afraid I have to say that at times I was putting my foot to the ground…but as I say to have to hop at my age is a very hard thing to do…and errr I really struggled with that."*

Simon: "*I also kept thinking I need to…to rest. I need to focus on the bone healing and…and my body repairing and I can't be wandering around and being…that's gonna slow the recovery process down as well so."*

Despite their many frustrations, many people acknowledged the walking aids as a necessary inconvenience, accepting they were essential in the earlier stages of recovery. People also acknowledged the kindness and consideration from others whilst they were using walking aids. Some participants acquired wheelchairs to help with getting out and about and those who did described how the ability to get out and about overrode any feelings of self-consciousness about using a wheelchair. The use of a wheelchair, however, also produced other difficulties, such as the discomfort of being pushed in a wheelchair by someone else and subsequently not having control of your own movements.

Mary: *"Very frustrating. Erm I used to wake up in the morning and look at this frame and think "Grrrr" [grumbling noise] [laughs]. Erm I couldn't have managed without them though…I just couldn't…no."*

Sue: "*Errm people are very kind to you when you're in a wheelchair or on crutches. So…you know people are very considerate…"*

Sue: *"And the…the thing when you're in a wheelchair as well…people have a tendency to push you and perhaps leave you facing a wall or something [laughs]…'please turn me round so I can see what's going on'."*

Once walking independently without weight-bearing restrictions, several people described issues with altered walking patterns and limps. Many people felt they should make conscious effort to eliminate their limp, describing a need to "think more about walking" or "get out of the habit" of limping. Others were more reluctant to do so, with one individual feeling an attachment to their limp as a symbol of their injury. Several people described a dislike or avoidance of uneven surfaces, some attributed this to pain and others to a feeling of instability or worry that they might "go over" on their ankle. With this, many described a sense of heightened awareness or vigilance they had when walking outdoors, describing how they were always assessing the ground for uneven areas. People also described a fear of falling, a concept only discussed by the older participants in this study. Some participants described falls that they had during their ankle fracture recovery which often caused fears of further damage of the ankle. Some of the older individuals described how they always wanted to have someone with them when they walked or went outside, due to a fear of falling or not being able to get back indoors.

Sue: "*It's still not right erm I'm constantly trying not to limp. I was limping for so long that it sort of becomes a habit…."*

Katie: *"But in terms of losing the limp part of me is probably a little bit psychologically attached to the limp…to say 'I'm not fixed yet it's still hurts'."*

Jenny: *"I think you're very aware now of uneven ground…so if it's kind of tarmac I'm absolutely fine…you know you worry about a divot in the grass or whatever so I think it just heightens your awareness now and your more aware of things…you're a bit nervous you just want sort of like flat areas to walk on."*

Mary: *"I was very dubious about [pause] going out there…suppose I go out there and fall and can't get back in you know?"*

People described a limited walking distance, initially because of the crutches causing hand pain, but also during the later stages of recovery due to limitations coming from the ankle, usually pain. Some also discussed a limited walking speed and difficulty keeping up when walking in a group. Several people discussed the use of stairs, escalators and lifts whilst out and about. Individuals described how they found it easier to use the stairs either on their knees or shuffling on their bottom. One individual avoided using the stairs altogether and stayed upstairs full time as he was reluctant to use his crutches on the stairs. Several people explained how going down stairs was more difficult for their ankle than going up.

Mary: *"I can't walk very far or very fast."*

Jenny: *"…you find you're not quite as fast…so like you know we're in like a big group and then suddenly you find yourself at the back of the group."*

Mary: *"Coming down [stairs] is hard. Erm…to walk down…because this foot doesn't bend…I can do it sideways. But I still can't do it…straight on….no."*

Mary: *"I went to [shopping centre] erm I was OK in Marks and Spencer's going up the escalator but coming down I just looked at the escalator and thought 'I can't do that'. Erm it seemed to be going so fast and I thought 'if I fall'…so I had to go down in the lift and I hate lifts but I just thought you know 'I can't do that'."*

Some individuals spoke of the loss of spontaneity in getting out and about, citing the annoyance of having to pre-plan everything more than usual. Others described how much longer walking took when on crutches with weight bearing restrictions. Getting out and about was often restricted by participants' inability to drive during the earlier stages of their ankle fracture recovery and for many people this was a source of great frustration and inconvenience. Some people spoke of this as being one of the most important factors in their recovery, describing defining progress in recovery when they were able to return to driving.

Adam: *"You've always got to have a plan in your head and pre-plan everything. You can't really do anything spontaneously."*

Mike: *"Erm the main thing was errm that I noticed was how long it takes to do anything."*

Jenny: *"…the worst was when I wasn't able to drive; once I was able to drive again I think that was a turning point."*

Loss of independence

All of the participants spoke about their inability to complete their usual activities independently and the help they required from others to complete various tasks during their recovery including meal preparation, cleaning, washing and dressing and transportation for themselves and/or dependents. People described how they adapted the way they would complete tasks in order to retain independence as much as possible.

Mike: *"I did get help with things like washing cos' obviously that was just difficult to carry a bag into the kitchen…well it was just impossible so…so yeah thankfully I had sort of help there."*

Mary: *"They were bringing meals in. My daughter in law….she was brilliant…she'd sort of say 'right that bed needs changing…I'll go and do that' and she'd take the washing off and do it and you know bring it back…which was brilliant."*

Peter: *"Quite regularly every other day they'd ask 'when are you going [fishing] next?' And they'd say 'Well whenever you're ready we'll take you yeah' so yeah."*

For many participants, the loss of independence and subsequent need to rely on others was a source of significant frustration for them. This was expressed more strongly in the female individuals and the older individuals in this study sample. Other people acknowledged that they needed to rely on others to complete tasks but displayed less, if any, frustration with this. These individuals often expressed concerns about how hard it must have been for their partner or families to take up the majority of house work, rather than for them to accept the help they needed.

Mary: "*It was extremely difficult…I mean it just stopped my life. I was very active before erm very independent. I'm on my own so you know I have to do everything for myself. That…it was very very difficult. I think I was in shock for at least a month.*"

Joan: *"Hated it. Really hated it because I'm a very active person and I don't like people doing things for me I'd rather be doing them myself so yeah."*

Adam: *"They [friends] were really good they offered but by the end it was just a bit [sighs]…I don't know. It was a bit like having my mum around really. It just got a bit…I dunno you just get a bit claustrophobic everyone doing everything for you."*

Richard: *"…yeah mentally I was fine erm for me at least [laughs] I don't know how my wife coped. Errm it was a lot of work for her…if you were interviewing her she'd say it was a big deal. But you know I spent a lot of time sitting on my bum so it couldn't have been that bad."*

Several people found it very frustrating when tasks that others completed for them were not done in the way they would have liked or the way they would personally do them. Not only were these individuals frustrated because other people needed to help, they also felt frustrated because the tasks were not completed the way they wanted them to be done. Some interviewees linked their frustration at relying on others to feelings of being a burden, feeling like a nuisance or not wanting to feel like another person's responsibility.

Mary: *"…and peop-…you know they did housework and [pause] everything really. Again that was quite hard…that's not the way I do that…but I can't say anything because….the day my grand-daughter had got…bless her…oven cleaner and putting it on my…top of my stainless steel hob I did shout 'No!' Erm…it was quite difficult seeing other people do things and…and thinking 'No…no…no don't do that' but you couldn't say a word. You just had to accept it."*

Katie: *"I would sit and stare at the washing up and I couldn't do it and it was so frustrating because I like a clean house. He's not so bothered about those sorts of things and I had to sit and look at the washing up and look at the wash bin you know over flowing and I just couldn't do it and it was really frustrating [laughs]."*

Sue: *"You feel as though you're being a nuisance all the while."*

Some people spoke about a general loss of agency but didn't relate this to any specific task or activity, using terms like wanting to "take charge of my life" or the desire to do things "how and when I want". Some individuals directly attributed their lack of independence to a decline in their mental wellbeing and some cited the lack of independence as one of the worst parts of their experience of ankle fracture recovery.

Sue: *"Yeah I think for peace…for mental reasons it's good to get back to normal errm that was quite important for me to feel as though I was able to take charge of my own life again. Erm yeah rather than relying on other people."*

Jenny: *"I think that's the thing that got me down the most having to rely on others to come and take you out."*

Interviewer: *"OK. Erm and what was the worst part of having an ankle fracture for you?"* Mary: *"Not being independent. Yeah…it was."*

Interviewer: *"OK. Erm and what was most important to you in your recovery?* Katie: *"I would say it would have been my independence. The day…even when I was in the cast the day that I could get up have a shower get dressed get out get to the station get on a train go and meet a friend have a couple of pints get on the train home go out somewhere else; that was when it really turned around for me. The worst part was not being able to do things when I wanted how I wanted that was the worst definitely."*

Sue: *"To be able to be my independent self. Erm not relying on other people. To be able to drive again erm…just getting my independence back I think….erm…being able to walk about again [laughs] it's funny how we take so many things for granted and yet when you've got a broken ankle there's so much you can't do because you just can't move about. Everything was brought to me or I was…and [husband]…I would start getting up in my chair and he would say 'what do you want? I'll get it.'…'I've got to get up and move'."*

Sue: *"What bothered me most? The lack of independence."*

Alongside the many frustrations that relying on others caused, participants countered these expressions with a profound gratitude and sense of appreciation for the help others provided. Many people declared that they don't know how they would have coped without them. Some interviewees communicated an emotional conflict between the frustration of needing to rely on others and gratitude that others were helping them, with one individual describing a "mental anguish" associated with this.

Mary: *"I couldn't have managed without them…those people...I really couldn't. They were brilliant…they really were."*

Sue: *"Erm, it is very difficult to errr put into words what's going on in your head. Because you're grateful to people for helping you to get out and about erm but on the other hand [long pause][sighs] I don't know. There's quite a lot of mental anguish I think."*

Ankle symptoms

All participants discussed the ankle symptoms resulting from their fracture during the interview. These included pain, stiffness, scar concerns, altered sensations,

swelling, bruising and muscle wasting. Pain was a common complaint in individual's recovery and people spoke of the most acute pain from their injury as being the worst in most cases. Despite this, once the acute phase of pain had subsided, many of the participants indicated that the pain was manageable and never stopped them from doing anything. All individuals interviewed took painkillers in some form to manage their pain, with others also using techniques such as ice, distracting with activities, removing their functional brace rest and elevation. There was one case of the ten participants who described having no pain following their initial injury pain and once she had the cast applied, there was no further pain in the ankle throughout the recovery.

Jenny: *"Pain wise paracetamol dealt with it...it wasn't anything that was…you know I was never sort of laid up not being able to do anything because of it sort of thing so…"*

Richard: "*Erm during the day when I was taking paracetamol and kind of busy with other things and it was noticeable but it wasn't debilitating."*

Mike: *"I still carried on. I mean really it was not…on a scale of one to ten…it was never more than a sort of four or a five to be honest."*

Joan: *"I can't really remember that I had any pain. On…on the day that I got to the hospital; yes it started then to be painful. But as soon as the cast was put on it was fine. And afterwards I didn't have any pain with it."*

Many people spoke about the pain changing throughout the recovery period, usually discussed within the context of changing immobilisation method, weight-bearing status or increasing their activity levels. Some people were very different in the way they related to their pain. Some used the terms "pushing through the pain" and were willing to feel some pain through their recovery, often guided by advice and reassurance from clinicians. Others displayed a reluctance to push into any pain or a fear of feeling any pain which limited their movement or willingness to remove the immobilisation.

Mary: *"…as I say it is very painful now but I've…the physios keep telling me it's quite normal…it will be painful….you've just gotta keep pushing yourself."*

Simon: *"I was anxious…that's the answer you're after. It was just so…I just didn't wanna take it off 'cos it hurt…I remember thinking…I was…I was quite frightened 'cos it was really painful [pause] …when you're lying in bed and thinking about*

taking the Aircast off and you just bend over just to do…undo one strap and you get a shooting pain in your ankle your thinking 'Nah I'm gonna leave this'."

People interviewed also spoke about swelling in their ankle. Some people linked the swelling to a discomfort in the ankle, sometimes describing this as a throbbing feeling. Others explained how they struggled to put on shoes or trouser legs because of the swelling. One individual explained they had no swelling, but the majority of people did and they described it as fluctuating throughout the day. People used rest, elevation and ice packs to help manage the swelling. Some individuals expressed concerns of discomfort from their operative wounds, pressure sores, bruising or skin issues. Mainly people spoke of concerns or anxiety regarding possible wound infections, although no one interviewed had experienced one. One individual had a specific anxiety about getting a plaster sore or ulcer from their cast. Several people described dry skin which caused them irritation and they usually attributed this to a period of wearing a plaster cast. Others who had an operation described altered sensations, usually numbness or oversensitivity around the wound area.

Adam: "*I just remember it used to throb and then it used to swell up because it was throbbing and then it would just push on the edge of the cast. So that was quite…it was more of a discomfort than pain."*

Mary: *"Well it's mainly not being able to get shoes on erm that's the big thing…I mean initially I was going out somewhere erm and I'd got some trousers which were quite narrow."*

Jenny: *"…and like I say of an evening if I…to come in from work not feel that it's swollen and not feel that it's uncomfortable erm then I think that perhaps that would be a hundred percent recovery."*

Jenny: *"Erm I was quite paranoid about erm…my erm…about the scar and it getting infected and erm but again I suppose it was quite good having the boot because at least the you could erm keep checking on it and… erm and making sure that it was clean and tidy and whatever"*

Katie: *"Err that…the first time I think I was really concerned about the heel was probably just after surgery. And I'd had erm a block for the surgery so it was all numb. But I was convinced it was numb because the cast was too tight and the fact that I couldn't look at it…I couldn't get to it…I couldn't see it. Errm I was convinced absolutely convinced that I was getting all sorts of you know pressure sores blisters."*

Adam: *"My skin with the ankle is still really dry it's really crumbly."*

Several of the participants spoke of ankle stiffness or lack of movement in their ankle, some of whom also stated this tended to be worse in the morning, in cold weather or after resting. Some spoke of reduced strength around the ankle muscles and others spoke about muscle wastage, or losing weight on their leg.

Adam: *"I can't really…I don't have anywhere near the movement I do in that one in this one and if I try to…yeah it does just hurt quite a lot."*

Mary: *"But err [pause] you know when I get out of bed first thing in the morning it is really stiff…it really is sort of…I'm sort of hanging on to things."*

Sue: *"I've got very good flexibility in my ankle but not much strength. So that's what I'm working on."*

Peter: *"Erm I was surprised how much weight I'd lost on it and why I'd lost it. It's probably because I haven't used it I don't know that…."*

Katie: *"I lost loads of muscle in my leg. Erm I could see it in the cast my leg was just getting smaller and I was just thinking 'Oh my god it took me twenty nine years to get that muscle how long is it going to take me to get it back' [laughs]"*

Sleep disturbance

Several participants discussed issues with their sleep during their recovery, with a few individuals citing the difficulty sleeping as one of the worst parts of their recovery for them. Some attributed their sleep issues to the pain they were experiencing in their ankle, whilst others felt that their sleep was impacted more by the inability to get comfortable in bed whilst wearing a plaster cast. Some people discussed their sleeping issues in the context of pain medications they were taking or even taking painkillers to aid with sleep. One participant described a wariness of the addictive nature of the painkillers and being wary of the reasons they were taking the tablets. Some individuals attributed mood changes to their disturbed sleep.

Richard: *"I think the main thing I remember was not being able to sleep for the pain."*

Katie: *"Erm and I was quite wary of the fact that I like codeine and it's a very…for me it's a very good drug and I was very aware of how much I was taking erm taking it to sleep at night became a big thing as well and I was quite wary of why I was*

taking the painkillers towards the kind of week five-ish of saying 'OK am I taking this so that I can sleep or am I taking this because I'm in pain'?"

Simon: *"at night I…I just wedge it within pillows and try and lie on me side and it…it was comfy but…and…and then you'd just have to make sure to take enough codeine so you wouldn't be disturbed."*

Peter: *"I think it stopped me from…yeah I think it stopped me from sleeping so much…But it's taken a bit of getting used to…getting it comfortable before I can go off. I just couldn't get it comfortable at all…it's when it was in the cast was the main thing, getting it comfortable in the cast... I have been a bit moody in the mornings as I haven't had enough sleep"*

Richard: *"Erm I think the loss of sleep was the worst. Yeah…it wasn't it wasn't even so much the pain itself as the fact that I wasn't sleeping properly and I was tired all the time for a few weeks I think that was the worst."*

Activities of daily living

Participants discussed their limitations to usual activities, including personal care, household tasks, activities of leisure and work. Many spoke of either being unable to do certain tasks, adapting ways of doing things, or having help from others to complete tasks. When discussing how they were unable to complete activities independently, many also spoke of the frustration associated with needing to rely on others, discussed further in the previous section on loss of independence. When discussing washing and dressing, many individuals spoke of finding new routines or ways of doing things in order to complete these tasks independently and adapting to new routines and behaviours. People described how walking aids, weight bearing restrictions and immobilisation methods all made the washing and dressing process slower than usual. Many referenced the hot summer during which their recovery took place and explained that this made dressing easier as they were able to wear shorts. Those with a plaster cast discussed the use of a cast cover and those with a functional braces spoke of the convenience of being able to remove this for showering. One individual discussed their frustration with not being able to shave their affected leg whilst in the plaster cast.

Mary: *"Erm [pause] you know either you…sort of wash yourself all over with a flannel or something and I just didn't feel that clean. I then started sitting on the edge of the bath which…you know very gingerly…very carefully."*

Adam: *"I did try once to take the crutches into the shower but then they start to get rusty so I had to get a new set of crutches and all that malarkey."*

Richard: *"Getting in the shower was a pain I would have to kind of erm I'd have to prop up my crutches somewhere accessible and then hop to the shower turn it on and then undress erm and then put the cover on the cast and then kind of hop into the shower and then by then the floors really slippy so I'm kind of hopping over a slippy floor in the shower erm with a weird pose to make sure I'm not putting too much weight on my leg. Erm so it took much longer to shower."*

Katie: *"Erm I…yeah I was super excited to get it [plaster cast] off. I wanted to air my leg…I wanted to shave my leg…I wanted to wash it properly…the first thing I did when I got home I took it off had a shower shaved my leg it was disgusting. My other half nearly divorced me over the state of my leg! [laughs]"*

Household chores and meal preparation were also affected. Participants usually spoke about this in the context of the inability to carry things whilst walking due to walking aids but also the inability to perform heavier housework, such as cleaning floors and dusting in high places. Many people stated that they accepted that heavier housework would just have to wait. Some described adaptations they used to complete as many tasks as they could, for example completing household cleaning on their hands and knees. Many people described how other people took on most of the essential housework and meal preparation, as described in the previous section on loss of independence. Several people spoke of their inability to garden and how it hard to watch the garden grow out of control without being able to do anything about it.

Katie: *"I didn't manage to do much in the kitchen for [pause] I…I would say probably four of five weeks. I did try one of the evenings my other half went out and I tried to cook and the effort of it nearly killed me."*

Adam: *"I remember the most frustrating thing…it was just the little things just having a big bag of washing on the top floor and I couldn't put it on the handle of the crutch and peg it down so I had to call someone up and they had to come and cart it down for me."*

Richard: *"Yeah well I was lucky my wife basically took over everything. We try and share most things and erm I was…I was totally just not contributing at all."*

Sue*: "Well things just had to wait…I hate to think what it would be like for somebody that was on their own. They just wouldn't cope on their own erm because initially just you can't carry a hot drink…if I needed a hot drink and [husband] wasn't here I would have had to get myself into the kitchen and stand to drink it there you know on my crutches or…or whatever. Erm so everything within the home is impacted because you are [pause] like a baby really…you can't do anything..."*

Mary: *"Erm [pause] yes I discovered that if I got on my knees…hands and knees [pause] I could actually…not hoover…but I could get a stick brush and sweep the carpet. Erm I did that for quite a while. And then you had to make sure that you were near something so that I could get back up again. I did do that. I also did gardening on my hands and knees because it was…to see it just going…it was heart-breaking. And I thought 'right if I could get out there I'm sure if I'd got something to kneel on I can actually do that' and I did."*

Many people adapted their use of leisure time for activities which were more achievable with their injury, usually more sedentary activities. This was acceptable for some, but for others was a source of frustration, describing their dislike for resting or "sitting around all the time". For participants who would usually participate in exercise and active leisure pursuits, they described how these had to stop and also discussed their return to these activities, often accompanied with a level of caution or sometimes anxiety about returning to exercise. Some people described the psychological effects of feeling frustrated or annoyed because they were not able to do their usual leisure activities. Several people explained how they missed trips or holidays, which was a source of sadness for many and for one individual made her feel guilty as her friends subsequently cancelled their trip as they chose not to go without her. For those with more sedentary hobbies, they reported no impact or minimal impact on these due to their injury.

Mike: *"Errm errrm but apart from that it was obviously a massive limitation on doing all the sport I like to do. Errm so that was….it sort of drove me a bit mad."*

Sue: *"…you know when you've got the opportunity maybe to sit and watch television all day or read all day or whatever but you don't…mentally you don't feel like doing that either?"*

Katie: *"I think for me the long term was all about getting back to gymnastics… When they told me I'd have three months off work I was like 'fine'. When they told me I'd have a year off gymnastics I cried my eyes out."*

Mary: *"I mean I do sewing…I make bags and that…well of course I just hadn't got the strength in that foot to press the pedal down on the machine [pause] so I couldn't do…yeah anything like that. Erm [pause] you know and I felt I needed to be doing things…well a friend and I organise a craft market but of course this year I could-…I couldn't do anything. I just could not do anything. Erm but that really you know did hit me the fact that I'd left my friend to do it all you know and I couldn't take part sort of."*

Adam: *"I had a ski trip booked at Christmas erm so I had to cancel that and everything and all of my friends went on the same one. Errm so that was a bit rubbish being at home for that."*

Richard: *"I mean most of our hobbies involve sitting with children and doing activities with them so most…for the most part it wasn't too bad."*

Several of the participants were working or studying at the time of their injury and discussed the impact of their injury on this aspect of their life. People spoke of being signed off from work due to their injury, viewed by some positively and some negatively. Many people saw this as a mixed blessing, as it was enjoyable to begin with but then became more difficult. Many people blamed this on a loss of purpose and gratification of going to work, a loss of the social aspect of work, boredom and financial reasons. Some participant's had previous experience of being signed off work so knew what to expect. Some people took some time working from home and mentioned how because their role is sedentary they did not need to take time off, citing that the main issue was getting to and from work or university rather than the work itself. Two individuals, who both worked in the education sector, said that the summer period is a relatively quiet time at work so they felt lucky that they were able to take time if needed. Some people said it was a bit annoying to need to take time off work to attend medical appointments and others were unable to attend rehabilitation classes because of work. Some people required adaptation to their work role, however others did not require this once they were back at work. A couple of people were retired so this did not affect them. One individual did voluntary work which she was still able to do as her husband drove her to this. Many people praised their supportive employer or university for their assistance and understanding

throughout their injury recovery. Some people spoke of the effect of pain on their work and the need to take painkillers to complete work, particularly those whose jobs required them to be on their feet for most of the day.

Katie: *"I've had long term sick before erm so I knew what to expect in terms of you know a bit of isolation and not having that gratification of going to work purpose and all those sorts of things I was sort of mentally prepared for that because I've done it before."*

Peter: *"Not being at work didn't help. Being bored. Watching loadsa' telly."*

Joan: *"I'm a volunteer for [charity name] and I go visit and elderly lady erm for an hour. Erm but no it didn't impact…impact on that. I still went cos' he [partner] was driving me you see."*

Richard: *"Erm but because my work is quite sedentary and it was the summer when not so much was going on…my work wasn't really that affected actually I think I got quite lucky with the…at least in terms of the time it happened. I was still working but I could set my own hours and I it was very flexible about when I had to be in the office or not…So erm you know sitting at my desk and actually physically doing the work was fine that was not too affected. The main issue was getting to and from the office."*

Simon: *"And…and I've got an employer who's very very sympathetic and very very people focussed in terms of errr absence from work and returning to work and…and that side of rehabilitation."*

Adam: *"Uni were really accommodating actually so I got erm extenuating circumstances for a few deadlines so they were…those got pushed back while I was in hospital. So it's just generally the getting about from building to building rather than actually the uni bit. Erm but no once you were there it was alright. It just took a bit of planning to get there."*

Katie: *"I have done some shifts which are twelve and a half hours. Erm it hasn't really affected it that badly…when I know that I'm doing a shift I will take paracetamol and ibuprofen alternately throughout the whole shift to make sure that I can work."*

Psychological effects

Participants spoke about the psychological impact of their injury and strategies to maintain mental wellbeing in varying degrees. Some individuals, all of whom were male in this study, indicated that they experienced no mood changes or psychological impact during their recovery or as a result of their injury.

Richard: "*Erm I mean mentally it was ok…some of the questionnaire questions kind of hinting at you know 'was this making you depressed?' and things. But it was nothing like that at all."*

Mike: *"No…I wouldn't honestly say it affected my mood erm you know as I say I just had to get on and deal with it and accept my lot…"*

Others described a range of psychological and emotional effects attributable to their injury. The most commonly described emotion was frustration, usually borne from participants' inability to do their usual tasks, leisure and exercise or their lack of independence. Some indicated a frustration with a perceived lack of or slow progress with their recovery, mobility, sleep quality or progress with rehabilitation.

Simon: *"there was a lot of frustration as well because it's just…I'd rather be up and about and…and doing jobs and bits and pieces erm so psychologically it's just frustrating."*

Sue: *"Frustration. Not being able to do things for yourself."*

Adam: *"I think I watched…I went out to the rugby club a few times to watch them play but it's just a bit crap being on the side lines but no…nothing major at all."*

Other emotions conveyed were feeling down, depressed, emotionally labile and resentful. Many people, when discussing feeling down, reported that it wasn't as severe as depression but described a low mood which was attributable to their injury. Others described their quickly changing emotions, using terms such as "ups and downs" or "good days and bad days". Often people were unable to attribute the exact cause of the bad days. One individual described an emotional lability whereby they would cry more frequently and readily than prior to their injury. Many people attributed their mood changes to the need to rely on others and not being able to do usual activities. Some also said that they avoided social situations and felt they did not want to go out and do things because of their mood.

Jenny: *"You do have ups and downs I think and you do get a bit…yeah I mean I think that anybody who has an injury like that I think would get like that wouldn't they?"*

Sue: *"Yeah. Tearful…It's just a horrible feeling…I don't…it's so difficult to be able to pin point exactly what it is that actually brings the tears on. Maybe it's something on the television that's a bit sad and whereas you'd normally think 'aww [sympathetic noise]' suddenly…[whooshing noise] yeah…you start…start crying."*

Sue: *"And that makes you feel really erm…I wouldn't want other people to have that responsibility of looking after me. Cos I'd have somebody behind me going like this [imitates holding on to waist from behind whilst walking up stairs] and somebody sort of in front saying 'are you ok? Are you ok?' but maybe it's just me…people didn't mind but I felt very [pause] oh I don't know. I was low…very low."*

Joan: *"Errr yes a little depressed. Yeah yes because as I say I'm an active person and then it comes back on you then you know that you can't do things and [pause] it was horrible to have to accept it. Yeah but it did make me depressed at times yeah."*

Some people discussed anxiety about their ankle fracture and potential complications, as well as fears of doing further damage to their ankle. Female participants in the over 50 years of age category expressed concerns about possible osteoporosis. People, usually older participants, also communicated anxiety about losing confidence with walking and driving, sometimes cited as a reason for individuals to push themselves with mobility or to get back to driving. Older participants also described a fear of falling which was not evident in the younger population. Younger participants were more likely to have anxieties about possible complications and the longer term recovery of function at a desired function level for sports and leisure activities. Others had a specific worry about undergoing the operation they needed to fix their fracture.

Katie: *"Any anxieties I had at that time were around long term recovery…erm but in terms of actual mood I mean I probably could have been happier but I wouldn't have gone so far as to say I was actually depressed."*

Mary: *"I think if I'm not careful it would be so easy to lose your confidence and think 'I can't do this'."*

Sue: *"It's probably…probably because in case I overbalanced again [laughs] I had an osteoporosis test 'cos I didn't know whether it was…I'd got weak ankles or*

whether it was erm [pause] well as they said the osteoporosis test came back…I haven't got osteoporosis she said 'you just seem to be really unlucky' [laughs]."

Simon: *"But I know I…I…I was in….I was anxious…that's the answer you're after. It was just so….I just didn't wanna take it off 'cos it hurt. I mean I was…it was…horrible going to the X-Rays because you had to take it off. Yeah but it's just it…well it's alright now but at the time I remember thinking…I was…I was quite frightened 'cos it was really painful."*

Mike: *"I s'pose what I'm saying is I was more conscious about not wanting to do anything stupid so I was probably a bit more cautious on the trails and slightly less reckless than I would have been before. But I think that's more a psychological thing…"*

For some, body image was an issue, with concerns about inactivity could cause weight gain. One individual attributed her loss of appetite and subsequent weight loss to a state of shock following her injury. Others also discussed and referenced the ageing process, explaining that their injury made them "feel old".

Katie: "*Erm and that is so important to me to be able to get back to that to be fit you know so I'm quite weight conscious I'm very conscious of the fact that I don't want to get fat sitting around and not doing anything [laughs] and also not being able to exercise is not a good combo for weight gain."*

Mary: *"And then people came in and put meals in front of me and I used to think 'Ohh I can't eat that'. Erm I did feel quite ill to begin with…yes. They kept telling me it was shock. You know, I'd lost a stone and a half. And they kept saying 'its shock'."*

Joan: *"Errm one of these elderly people's push about things in the house [pause] and I just get it into my mind that I'm not old enough for one of them yet [laughs] so that was very disheartening yeah to use that."*

In one specific incident, an individual described the severe distress caused by the communication from their clinicians regarding the weight bearing restrictions they were under. This particular case, the individuals had been told it was six weeks non-weight bearing, then to go back and be told at six weeks that it would be a further four weeks. This caused significant emotional distress and a sense of "moving goal posts" during their recovery.

Mary: *"Erm I have…you see…I think one of the hardest things about this has been…the day after they operated the consultant came to me and he said 'six weeks non-weight bearing'. I thought 'that's not too bad then…June…July…that will be the beginning of August…that's not too bad'. Following day another doctor came and said 'eight weeks non-weight bearing.' And I thought 'Mmm.' Then we went to the clinic…it was ten weeks. Erm 'Yep you come back it'll be ten weeks…we'll start you walking.' And I was just living for that day. It was the twenty-eighth of August. Of course we went back to another doctor and he said 'Oh no…no no no another four weeks at least.' That nearly destroyed me. That was really [pause] very difficult. Erm [pause] you know I don't know why they do it…I think it's even harder when they keep push-…moving the goal posts all the time."*

Mary: *"Yes…I think if they'd told me that in the first place…yes it would have been 'Oh gosh'…but I'd have gotten used to it. Instead of that it was going back and…you know 'ooh yeah…can I start walking soon?' you know or 'can I start physio?' and 'No…no…no…not yet. No you can't'. You were just knocked back again. You know [sighs]."*

People described their coping strategies to deal with these psychological impacts which were relationships with people and pets, using activities or work as distraction and being in good weather. Many people sought comfort in setting and achieving goals and always wanting to push themselves to progress. Other people described feelings of gratitude around the idea that "things could have been worse", explaining that this reminder gave them strength when things felt difficult. Almost all participants referenced their understanding of the impermanence of the injury, the understanding that it would heal over time and this situation was temporary, which seemed to be a comfort. People spoke of an acceptance of their injury, often using terms like you just have "got to get on with it".

Mary: *"It was very hard…very hard…there were [pause] [sigh] there were times when I thought 'Ohh can I do this?' and then I thought 'No…I've got to…that dog's relying on me' which probably sounds quite silly but you know…I had to."*

Joan: *"He [partner] was the perfect one. He would take my mind off of it straight away to do something. I'm afraid at times [pause] he's a person that always makes me laugh when I don't want to sometimes [laughs]. So he was the perfect help…yep."*

Katie: *"…there wasn't really any one thing that I did to kind of make it better. Erm it was all about kind of focussing on 'well last week I couldn't do this and now I can do…you know I can make a piece of toast or I can make a drink and stand and drink it at the sink.' I [laughs] became very good at throwing things...erm but in terms of actually coping it was just counting down the days and also just distracting as well. I was trying to set myself goals to do one thing a day you know 'today I'm gonna go…I'm gonna get my mum to take me for lunch' and 'tomorrow I'm gonna meet somebody else and then on Thursday I'll do something else' so having that one goal and saying yeah ok recognising that win."*

Mike: *"So I was pretty erm [paused] gutted but resigned I thought well 'you know it won't be forever I'll just have to deal with it.' You know."*

Mary: *"And I haven't really got that down erm I'm not that sort of [pause] a person but there've been odd days. But when you wake up and the sun's shining…you know that summer was…I mean I didn't really see much of it but I…it just…You know when I could get to a point where I could go outside and just sit outside like with a book even for just ten minutes or something erm it was fantastic. It really was."*

Joan: *"Yeah yes yes. Yeah it was. You see as I say when you're elderly you can't do things like you youngsters can do. I'd like to!"*

Mary: *"you just have to get on with it don't you? Erm no it hasn't stopped me. Erm [pause] as I say it is very painful now but I've…the physio's keep telling me it's quite normal…it will be painful….you've just gotta keep pushing yourself. So that's what I keep telling myself."*

Family and social life

Social and family lives were discussed as an area of impact during the recovery period. People also acknowledged that whilst their social life didn't change in terms of how often they saw friends, that the nature of the meetings were affected, with friends coming to visit them at home, rather than going out to meet people. Others described how plans didn't change, but just the way they got there or logistics of getting out to meet people needed to be considered more than usual. Some individuals described how they saw people more often or had a better social life because they always had visitors and people checking in on them during the earlier stages of their recovery. People spoke about changes in their drinking habits as they would often meet friends at the pub which was not as easily achievable to get to.

Drinking habits were also affected by advice from clinicians to reduce alcohol intake whilst taking painkillers. Many people acknowledged the helpfulness of their friends during the initial stages, particularly friends who would visit them whilst less mobile, drive them so they were able to get out and about, helped with household tasks. Some people spoke of the increased interest and sympathy from others they acknowledged receiving with their injury.

Richard: *"I don't have much of a social life at the moment [laughs]. Erm [pause] so I was advised not to drink and because I was taking painkillers I….there were a few outings to the pub I missed but to be honest again because of my lack of social life at the moment I don't think it was too bad."*

Adam: *"Erm in a respect it [social life] got bigger…in at least bigger in the fact that you had to be around people all the time and you relied on people a bit more. And therefore you just naturally see people more. Erm but no it didn't really affect…plans didn't change or anything it's just how you got there changed."*

Simon: *"Erm I probably…I drank a lot less alcohol when I had the boot on erm 'cos I was taking the painkillers erm so…and…and a lot of my social activities revolve around food and drink so the food and drinks that….but that didn't really change much it just…still….I was still doing social things but I just didn't go out to the pub."*

Sue: *"I had loads and loads of visitors. Constantly….it looked like a florist shop in here."*

Mary: *"Errm you know they [friends] were in and out all the time. Erm [pause] you know doing things. I have one friend…bless her…she has been and walked [dog's name] every single day. She was fantastic, bless her."*

One individual did explain how her mental state and mood contributed to her not wanting to go out and socialise, despite encouragement from her friends. When questioned about the cause of this, she clarified it was her psychological state that she felt prevented her from socialising as much as she normally would, describing how her and her husband would "force themselves" to go out and be sociable.

Sue: *"…but again you don't feel like doing the things. And that's what's different I think. Erm people would say 'Ooh you know come and do this.' And you think 'Oh I don't want to be out in company' but we did I mean we forced ourselves to… Erm [pause] so yes it did impact on it but only probably because we allowed it to in the sense that we didn't want to go anywhere."*

Individuals spoke of the impact of their fracture upon their family, family roles and family activities. People spoke of needing to adapt family activities in light of the injury, usually doing more sedentary activities rather than active ones for example. Those participants with children to care for spoke about adapting family and caring roles with their partner in light of their ankle fracture. One participant discussed how she was unable to support her son in the usual way at a significant time in his life during school exams, discussing the psychological impact and frustration of not being able to perform her usual role for dependents due to the fracture. Participants with adult children who were self-sufficient usually described this as "fortunate" because they were not required to care for anyone else during their ankle fracture recovery. Those with adult children spoke of how they saw their children more throughout the recovery because of the help they provided. One individual described that their injury and accident had brought them closer as a family and attributed that to the shared trauma of the accident. Many individuals spoke of the possible impact of the injury on their family, describing an awareness of their family member's emotions regarding their injury, such as anxiety about them walking on their own and falling, being angry about how the injury occurred and struggling with medical issues and equipment. Others described an awareness of how difficult their injury must have been for their partner, who needed to take up most of the household tasks or were disturbed at night when the individual was unable to sleep. Some individuals described arguments with family members attributed to spending more time with each other. As described in the previous theme on reliance on others, many participants acknowledged the gratitude they had for their family and friends, explaining frequently how they would not have coped without their support.

Peter: *"Well I can't play with them [grandchildren] out in the garden…with the cast on but in here you can still play games with them yeah stuff over there [points to cupboard] snakes and ladders and scrabble could do them no problem at all 'cos with me leg up in the air."*

Jenny: *"I suppose the most frustrating thing for me really was that erm [name of child] my eldest erm was doing his A-Levels and so erm I was unable to take him backwards and forwards because, he was…he was at [school name], I was unable to drive him to his exams so my dad he was brilliant he stepped in but…just little things like that it just makes you feel a little bit [pause]…you….psychologically you*

feel like you're not able to perform the role that you normally perform sort of thing so…"

Simon: *"So there was no…no I didn't need to look after any children…I mean fortunately."*

Simon: *"It did bring us closer together. It's a bit weird but erm…I didn't think we could get any closer but that…this…the….the last four five months has just definitely brought the…the closer ones to us closer. Yeah and then…it's…I guess it's a shared trauma isn't it?"*

Sue: *"I had to get up in the night and [husband] said 'Call me!' and I said 'I'll be alright!'…'cos he was worried I was gonna fall again and actually I did fall off the crutches once trying to open the spare bedroom door it's quite stiff on the carpet and I pushed it and all of a sudden went and tried to save myself but I mean…[sighs] oh he went mad [laughs] absolutely…'what are you trying to do'?"*

Jenny: *"my youngest is…he struggles a little bit with anything to do with illnesses and things and I think he struggled with it; he didn't like the wheelchair, he didn't like the crutches."*

Adam: *"Yeah I think erm Mum was a bit angry when I did it [the injury] but I think that's just because I was drunk and a few things have happened like that before [laughs]."*

Peter: *"I fear I might have kept the young lady awake at night, turning and trying to get it comfortable."*

Katie: *"Having to spend so much time with my mum that we started to grate on each other's nerves after a while."*

Experience of healthcare

Participants for the study described the healthcare service experiences, from paramedics, accident and emergency, inpatient stays, fracture clinic and appointments with physiotherapists. Many people commended the NHS services and staff they experienced whilst seeking diagnosis and treatment for their fracture. Participants who did require an inpatient stay described this as lonely or difficult. Many had positive experiences of fracture clinic follow up, but some did express frustration at the lack of wheelchairs to get in and out of the hospital, for the clinician communication regarding weight bearing and also for seeing a different clinician to

the one who performed their fracture fixation operation. Those who received physiotherapy recounted this as a positive and solution focussed experiences with the benefit of providing education and reassurance on progress. Some described the rehabilitation process as tedious, hard work or slow. Those who were not offered physiotherapy explained how they felt they would have benefitted from this, or just some simple exercises to complete when the immobilisation had been removed.

Katie: *"Erm I was really well looked after in hospital so that was fine errr a bit lonely."*

Mary: *"It happened on the Tuesday and I was in there until the Friday. Well I…I…I won't say I discharged myself…I did say to them 'I can't do this anymore.' I was in a four bedded ward and the other three ladies…bless them…had all got dementia. So I wasn't getting any sleep and I just thought 'God I can't do it'."*

Peter: "*Well when they took the cast off the only thing that did concern me; I didn't have the same surgeon looking to say 'you're alright [inaudible] everything looks okay we're quite happy with it'. I had a different surge…different doctor to look at it. Usually you get the same surgeon all the way through the same doctor all the way through."*

Jenny: *"One of the most frustrating and upsetting things was after I'd very first broken the ankle, going to the hospital and getting from the car park to the fracture clinic; no wheelchairs and there's just no wheelchairs to be found anywhere…"*

Jenny: *"She [physiotherapist] was absolutely brilliant really, really good and I think that just helped as well just having somebody you know showing you the things you could do but also say… you could just give you the confidence that yes you are fine to walk on it it's not going to [pause] do anything or whatever so that...I think…I think the physio is really important."*

Peter: "…but now he's [physiotherapist] given me all these exercises to do and to stretch them ligaments again and that's probably why I'm better than most because I've done what he's asked with it…without their information I would have known nothing."

Mary: *"The physiotherapy is fantastic. Erm you see the same girl every week. And she's just brilliant. You know it…it was a breath of fresh air…you know because all I kept getting from the doctor was 'well you'll always have pain there…well you'll have a limp because the heels much shorter than the other one…and….and at your*

age…' it really was…whereas she was 'right [pause] we're gonna get you back walking…we're gonna get you…get you…get your life back for you'."

Joan: *"the only thing I was very surprised at was the doctor turned round and said 'Well that's it.' [pause] And I had to come away and I thought 'surely you have exercises or something like that?'…not that I couldn't walk I could. But I just thought it strange and I thought perhaps there would be one or two exercises they would tell you to do."*

Adam: *"No that's the only thing really they never offered me physio. Erm whether they felt I didn't need it or anything I'm not sure. Or whether that's something you have to pursue privately I don't know. But no they never really offered it or spoke about it. Erm but my Mum said 'you should probably get some' but I just never really followed it up."*

For others, the radiographic outcome was of great importance to them and understanding that the bone break had healed on the images was a source of concern for one individual.

Interviewer: *"When you were recovering from your ankle fracture what was the most important aspect for you?"*

Simon: *"Oh seeing the radiographs 'cos I wanted to see the bone…solid again. That…that's what I wanted to see anyway. And every time I kept going back and just seeing the gap…'cos it wasn't pinned it was just and that goal…that was the target I wanted to see a radiograph that looked like it didn't have a split down the bone…Well it was yeah 'cos I…that's the…that's the measure for me…you can quantify the repair 'cos you can see it you know if it's not cracked your bone's fixed and if it's cracked it's not fixed isn't it?…But that was the initial thought it was like let's see the radiographs…when that crack's gone I'm…I'm well again. But yeah so that was…that was the measure really I just wanted to see it….see it fixed but I'm not gonna see that because he's….he's discharged me…I've got the last radiograph and it's not solid erm but he doesn't want to see me again. Which is fine. I'd like to go back and have an X-Ray in six months and see if…and see it normal again. But you can't do that can you?"*

Participants discussed their allocated interventions within the trial and some experienced more than one of the interventions. When describing the functional brace, participants were generally pleased with the ability to remove it for washing,

discussed as the biggest advantage for many. This was also an advantage to air their leg in hot weather, for shaving their leg or checking on surgical wounds if applicable. The negatives of the functional brace included feeling heavy to walk in. People also spoke about the lack of information or guidance about how tight it should be or how much to inflate the air pockets within the brace.

Others enjoyed the adjustable features for comfort purposes with swelling. One individual was frustrated that it couldn't be recycled as the hospitals wouldn't take it back. Participants also described how walking was made easier with the boot due to the ergonomic sole – one individual who was meant to be non-weight bearing in the boot described this as a difficulty as the sole made her more tempted to walk on it. Those with experience of walking on a plaster shoe felt the boot was easier to walk on because of the design of the sole. Some described the functional brace as very secure and reassuring, with some describing an attachment to it, likening it to a "comfort blanket".

Jenny: *"I was quite paranoid about erm …about the scar and it getting infected and erm but again I suppose it was quite good having the boot because at least the you could erm keep checking on it and… erm and making sure that it was clean and tidy."*

Sue: *"I say that's the biggest benefit it is removable erm whereas the plaster cast you can't."*

Richard: *"And the boot was helpful as well because it was…because it had a sole and it was more s….erm…and it was quite well designed it was easier to get around I think."*

Simon: *"I wasn't overly… I wasn't think 'Oh I can't wait to get the boot off… can't wait to get the boot off.' 'Cos the boot was making me secure…it was [laughs] it was like my…my comfort blanket."*

The positives of the plaster cast as described by participants included the lightness in weight, the choice of colours of plaster and the feeling of security and stability. People spoke of their casts as necessary for healing and often described how they weren't as bad as they expected them to be. Some individuals who were randomised to the cast explained that they felt a sense of disappointment upon receiving the news of this allocation, but that in retrospect the experience was not as bad as they expected. Some individuals spoke of their previous experience with a plaster cast or

knowing someone with previous experience of a plaster cast which framed their expectations. All those who received the plaster cast described the use of a plaster cover to use whilst washing. The main negative aspects of the plaster cast that people spoke about were the lack of ability to remove it, itchiness, difficulty using a cast shoe to weight bear, anxiety over skin condition or pressure sores, dry skin or not being able to shave. People sometimes described a feeling being "trapped" or "constrained" in the cast and others simply described the cast as "clumsy" in comparison to the more "elegant" boot.

Joan: *"….and a choice of colours! [laughs]"*

Adam: *"Erm I think it was (cast) necessary though for the first few weeks because it isolated it so much you couldn't really move it even if you wanted to…"*

Peter: *"Erm it was I dunno itchy that's a fact I always got a little needle…a little knitting needle to giving it a good scratch all the time."*

Richard: *"Getting around with the boot was much easier. Having a proper shoe sole on the bottom I think was the main difference. Erm when I was in the cast I had like a Velcro sandal that came with it to go outside and that was rubbish it kept slipping off and it was a bit narrow for my foot so I was kind of…it was kind of bending round the side a bit. But that was a bit….that wasn't as good as the boot the boot felt like I was wearing a normal shoe and I could get around much better."*

Katie: *"I really struggled with actually having the cast on and being kind of enclosed and I was convinced that I had problems with my heel that it was going numb that I was getting all sorts of pressure damage I was convinced even thought there was actually nothing wrong with it. I was convinced it was numb because the cast was too tight and the fact that I couldn't look at it…I couldn't get to it…I couldn't see it. Errm I was convinced absolutely convinced that I was getting all sorts of you know pressure sores blisters…it would wake me up at night I would be moving the cast around. It was awful."*

This chapter has outlined the methods of the qualitative work completed here and in the subsequent chapter, as well as the results for the initial part of the qualitative work package examining the patient experience of ankle fracture and the factors of

most importance to them in recovery. The next chapter will outline the results of the second part of the qualitative project, which is concerned with the relevance and comprehensiveness of the OMAS questionnaire content. Finally, the results of both qualitative parts will be discussed at the end of the subsequent chapter.

5. The relevance, comprehensiveness and acceptability of the Olerud Molander Ankle Score

Aim: As outlined in the previous chapter, this work details the results of the second part of the qualitative study. The aim of this section was to evaluate the relevance, comprehensiveness and acceptability of OMAS in a sample of individuals taking part in the AIR trial.

Methods: This purposive sample is the same as used in chapter 4 and this work formed the second part of the interview with these individuals. Data were analysed using reflexive thematic analysis and analysed using NVivo software.

Results: Three themes were constructed from the data in this second part of the project. Themes were relevance, comprehensiveness and appropriateness of response options. The items which were deemed relevant by most individuals in this study were pain, stiffness and work/activities of daily life. The items deemed not relevant by most people were running and jumping. Of the sample, three people felt OMAS missed items out, such as wound complications, sleep concerns, mental health and return to driving. Five people felt the OMAS was comprehensive. Some individuals described inappropriateness of response options and these issues were having only binary response options available and response options which were not reflective of their experience.

Conclusion: Not all individuals who have sustained and are recovering from an ankle fracture find the OMAS to be relevant and comprehensive in reflection of their experience. Furthermore, there are some issues with response options of the questions included in the score. This highlights the importance of involvement of patients when developing PROMs.

This work has been accepted for presentation at an international conference:

Title: A qualitative study of the relevance, comprehensiveness and comprehensibility of the Olerud Molander Ankle Score in adults with an ankle fracture (poster presentation) **Event:** European Federation of National Associations of Orthopaedics and Traumatology (EFORT) Congress, June 2020 Vienna, Austria. (accepted but not presented due to COVID-19 pandemic)

5.1 Results: relevance and comprehensiveness of OMAS

During this section of the interview and analysis, three themes were generated from the data; relevance, appropriateness of response options and comprehensiveness. I have displayed the results in tables corresponding to how many people expressed the different views on the questionnaire, as well as demonstrating when the outcome was unclear. The purpose of these tables is for visual display of the results only and I will not be completing formal analysis of the frequencies of responses due to this being a qualitative research project.

5.1.2 Relevance

Table 50 - Relevance of items included in OMAS

Questionnaire Item	Relevant	Not relevant	Relevance unclear
Pain			
Stiffness			
Swelling			
Stair climbing			
Running			
Jumping			
Squatting			
Supports			

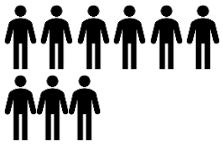
Work and activities of daily living			

Table 50 shows the responses from participants regarding the relevance of each item included in OMAS. When discussing the relevance of the questionnaire item of pain in their recovery, all participants acknowledged that it was very important for them to either not be in pain or to be able to control their pain throughout their ankle fracture recovery. Participant's said that they felt this was an appropriate and relevant aspect to be asked about in the questionnaire. Despite this, some people reported that pain was either not a concern for them or always manageable and not the most troublesome aspect of their fracture for them, but that asking about pain was appropriate.

Simon: *"Hundred percent relevant. That was…there's nothing more important than being pain free."*

Mike: *"Oh I see… but for me pain was not really an issue to be honest…erm and there was some but it…easily err treatable with the…you know ibuprofen helped and it was actually a fairly short duration so not …I was just lucky not to have had it...much. But yeah I think it's an important factor yeah…to include…for sure."*

The relevance of stiffness as an item in the OMAS was evident for several participants in this study with most people relating their ankle stiffness to certain functional activities like walking or standing on tip toes. For some, however, the relevance of stiffness was not apparent as they did not experience stiffness in their ankle at all. Some felt that whilst the aspect of stiffness was relevant, it was something that they got used to or accepted as a consequence of the injury.

Katie: *"Erm to the…to the point where I could walk without a limp that would be quite nice because that's part of the stiffness. Erm you know I'll never be able to point my toes like I used to [pause] I'm fine with that sort of stiffness but to be able to walk without a limp and to be able to run and do things that would all be attributable to the stiffness."*

Adam: *"Erm it was important but then I kind of realised it comes with the fact that you've broken your ankle I think and it will just take a bit of time to get that stiffness to go away."*

Joan: *"No. No never. I had no stiffness."*

The item relevance of swelling was mixed or unclear for some participants. Several individuals described how they experienced minimal or no swelling which didn't bother them, indicating that this item may not be relevant for these participants. Some individuals referenced the swelling as always being present alongside other symptoms such as pain or stiffness or causing issues with footwear. Therefore detangling their concerns regarding the swelling alone was difficult as they felt this was inextricably linked with other symptoms. Others described how their swelling fluctuated or changed over time so there was a transient nature in the way that individual's described the swelling, making it rather unclear how relevant this item was for them at different times throughout recovery.

Joan: *"No. Never had that [swelling] at all. So that was good I expected to but no."*

Katie: *"The swelling just had…it hasn't even been a thing for me because it hasn't been related to the pain obviously it doesn't look very nice but it hasn't hung around for very long so it hasn't bothered me at all really."*

Mary: *"Erm well it is a nuisance because I can't get shoes on…I mean it is really constant."*

Stair climbing was of mixed applicability to individuals, usually related to whether they needed to do stairs at home. Those who didn't still felt this was an important issue when out and about. Some acknowledged more difficulty in going down stairs than going up but it wasn't clear from the data collected here whether they perceived the item in the questionnaire as going up or going down stairs (or both). For one individuals who did have stairs at home, he still felt that stairs were not a priority to get back to or not the main focus of his recovery. One individuals felt that doing the stairs with crutches wasn't bothersome for him once he knew how to do it.

Jenny: *"Erm again I suppose with stair climbing I fine when I'm going up but not so good on the going down."*

Katie: *"Well without it I wouldn't have been able to get home! [laughs] So it was hugely important to feel safe when I first got home."*

Adam: *"Erm I think once I first did it they gave me crutches the hospital didn't show me how to do stairs properly. Erm so I was just kind of hobbling up them any way I*

could but then once the physio took me to some stairs and showed me how to do them properly I was fine. Erm yeah it's not too bad at all once you know how to do it."

Simon: *"Erm it [stairs] wasn't…that wasn't one of the priorities."*

Running was mostly irrelevant to the participants included in this study, with most individuals stating that it was not something they did regularly prior to their fracture. One of these individuals was keen to get back to this level of mobility where she would be able to run if required (e.g. for a bus). Another individual explained, however, that she could live without running again but would like to be able to walk faster. One individual explained how he did not run regularly prior to his fracture but also that he found it difficult to answer this question as he did not "test" running during his recovery. He also stated how, whilst it wasn't important for him to get back to running, he accepted this as a reasonable question to be asked as a sensible objective measure of his progress, which made the true relevance unclear.

Jenny: *"Not really because I'm not someone that runs anywhere anyway."*

Joan: *"[laughs] Can you expect a 75 year old to run? No! I wouldn't even try without it, you know?"*

Mary: *"Running is impossible…and I can live without that."*

Richard: *"Errrm...yeah again well so I couldn't run for a very long time I probably still can't run very well I don't do very much running so I wasn't testing myself. But I know for sure it would have been impossible for me to run…erm so…in terms of importance for my recovery not that important but as a measure of how well I am recovering it seems like a sensible question."*

Mike: *"Yes it was that was pretty important but it was impossible obviously. So err yeah [pause] definitely an important one for me."*

When asked about the relevance of jumping as a questionnaire item, the majority of individuals explained that this was not relevant to them in their recovery and not something they normally do in life. Several people acknowledged that jumping was something that they struggled with and identified as "the higher level of recovery" by some, acknowledging that clinicians would use it as an objective measure and therefore it seemed easily measurable and sensible to measure as part of a questionnaire, even if they did not do it in their day to day lives. Others felt more strongly that it was a silly question as people don't usually jump in their day to day

life. One individuals said she wouldn't even attempt to jump again due to fears of sustaining another fracture.

Peter: *"I've no reason to jump….so running and jumping wouldn't have made any difference at all."*

Richard: *"No not really I don't really do sports so except for some cycling so it wasn't important to me erm although it would be easy to measure."*

Mike: *"The same for that really yeah jumping's pretty important"*

Mary: *"I don't think I would attempt. 'Cos I…you know I do not want to break my heel again so I don't think I would do that again."*

Simon: *"Yes…yes. Yeah…who jumps in their day to day life…I don't know anybody that goes round 'oh just going out jumping.' Or 'oh I can't wait to go jumping this weekend.' It's just a bit odd."*

Individuals had mixed responses on the relevance of the squatting item. Some people felt that squatting was not an issue at any time of their recovery. Others explained that it was not an issue as they used their knees and adapted how they would squat to get to the floor or complete functional tasks. There seemed to be differences in the way this question was perceived, with some discussing the action of squatting to the floor to pick something up and others as the purposeful exercise of completing squats in the gym. Other explained how they could adapt to kneel instead of squat and therefore it wasn't that important.

Richard: *"I think again squatting; easy for me to answer the question certainly it was very difficult for a while I can do it now erm I…it wasn't vital as part of my recovery I was squatting a lot picking up the baby and changing nappies and things so it was…it was noticeable but errm not quite as important as some of the things higher up on this list."*

Katie: *"I mean it's kind of mildly important because I would…I was doing kind of fifty squats at the gym at the weights gym with the weights bar over my shoulders before I did this. Now I just don't do that. It doesn't really matter I can do something else."*

Mary: *"Not particularly erm [pause] again you know if I wanted to get down I'd probably get down on my knees rather than squat down…yeah."*

Whilst a lot of individuals spoke of the desire to no longer need walking aids and supports, many people also stated how they wouldn't have coped without them throughout the initial stages of their injuries. Some people said that the crutches were more of a nuisance than the cast or boot and they were not described as equal in importance when discussing terminating use of supports. Katie also mentioned how having crutches was useful when out and about using public transport so people knew she needed a seat so this was a reason for not being of highest relevance. Subsequently, the relevance of this item was unclear in most cases as the majority of people acknowledged that they wouldn't have coped without the walking aids.

Jenny: "*Oh no I was…I was glad to get rid of the crutches definitely but however at the time they were very, very good to have."*

Richard: *"I think so as much as I could....well I needed the crutches right erm in as much as I couldn't get around anyway I was keen to get off the crutches but it ended up being quite a gradual process erm but I mean I guess there was a day when I decided that I will not use the crutches anymore so to answer this question it was ok and that was a nice thing."*

Katie: *"Getting rid of the crutch is super important erm but on the other side of that it was very helpful if I was on the train or on a bus to be able to say…or sort of for people to notice and say 'would you like a seat?' 'Cos actually on a moving platform that is something that's very painful [laughs]. Erm but in terms of it being the bane of my existence I wanted to get rid of it erm so yeah. It was a good thing."*

Mike: *"Erm not really…I mean the crutches were a faff yeah because so it was important to get rid of the crutches really and I…when I went to use one crutch it helped. But I wasn't overly…I wasn't overly…I wasn't think 'Oh I can't wait to get the boot off… can't wait to get the boot off.' 'Cos the boot was making me secure."*

Generally all participants felt that the final item of OMAS regarding work and activities of daily life was relevant and important to them. The item in question included activities of daily life, therefore those participants who didn't work also found this a relevant question citing the impact that their injury had on their usual activities.

Katie: *"Oh massively, massively important."*

Mike: *"Yeah really important I mean that's what I was sort of living for really."*

Sue: *"Erm yeah I think for peace…for mental reasons it's good to get back to normal errm that was quite important for me to feel as though I was able to take charge of my own life again."*

Mary: *"Very important. Erm [pause] I mean I'm still not there…yet."*

5.1.3 Comprehensiveness

Table 51 - Comprehensiveness of OMAS

	OMAS comprehensive (not missing anything)	**OMAS not comprehensive (misses items out)**	**Comprehensiveness not clear**
Number of individuals			

Table 51 shows the results for comprehensiveness of OMAS. Several individuals said that they felt the questionnaire was comprehensive and didn't miss anything out which they perceived as important during their recovery. The individuals who said the OMAS wasn't comprehensive stated that the questionnaire missed the following items out: operative wound, returning to driving, having other injures which might affect outcome and mental health throughout recovery. For one individual, it was difficult to ascertain comprehensiveness; he stated that the main things he was concerned with was sleeping and being able to carry things, which he felt were covered in OMAS, although not explicitly.

Richard: *"I think…the kinds of things I was thinking about 'can I do or not?' was sleeping and carrying things. They were the main two so they…they fall in here somewhat but they're not explicit."*

Katie: *"I think the only thing that I was surprised by was the numbness that I had the sensation around the scars as well. I know there are other questionnaires that that is featured in but that is something that I found quite…I wasn't expecting that."*

Interviewer: "*And if there was something that you could add into this questionnaire that was important in your ankle fracture recovery but not included here…is there anything you feel that this misses out?"*

Sue: "*Mental health. Yeah. Because I think that's the…frustration which causes depression really I think. I was depressed definitely.*"

5.1.4 Appropriateness of response options

Table 52 - Appropriateness of response options in OMAS

Questionnaire Item	Response options appropriate	Response options not appropriate	Appropriateness of response options unclear
Pain			
Stiffness			
Swelling			
Stair climbing			
Running			
Jumping			
Squatting			
Supports			
Work/activities of daily living			

Table 52 shows the responses from participants regarding the appropriateness of response options available in each question of OMAS. The two main issues with response options were issues with the description of response options as well as the number of response options available for some questions.

In order to visual present the results, I have listed each OMAS item as I dicuss them along with the response options in tables 53 to 61. These are all taken from the original article describing the OMAS [123].

When discussing the response options for pain, which are shown in table 53, some people felt that the response options weren't suitable as they all discussed pain whilst walking. Some highlights that they didn't know how to answer this question during

the period they were not walking yet and others suggested that the even-ness of the surface doesn't affect their pain. Some individuals felt that these response options were suitable to be able to answer the question.

Table 53 - Response options of pain question in OMAS

Pain	None
	While walking on uneven surface
	While walking on even surface outdoors
	While walking indoors
	Constant and severe

Richard: *"I wouldn't say the even-ness of the surface was a big factor in the effect on the pain. Erm so I didn't quite get those options erm if I was asked…I don't remember the even-ness of surfaces being a big issue."*

Katie: *"Yeah I feel like erm [pause] it kind of covers everything from erm…as I look at it that's the way I went through it…it was constant and severe and then it was you know it got better when I was walking indoors and then it got better as I was walking outdoors."*

The response options for the question on stiffness are show in table 54. The main issue people described with this question was the binary response, with some individuals saying that they feel this question would not show changes or improvements in their ankle stiffness. Others, however, felt that a binary response was appropriate here.

Table 54 - Response options of stiffness question in OMAS

Stiffness	None
	Stiffness

Sue: *"Hmm…maybe erm something in between 'none' and 'stiffness'? You know how bad is the stiffness? Maybe there could be something…'slight stiffness' or 'stiffness first thing in the morning' or…you know a bit more qualification of that…I'd have to say stiffness. Because it's not none."*

Richard: *"Erm [pause] I think…I mean…if I…it's an obvious way to do a binary choice I'm no questionnaire designer but I would've expected more than two options."*

Adam: *"Yeah…although [sighs] you see it's not stiff all the time but there is some there still. Again maybe just one in the middle possibly?"*

Katie: *"Yes because it's very simple. Have you got it or not? Yes I have so…"*

The response options for swelling are show in table 55. Some individuals found these fit for purposes, noting that they experienced evening ankle swelling during their recovery. However, others felt that the wording of the middle category was restrictive and not representative of the swelling they experienced, which sometimes was the worst at other times of the day rather than the evenings. Some felt they were unable to attribute their swelling to a particular time of day or that the swelling changed from day to day. For some this was unclear as they found it difficult to recall.

Table 55 - Response options of swelling question in OMAS

Swelling	None
	Only evenings
	Constant

Richard: *"Yeah I had a lot of swelling initially. It came down quite quickly when I…as soon as I had the cast. Certainly I didn't notice it changing in the evening as the question suggests. So actually I…I had a bit of trouble answering this for the purposes of the questionnaire because although there was swelling it was very minor and I think if I'd ticked 'I have constant swelling' it would feel to me like I'm saying something's worse than it actually is."*

Peter: *"Swelling, now the questions there....I wouldn't…I wouldn't agree with them cos' it's not only evenings, it's all day that the swelling's there. It's not constant though as in the afternoons when I've got it up the swelling goes down, so."*

Mary: *"Well I…I 'spose [pause]…it's not only in the evenings [pause] it…but then it you know it's…I haven't got it at…at night so I don't…again it almost needs another one…doesn't it?"*

For stairs the response options are show in table 56. The response options for this item were reasonable for the majority of individuals asked about this. Many acknowledged the presence of the intermediate option of "impaired" as being useful for answering this to demonstrate improvements in function. Several individuals described how stairs were never impossible and that they were always able to complete them one way or another.

Table 56 - Response options of stairs question in OMAS

Stairs	No problems
	Impaired
	Impossible

Simon: *"Well they…they…they're alright you can either get upstairs or you can't get upstairs or you can get upstairs with some difficulty. So that's fine."*

Katie: *"I think they're really appropriate because at first they are impossible and then I would say I'm just about in the…the mild end of impaired. I'm still having to use a bannister because I haven't got that motion and stuff."*

The response options for running are found in table 57. Despite this being a binary response option, the appropriateness was more unclear than the stiffness question. Some individuals felt that the binary response options provided for these question was appropriate, explaining that running is either something you can or can't do. Some individuals, however, explained that they were at a stage where they could probably run in some form if needed, but it would certainly be impaired or might be painful and therefore they advised that a middle option would be suitable in this instance.

Table 57 - Response options of running question in OMAS

Running	Possible
	Impossible

Mary: *"I think so aren't they…yes I mean you either can or you can't."*

Simon: *"Erm but that's not on there is it? It's…again that's a bit too binary. 'Cos you could…you could …you could possibly run with sort of a little bit of pain."*

Adam: *"Yeah. Yeah I could run to the bus stop or probably just about but you wouldn't wanna run any…any more than that."*

Table 58 shows the response options for jumping. Similarly to running, there was mixed response with the appropriateness of the options given for the jumping item. Some felt that the binary response choices were appropriate, again explaining that jumping is either something you can or can't do. Others, however, said the term impossible felt extreme, with Katie saying that she probably could jump if needed to.

Table 58 - Response options of jumping question in OMAS

Jumping	Possible
	Impossible

Katie: *"Erm a bit like running I feel like impossible is a really extreme term. Because like I say I mean if my life depended on it or if I had to jump out of a burning building or something I probably could jump."*

Adam: *"Yeah they're probably about right. You can either jump or you can't really. There's probably no in between."*

The response options for squatting are found in table 59. Again, there were several people who felt these response options weren't appropriate. The main reason for this was the binary nature of the response, many indicating that they would have preferred an "impaired" or middle ground option. Simon described this period of being impaired at an activity such as squatting as essential to the process of becoming fully functional at that activity again and how it is necessary to go through a period of being impaired to get to full recovery.

Table 59 - Response options of squatting question in OMAS

Squatting	No problems
	Impossible

Katie: *"I mean yeah again the same as running and jumping and squatting its two kind of polar opposites sometimes there should be something in the middle."*

Simon: *"Well 'cos you…you have got that transitional bit where you can do it but it just does hurt a little bit but you have to go through that…you can't just say 'oh I can't do it I'm not doing it' you've gotta push it… to be able to go from impossible to possible you're going to have to go through a phase of being impaired slightly…it's part of recovery."*

Adam: *"Yeah probably about right."*

The response options for supports are found in table 60. The majority of individuals were unclear on whether these response options were appropriate for them. In some cases the response options to this question seemed to cause some confusion, mainly with the middle response option of "taping, wrapping". Some people suggested that they didn't understand this term and others explicitly stated that they didn't. Individuals also felt that none of the options were suitable to suggest they were wearing their functional brace, so felt there was not a suitable response option for them here.

Table 60 - Response options of supports question in OMAS

Supports	None
	Taping, wrapping
	Stick or crutch

Mary: *"Err…supports…well as I say I'm just wearing a sort of Tubigrip at the moment…yeah. Which…you know…do they call that taping, wrapping? I don't know."*

Simon: *"So it's no supports…well where's the air…aircast boot?...Erm but the option for me was either the aircast boot support or the cast…and that's not on there. So I just…err I mean I would have answered that just to say…when I was using my crutches but even with…I don't know I can't remember what I put when I had the boot on. I'd have probably put none 'cos the boot's not an option."*

Table 61 shows the response options for the work and activities of daily living question on OMAS. There was mixed acceptability of these response options. One individual had difficulty with the term "loss of tempo" citing how he felt it wasn't clear and possibly could be interpreted in different ways. Another individual commented on how appropriate that term was to describe how his ankle affected his

work and activities of daily living. Mary, however, could not find a response option that suited her situation of being off work still from her usual part-time job. Simon also struggled citing that his job just has less hours but it's not simpler (i.e. hasn't changed in nature). Mike noted how two of the response options only relate to work, which was not applicable to him as he is retired.

Table 61 - Response options of work, activities of daily life question in OMAS

Work, activities of daily life	Same as before injury
	Loss of tempo
	Change to a simpler job/part-time work
	Severely impaired work capacity

Richard: *"Yeah so erm work was the same as before the injury throughout more or less. General activities were somewhat affected and I think the answer 'loss of tempo' on this questionnaire kind of captures that fine."*

Mike: *"I suppose not observation of [inaudible] but 'loss of tempo'…not quite sure what that means? Erm there may be a more plain English way to sort of express that. Erm cos' I think it [pause]…you know people would interpret that possibly in different ways."*

Mike: *"Err and obviously two of them relate to work that erm for people like me wouldn't be relevant."*

Mary: *"Same as before injury….well no I'm not back to where I was. Loss of tempo [pause] well yes 'cos I can't do what I did can I? It's…you know. Changed to a simpler part time job…well I only had a part time job anyway and hopefully I will go back to that."*

Simon: *"Yeah they're…if I'm over picky I don't think I …I could tick any of them because it's not the 'same as before'…yet. I don't think I've lost tempo. I've slowed down a little bit but that's just because I'm a bit more reflective. I'm…losing tempo is a bit more something you'd say that was indicative of just someone [pause] I dunno it's a bit more of a negative connotation to it. Erm I've not 'changed to a simpler job' although I am working part time at the moment but working part time shouldn't be…simpler job slash part-time is not the same thing is it? I'm back*

to…back to my normal job which definitely isn't a simple job but I'm am working rehabilitation hours erm and I will be just 'til the end of the month so. That…the…the last ones alright 'severely impaired work capacity' but they're just a bit…just a bit odd isn't it? Just how I interpret it."

5.2 Discussion

This qualitative study demonstrates the wide-ranging impact of ankle fractures, which cover further issues beyond the commonly understood symptoms of ankle fracture such a pain and swelling. When identifying items which were most important to individuals, many spoke of regaining their independence, through activities of daily living, walking unaided, transport and driving or usual caring roles and responsibilities. Other participants spoke about how radiological union visible to them on X-Ray was an important indicator of treatment success and a positive outcome for them. Individuals interviewed here spoke of the psychological effects of their injury, which seemed to affect the older participants more severely in this sample of individuals. Many described frustration throughout their recovery, as well describing emotional lability, depression or low mood.

Research into factors affecting outcome following orthopaedic trauma has demonstrated that the psychological symptoms such as anxiety and depression were predictors of poorer outcome, along with older age [212]. An observational study showed the prevalence of psychological distress that individuals recovering from orthopaedic trauma encountered, finding that as many as 22% of participants fulfilled the criteria for psychological illness during their recovery [213].

A study completed by Gong, et al. [214] studied the rates of depression in patients following either ORIF or cast immobilisation for a distal radius fracture. The authors concluded that there was no difference in depressive symptoms between those treated with an ORIF compared to cast immobilisation. Authors also found that 47% of the sample overall displayed symptoms of major depressive disorder two weeks following their injury. Another study by Crichlow, et al. [215] looked at the prevalence and severity of depression in orthopaedic trauma patients. The results found that of the sample included in the study, 45% displayed moderate, moderate-to-severe or severe depression. The authors found no association between the fracture severity and the incidence of depression, but did not an association between depression and open fractures. These articles show similar findings to the results presented here regarding the possible psychological and emotional effects of trauma and fractures.

The development of prognostic models which could pre-empt likely recovery trajectories and therefore identify individuals requiring more input would be advantageous. This has been research by Keene, et al. [216], who aimed to develop a prognostic model based on injury information, functional and lifestyle attributes. The research team found very little accuracy in using their four prognostic factors to predict outcome and they concluded that psychosocial factors may provide the most accurate predictors of outcome in ankle fracture. Evidence from other research articles suggests that factors such as pain related fear, catastrophizing and self-efficacy levels are moderate predictors of outcome following fractures [217] [218]. With the occurrence of adverse psychological effects so common and evidence that these factors may have a negative effect on outcomes, it's important that individuals recovering from orthopaedic trauma who may be at higher risk are provided with additional support and rehabilitation.

In the sample of individuals interviewed here, I also noted differences between older and younger participants in the way that the spoke of their injury recovery, particularly with regard to the difficulties in following weight-bearing restrictions and using walking aids. Older participants were most likely to display more of an aversion to walking aids and describe a difficulty or lack of ability to follow weight bearing restrictions. The notable psychological distress described by one individual regarding the communication of prolonged weight-bearing restrictions demonstrates the severe impact that these restrictions had on this person's life. Further research should aim to evaluate the most appropriate weight bearing regimes, particularly for operatively managed fractures, where audit data in the UK has demonstrated that practice can be varied and inconsistent [53]. It may be that, for some individuals, the benefits of a full weight bearing protocol after ankle fracture fixation outweigh the risks associated with it.

The results found here for patient experiences of ankle fracture corresponds to those found in a similar article by McPhail, et al. [173], who completed interviews with both patients and clinicians regarding the experiences of ankle fracture. They identified similar themes in the data they collected, such as physical, psychological, social, occupational and domestic issues. Another article by van der Sluis, et al. [219] evaluated the outcome of individuals using a postal questionnaire 6 months following an ankle fracture and found a range of physical, social and psychological effects including issues with sleep and fatigue, which were attributed to the injury. Indeed, 15% of the

105 included in the study reported symptoms of depression and 16% reported anxiety, which is consistent with findings here. Research completed by Pinsker, et al. [220] evaluated the experiences of individuals recovering from ankle reconstruction and authors identified a central theme of vigilance of the participants affected ankle. This is similar to the individuals' description of a heightened awareness of their ankle included in this study.

The patient views on the OMAS questionnaire were varied and somewhat difficult to interpret. Despite these variations between individuals, it is clear that OMAS may not be relevant and comprehensive for all individuals. Running and jumping were the items which most individuals here described as not relevant. The relevance of items such as squatting and supports were unclear. The majority of individuals felt that OMAS was comprehensive and captured everything of importance to them. Some individuals, however, identified missing aspects including mental wellbeing, impact of other injuries, returning to driving and concerns with the operative wound. The inclusion of concerns related to the operative wound indicates that the OMAS may not be suitable for both operatively and non-operatively managed individual. Furthermore, the inclusion of mental wellbeing as being a missing item in the questionnaire further acknowledges the impact that this injury can have on individual's psychological wellbeing as previously discussed.

The appropriateness of the response options available for OMAS was generated as a theme from the data collected in this study and some individuals felt the response options in the questionnaire were not suitable for answering the questionnaire items at all stages of their recovery. Some found difficulty in interpreting the description of response options and others felt binary response options were inappropriate to answer certain questions. It may be that if response options aren't appropriate, individuals may simply not answer the question. This causes significant issues with missing data for primary end point analysis when using a weighted score such as OMAS, where mean imputation of data is not feasible. Further research into missing data of OMAS within clinical trials is warranted to understand whether issues identified here could be contributing to missing data in clinical trials.

Results of the analysis of the OMAS are similar to those found by Wylde, et al. [221], who studied the patient perspective of a commonly used hip PROM, the Oxford Hip Score (OHS). This author team reviewed the comments left by participants on the questionnaires and analysed the themes of the annotations according to the content

of the issue described by participants who had undergone a total hip replacement in the past 24 months. Results were sorted into categories of annotation, which were issues with question clarity, measurement of pain, the use of restrictive or irrelevant questions, to explain the influence of co-morbidities on an answer and to answer a question with two or more answers. Whilst no patients here explained any issues with clarity of questions, there were some issues with the clarity of response options. The restrictive and irrelevant questions found by the author team mirrors the way in which some spoke of items in the OMAS. Several individuals stating how there wasn't an option which they could categorise themselves into, much like the OHS study results, whereby participants drew new boxes to demonstrate where they felt another response should be which was more appropriate.

There are some limitations to the research presented here. Many argue that information redundancy (also known as data saturation) is impossible to reach, because there would always be the possibility of finding new themes and ideas with every person you interviewed [206]. I accept that further themes could have been raised if I have interviewed more people for this study. However, the aim of qualitative research is depth rather than breadth [222] and upon transcribing the data, I felt that I had sufficient rich data for the analysis. Another limitation is the sample could have been more diverse in relation to ethnic background, as all the participants included in this study were from a white, British background. A variable of ethnicity was initially used within the purposive sampling framework, however, the individuals within the appropriate time frame for the study was not able to be controlled. The individuals I approached from other ethnic backgrounds declined or were not able to be contacted.

As is the case with all qualitative research, I acknowledge that my own background, beliefs, personal and professional experience will have influenced the way in which I approach data collection, analysis and interpretation [204]. The fact that I am a white female with a professional background in physiotherapy will have influenced the way in which data collection and analysis was approached. For example, in the case of data collection during the interview, my gender may have influenced the way in which I ask questions and the way in which respondents answer, based upon gender roles and norms [223]. Whilst this is less likely to be as evident when discussing things which are sensitive in nature, the influence of gender and other factors such as race

and socio-economic background will have influenced the way in which data was collected and analysed.

One instance where I felt gender roles may have influenced the data collected was when I reflected on the first interviews and noted that males were not talking about psychological effects of their injury as much as females. There are three possible reasons for this; because males tend to not experience these issues much as females. Secondly, because I did not ask the questions in an appropriate way to males or give them enough time to consider the question compared to female participants; it's possible that subconsciously I assumed these were not "male issues" or felt more difficult asking questions surrounding this subject. Finally, this could have happened because the individuals did not feel able to talk about these particular issues with me, possibly due to the gender roles and norms which are at play in the social world. When I realised this after a couple of interviews, I noted this in my reflexive diary and examined why I felt this was the case. I ensured from then that I would try to be more open in discussing these issues with participants regardless of gender, to ensure I was giving all participants space to talk freely and openly about these issues. I made every effort to continually examine, acknowledge and reflect of my beliefs and assumptions which fed into this process through the use of reflexivity. This ensured that I understood the role I played in the creation of the knowledge presented here, rather than believing I was completely objective to the research [204 205].

The fact that the independent researcher who completed the duplicate data analysis was also a physiotherapist is also a limitation. It would have been more beneficial to have a broad range of professional backgrounds involved in the analysis. Resource availability and suitable training and experience meant that someone from a different professional background wasn't available at the time required, which made this not feasible.

Despite issues identified with regarding the relevance and comprehensiveness of OMAS, research detailed in chapter 2 of this thesis shows that the instrument is the most commonly used in trials of interventions for ankle fracture. Furthermore, the analysis of registered but unpublished studies shows that OMAS is currently being used in ongoing studies of this nature. In order to gain a complete understanding of the measurement properties of the outcome measure to inform its ongoing use, it's pertinent to assess the remaining measurement properties of this widely used outcome measure, such as construct validity, reliability, responsiveness and aspects

of interpretability. Whilst studies have completed some of these assessments in translated versions of the questionnaire [168-170], there is currently no research assessing the measurement properties of the English version of the PROM. The next work project for this thesis will aim to assess the measurement properties of OMAS to gain a complete picture of the performance of this PROM within the context of measuring primary end points in clinical effectiveness trials.

The previous two chapters have outlined the qualitative work completed as part of this thesis, which examined the patient experience of ankle fracture and the factors of most importance to them in recovery. The results of this enabled a greater understanding of the construct of interest for these individuals. The subsequent chapter outlined the results for the relevance and comprehensiveness (content validity) of OMAS, whereby individuals were asked to comment on the content of the OMAS and the questionnaire. The next chapter will focus on different measurement properties of OMAS, including the construct validity, structural validity, internal consistency and interpretability of the score in a sample of adults with ankle fracture.

6. A study of the measurement properties of the Olerud Molander Ankle Score in adults with ankle fracture

Aim: The aim of this project was to explore the validity, internal consistency, responsiveness and interpretability of the OMAS in a sample of adults with an ankle fracture within the context of a randomised controlled trial.

Methods: The project was embedded in the AIR trial. An RCT comparing immobilisation methods for adults with ankle fracture. Data were collected at baseline, six, 10 and 16 weeks post injury. Outcomes collected were the OMAS, EQ-5D, DRI, MOXFQ and a global impression of change score. Structural validity was explored using exploratory factor analysis. Construct validity was assessed using hypothesis testing, Cronbach's alpha was calculated to determine the internal consistency and responsiveness was assessed through assessment of change in OMAS between the global impression of change score groups. Interpretability was assessed through assessment of edge effects and an estimation of the minimally important change was made using an anchor based method. Measurement properties were explored in subgroups determined by patient age to assess the functioning of the OMAS in these groups.

Results: Exploratory factor analysis showed two subscales within the score. Internal consistency of the score was acceptable at $\alpha = 0.76$. The majority of the hypotheses were met for construct validity, with the exception of the EQ-5D anxiety and depression domain. The responsiveness was shown using boxplots, which showed good visual changes in scores along with subjective change on the global impression of change score. There were no edge effects at any time points or overall and the minimally important change was estimated at 9.7 points.

Conclusion: Results presented here showed that the OMAS had sufficient construct validity and internal consistency in the patient population in the context of an RCT. The score demonstrated two distinct subscales and no edge effects were identified. These results provide useful information for individuals planning research in this patient population using this outcome measure.

This work has been submitted to a peer-reviewed journal:

McKeown, R., Parsons, H., Ellard, D. R and Kearney, R. S. (2020) A study of the measurement properties of the Olerud Molander Ankle Score in adults with an ankle fracture. *Bone and Joint Research* (submitted for publication).

This work has been accepted and submitted to national and international conferences:

Title: A study of the construct validity, internal consistency and interpretability of the Olerud Molander Ankle Score in adults with ankle fracture. (poster presentation) **Event:** EFORT Congress, June 2020, Vienna, Austria (accepted but not presented due to the COVID-19 pandemic)

Title: A study of the validity, reliability and interpretability of the Olerud Molander Ankle Score in adults with ankle fracture. (submitted abstract) **Event:** BOA Annual Congress, September 2020 (abstract submitted but conference cancelled prior to abstract selection due to the COVID-19 pandemic - awaiting to receive outcome of application for presentation at virtual congress)

6.1 Introduction

As discussed in previous sections of this thesis, the use of PROMs with adequate measurement properties is of paramount importance in randomised controlled trials [82]. I found in chapter two of this thesis that OMAS is the most commonly collected outcome measure in clinical trials which evaluate individuals who have sustained an ankle fracture. In chapter three, I completed a systematic review which demonstrated that the PROMs collected in this type of study, including the OMAS, do not have adequate evidence of sufficient measurement properties to enable recommendation for use. During this chapter I also found that, whilst there were four articles on the OMAS; one development paper and three validation papers, all of these had been completed in different languages (Swedish, Norwegian and Turkish). There is currently no validation evidence for OMAS in the English version of the questionnaire, despite it being previously and currently used in several trials in an English population [41 44 54 72].Whilst there is a lack of evidence to support the use of these PROMs, there is also no evidence to demonstrate insufficient measurement properties and the review highlighted several evidence gaps in the measurement properties of these instruments.

In chapter four, I explored the construct of interest in individuals with ankle fracture. I used qualitative research methods to explore the patient experience of ankle fracture recovery using semi-structured interviews. Throughout the interviews, I also explored the relevance and comprehensiveness of the OMAS questionnaire in the same sample of participants. As discussed in chapter four, not all items were thought to be relevant by the individuals interviewed. These included the running and jumping items. Several individuals had suggestions for further items to be included which were currently missing from the questionnaire. These included psychological wellbeing, return to driving and wound issues. Finally, several people indicated that the response options may not be fully acceptable to them in answering the questions, such as issues with binary response options and with inappropriate response options which did not reflect the individuals' status.

OMAS remains the most commonly used questionnaire in trials for people recovering from an ankle fracture. Assessing the measurement properties of this commonly used score is essential in understanding whether the use of this questionnaire should be continued in this study type in the population of adults with an ankle fracture. The aim of this research project is to assess the measurement

properties of the scores of OMAS in a sample of adults recovering from an ankle fracture, adhering to the COSMIN taxonomy of definitions [93].

Figure 14 shows the COSMIN taxonomy of measurement property definitions.

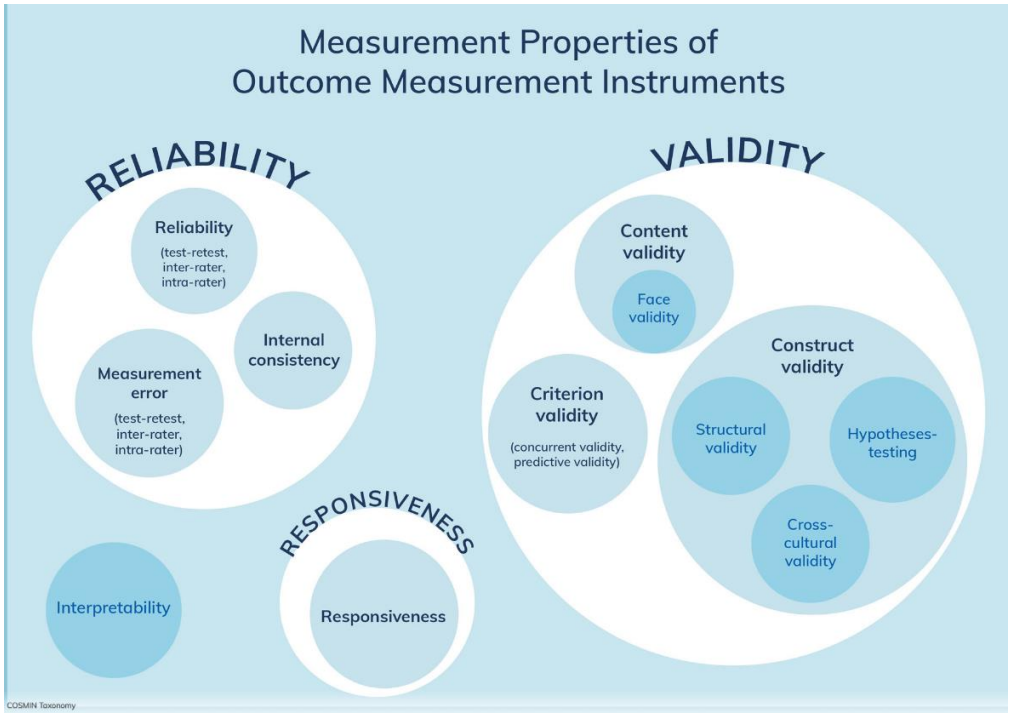


Figure 14 - COSMIN taxonomy of measurement properties [224]

6.1.1 Measurement theory

There are two main measurement theories used in the development and evaluation of PROMs used to measure latent variables; classical test theory (CTT) and item-response theory (IRT) [225]. CTT is a more traditional theory of approaching evaluation of measurement scales by determining how well the latent variables are represented by the observable items on the scale [226]. CTT is founded on the assumption that there are three components to PROM scores; the observable score (X), the true score (T) and the error score (E) and this is depicted in the formula X = T + E [226 227]. CTT is described as "test based" and is most concerned with the functioning of the test as a whole at a population level [228].

The advantages of CTT are that it's familiar with a wide range of individuals and therefore it's widely used and understood [226]. It's also relatively straightforward to complete with routinely used statistical packages [226]. The disadvantages are that it is less sensitive to understanding the issues with specific items within a questionnaire, which can distort the differences in scores obtained [226]. There are also limited

options for assessing item-level bias in different groups of individuals and item properties within a scale are based upon correlations of the sample, and therefore these cannot easily be compared with samples of different populations [226 229].

IRT refers to a more modern measurement theory of using mathematical models to predict probability of individuals' responses to PROM items based upon their level of the latent variable being assessed by the scale [225]. In PROMs which are based on an IRT model, the items are able to be placed in a hierarchical order of easy to difficult in relation to the latent trait being measured [230]. IRT is described as "item based" and is most concerned with the individual items within the test [228].

The advantages of IRT are the hierarchical nature of the items are best suited to the measurement of a unidimensional construct, that is a construct which measure one underlying variable or trait [230]. IRT also allows for a much more robust method of assessing possible biases which could occur at an item level due to socio-cultural difference between groups, known as differential item functioning [229]. This theory, however, requires specialist statistical software to do these analyses [231].

Here, CTT is the most appropriate measurement theory because there is no evidence that OMAS was developed using IRT and testing the PROM based upon this theory is therefore not appropriate [123 232]. Whilst assessing the differential item functioning is important in the cross-cultural validation of a PROM, this is not a research aim of this project as we do not have the data from different populations to do this analysis. Furthermore, because outputs of this project will be disseminated to clinicians regarding the suitability of use of OMAS within clinical trials, it's important that the psychometric testing methods used are accessible for all, not just those with a specialist interests in measurement theory. Because CTT is the most universally used and understood, it is best suited for this specific audience [226]. Finally, using CTT methods to assess PROMs is known to be more accessible for researchers who are new to this particular research methodology [226] in comparison to the practical issues of the statistical software and skills required for IRT analysis [231].

6.1.2 Aim

The aim of this research project is to explore the validity, internal consistency, responsiveness and interpretability of the OMAS in a sample of adults with an ankle fracture, within the context of a clinical trial.

6.1.3 Research question

What is the validity, internal consistency, responsiveness and interpretability of the OMAS in a sample of adults recovering from an ankle fracture?

6.1.4 Objectives

The following objectives will be achieved in a sample of adults with ankle fracture:

1. To assess the convergent validity of the scores of OMAS.
2. To assess the discriminative validity of the scores of OMAS.
3. To determine the internal structure of OMAS scores.
4. To assess the internal consistency and per item redundancy of OMAS scores.
5. To assess the responsiveness of OMAS.
6. To assess whether there are any floor and ceiling effects present in scores of the OMAS.
7. To explore the minimally important change in scores of OMAS.
8. To describe the above measurement properties of OMAS in subgroups according to age.

6.2 Methods

6.2.1 Ethical approval

This project gained favourable ethical approval from the West Midlands Edgbaston NHS Research Ethics Committee (Reference 17/WM/0239) on 30/04/2019 through a substantial amendment to the AIR trial protocol [72].

6.2.2 Participants

This project used data collected in the AIR trial and my active role in the set up and ongoing management of the trial is outlined in further detail in the declaration section of this thesis. Individuals included had closed ankle fractures and were aged ≥18years of age. The eligibility criteria for AIR can be found in section 1.5.2.

6.2.3 Sample size

There is limited guidance and published literature into the sample size calculations for studies of the measurement properties of PROMs. A systematic review into sample size justification for this type of study has shown that practice is varied and highlights a lack of guidelines in the formulation of sample sizes for validation studies [233]. de Vet HCW [94] recommends a minimum of 50 patients for studies of

construct validity, but ideally over 100 patients, with a minimum of 50 patients per subgroup for known groups validation.

For exploratory factor analysis, some authors describe a subject to item ratio of 10:1, meaning for every item in a score, ten participants are required to test the score [234]. Other researchers recommend a definitive minimum sample, which range from 50 to 400 [235]. For factor analysis de Vet HCW [94] recommend having a sample of 4-10 patients per item in the questionnaire under assessment, or ≥100 participants, whichever number is higher. In the COSMIN risk of bias checklist the criteria to reach a score of "very good" on the structural validity section is a sample size of seven times the number of items in the questionnaire and ≥100 patients [164].

OMAS contains 9 items, a minimum of 63 patients is required to achieve this criteria. Aiming for the higher level of 10 patients per item, this would be 90 patients. As both of these figures are less than the minimum of 100, I aimed to ensure that data for a minimum of 100 patients were retrieved for completing the factor analysis section of this particular project.

At the time of the data extraction for this project, the total number of patients recruited was approximately 650 participants. Therefore prior to retrieving the data from the online database I was confident that a sufficient amount of data were available for assessing the measurement properties of scores of OMAS using both the minimum absolute sample of 100 described by the COSMIN group and the subject to item ratios as described by other researchers for factor analysis [164 235]. With regards to discriminative validation, it should be possible to ensure that there are 50 individuals in each group based up on the numbers of participants in the trial so far, however until the data is retrieved and examined, this is not known.

6.2.4 Data collection

Baseline questionnaires were completed in fracture clinic at the time of recruitment. The remaining follow up questionnaires were collected by post in the first instance. In situations where this was not achieved (i.e. participants did not return the postal questionnaires), the participants were telephoned to obtain the questionnaire data over the phone. If both of these were unsuccessful within the appropriate timeframe of the follow up period then this was assigned as missing data and recorded in the follow up rates of the trial. Appendix 7 shows the AIR study documents, which included the consent form and a sample questionnaires. My involvement in the

collaborative design of these documents is outlined in the declaration section of this thesis.

The AIR CI (RSK) and TMG reviewed the protocol and gave permission for the data to be provided for purposes of the project. The AIR statistician, Helen Parsons (HP), accessed the statistics view of the AIR online application and retrieved the necessary data for the project on 24/07/2019. All data were pseudonymised by trial ID number and no information on the trial treatment allocation, randomising hospital site or personal information or contact details were included in the extracted dataset. Prior to saving the dataset, all dates of birth were converted to age in years to maintain participant confidentiality. The data were presented in Microsoft Excel format and saved to University of Warwick secure servers on files accessible only to authorised trial personnel. All data was handled in accordance with the Data Protection Act (2018) and with reference to WCTU SOPs regarding Data management and security.

6.2.5 Missing data

There are two types of missing data possible in this project; unit non-response and item non-response [236]. Unit non-response refers to a situation whereby an individual has not responded to the whole questionnaire and the most common reasons are that they have not returned their questionnaire within the time frame or they have withdrawn from the study. Item non-response refers to the situation whereby a participant has missed one or more items in the questionnaire which they return to the trial office, resulting in a partially completed questionnaire [236]. The absolute unit non-response rates were not explored here but in the final results article for the clinical trial. I recorded the follow up rates of the trial at the time of data retrieval for this project, indicating the number of participants who had responded and the number of participants who had reached that time point, which is also displayed as a percentage.

There are no guidelines outlined for dealing with missing items within the OMAS [123]. In order to deal with item non-response in this project, I used complete case analysis, whereby only questionnaires which are complete were included and those which had one or more item of missing data were excluded from analysis [236]. This disadvantages of using complete case analysis are that we cannot be sure that the items of missing data are missing completely at random (MCAR) [237]. This means that there are likely to be specific reasons why items are missing and this may mean

there are important differences between individuals who complete the questionnaire fully and those who do not [238].

Another option in dealing with missing data would be to use a method of mean imputation to assign missing values of OMAS [237]. This is, however, not possible here because OMAS items are weighted and different amounts of scores are given for different questions. Using the last value carried forward is also an approach I could have used [237]. Whilst the approach of complete case analysis does have disadvantages, as outlined above, this is a study primarily of the performance of a PROM in terms of the measurement properties of its scores. By making assumptions about those scores through imputation, it defeats the object of these analyses with regard to an understanding of the true measurement properties of the scores included here. One of the issues of using complete case analysis in clinical trials is that it reduces the amount of available data and therefore power in order to answer the question of which intervention is best for the population [238]. This is not an issue for this particular exploratory study into the measurement properties of a PROM.

For purposes of consistency across the project, all comparator questionnaires which contained missing items were excluded from the analysis (e.g. MOXFQ, EQ-5D-5L and DRI). For the EQ-5D, only the first five multiple choice questions are involved in the calculation of the value set. The overall health state question at the end does not contribute to the overall value set. Therefore, questionnaires will only be excluded in instances whereby one of the five multiple choice questionnaires have been omitted. In instances whereby the overall health state have been missed out but the five multiple choice questions have been answered, these were included in the analyses which require a total EQ-5D value set and excluded from any analyses which require the health thermometer score.

Information on missing data was presented by item for OMAS. For all questionnaires, I report the total number of questionnaires which we would expect if we had all the data for all participants included in this study (i.e. total number possible). I will also report the number of questionnaires we have with at least one item completed (i.e. number of questionnaires received). I also report the number of these questionnaires which have at least one item missing from it (i.e. partially completed questionnaires), which were excluded from the analyses based upon the complete case analysis approach.

In instances whereby subscales or domains were analysed (for example in the convergent validity analysis), I utilised all possible data. This means that, if a participant had only omitted a data point from one subscale, but the remaining subscales are complete, these data were used for the complete subscales, where appropriate. This maximised the amount of valid data able to be used in this project.

6.2.6 Outcome measures

All outcome measures described below are found in the AIR case report form (CRF) in appendix 7.

The OMAS is a nine-item questionnaire to measure outcome for people who have sustained an ankle fracture [123]. It consists of items of pain, stiffness, swelling, stair climbing, jumping, running, squatting and activities of daily living in a multiple choice, single response answer format. The questionnaire is scored out of 100 with higher values indicating a more favourable outcome for individuals. The scores for each question are weighted so each question is worth a varying number of points in total. The OMAS was collected at baseline, six, 10 and 16 weeks in the AIR trial.

The EQ-5D-5L score was developed by EuroQol and is a measure of health related quality of life designed specifically to measure cost utility and calculate quality adjusted life years [91]. It consists of five domains relating to mobility, washing and dressing, work and activities of daily living, pain and discomfort and anxiety and depression. Each of these questions has a total of five response options ranging from no problems at all to unable to complete the task completely. Respondents answer each question by selecting a single response and the results of these five scores are calculated to form a value set, which ranges from -0.594 – 1 in the case of the crosswalk value set (explained in further detail below). The final question is a health thermometer, asking the respondent to rate their health out of 100 on a thermometer diagram, with 100 being the best quality health and 0 the worst.

The National Institute for Health and Care Excellence (NICE) have completed an extensive review of value sets for calculating EQ-5D scores in England [239]. This review has recommended against use of the value set developed originally for the UK by Devlin, et al. [91] until more appropriate ones are developed. Therefore, NICE recommend the use of the mapping function, known as crosswalk value sets, developed by van Hout, et al. [240] for use in the interim period. Therefore, the crosswalk value set was used to calculate scores of the EQ-5D-5L for this project in

the absence of a newly developed value set. The EQ-5D-5L questionnaire was collected at baseline, six, 10 and 16 weeks in the AIR trial.

The DRI is a 12-item tool to assess disability related with a condition or injury of the lower limb [241]. The questionnaire has Likert scale questions asking respondents to rate how easy or difficult a particular task is by marking a vertical line on a 10cm line with a vertical line. Items included in the score range from walking, carrying a bag, making a bed to running and high level sport activities. Scores range from 0 to 100 with higher scores indicating higher levels of disability for the individuals completing the questionnaire. Scores are calculated by measuring the distance from the end of the line to the marked vertical line in millimetres, recording the measurement, adding all measures up and dividing by 12 to generate the mean disability score. The DRI was collected at baseline, six, 10 and 16 weeks in the AIR trial.

The MOXFQ is a 16-item questionnaire assessing outcome following foot and ankle disorders [124]. The original version is a questionnaire related to only foot disorders but the foot and ankle version covers disorders of both anatomical areas. The questions are single answer multiple choice questions and ask patients to recall the past four weeks. The questionnaire consists of three domains; pain, walking-standing and social interaction. The scores range from 0 to 100 with lower scores indicating a better outcome. The MOXFQ was collected at baseline and 16 weeks in the AIR trial.

The global impression of change (GIC) score was used to estimate the responsiveness and minimally important change of the OMAS in the patient population. This questionnaire was adapted from the HTA report on PROMs in clinical trials by Fitzpatrick, et al. [82], the score asks the participant to rate how much their ankle has improved since the last time point. In this instance the GIC score was collected at the 16 week follow up time-point only and asked patients to indicate how their ankle is since their last questionnaire at 10 weeks.

6.2.7 Data analysis

Data for this project were analysed using SPSS (IBM version 24). The significance level was set at 0.05. Descriptive statistics were used to describe the sample of participants included in this study. Means and standard deviations were calculated for continuous variables. Descriptors presented as part of this project were demographic information. These were age, gender, fracture management, fracture

classifications and mechanism of injury. These were shown as a whole sample and also by the specified subgroups outlined in section 6.2.8.

Descriptive statistics of the outcomes were reported for both the total sample and the subgroups. The distribution, kurtosis, skewness, range, mean and standard deviation of OMAS scores were calculated. Box plots of the OMAS data for each time point were made to give a visual representation of the data. An assessment of missing items present in the OMAS was calculated using simple descriptive statistics, to describe the number of times each item in the scores was missed by participants. The distribution of the PROMs used in this project (OMAS, EQ-5D, MOXFQ, GIC score and DRI) were analysed using histograms to give a visual representation of the data.

The techniques of analysis for each measurement property assessed are outlined under the appropriate headings below.

6.2.7.a Construct validity

Construct validity is defined as the degree to which the scores of a PROM are consistent with hypotheses set concerning internal relationships, relationships with other scores or differences between relevant groups [94].

Construct validity of OMAS was assessed by determining the convergent validity, discriminative validity and structural validity of the scores. Each of these measurement properties are described below, along with an explanation of the methods used here.

Convergent validity

Convergent validity is a subcategory of construct validity which relates to the correlation of scores of a PROM with scores of comparator instruments measuring the same or related constructs of interest [94]. To assess the convergent validity, OMAS scores were correlated with scores of the three other PROMs collected as part of the AIR trial, which are EQ-5D, MOXFQ and DRI. The scores of OMAS were also correlated with the scores of the available domains of these PROMs. This included the six domains of the EQ-5D (mobility, pain, anxiety and depression, usual activities, self-care and overall health state) and the three domains of the MOXFQ (pain, social interaction and walking standing). This totalled 12 different correlations; three of OMAS and the comparator instruments (MOXFQ, EQ-5D and

DRI) and nine of OMAS and the scores of the subdomains of comparator instruments (six for EQ-5D and three for MOXFQ).

As previously discussed, convergent validity was assessed through hypotheses testing, whereby a set of hypotheses are set to establish expected associations with scores of the comparator PROMs and subscales of these PROMs where appropriate. Abma, et al. [167] suggest using a cut off of 75% for sufficient construct validity using hypotheses testing. Therefore, for OMAS to be considered construct valid against the hypotheses set a priori, nine of the 12 hypotheses were required to be met with regard to the magnitude of correlations between scores of instruments/domains.

To ensure that data included in the analysis was independent and did not display repeated measures effects, each patient will be entered into the analysis once. To achieve this, I assigned a single time point for each participant using a random number generator and this time point was included in the analysis. Any time points for which a questionnaire does not exist (i.e. the patient did not return their questionnaire for this particular selected time point) was excluded from analysis. For the subgroup analysis, this random sample was split into subgroups by age (as outlined in section 6.2.8).

A visual check on the normality of the data was completed prior to this analysis, using a quantile-quantile (QQ) plot and histogram. Assuming the data are normally distributed, Pearson's correlation coefficient was used to compute the correlations. If the data had not normally distributed Spearman's correlation coefficient would have been used. Correlations are displayed in table format as well as using scatter plots.

Null Hypothesis: There will be a low correlation between the OMAS scores and scores of the MOXFQ, DRI and EQ-5D-5L and domains of scores, defined as $r < 0.4$.

Alternative Hypotheses:

1. OMAS scores will correlate highly negatively ($r \leq -0.7$) to the MOXFQ and the three MOXFQ domains of walking standing, pain and social interaction.
2. OMAS scores will correlate moderately negatively ($r \leq -0.5$) to DRI scores
3. OMAS scores will correlate moderately positively ($r \geq 0.5$) to EQ-5D scores and the EQ-5D overall health state domain and moderately negatively ($r \leq -0.5$) to the EQ-5D mobility, pain & discomfort, anxiety & depression, self-care and usual activities domains.

Because both MOXFQ and OMAS are ankle specific outcome measures and therefore measure a similar construct, thus in theory the scores of these PROMS were expected to correlate highly. Therefore the hypothesis for the expected magnitudes of these correlations were set at the higher level [167].

The DRI and EQ-5D are more general measures of disability and health related quality of life respectively, therefore I hypothesised that the scores of these instruments would be less highly correlated with scores of OMAS. However, it was predicted that these scores would still demonstrate a moderate association with scores of OMAS due to the known impact of ankle fractures on individuals' levels of health related quality of life and disability. Therefore, the hypothesis was that there would be moderate correlation between scores of OMAS and scores of the EQ-5D, domains of EQ-5D and the DRI.

Discriminative validity

Discriminative validity is a facet of construct validity and is defined as the ability of a PROM to differentiate between clinically known subgroups, such as those with severe and mild disease or impairment [94]. There is very limited research regarding what constitutes a severe ankle fracture and therefore several groups will be explored here. The subgroups of interest were the management of the fracture (operative or non-operative), the Weber classification, and the number of malleoli involved in the fracture. These were used as surrogates for injury severity in this project.

Two groups based upon the operative management of the fractures were used as a surrogate for injury severity. This is because ankle fractures which require operative intervention could be argued to be more severe than those which don't because the fracture is unstable, normal joint lines are disrupted and operative intervention is required to provide stability and alignment at the joint [242]. The severity of associated soft tissues injury is also likely to be more significant in injuries which require operative intervention [9 243]. For example, a study assessing factors which could predict outcome in individuals with an ankle fracture found that individuals which had non-operative management of their ankle fracture had better PROM scores and ROM outcomes than those who had surgery [244]. For these reasons, I hypothesised that the mean OMAS scores will be lower in patients who require operative fixation for their ankle fracture (i.e. they will have poorer outcomes) than those who don't require operative fixation.

There are, however, exceptions to this. For example, a large RCT has shown that, in patients over the age of 65, non-operative intervention of unstable ankle fractures provides outcomes of a similar quality to non-operative [41]. Results of this trial show that clinicians may be re-evaluating the decision making criteria on whether to operate on an ankle fracture [42]. This question is currently being answered in a younger population of patients through an RCT [44]. Despite these exceptions, in general terms the majority of clinicians utilise more traditional criteria for deciding on how to manage unstable ankle fractures. Furthermore, if patients require close contact casting, they were excluded from AIR, therefore would not be included here.

The Weber classification system is a commonly used system in clinical practice for categorising ankle fractures and is outlined in section 1.3.3 of this thesis. The Weber classification has been used as a surrogate for injury severity due to the differing injury profiles of each of these injuries which could feasibly contribute to differing outcomes. From the descriptions of each classification, it's apparent that the greater the Weber classification, the more soft tissue disruption that occurs, therefore the higher the Weber classification, the more severe the soft tissue disruption and therefore injury would be [245]. A study completed by Kennedy, et al. [246] showed that, in the case of unimalleolar injuries, the Weber system is a good prognostic indicator of outcome in individuals with an ankle fracture. Additionally, there is also evidence to suggest that ankle fractures with concurrent syndesmosis injuries have poorer functional outcomes [247].

Therefore, for this reason, I hypothesised that the mean OMAS scores would be higher (i.e. patients score better) in individuals with a Weber A, compared to Weber B and higher in Weber B compared to Weber C. The individuals with fractures which are not classifiable under the Weber classification system (identified as Weber N/A in this project) will be excluded from this analysis. This is because the only information we have about these fractures are that the lateral malleolus is not fractured and therefore we cannot be sure that this group contains similar types or severity of fractures. The numbers of patients in the Weber N/A group will be presented in the descriptive statistics section but will not be included in the analysis.

There are some fractures which are unable to be classified under the Weber classification, as described above. Therefore, one additional analysis was included here regarding the number of malleoli which are involved in the fracture. As discussed in section 1.3.3, there are three possibilities for this classification system

with regards to an individuals' ankle injury; one malleoli involved in fracture (unimalleolar), two malleoli involved in the fracture (bimalleolar) or three malleoli involved in the fracture (trimalleolar). The basis of this analysis is the hypothesis that the more bones involved in a fracture, the more significant the injury is likely to be. There has been some research done to assess fracture affecting a differing number of malleoli, and there is evidence to support the hypothesis that fracture involving more malleoli are associated with a poorer prognosis than isolated unimalleolar fractures [244 248 249]. Furthermore, there is evidence to suggest that fractures which involve the posterior malleolus are associated with poorer outcomes for individuals [250-253]. Therefore, one would expect that the mean OMAS scores of individuals with trimalleolar fractures to be lower than those with bimalleolar fracture and individuals with bimalleolar fractures would be lower than those with unimalleolar fractures because a bimalleolar fracture involves more fracture points than a unimalleolar.

Null hypothesis: The mean of OMAS scores will not be statistically significantly different between pre-specified subgroups of injury severity (i.e. operatively or non-operatively managed individuals, Weber classification and number of malleoli involved in fracture).

Alternative Hypotheses:

1. The mean OMAS scores of non-operatively managed patients will be statistically significantly higher than those of operatively managed patients at baseline, six, 10 and 16 weeks.
2. The mean OMAS scores of individuals with Weber B fractures will be statistically significantly higher than those with Weber C fractures at baseline, six, 10 and 16 weeks.
3. The mean OMAS scores of individuals with Weber A fractures will be statistically significantly higher than those with Weber B fractures at baseline, six, 10 and 16 weeks.
4. The mean OMAS scores of individuals with bimalleolar fractures will be statistically significantly higher than those with trimalleolar fractures at baseline, six, 10 and 16 weeks.
5. The mean OMAS scores of individuals with unimalleolar fractures will be statistically significantly higher than those with bimalleolar fractures at baseline, six, 10 and 16 weeks

As each of these hypotheses contain four time points, each time point as regarded as a separate hypothesis. Therefore, there were 20 hypotheses in total. In order to reach acceptable discriminative validity as per guidance by Abma, et al. [167], 75% of these would need to be met to regard OMAS as having sufficient discriminative validity, which is 15 of the 20. To test the hypothesis regarding operatively and non-operatively managed patients, Student's t-test was used to identify whether these means are significantly different. Prior to conducting the t-test, Levene's test was used to determine the homogeneity of variance of the groups.

To determine pairwise differences between fracture types using the Weber classification and number of malleoli involved in the fracture, analysis of variation (ANOVA) was utilised. Where difference was found to exist between these groups, I used post-hoc testing to determine the pairwise differences. I used Tukey's honest significant difference (HSD) test as a post-hoc test.

Internal structure

Structural validity refers to the degree to which the scores of a PROM are an adequate reflection of the dimensionality of the construct to be measured [94]. Because there is limited information available on the internal structure of the OMAS in the development paper or subsequent articles published on the performance of the measure, I used exploratory factor analysis to explore the internal structure of the score. Principal component analysis was used with orthogonal rotation using a varimax procedure [254]. A scree plot was formulated and the eigenvalues of factors were examined. Components which were found to have eigenvalues greater than one were extracted [255]. Item loadings of >0.45 were considered sufficient to be included within a subscale, which is a threshold used by other researchers completing similar analyses [241 256-258].

To ensure that data included in the analysis are independent with no repeated measures effects, each individuals were included in the analysis once. To achieve this, I used the process to selected a single time point as was described in the methods for the analysis of convergent validity. This involves allocating a time point for each participant using a random number generator and that time point is the one included in the analysis. Any time points for which a questionnaire does not exist (i.e. the patient did not return their questionnaire) meant that this individuals data was not included in the analysis.

6.2.7.b Reliability

Reliability is defined as the extent to which a PROM is free from measurement error [95]. To assess reliability, the internal consistency of the measure was calculated, which is a facet of reliability [93]. Ideally, I would have been able to assess the stability of the OMAS, testing the test-retest reliability of the score in patients whose ankle status has not changed [101]. Here, I felt this was not feasible or ethical as the sample used was those participating in a clinical trial. These individuals were already completing eight questionnaires as part of the trial follow up schedule and therefore adding a further questionnaire at an additional time point would be too burdensome for the participants. Therefore, to assess reliability, I completed an assessment of internal reliability only.

Internal consistency

Internal consistency measures the degree of interrelatedness of the items included in a questionnaire [93]. When items are combined to form a scale, internal consistency is important to ensure that they all contribute to measurement of the same construct and therefore should correlate with one another [259].

Internal consistency was assessed using Cronbach's alpha. The item redundancy of items within the OMAS was also explored, using the function of Cronbach's alpha if item deleted, to assess any items which are not contributing to the internal consistency of the score. An item was defined as redundant either if the Cronbach's alpha increases or remains constant when the item is removed from the analysis.

Hypothesis: OMAS will demonstrate sufficient levels of internal consistency with Cronbach's alpha of $\alpha \geq 0.70$ and $\alpha < 0.95$. No item redundancy will be observed in the items, which defined as the alpha remaining constant or decreasing when the item is removed from the analysis.

6.2.7.c Responsiveness

Responsiveness is the ability of a PROM to detect change in an individuals' condition when a one actually occurs [93]. The responsiveness of the OMAS was assessed using an anchor based method by calculating the change in OMAS score and plotting this against the GIC score groups. The GIC score was asked to participants at 16 weeks and can be found in appendix 7. It asks participants to rate how much their ankle has improved since their last questionnaire (10 weeks). The

mean changes in OMAS scores between 10 weeks and 16 weeks were calculated and plotted against each of the GIC score responses using a boxplot. The graph was visually inspected for changes in the distribution of OMAS along with changes in the GIC response.

Aim: To explore the changes in OMAS scores between 10 and 16 weeks in relation to the GIC score regarding individual's subjective rating of change during the same time period.

6.2.7.d Interpretability

Interpretability is the ability to assign qualitative or clinical understanding to scores or changes in scores of a PROM [93]. Interpretability of scores of OMAS was assessed by understanding if floor and ceiling effects (edge effects) are present in the scores of the sample studied here. An estimate of the minimally important change (MIC) of the score in this sample was also completed.

Floor and ceiling effects

Floor and ceiling effects are the proportion of scores of a PROM which are at the lowest or highest possible level [175]. It is undesirable to have a high proportion of scores which are at the highest or lowest level within a population because it means the score is unable to differentiate between individuals at either ends of the scale [176]. To assess whether OMAS demonstrates edge effects, I calculated the percentage of scores which were either 0 (minimum OMAS score) or 100 (maximum OMAS score). This was calculated at each time point (baseline, six, 10 and 16 weeks) as well as overall. Floor or ceiling effects are defined here as ≥15% of the scores being at the minimum or maximum score respectively. The parameter of 15% has been recommended by COSMIN [108], other researchers [176] and also utilised in other studies assessing edge effects of PROMs [175 177 260 261].

Minimally important change

MIC is defined as the smallest level of change in a PROM score which is perceived as relevant to the patient [97]. An anchor based method was used to make an estimation of the MIC using the data provided form the GIC score. The mean and standard deviation of change in OMAS score was calculated for participants in each GIC score response group. The mean change in OMAS score for the group who responded "minimally improved" between the 10 and 16 week time points was used as an

estimate for the MIC in the OMAS score. The mean changes for all GIC score groups was calculated for completeness.

6.2.8 Subgroup analyses

As well as assessing the above measurement properties in the overall sample of adults with a closed ankle fracture, I also assessed the functioning of OMAS in by age. As previously discussed, the literature surrounding epidemiology of ankle fractures suggests there is a bimodal distribution of ankle fractures, being most prevalent in younger males and older females [26 262].

There are several reasons for choosing to perform a subgroup analysis by age. Firstly, the work completed during chapter 4 explored the patient experience of ankle fractures and one of the findings was that older participants reported more substantial difficulties with weight bearing restrictions, use of walking aids and a more significant impact on their psychological wellbeing, usually attributed to a feeling of losing independence as a result of their injury. Additionally, the results outlined in chapter 5 highlighted that some of the items on OMAS, most commonly the items of running and jumping, were not perceived to be of importance to all individuals and this was more usually the case for the older participants of the study who did not perform these functional activities prior to their injury. There is indication from this work that the injury experience of younger individuals and older individuals could be substantially different, possibly meaning the performance of PROMs might differ between these groups. Therefore, the functioning of OMAS was assessed in subgroups according to age for these reasons.

The subgroups were created using the median age of the sample, which was be calculated following retrieval of the data and the sample was split into two, using the median age as the cut off point for the two subgroups. I chose to use the median as a cut off point for the subgroups for several reasons. Many trials of orthopaedic fixation techniques use ages as a surrogate for bone density, on the premise that surgical fixation quality maybe affected by the density of the bone. In studies of bone density loss over age, bone density loss has been estimated to occur more rapidly from age 50 onwards [263 264]. In studies assessing rehabilitation interventions, however, bone density is of lesser importance than other factors such as pre-injury activity levels, frailty and co-morbidities, which would have a greater effect on the outcome in studies of this nature. Therefore, I felt that to use age as a surrogate measure in this evaluation of measurement properties of OMAS was not appropriate.

Instead, the use of the median allowed for the sample to be split in two subgroups of older and younger patients using the mid-point of the ages included in this sample of patients.

The measurement properties of OMAS described above was analysed in both subgroups, where appropriate. The only measurement property which was not assessed in the subgroups was the internal structure of OMAS because this would not add any additional information when principal component analysis is completed by subgroup.

6.3 Results

6.3.1 Descriptive statistics of total sample

6.3.1a Demographic and injury data of total sample

Data for this project was retrieved on 26/07/2019 during which the trial was still in the recruitment phase. The total number of participants entered into the trial on that date was 638. Eighteen of these had been randomised but the trial office had not yet received their baseline data (i.e. demographic, injury and outcome data) therefore these were excluded from the analysis. Therefore, the sample size for this project is 620 participants.

The follow up rates for the trial at the time of the data extraction are outlined in table 62. At the time of data retrieval, 11 participants had been randomised but not yet reached the six week time point.

Table 62 - Follow up rates of AIR trial at time of data extraction

Time point	Follow up (n=receieved/n=reached time point)	Percentage follow up (%)
6 weeks	452/609	74
10 weeks	415/599	70
16 weeks	434/571	76

Table 63 - Participant age

Age (years)	
Number	620
Range	76
Minimum	18
Maximum	94
Mean	45.96
Std. deviation	16.73
Median	47
Mode	21

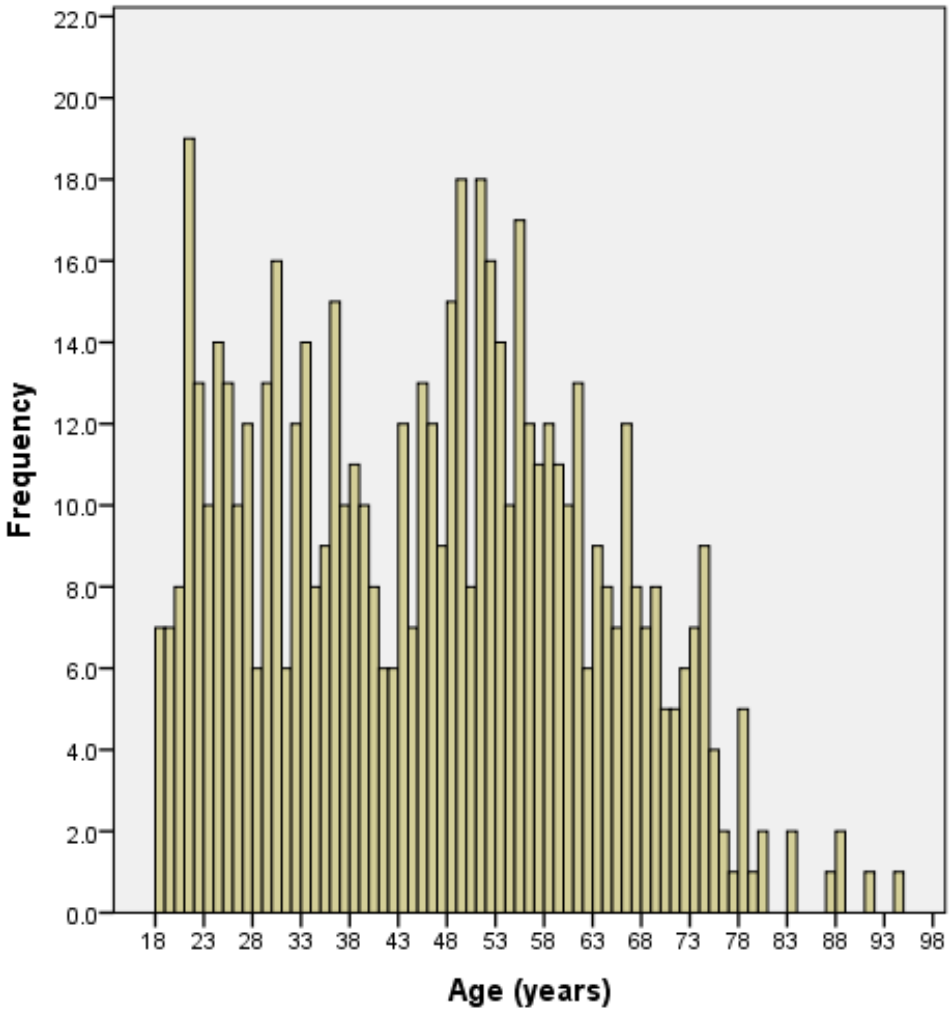


Figure 15 - Histogram showing participant age

The age range of participants in this sample of individuals was 18-94 years, with a mean of 45.96, these are shown in table 63. The median age was 47 years and therefore the groups for the subgroup analyses used during this project will be 47 years (yrs) and under and 48 years (yrs) and older. These will be shortened to ≤47yrs and ≥48yrs groups. A histogram for the ages of the sample is shown in figure 15.

Table 64 - Participant gender

Gender	Frequency (n=)	Percent (%)
Female	345	55.6
Male	275	44.4
Total number	620	100

Table 65 - Participant ethnicity

Ethnicity	Frequency (n=)	Percent (%)
White - English/Welsh/Scottish/Northern Irish/British, Irish, Gypsy, traveller, or any other white background	526	84.8
Mixed/multiple ethnic groups	13	2.1
Arab or any other ethnic group	5	0.8
Asian/Asian British - Indian, Pakistani, Bangladeshi, Chinese or any other Asian background	27	4.4
Black/African/Caribbean/Black British	20	3.2
Other	9	1.5
Missing data	20	3.2
Total	620	100

Table 66 - Side of injury of included participants

Side of injury	Frequency (n=)	Percent (%)
Right	298	48.1
Left	303	48.9
Missing	19	3.1
Total	620	100

Table 67 - Fracture management of included participants

Fracture management	Frequency (n=)	Percent (%)
Operative	331	53.4
Non-operative	289	46.6
Total	620	100

Table 68 - Mechanism of injury of included participants

Mechanism of injury	Frequency (n=)	Percent (%)
Road traffic accident	24	3.9
Crush injury	2	0.3
Contact sports injury	46	7.4
Low energy fall	381	61.5
High energy fall	93	15
Other	52	8.4
Missing	22	3.5
Total	620	100

Table 69 - Weber classification of included participants

Weber classification	Frequency (n=)	Percent (%)
Weber A	38	6.1
Weber B	409	66
Weber C	104	16.8
Weber N/A (no lateral malleolar fracture)	25	4
Missing	44	7.1
Total	620	100

Table 70 - Number of malleoli involved in fractures of included participants

Number of malleoli involved in fracture	Frequency (n=)	Percent (%)
One (unimalleolar)	406	65.5
Two (bimalleolar)	123	19.8
Three (trimalleolar)	65	10.5
Missing	26	4.2
Total	620	100

Of the 620 participants in this study, 345 (55.6%) were female, as shown in table 64 and 526 (84.8%) were from a white British background. Twenty-seven Participants (4.4%) were from an Asian/Asian British background, 20 (3.2%) were from a black/African/Caribbean/Black British background and 13 (2.1%) were from multiple or mixed ethnic groups. Five participants (0.8%) were from an Arab or any other ethnic background and 9 (1.5%) were from another background, with data for 20 participants missing for this information. These are shown in table 65.

Of the sample included here, 298 (48.1%) participants sustained an injury to their right ankle, with missing data for this item in 19 cases (3.1%). These data are shown in table 66. As shown in table 67, of the 620 participants, 331 individuals (53.4%) had operative fixation of their ankle fracture. A total of 381 individuals in the study (61.5%) sustained their injury in a low energy fall, defined as a fall from standing height or less. A high energy fall was the cause of 93 (15%) of the injuries, defined as a fall from a height larger than standing height or whilst doing an activity such as running or jumping. Forty-six individuals (7.4%) sustained their injury through contact sport injury and 24 (3.9%) individuals sustained their fracture in a road traffic accident. Two individuals (0.3%) sustained their fracture from a crush injury and 52 individuals (8.4%) sustained their fractures another way. Data for the mechanism of injury for 22 participants (3.5%) were missing and these data are shown in table 68.

The majority of fractures in the sample were Weber B, accounting for 409 fractures (66%). There were 104 (16.8%) Weber C fractures and 38 (6.1%) were Weber A fractures. A further 25 individuals (4%) had fractures which were not classifiable under the Weber classification system as they did not sustain a lateral malleolar fracture (Weber N/A). Data for Weber classification of 44 (7.1%) of the individuals

included in this study were missing. Of the 620 participants in the study, 406 (65.5%) had unimalleolar fractures, 123 (19.8%) had bimalleolar fractures and 65 (10.5%) had trimalleolar fractures, with data for 26 participants (4.2%) missing. Data regarding the fracture classifications of the sample are shown in tables 69 and 70.

6.3.1b Outcome data of total sample

There were 620 participants included in this study. Therefore, we would expect 620 multiplied by number of time points included in this study, which is four. That means that a total of 2480 OMAS questionnaires. There will, however, be some questionnaires missing for several possible reasons. Firstly some participants in this study will not have reached some of the time points (as shown in table 62 showing follow up rates of trial at the time of data retrieval). Some will have withdrawn from the study and some will have not returned questionnaires at the specified time points. For the purposes of this validation study, the reasons for each unit non-response was not recorded.

Table 71 shows the item non-response in the OMAS questionnaire results. Of the total expected number of OMAS responses, there were a total of 1873 responses of OMAS in the data retrieved, and of these questionnaires there were 106 instances where at least one item was missed on OMAS, leaving a total of 1767 fully completed OMAS scores.

Table 71 - Item non-response in OMAS answers

Questionnaire item	Number of responses	Missing	
	n=	Count (n=)	Percent (%)
Pain	1844	29	1.5
Stiffness	1863	10	0.5
Swelling	1852	21	1.1
Climbing stairs	1861	12	0.6
Running	1866	7	0.4
Jumping	1869	4	0.2
Squatting	1854	19	1
Supports	1862	11	0.6

Work, activities of daily life	1858	15	0.8

Table 71 shows the rates of item non-response in OMAS by question. The most commonly missed item is the first question concerning pain in 29 cases (1.5%), followed by swelling in 21 cases (1.1%) and squatting in 19 cases (1%). The least frequently missed questions were those concerning jumping, missed in 4 (0.2%) cases, running missed in 7 cases (0.4%) and stiffness, missed in 10 cases (0.5%).

MOXFQ was collected at baseline and 16 weeks, therefore we would expect that a total of 1240 completed MOXFQ questionnaires would be included here. Accounting for missing total questionnaires, there total of MOXFQ questionnaires received were 997. Of these, there were 22 instances where a questionnaire was missing at least one missing item, therefore these were excluded from analyses which required a total MOXFQ score, leaving a total of 975 completed scores.

DRI was completed at all four time points, therefore we would expect 2480 completed questionnaires. Here, we obtained data for 1799 DRI questionnaires. Out of this total, there were 49 instances of a questionnaires at least one item, therefore were excluded and left a total of 1750 valid responses for the DRI.

EQ-5D was also completed at all four time points, therefore we would expect 2480 completed questionnaires. From the total of 1870 of EQ-5D scores we received by the time of data retrieval, there were five instances of one or more item missing in the first five multiple choice questions, therefore these were excluded from the analyses which required a total EQ-5D value set. This left 1865 total EQ-5D value sets included in the analyses here. There were a further six instances whereby the individual had omitted the overall health state questionnaire. As discussed in the methods, these will be included in the analyses which required an overall EQ-5D value set but will be excluded in any analysis which looks at the overall health state alone (e.g. in the convergent validity analysis).

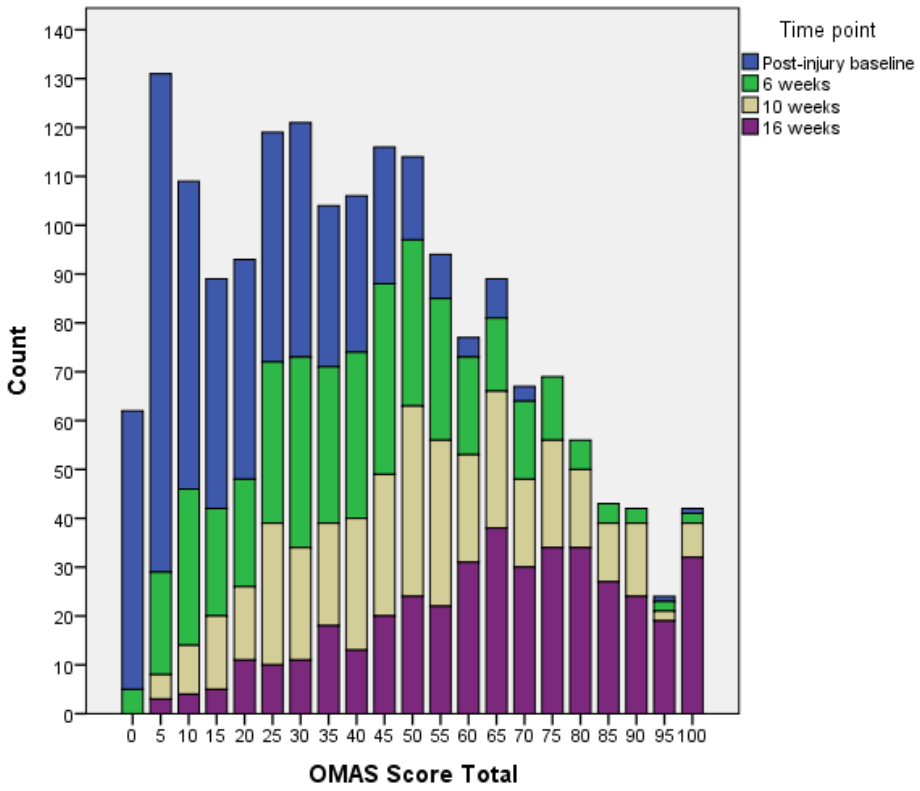


Figure 16 - Histogram of OMAS scores by time point

Table 72 - Descriptive statistics of OMAS scores overall

Number of OMAS responses (n=)	**1767**
Mean	41.98
Median	40
Standard deviation	26.68
Skewness	0.31
Kurtosis	-0.82
Range	100

Table 73 - Descriptive statistics of OMAS scores by time point

	Number of responses (n=)	Range	Mean	Standard deviation
Baseline	545	100	21.22	17.39
6 weeks	423	100	38.94	21.43
10 weeks	389	95	50.37	22.61
16 weeks	410	95	64.76	23.28

Figure 16 shows a histogram of OMAS scores by time point. Descriptive statistics for the OMAS score in the sample are shown in table 72 and 73. The mean OMAS score across all time points was 41.98(±26.68) and a median of 40. The means and standard deviations at each time point were; baseline 21.22(±17.39), six weeks 38.94(±21.43), 10 weeks 50.37(±22.61) and 16 weeks 64.76(±23.28). There were 1767 fully completed OMAS questionnaires in the dataset overall and a further 106 questionnaires which were missing at least one item.

The boxplot in figure 17 shows a visual representation of the OMAS scores across each time point. Figure 18 shows a QQ plot for OMAS scores.

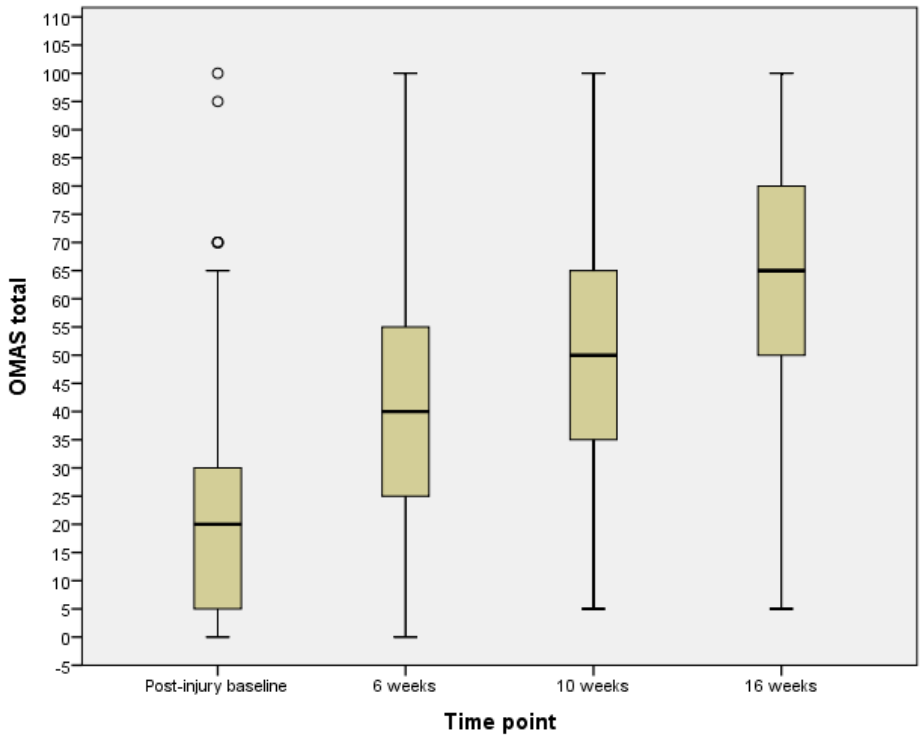


Figure 17 - Boxplot of OMAS scores by time point

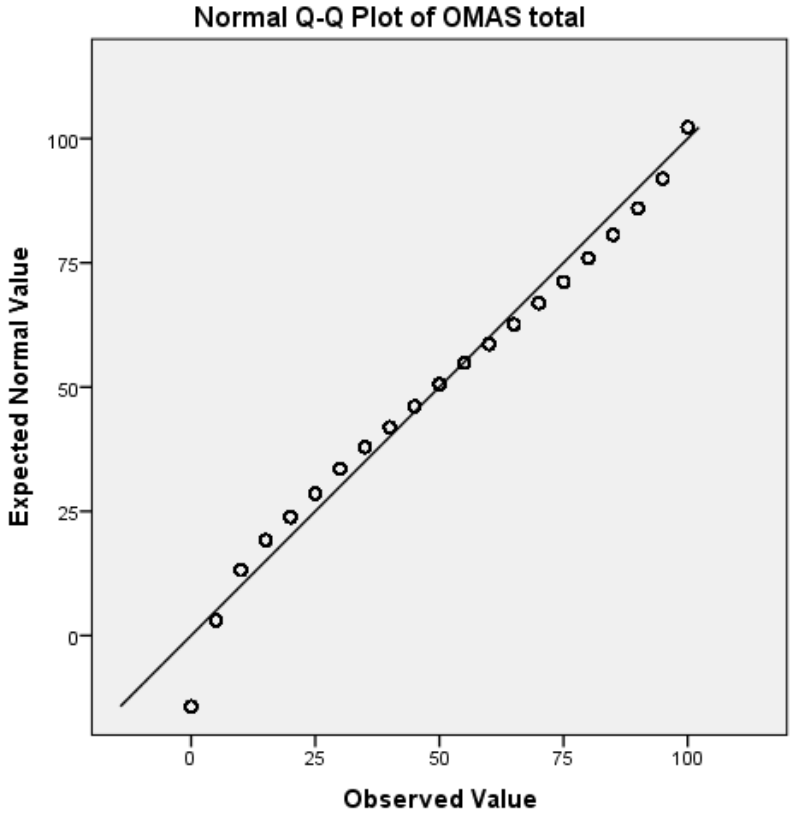


Figure 18 - QQ-plot of OMAS scores overall

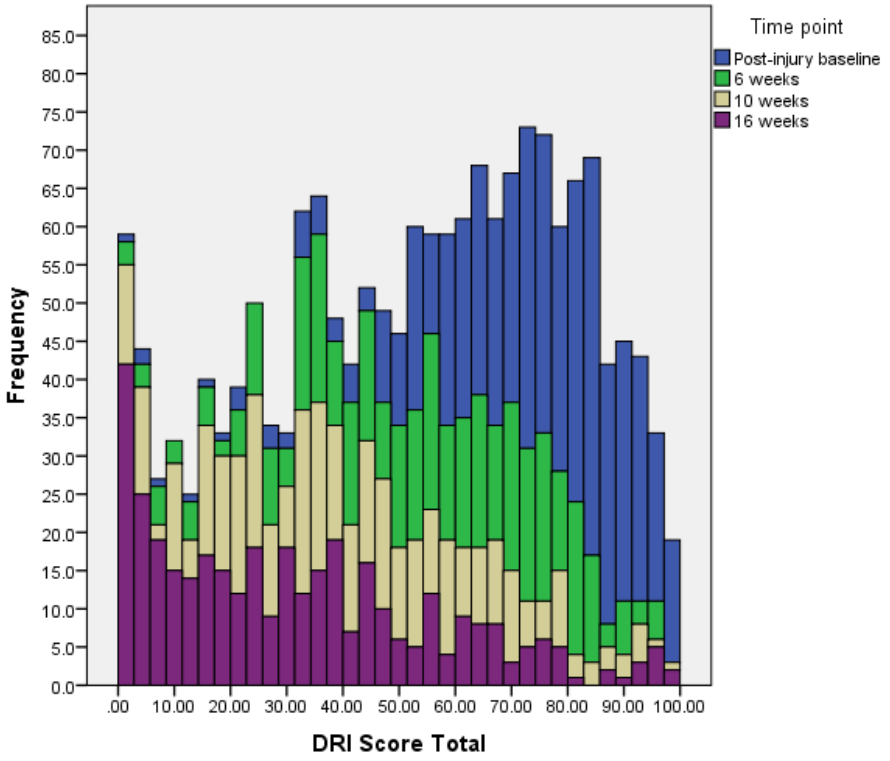


Figure 19 - Histogram of DRI scores by time point

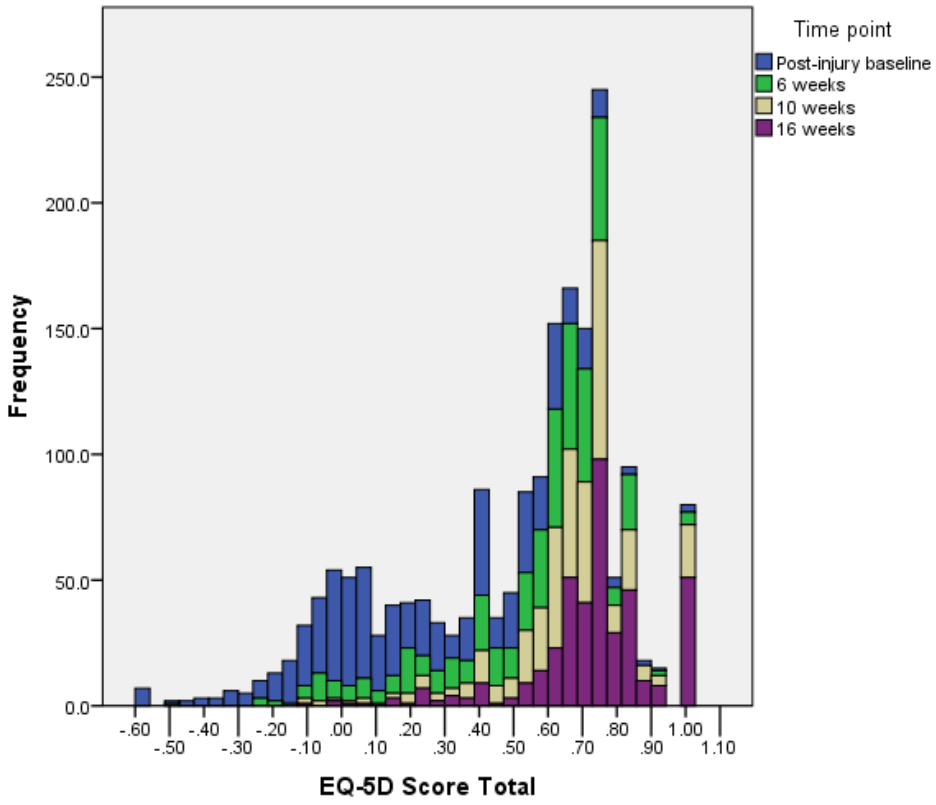


Figure 20 - Histogram of EQ-5D scores by time point

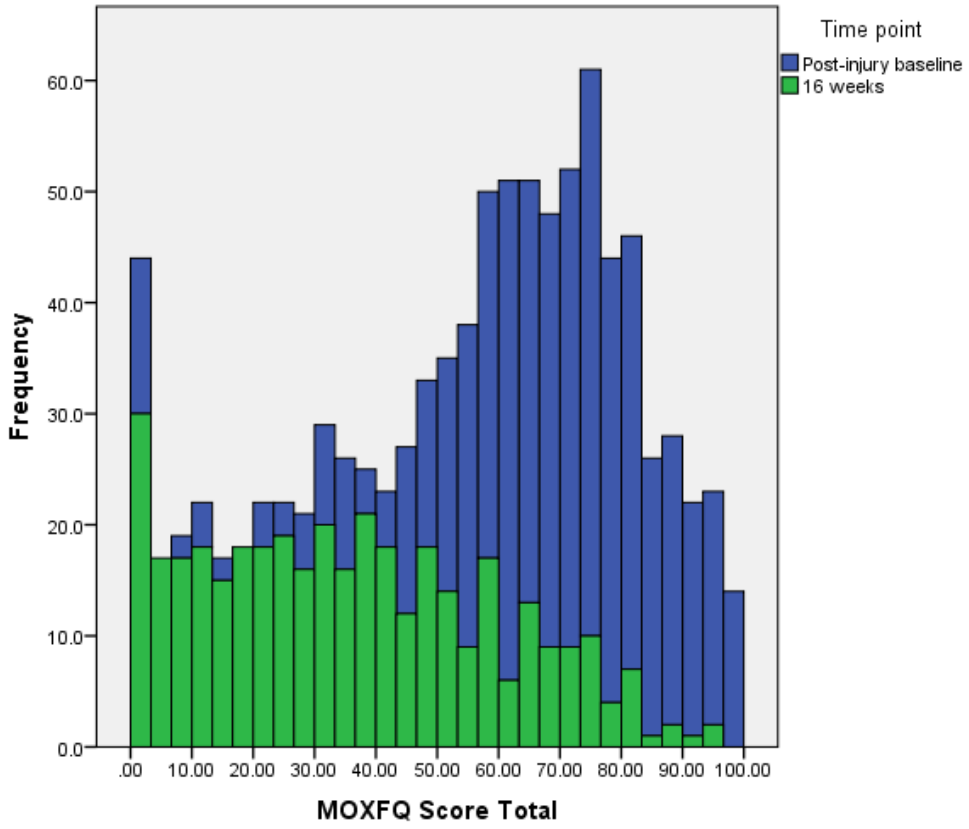


Figure 21 - Histogram of MOXFQ scores by time point

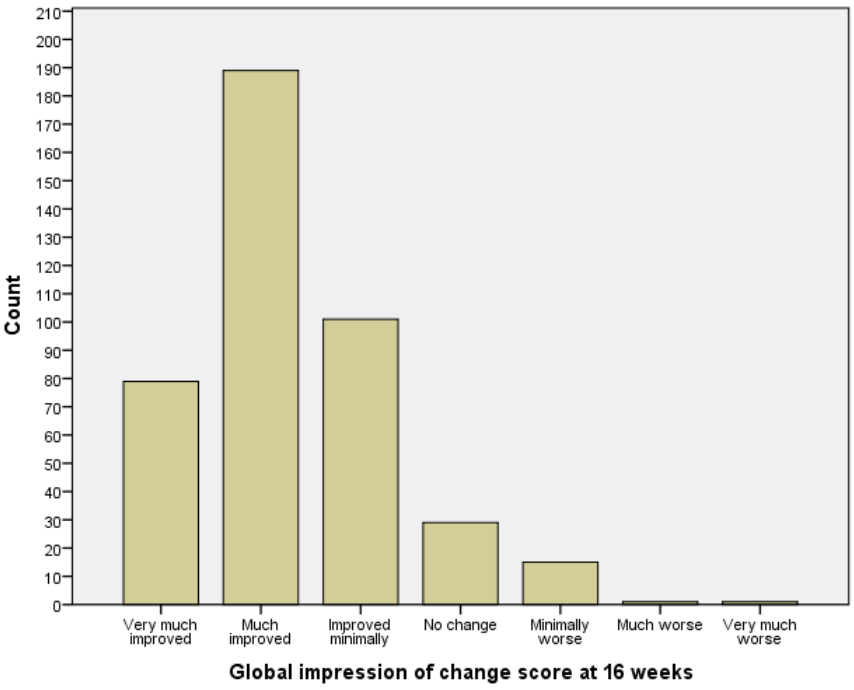


Figure 22 - Histogram of GIC score responses

Figures 19-22 show histograms for each comparator outcome measures (DRI, EQ-5D, MOXFQ and GIC score) showing results of reach time point. Table 74 shows the descriptive statistics for the GIC response groups by frequency and percent. From this, we can see that the majority of participants who responded to this question stated that their ankle improved in some way between 10 and 16 weeks. Seventy-nine patients (12.7%) stated it was very much improved, 189 (30.5%) reported it was much improved and 101 (16.3%) reported it had improved minimally. In contrast, 29 (4.7%) stated there was no change, 15 (2.4%) were minimally worse and there was one person (0.2%) who said they had either got much worse or very much worse respectively. Data for 205 (33.1%) people were missing.

Table 74 - GIC score responses

GIC response at 16 weeks	Frequency (n=)	Percent (%)
Very much improved	79	12.7
Much improved	189	30.5
Improved minimally	101	16.3
No change	29	4.7
Minimally worse	15	2.4
Much worse	1	0.2
Very much worse	1	0.2
Missing	205	33.1
Total	620	100

6.3.2 Descriptive statistics of subgroups

6.3.2a Demographic and injury data of subgroups

Table 75 - Participant age by subgroup

Age group	n=	Range	Minimum	Maximum	Mean	Standard deviation
≤47yrs	316	29	18	47	32.1	8.49
≥48yrs	304	46	48	94	60.4	9.37

Table 76 - Participant gender by subgroup

Age group	Gender	Frequency (n=)	Percent (%)
≤47yrs	Female	148	46.8
	Male	168	53.2
	Total	316	100
≥48yrs	Female	197	64.8
	Male	107	35.2
	Total	304	100

Table 77 - Fracture management by subgroup

Age group	Fracture Management	Frequency (n=)	Percent (%)
≤47yrs	Operative	183	57.9
	Non-operative	133	42.1
	Total	316	100
≥48yrs	Operative	148	48.7
	Non-operative	156	51.3
	Total	304	100

Tables 75-77 show the demographic and injury data for the subgroups analysed in this project. The total sample size for each subgroup is 316 in the ≤47yrs group and 304 in the ≥48yrs group. The mean age of the ≤47yrs group was 32.1 years (±8.49) with a range of 29 and the mean age of the ≥48yrs group was 60.4 years (±9.37) with a range of 46. There were 148 (46.8%) females in the ≤47yrs group and 197 females (64.8%) in the ≥48yrs group. In the ≤47yrs group, there were 183 (57.9%) individuals who received operative treatment for their fracture, compared to 148 individuals (48.7%) in the ≥48yrs group.

Table 78 - Mechanism of injury by subgroup

Age group	Mechanism of injury	Frequency (n=)	Percent (%)
≤47yrs	Road traffic accident	15	4.7
	Crush injury	0	0
	Contact sports injury	41	13
	Low energy fall	146	46.2
	High energy fall	66	20.9
	Other	34	10.8
	Missing	14	4.4
	Total	316	100
≥48yrs	Road traffic accident	9	3
	Crush injury	2	0.7
	Contact sports injury	5	1.6
	High energy fall	27	8.9
	Low energy fall	235	77.3
	Other	18	5.9
	Missing	8	2.6
	Total	304	100

Table 79 - Weber classification of fractures by subgroup

Age group	Weber classification	Frequency (n=)	Percent (%)
≤47yrs	Weber A	20	6.3
	Weber B	190	60.1
	Weber C	65	20.6
	Weber N/A (no lateral malleolar fracture)	16	5.1
	Missing	25	2.9
	Total	316	100
≥48yrs	Weber A	18	5.9
	Weber B	219	72
	Weber C	39	12.8
	Weber N/A (no lateral malleolar fracture)	9	3
	Missing	19	6.3
	Total	304	100

Table 80 - Number of malleoli involved in fractures by subgroup

Age group	Number of malleoli involved in fracture	Frequency (n=)	Percent (%)
≤47yrs	One (unimalleolar)	210	66.5
	Two (bimalleolar)	61	19.3
	Three (trimalleolar)	29	9.2
	Missing	16	5.1
	Total	316	100
≥48yrs	One (unimalleolar)	196	64.5
	Two (bimalleolar)	62	20.4
	Three (trimalleolar)	36	11.8

	Missing	10	3.3
	Total	304	100

Table 78 shows the mechanism of injury by subgroup. In both subgroups, the most common mechanism of injury was a low energy fall, accounting for 146 (46.2%) of the injuries in the ≤47yrs group and 235 (77.3%) of the injuries in the ≥48yrs group. In the ≤47yrs group, 66 fractures (20.9%) were caused by a high energy fall, 41 (13%) by contact sports injury, 34 (10.8%) were caused by an other mechanism and 15 (4.7%) in a road traffic accident. In the ≥48yrs group 27 fracture (8.9%) were caused by a high energy fall, an other mechanism of injury in 18 cases (5.9%) and a road traffic accident in 9 cases (3%).

Tables 79 and 80 show the fracture classifications for each subgroup. There were 190 (60.1%) Weber A fractures in the ≤47yrs group, 65 (20.6%) Weber C fractures and 20 (6.3%) Weber A fractures, with 16 (5.1%) fractures which were not classifiable under Weber classification in this subgroup. In the ≥48yrs group, 219 (72%) of injuries were classified as Weber B, 39 (12.8%) were Weber C and 18 (5.9%) were Weber A. Nine fractures (3%) in this group were not classifiable under Weber classification and data for 19 (6.3%) of this group were missing. In both subgroups the majority of fractures were unimalleolar, at 210 (66.5%) in the ≤47yrs group and 196 (64.5%) in the ≥48yrs group. There were 61 (19.3%) bimalleolar fractures in the ≤47yrs group and 61 (20.4%) in the ≥48yrs group. There were 29 (9.2%) trimalleolar fractures in the ≤47yrs group and 36 (11.8%) in the ≥48yrs group. Data were missing for 16 (5.1%) of individuals in the ≤47yrs group and 10 (3.3%) of the ≥48yrs group.

6.3.2b Outcome data of subgroups

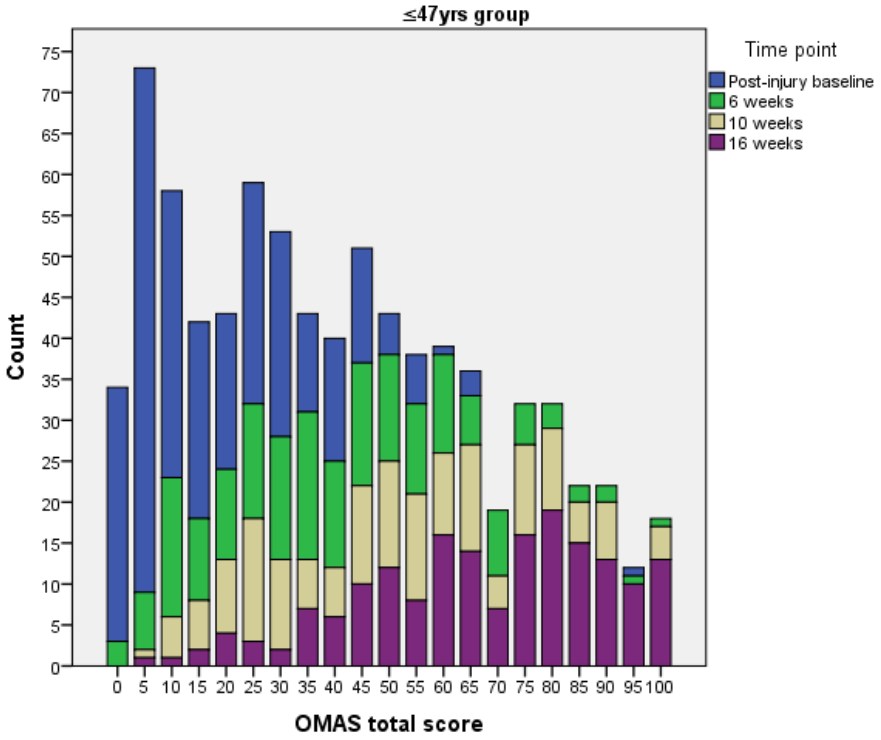


Figure 23 - Histogram of OMAS scores by time point for ≤47years group

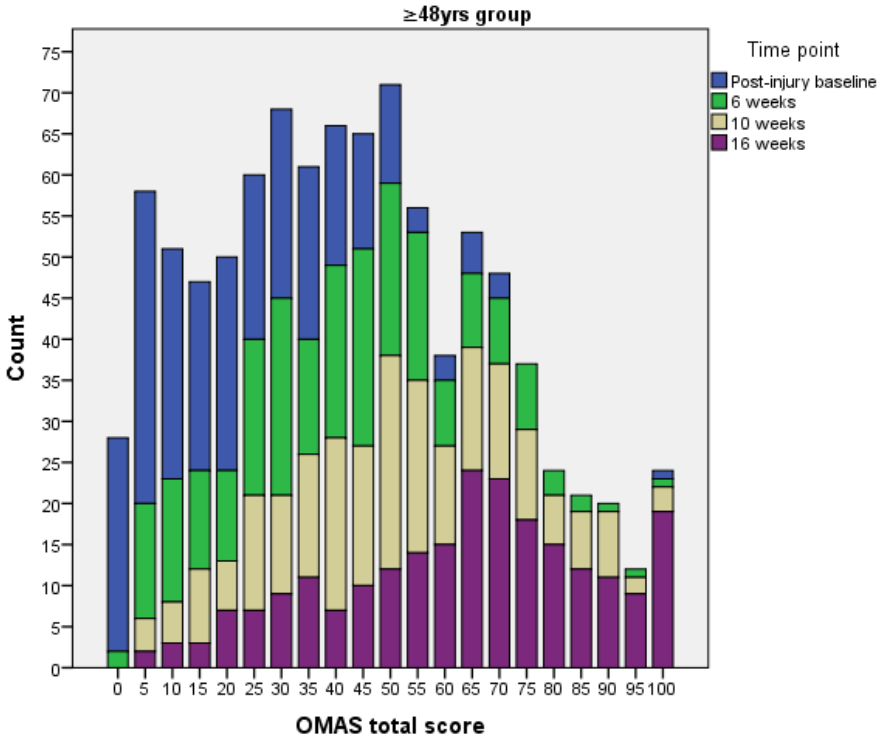


Figure 24 - Histogram of OMAS score by time point for ≥48years group

Figures 23 and 24 show histograms for OMAS scores by time point for each of the subgroups analysed here.

Table 81 - Descriptive statistics of OMAS scores by subgroup

Age group	**≤47yrs**	**≥48yrs**
Number of patients in group (n=)	316	304
Number of OMAS responses (n=)	809	958
Mean	40.67	43.09
Median	35	40
Mode	5	50
Standard deviation	27.68	25.77
Skewness	0.37	0.27
Kurtosis	-0.91	-0.73
Range	100	100

Table 82 - Descriptive statistics of OMAS scores for each time point by subgroup

Age group	**Time point**	**Number of OMAS scores (n=)**	**Range**	**Mean**	**Standard deviation**
≤47yrs	Baseline	282	95	19.2	16.5
	6 weeks	187	100	38.96	22.09
	10 weeks	161	95	50.87	23.77
	16 weeks	179	95	67.09	22.39
≥48yrs	Baseline	263	100	23.38	18.09
	6 weeks	236	100	38.92	20.94
	10 weeks	228	95	50.02	21.80
	16 weeks	231	95	62.94	23.84

Table 81 shows the descriptive statistics for OMAS scores by subgroup. The mean OMAS score was 40.67 (±27.68) in the ≤47yrs group and 43.09 (±25.77) in the ≥48yrs group. The median OMAS score was 35 in the ≤47yrs group and 40 in the ≥48yrs group. Table 82 shows the descriptive statistics for OMAS scores by time point for each subgroup. In the ≤47yrs group, mean OMAS scores were 19.2 (±16.5) at baseline, 38.96 (±22.09) at 6 weeks, 50.87 (±23.77) at 10 weeks and 67.09 (±22.39) at 16 weeks. In the ≥48yrs group mean OMAS scores were 23.3 (±18.09) at baseline, 38.92 (±20.94) at six weeks, 50.02 (±21.80) at 10 weeks and 62.94 (±23.84) at 16 weeks.

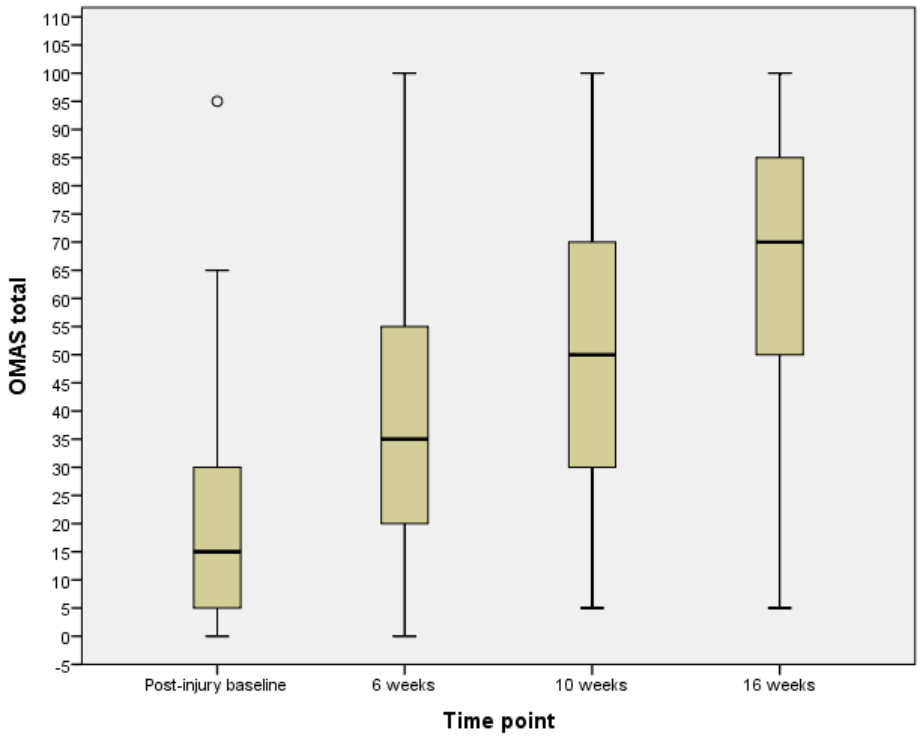


Figure 25 - Boxplot of OMAS scores by time point for ≤47years group

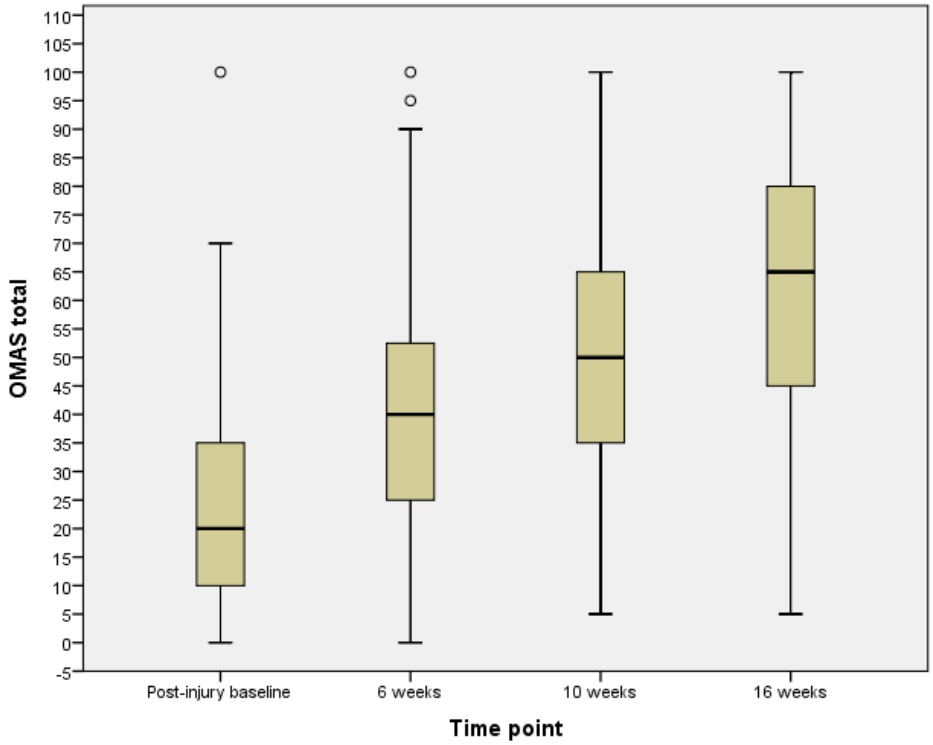


Figure 26 - Boxplot of OMAS scores by time point for ≥48years group

Figures 25 and 26 show boxplots of OMAS scores by time point for each subgroup to give a visual representation of the data across the time points of this study.

Figure 27 - QQ-plot of OMAS scores for ≤47years group

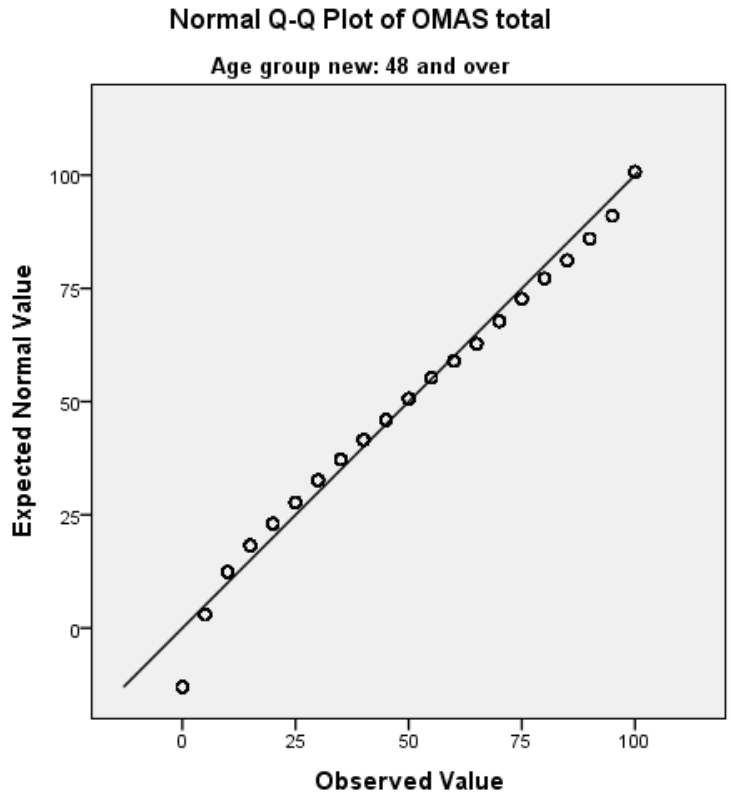


Figure 28 - QQ-plot of OMAS scores for ≥48years group

Figures 27 and 28 show QQ-plots for OMAS data by subgroup. Table 83 shows the GIC score responses for each subgroup.

Table 83 - GIC score responses by subgroup

Age group	GIC response	Frequency (n=)	Percent (%)
≤47years	Very much improved	34	10.8
	Much improved	91	28.8
	Improved minimally	38	12.0
	No change	12	3.8
	Minimally worse	9	2.8
	Much worse	1	0.3
	Very much worse	0	0
	Missing	131	41.5
	Total	316	100
≥48years	Very much improved	45	14.8
	Much improved	98	32.2
	Improved minimally	63	20.7
	No change	17	5.6
	Minimally worse	6	2.0
	Much worse	0	0
	Very much worse	1	0.3
	Missing	74	24.3
	Total	304	100

6.3.3 Measurement properties of OMAS in the total sample

6.3.3.a Convergent validity

Following assignment of a single time point per participant, and thus excluding any participants for whom a questionnaire did not exist for their selected time point, the sample sizes ranged from 223 to 442 for the assessment of convergent validity. Details of sample size for each correlation are found in table 84, along with the Pearson's correlation coefficients for each analysis. There are some samples which are notably lower, for example the MOXFQ and MOXFQ domains. This is because these questionnaires were only collected at baseline and 16 weeks, therefore there are less of these questionnaires available to include in the analyses.

Table 84 - Correlations of scores of OMAS with scores of comparator instruments and domains

Comparator instrument/domain	Sample size (n=)	Pearson's correlation with OMAS scores (r=)	Significance (2 tailed)
DRI	411	-0.77	<0.01
MOXFQ	223	-0.86	<0.01
MOXFQ-pain domain	225	-0.73	<0.01
MOXFQ-walking standing domain	223	-0.85	<0.01
MOXFQ-social interaction domain	224	-0.75	<0.01
EQ-5D value set	442	0.73	<0.01
EQ-5D-mobility domain	442	-0.78	<0.01
EQ-5D-pain & discomfort domain	442	-0.64	<0.01
EQ-5D-self-care domain	442	-0.59	<0.01
EQ-5D-anxiety & depression domain	442	-0.34	<0.01
EQ-5D-usual activities domain	442	-0.75	<0.01
EQ-5D-overall health domain	441	0.56	<0.01

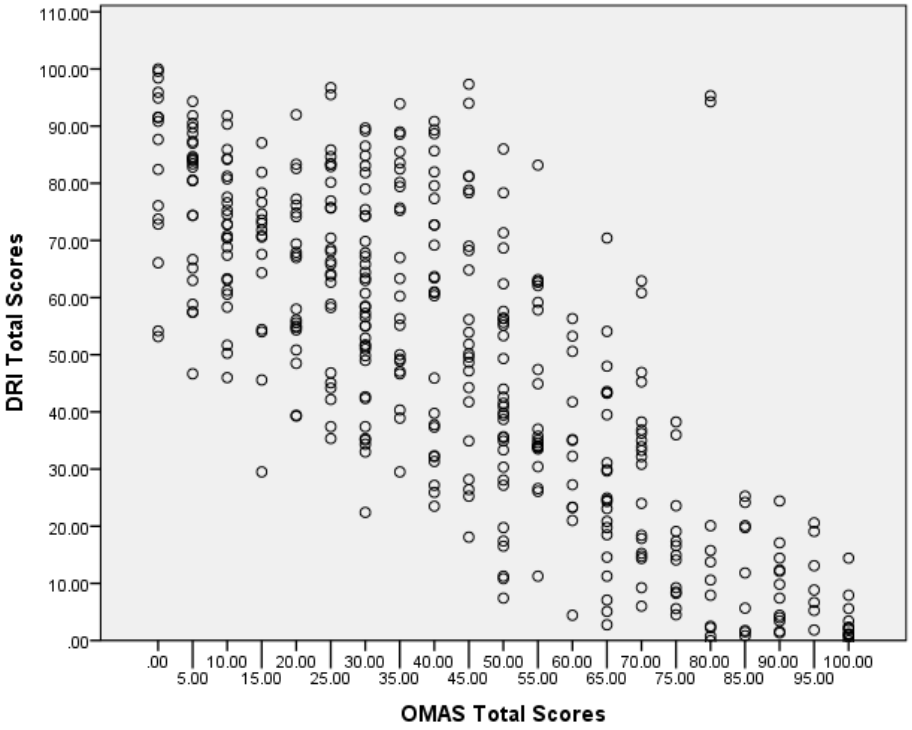


Figure 29 - Scatter plot of OMAS scores and DRI scores

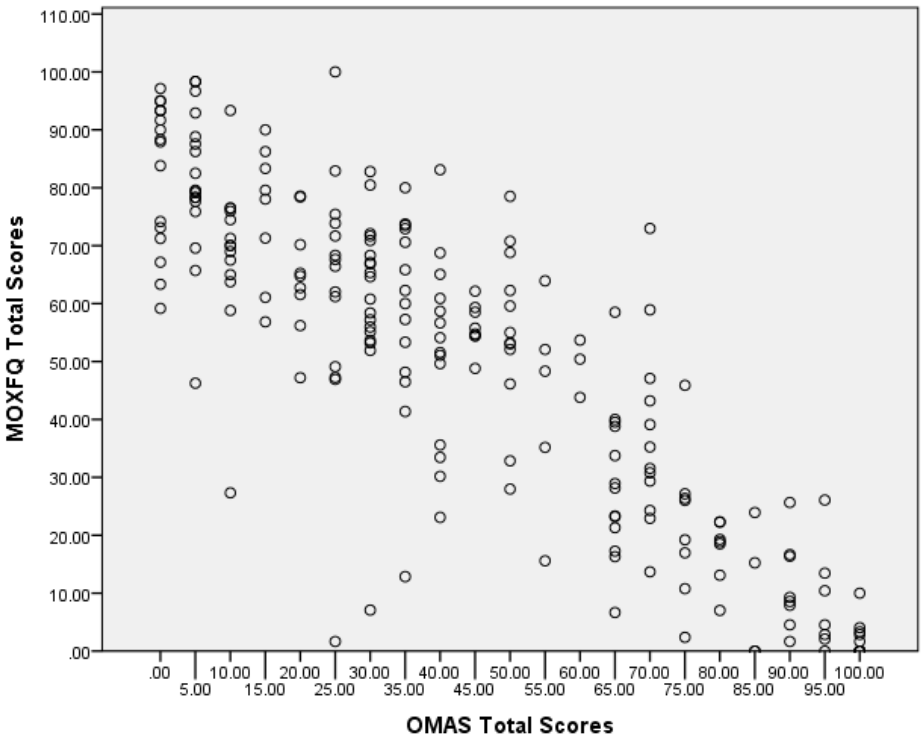


Figure 30 - Scatter plot of OMAS scores and MOXFQ scores

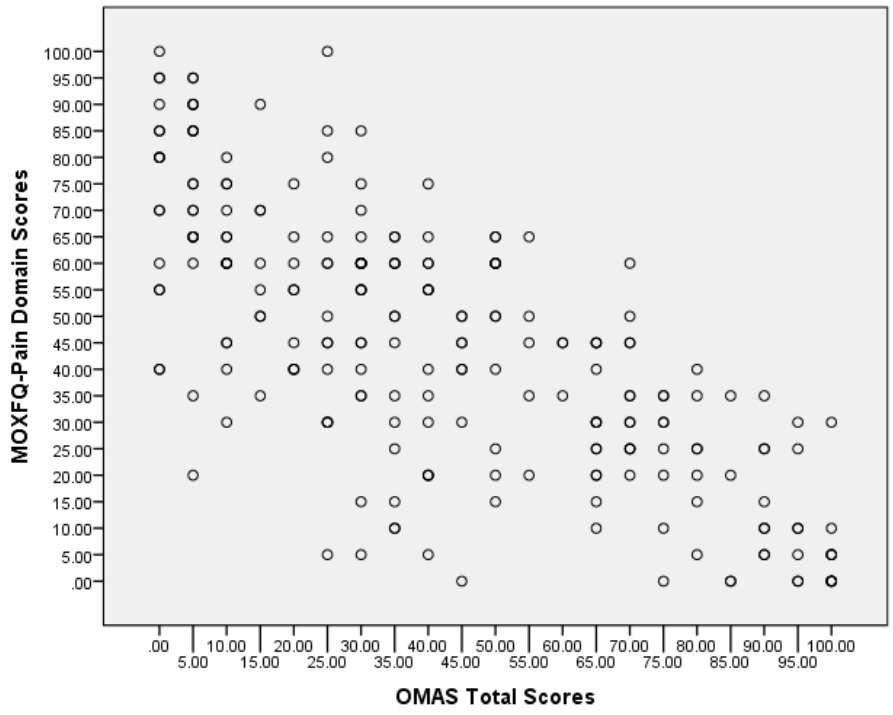


Figure 31 - Scatter plot of OMAS scores and MOXFQ-pain domain scores

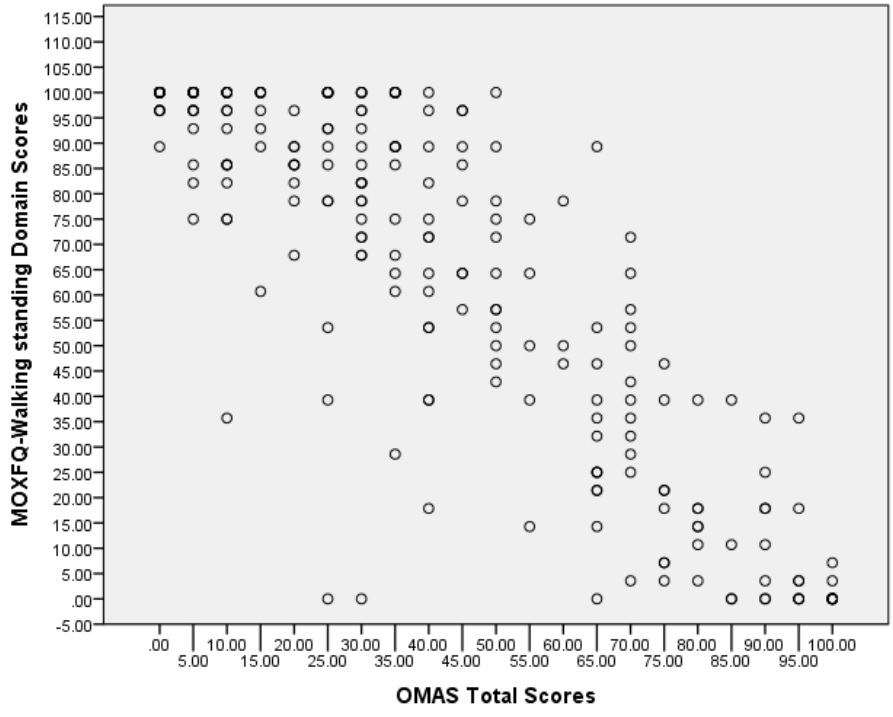


Figure 32 - Scatter plot of OMAS scores and MOXFQ-walking standing domain scores

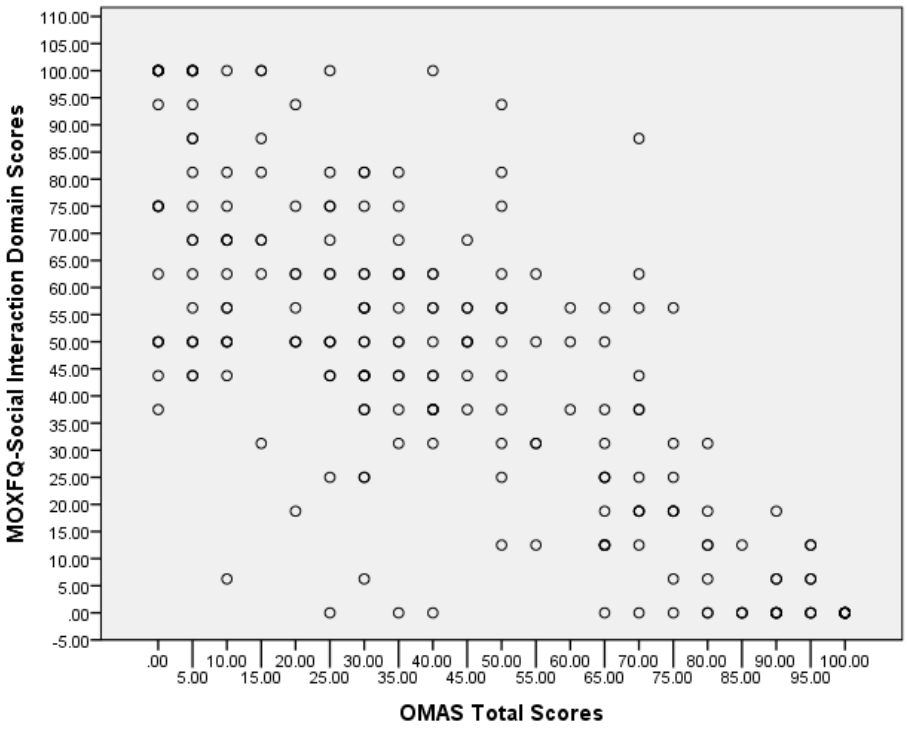


Figure 33 - Scatter plot of OMAS total and MOXFQ-social interaction domain scores

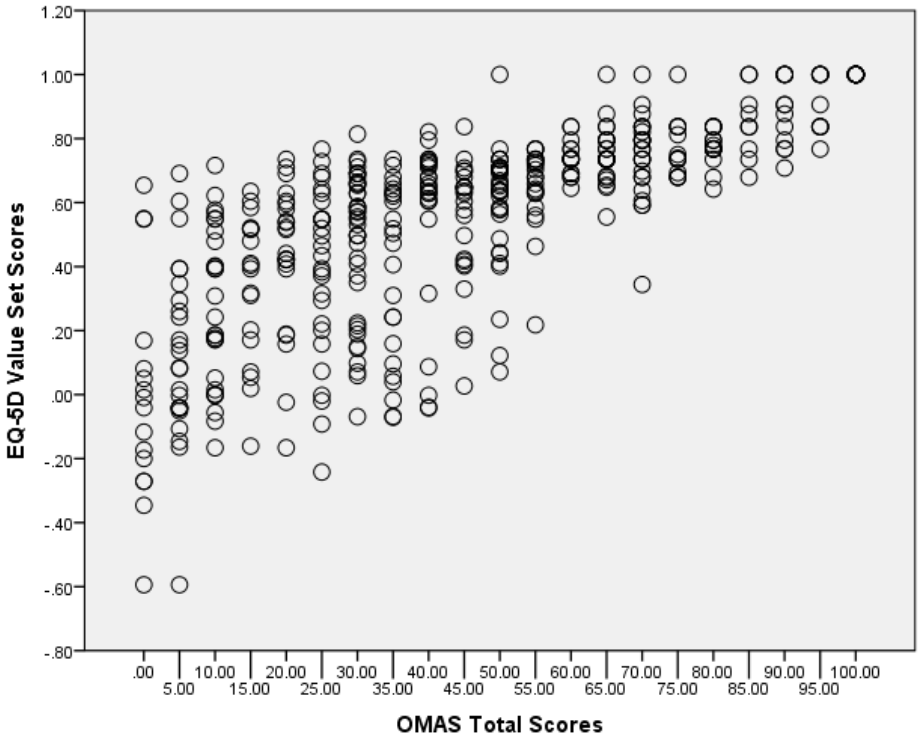


Figure 34 - Scatter plot of OMAS total scores and EQ-5D value set scores

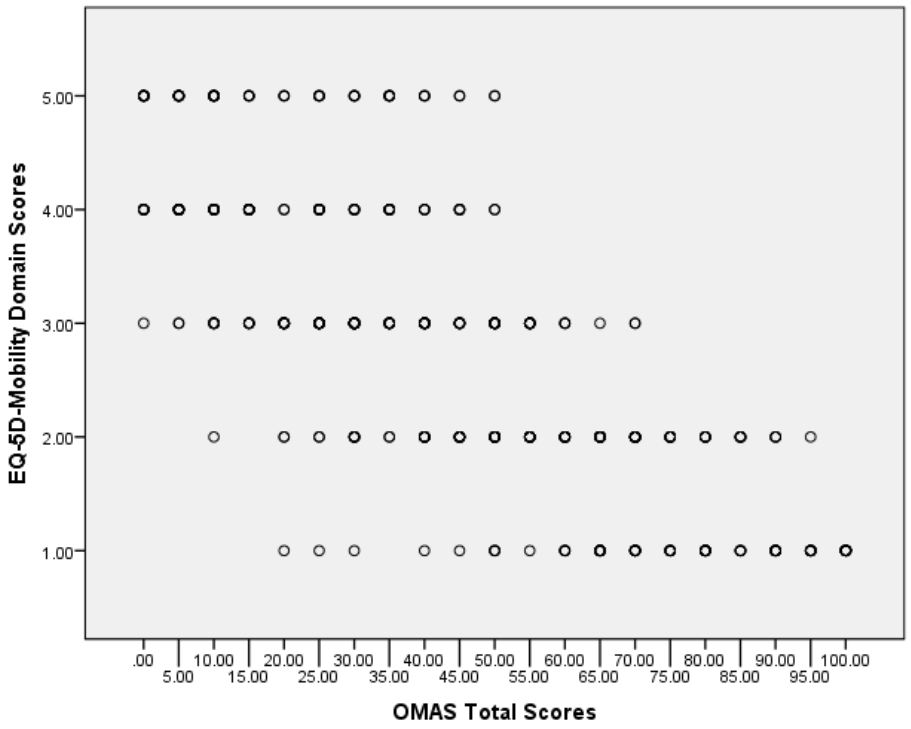


Figure 35 - Scatter plot of OMAS scores and EQ-5D-mobility domain scores

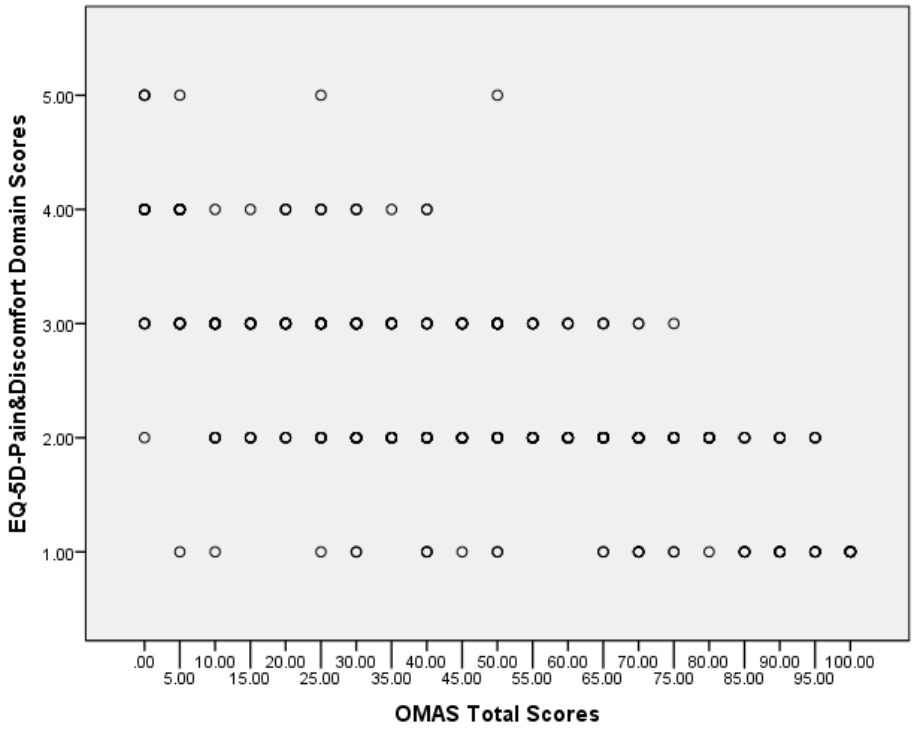


Figure 36 - Scatter plot of OMAS scores and EQ-5D-pain & discomfort domain scores

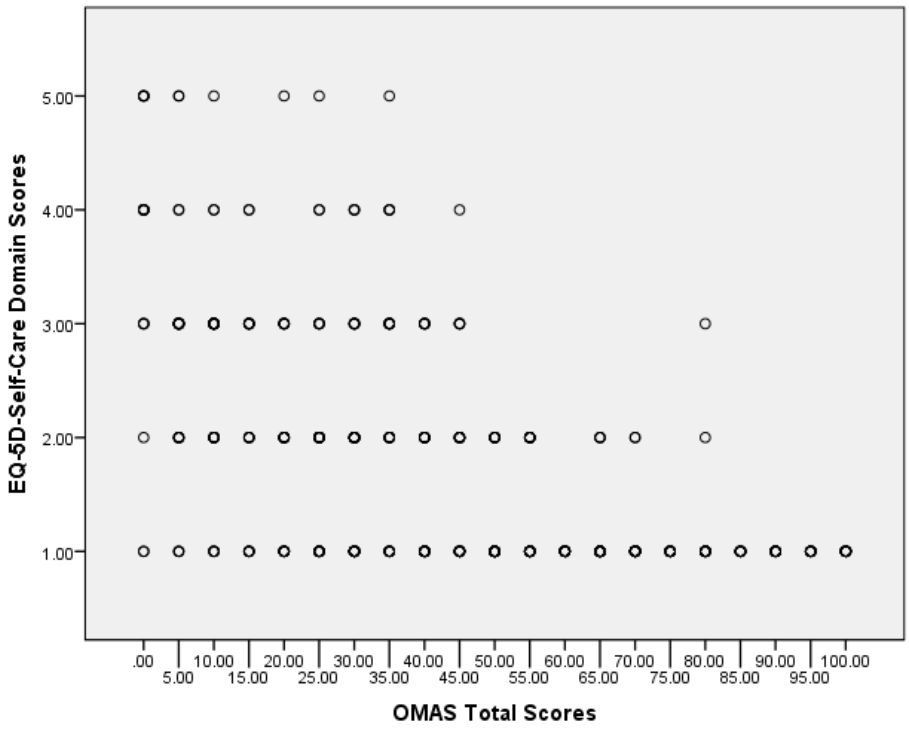


Figure 37 - Scatter plot of OMAS scores and EQ-5D-self-care domain scores

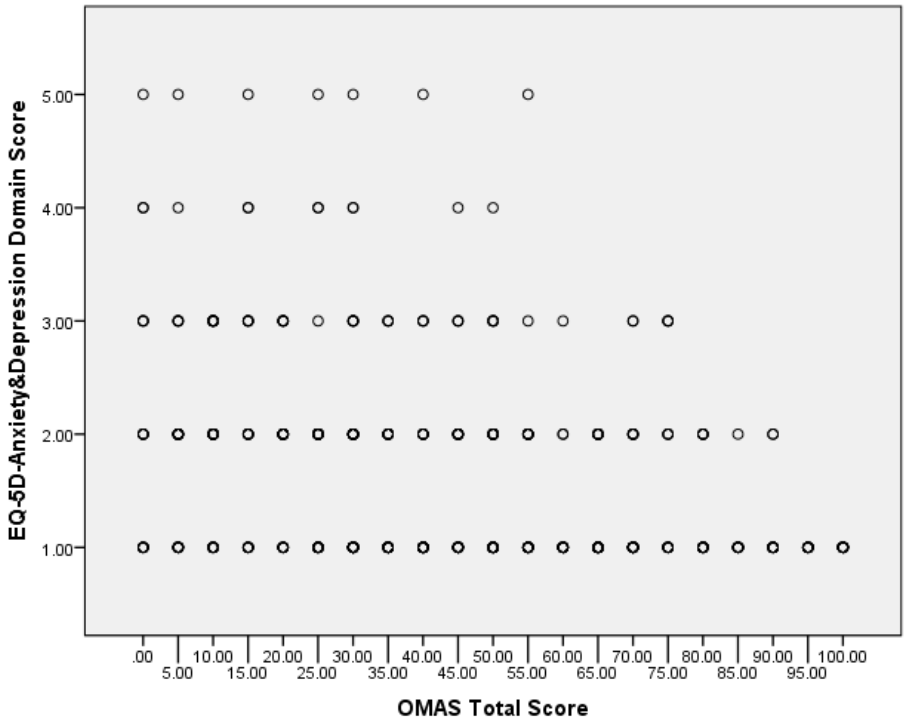


Figure 38 - Scatter plot of OMAS scores and EQ-5D-anxiety & depression domain scores

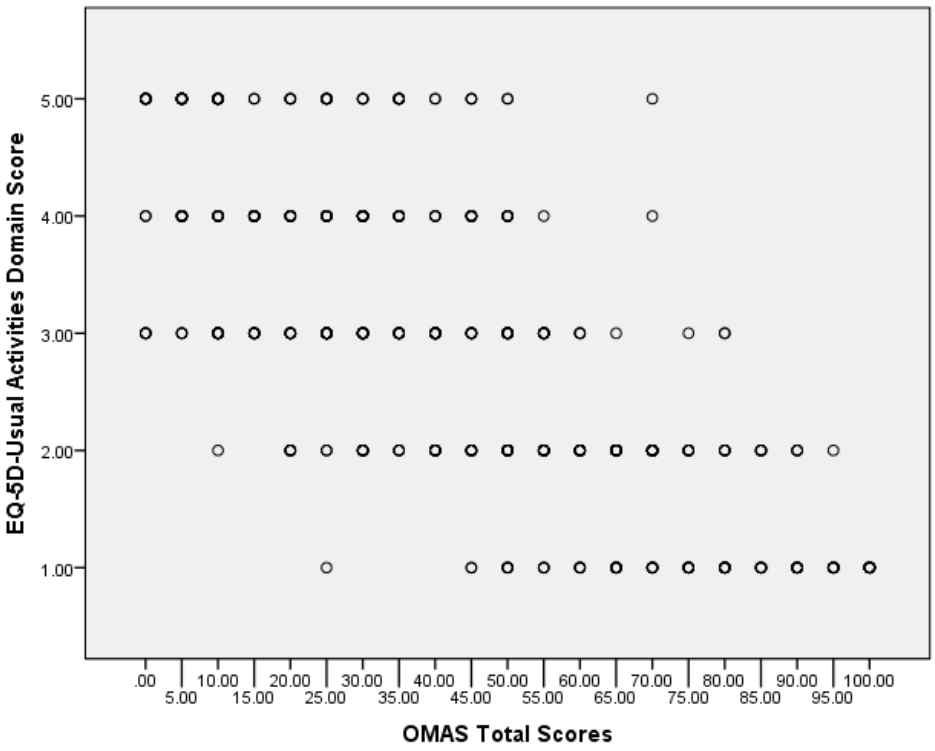


Figure 39 - Scatter plot of OMAS scores and EQ-5D-usual activities domain scores

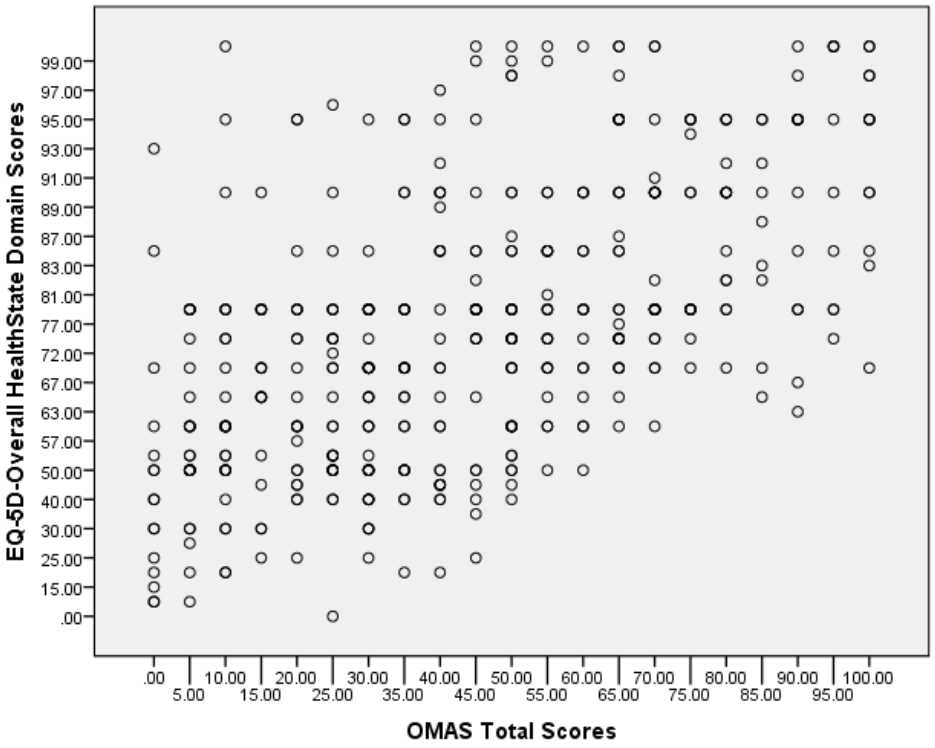


Figure 40 - Scatter plot of OMAS scores and EQ-5D-overall health domain scores

Table 84 shows the correlations of scores of OMAS with scores of the comparator instruments and domains, which are also displayed in the scatter plots in figures 29-40. The OMAS scores demonstrate a highly negative correlation with scores of the MOXFQ (r= -0.86) and the DRI (r= -0.77) and a highly positive correlation with scores of the EQ-5D instrument (r= 0.73) in this sample. The OMAS scores correlate highly negatively with the domains of the MOXFQ, with the scores of the walking standing domain being the most strongly associated with scores of OMAS (r= -0.85). The scores of MOXFQ domains of pain and social interaction were also highly negatively associated with scores of OMAS (r= -0.73 and r= -0.75 respectively).

The scores of mobility and usual activities domains of EQ-5D were highly negatively associated with score of OMAS (r= -0.78 and r= -0.75 respectively). The scores of EQ-5D domains of self-care and pain were moderately negatively associated with scores of OMAS (r= -0.59 and r= -0.64 respectively), and the overall health state domain of EQ-5D was moderately positively associated with OMAS (r= 0.56). The anxiety and depression domain of EQ-5D demonstrated a low negative correlation with scores of OMAS (r= -0.34).

Based on these results, 11 of 12 hypotheses regarding expected magnitudes of correlation with comparator instruments, including the EQ-5D, DRI and MOXFQ. The only hypothesis which was not met was that for the correlation between OMAS scores and scores of the EQ-5D anxiety and depression domain.

6.3.3.b Discriminative validity

Table 85 - Descriptive statistics for operatively and non-operatively managed patients by time point

Time point	Fracture management	Sample (n=)	Mean OMAS score	Standard deviation	Standard error of the mean
Baseline	Operative	298	21.74	18.05	1.05
	Non-operative	257	20.51	16.39	1.02
6 weeks	Operative	225	35.71	20.75	1.38
	Non-operative	198	42.60	21.65	1.54
	Operative	203	47.04	23.24	1.63

10 weeks	Non-operative	187	54.01	21.33	1.56
16 weeks	Operative	209	63.30	23.58	1.63
	Non-operative	201	66.27	22.93	1.62

Table 86 - Results of t-test for equality of means between operative and non-operatively managed patients by time point

Time point	**t-test for equality of means**						
	t	**df**	**Sig. (2-tailed)**	**Mean difference in OMAS**	**Std. error of the difference**	**95% confidence interval of the difference**	
						Lower	**Upper**
Baseline	0.84	553	0.401	1.24	1.47	-1.65	4.13
6 weeks	-3.34	421	0.001	-6.89	2.06	-10.95	-2.83
10 weeks	-3.08	388	0.002	-6.97	2.27	-11.42	-2.51
16 weeks	-1.29	408	0.20	-2.97	2.30	-7.49	1.55

The descriptive statistics for each group at each time point are found in table 85. Prior to conducting the t-test, Levene's test for equality of variances was completed. All tests were not significant (Baseline f=2.65 p=0.104, six weeks f=0.08 p=0.774, 10 weeks f=2.92 p=0.089 and 16 weeks f=1.11, p=0.293). Therefore, t-test assuming equal variances was subsequently used. Results of this t-test are shown in table 86. The means of the operatively and non-operatively managed patients were not statistically significantly different at baseline or 16 weeks (baseline t= 0.84 p=0.401, 16 weeks t= -1.29 p=0.20). At six and 10 weeks, however, the mean OMAS scores in the operatively and non-operatively managed groups were statistically significantly different, with scores in the operatively managed group lower than the non-operatively managed group at both time points (six weeks t= -3.34 p=0.001, 10

weeks t= -3.08 p=0.002). This shows that there is a statistically significant difference in the OMAS scores between operatively and non-operatively managed patients at six and 10 weeks (-6.89 and -6.97 respectively), but a difference is not observed at baseline and 16 weeks. Therefore, with regards to the hypotheses set a priori, two of the four hypotheses have been met for discriminative validity of OMAS.

Table 87 - Descriptive statistics of number of malleoli involved in fractures of included participants

Time point	Number of malleoli involved in fracture	Sample (n=)	Mean OMAS score	Standard deviation	Standard error of the mean
Baseline	Unimalleolar	364	21.36	17.07	0.90
	Bimalleolar	116	20.47	18.02	1.67
	Trimalleolar	53	19.91	18.44	2.53
6 weeks	Unimalleolar	278	40.86	21.65	1.30
	Bimalleolar	86	33.72	19.81	2.14
	Trimalleolar	44	35.8	18.89	2.85
10 weeks	Unimalleolar	257	52.33	21.82	1.36
	Bimalleolar	76	45.72	23.41	2.69
	Trimalleolar	42	47.26	24.67	3.81
16 weeks	Unimalleolar	279	66.22	23.80	1.43
	Bimalleolar	76	63.09	21.42	2.46
	Trimalleolar	42	58.10	23.61	3.64

Table 88 - ANOVA and Tukey HSD test for comparison of OMAS scores between groups of fracture types by number of malleoli involved in injury

Time point	F statistic	Significance	Malleoli involved in fracture (I)	Malleoli involved in fracture (J)	Mean difference (I-J)	Standard error	Significance	95% confidence intervals	
								Lower bound	Upper bound
Baseline	0.235	0.791	Unimalleolar	Bimalleolar	0.89	1.86	0.882	-3.48	5.25
				Trimalleolar	1.45	2.56	0.837	-4.57	7.47
			Bimalleolar	Unimalleolar	-0.89	1.86	0.882	-5.25	3.48
				Trimalleolar	0.57	2.89	0.979	-6.22	7.36
			Trimalleolar	Unimalleolar	-1.45	2.56	0.837	-7.47	4.57
				Bimalleolar	-0.57	2.89	0.979	-7.36	6.22
6 weeks	4.309	0.014	Unimalleolar	Bimalleolar	7.14*	2.59	0.017	1.05	13.24
				Trimalleolar	5.07	3.41	0.298	-2.95	13.08
			Bimalleolar	Unimalleolar	-7.14*	2.59	0.017	-13.24	-1.05
				Trimalleolar	-2.08	3.89	0.855	-11.23	7.08
			Trimalleolar	Unimalleolar	-5.07	3.41	0.298	-13.08	2.95

				Bimalleolar	2.08	3.89	0.855	-7.08	11.23
10 weeks	3.007	0.051	Unimalleolar	Bimalleolar	6.61	2.94	0.064	-0.29	13.52
				Trimalleolar	5.07	3.74	0.365	-3.73	13.87
			Bimalleolar	Unimalleolar	-6.61	2.94	0.064	-13.52	0.29
				Trimalleolar	-1.54	4.32	0.933	-11.71	8.63
			Trimalleolar	Unimalleolar	-5.07	3.74	0.365	-13.87	3.73
				Bimalleolar	1.54	4.32	0.933	-8.63	11.71
16 weeks	2.451	0.088	Unimalleolar	Bimalleolar	3.13	3.02	0.555	-3.98	10.23
				Trimalleolar	8.12	3.86	0.091	-0.97	17.21
			Bimalleolar	Unimalleolar	-3.13	3.02	0.555	-10.23	3.98
				Trimalleolar	4.99	4.49	0.506	-5.56	15.56
			Trimalleolar	Unimalleolar	-8.12	3.86	0.091	-17.21	0.97
				Bimalleolar	-4.99	4.49	0.506	-15.56	5.56

* The mean difference is significant at the 0.05 level.

Table 87 shows the descriptive statistics for the samples used in the discriminative validity analysis for number of malleoli involved in the fracture. There were a total of 364 unimalleolar fractures at baseline, 116 bimalleolar fractures at baseline and 53 trimalleolar fractures at baseline. Levene's test for equality of variance was completed prior to the ANOVA, and all results were not significant (baseline f= 1.18 p=0.308, six weeks f= 1.53 p= 0.218, 10 weeks f=0.82 p= 0.442 and 16 weeks f=0.65 p= 0.521). This means the null hypothesis can be accepted and homogeneity of variances exists between the groups, which means Tukey's post hoc test assumptions have been met. Therefore, ANOVA test was completed with Tukey's HSD test to determine pairwise differences.

Table 88 shows the results of ANOVA and Tukey's HSD. The results show that there are no statistically significant pairwise differences between any of the groups except for the difference between unimalleolar and bimalleolar fractures at six weeks. The mean difference in OMAS score between these two fracture types was 7.14 points, with scores higher in the unimalleolar group compared to bimalleolar group. The results of these analyses show that none of the four hypotheses have been met with regard to discriminative validity because the expected differences outlined in the hypotheses was that the scores would be higher in the bimalleolar group than the unimalleolar group.

Table 89 - Descriptive statistics for Weber classification of injuries of participants

Time point	Weber classification	Sample (n=)	Mean OMAS score	Standard deviation	Standard error
Baseline	Weber A	36	13.33	12.87	2.15
	Weber B	361	22.04	17.76	0.94
	Weber C	95	19.58	17.25	1.77
	Weber N/A	23	26.52	17.15	3.58
6 weeks	Weber A	27	40.74	21.56	4.15
	Weber B	281	39.79	20.84	1.24
	Weber C	72	35.63	22.01	2.59
	Weber N/A	14	41.43	26.20	7.00
	Weber A	23	53.48	23.91	4.99

10 weeks	Weber B	261	50.48	22.46	1.39
	Weber C	65	48.00	21.95	2.72
	Weber N/A	17	50.29	27.92	6.77
16 weeks	Weber A	25	68.40	21.25	4.25
	Weber B	278	64.71	23.32	1.40
	Weber C	70	62.21	25.50	3.05
	Weber N/A	17	70.29	23.01	5.58

Table 89 shows the descriptive statistics for the groups involved in the discriminative validity analysis for Weber classification. At baseline, there were 36 Weber A fractures, 361 Weber B fractures, 95 Weber C fractures and 23 fractures which were not classifiable under the Weber classification system. As previously discussed, the fractures which were not classifiable under the Weber system have not been included for further analysis.

Table 90 - Results of ANOVA and Tukey's HSD test

Time point	F statistic	Significance	Weber classification (I)	Weber classification (J)	Mean difference (I-J)	Standard error	Significance	95% confidence intervals	
								Lower bound	Upper bound
Baseline	3.74	0.011	Weber A	Weber B	-8.70*	3.03	0.022	-16.52	-0.89
				Weber C	-6.25	3.40	0.256	-15.00	2.51
			Weber B	Weber A	8.70*	3.03	0.022	0.89	16.52
				Weber C	2.46	2.00	0.609	-2.70	7.61
			Weber C	Weber A	6.25	3.40	0.256	-2.51	15.00
				Weber B	-2.46	2.00	0.609	-7.61	2.70
6 weeks	0.84	0.470	Weber A	Weber B	0.95	4.29	0.996	-10.12	12.03
				Weber C	5.12	4.81	0.712	-7.29	17.52
			Weber B	Weber A	-0.95	4.29	0.996	-12.03	10.12
				Weber C	4.16	2.81	0.451	-3.10	11.42
			Weber C	Weber A	-5.12	4.81	0.712	-17.52	7.29

				Weber B	-4.16	2.81	0.451	-11.42	3.10
10 weeks	0.38	0.771	Weber A	Weber B	3.00	4.94	0.930	-9.76	15.76
				Weber C	5.48	5.51	0.753	-8.75	19.71
			Weber B	Weber A	-3.00	4.94	0.930	-15.76	9.76
				Weber C	2.48	3.15	0.860	-5.65	10.61
			Weber C	Weber A	-5.48	5.51	0.753	-19.71	8.75
				Weber B	-2.48	3.15	0.860	-10.61	5.65
16 weeks	0.782	0.504	Weber A	Weber B	3.69	4.93	0.877	-9.02	16.40
				Weber C	6.19	5.50	0.674	-7.99	20.37
			Weber B	Weber A	-3.69	4.93	0.877	-16.40	9.02
				Weber C	2.50	3.15	0.858	-5.64	10.64
			Weber C	Weber A	-6.19	5.50	0.674	-20.37	7.99
				Weber B	-2.50	3.15	0.858	-10.64	5.64

* The mean difference is significant at the 0.05 level.

Prior to completing the ANOVA test, Levene's test for equality of variance was completed. All results were not statistically significant (baseline f= 2.40 p= 0.067, six weeks f= 0.47 p= 0.704, 10 weeks f= 1.55 p=0.202 and 16 weeks f= 0.76 p= 0.519), therefore the null hypothesis of equal variances of groups was accepted and Tukey's HSD was completed as a post hoc test.

The results of the ANOVA and Tukey's HSD is shown in table 90. Results of the ANOVA and post hoc test shows that the majority of differences between the means of each group are not statistically significant, apart from the difference between Weber A fractures and Weber B fractures at the baseline time point. The Weber B fracture group had a mean OMAS score of 8.7 points higher than the Weber A group. Therefore the results of these analyses show that one of the four hypotheses have been met with regards to discriminative validity using Weber classification of injuries.

Therefore, a total of three of 20 hypotheses were met for discriminative validity of OMAS scores with injury classifications and types of fracture management.

6.3.3.c Internal structure

Following assignment of a single time point per participant, and thus excluding any participants for whom a questionnaire did not exist for their selected time point, the sample size for this calculation was 438 (described as the principal component analysis (PCA) sample). Table 91 shows the descriptive statistics of the PCA sample. Table 92 shows a breakdown of the time points in the PCA as a result of the sampling of a randomly seclected time point per participant.

Table 91 - Descriptive statistics of PCA sample

	Sample (n=)	**Minimum OMAS scores**	**Maximum OMAS score**	**Mean OMAS score**	**Standard deviation of OMAS score**
OMAS (PCA sample)	438	0	100	43	26.36

Table 92 - Breakdown of the time points included in the PCA sample as a results of the random sampling

Time point	Number of participants (n=)	Percent (%)
Baseline	132	30.1
6 weeks	110	25.1
10 weeks	100	22.8
16 weeks	96	21.9
Total	438	100

Table 93 - Correlation matrix of PCA

Item	Pain	Stiffness	Swelling	Climbing stairs	Running	Jumping	Squatting	Supports	Work, activities of daily life
Pain	1	0.24	0.36	0.35	0.26	0.25	0.30	0.26	0.38
Stiffness	0.24	1	0.29	0.28	0.29	0.26	0.31	0.21	0.30
Swelling	0.36	0.29	1	0.29	0.26	0.23	0.23	0.26	0.27
Climbing stairs	0.35	0.28	0.29	1	0.46	0.47	0.54	0.51	0.48
Running	0.26	0.29	0.26	0.46	1	0.72	0.47	0.43	0.40
Jumping	0.25	0.26	0.22	0.47	0.72	1	0.56	0.47	0.43
Squatting	0.30	0.31	0.23	0.54	0.47	0.56	1	0.56	0.51
Supports	0.26	0.21	0.26	0.51	0.43	0.47	0.56	1	0.57
Work, activities of daily life	0.38	0.30	0.27	0.48	0.40	0.43	0.51	0.57	1

Table 94 - KMO and Bartlett's test

Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy	0.86	
Bartlett's test of sphericity	Approx. Chi-square	1351.56
	df	36
	Sig.	<0.0001

Table 93 shows the correlation matrix for the PCA and table 94 shows results of the Kaiser-Meyer-Olkin (KMO) test for sampling adequacy and Bartlett's test for sphericity. Dziuban and Shirkey [265] recommend that sampling adequacy is optimal when the output for KMO is greater than 0.7 and the higher the number the more adequate the sample is for completing factor analysis. As you can see from the result, this outcome shows that the sampling is adequate for the planned factor analysis. Barlett's test of sphericity is a measure of the degree to which the correlation matrix differs from the identity matrix and assesses whether the data reduction technique of factor analysis is appropriate for the data [266]. As the $p<0.05$, this means we can reject the null hypothesis and confirm that the variables are correlated sufficiently for the planned factor analysis.

Table 95 - Communalities

Item	**Initial**	**Extraction**
Pain	1	0.56
Stiffness	1	0.39
Swelling	1	0.62
Climbing stairs	1	0.56
Running	1	0.61
Jumping	1	0.69
Squatting	1	0.63
Supports	1	0.58
Work, activities of daily life	1	0.54

Table 96 - Total variance explained for each component

Component	Total	Initial Eigenvalues		Extraction sums of squared loadings			Rotation sums of squares loadings		
		% of variance	Cumulative %	Total	% of variance	Cumulative %	Total	% of variance	Cumulative %
1	4.08	45.32	45.32	4.08	45.32	45.32	3.33	37.05	37.045
2	1.09	12.16	57.48	1.09	12.16	57.48	1.84	20.43	57.48
3	0.83	9.20	66.68						
4	0.74	8.21	74.89						
5	0.64	7.06	81.95						
6	0.53	5.84	87.79						
7	0.45	5.01	92.79						
8	0.39	4.29	97.09						
9	0.26	2.92	100.00						

Table 95 shows the communalities in the score and table 96 shows the total variance explained in OMAS scores. We can see from here that factor 1 explains 45.32% of the total variance within the score. Factor 2 explains 12.16% of the variance.

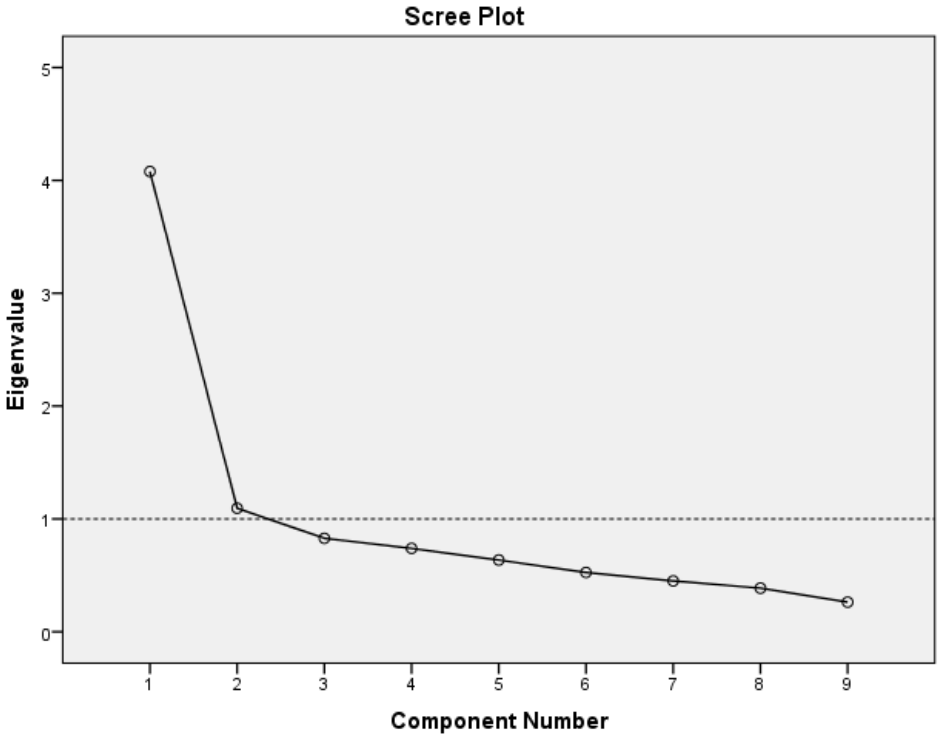


Figure 41 - Scree plot showing components extracted at an eigenvalue of greater than one

Figure 41 shows the scree plot of the components, along with a reference line at one eigenvalue. Table 97 shows the component matrix of the PCA.

Table 97 - Component matrix

Item	Component	
	1	2
Squatting	0.77	
Jumping	0.76	-0.35
Climbing stairs	0.75	
Running	0.74	
Work, activities of daily life	0.74	
Supports	0.73	
Stiffness	0.49	0.39
Swelling	0.48	0.63
Pain	0.53	0.53

Table 98 - Reproduced correlations

		Pain	Stiffness	Swelling	Climbing stairs	Running	Jumping	Squatting	Supports	Work, ADL
Reproduced correlation	**Pain**	0.56[a]	0.46	0.58	0.37	0.25	0.22	0.30	0.28	0.40
	Stiffness	0.46	0.39[a]	0.48	0.35	0.25	0.23	0.29	0.28	0.36
	Swelling	0.58	0.48	0.62[a]	0.33	0.19	0.15	0.24	0.23	0.36
	Climbing stairs	0.37	0.35	0.33	0.56[a]	0.56	0.58	0.58	0.56	0.55
	Running	0.25	0.25	0.19	0.56	0.61[a]	0.65	0.62	0.59	0.53
	Jumping	0.22	0.23	0.15	0.58	0.65	0.69[a]	0.65	0.63	0.55
	Squatting	0.30	0.29	0.24	0.58	0.62	0.65	0.63[a]	0.60	0.56
	Supports	0.28	0.28	0.23	0.56	0.59	0.63	0.60	0.58[a]	0.54
	Work, ADL	0.40	0.36	0.36	0.55	0.53	0.55	0.56	0.54	0.54[a]
Residual	**Pain**		-0.23	-0.22	-0.02	0.01	0.03	-0.01	-0.03	-0.01
	Stiffness	-0.02		-0.18	-0.06	0.03	0.03	0.02	-0.07	-0.06

	Swelling	-0.22	-0.18		-0.04	0.07	0.08	-0.02	0.03	-0.09
	Climbing stairs	-0.02	-0.06	-0.04		-0.09	-0.11	-0.05	-0.05	-0.07
	Running	0.01	0.031	0.07	-0.09		0.07	-0.15	-0.16	-0.13
	Jumping	0.03	0.026	0.08	-0.11	0.07		-0.09	-0.16	-0.12
	Squatting	-0.01	0.015	-0.02	-0.05	-0.15	-0.09		-0.04	-0.05
	Supports	-0.03	-0.07	0.03	-0.05	-0.16	-0.16	-0.04		0.03
	Work, ADL	-0.01	-0.06	-0.09	-0.07	-0.13	-0.12	-0.05	0.03	

a. Reproduced communalities

b. Residuals are computed between observed and reproduced correlations. There are 20 (55.0%) non-redundant residuals with absolute values greater than 0.05.

Table 99 - Rotated component matrix

Questionnaire item	Component	
	1	2
Jumping	0.83	
Running	0.77	
Squatting	0.77	
Supports	0.74	
Climbing stairs	0.67	0.34
Work, activities of daily life	0.63	0.38
Swelling		0.78
Pain		0.72
Stiffness		0.58

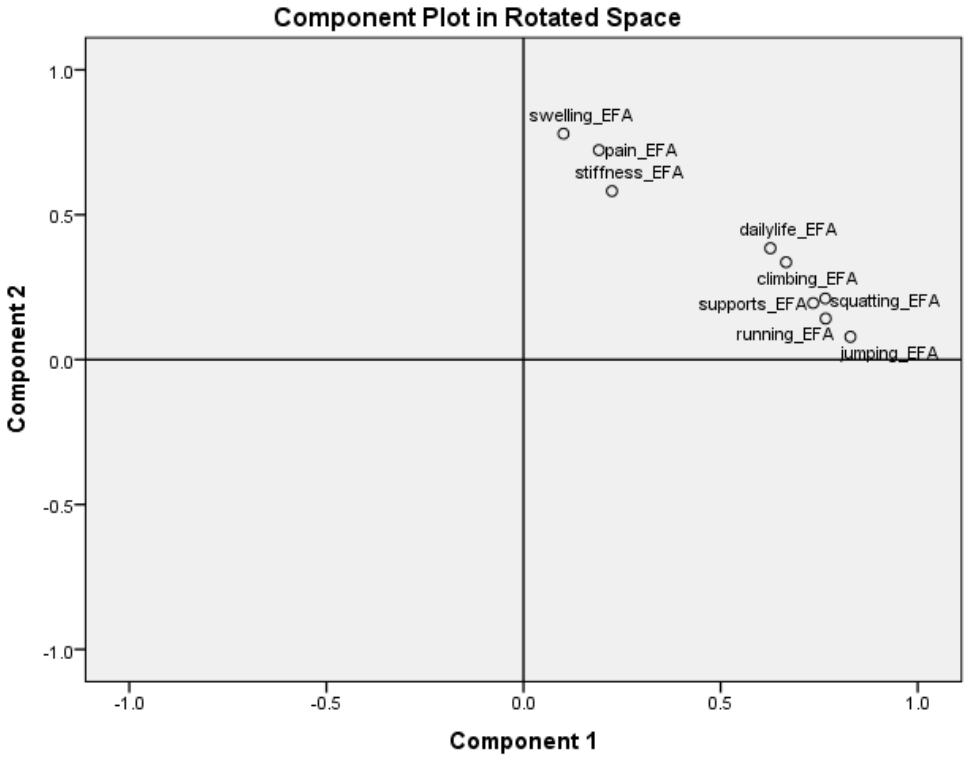


Figure 42 - Component plot in rotated space

Results of the exploratory factor analysis demonstrate that there are two underlying components present within OMAS. These are shown in the rotated component matrix in table 99. Component one comprises running, jumping, stair climbing, work & activities of daily living, supports and squatting. Component two comprises of pain, stiffness and swelling. Items in component one consist mainly of ankle functional items, whereas component two consists of ankle symptoms related to the injury. Therefore, these two sub-scales within OMAS have been identified the ankle symptom subscale and the ankle function subscale. As discussed in the methods section, the internal consistency of the overall score, as well as these two subscales will be analysed. Figure 42 shows the component plot in rotated space.

6.3.3.d Internal consistency

Table 100 - Cronbach's alpha result for scores of OMAS total

Cronbach's alpha	Cronbach's alpha based on standardized items
0.76	0.85

Table 100 shows the Cronbach's alpha of the OMAS using the sample of 1767 valid total scores. As mentioned in the descriptive statistics section, 106 questionnaires were excluded as they contained at least one item of missing data.

Table 101 - Cronbach's alpha if item deleted for scores of OMAS total

Item on questionnaire (total points available)	Mean	Standard deviation	Cronbach's if item deleted
Pain (25)	13.82	9.31	0.77
Stiffness (10)	2.37	4.25	0.75
Swelling (10)	3.72	3.79	0.74
Climbing stairs (10)	5.56	3.18	0.72
Running (5)	0.65	1.68	0.75
Jumping (5)	0.87	1.9	0.74
Squatting (5)	1.65	2.35	0.73
Supports (10)	4.13	4.78	0.71
Work, activities of daily life (20)	9.21	8.23	0.71

The results of Cronbach's alpha was α=0.76 and α=0.85 based upon standardized items for OMAS, which is within the range for acceptable internal consistency of the score. The Cronbach's alpha score reduces when all items are deleted, except for the pain item which increases to α=0.77, indicating item redundancy of this item in the questionnaire. The results for each item are shown in table 101. Table 102 shows the inter-item correlation matrix for the Cronbach's alpha analysis of OMAS scores.

Table 102 - Inter-item correlation matrix

	Pain	Stiffness	Swelling	Climbing stairs	Running	Jumping	Squatting	Supports	Work, activities of daily life
Pain	1	0.24	0.34	0.37	0.31	0.31	0.35	0.32	0.37
Stiffness	0.24	1	0.31	0.25	0.26	0.26	0.24	0.19	0.20
Swelling	0.34	0.31	1	0.28	0.26	0.26	0.29	0.24	0.29
Climbing stairs	0.37	0.25	0.28	1	0.46	0.47	0.54	0.53	0.51
Running	0.31	0.26	0.26	0.46	1	0.77	0.51	0.45	0.42
Jumping	0.31	0.26	0.23	0.47	0.77	1	0.58	0.50	0.45
Squatting	0.35	0.24	0.29	0.54	0.51	0.58	1	0.58	0.55
Supports	0.32	0.19	0.24	0.53	0.45	0.50	0.58	1	0.59
Work, activities of daily life	0.37	0.20	0.29	0.51	0.42	0.45	0.55	0.59	1

Table 103 - Cronbach's alpha analysis of ankle function subscale

Cronbach's alpha	Cronbach's alpha based on standardized items
0.76	0.87

There were six items included in the ankle function subscale: climbing stairs, running, jumping, squatting, supports and work/activities of daily life. The number of OMAS scores included in this analysis was n=1818, with 59 questionnaires excluded as they contained missing data. Table 103 shows the outcome of Cronbach's alpha for the ankle function subscale. This shows that α=0.76 with α=0.87 based upon standardized items for the ankle function subscale.

Table 104 - Cronbach's alpha analysis of ankle symptoms subscale

Cronbach's alpha	Cronbach's alpha based on standardized items
0.46	0.56

There were three items included in the ankle symptoms subscale: pain, stiffness and swelling. The number of OMAS scores included in this analysis was n=1873, with 54 questionnaires excluded because they contained missing data. Table 104 shows the outcome of the Cronbach's alpha for the ankle symptoms subscale. As the table shows, α=0.46 for the subscale and α=0.56 based upon standardized items for the subscale.

6.3.3.e Responsiveness

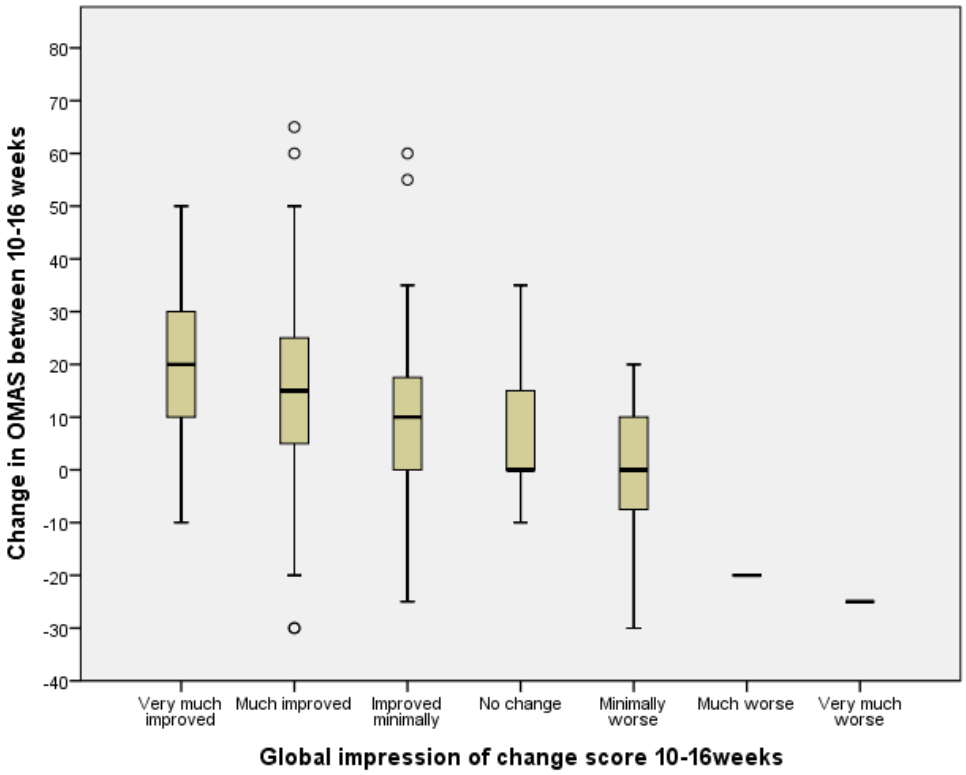


Figure 43 - Boxplot showing change in OMAS from 10-16 weeks by GIC respondent group

Figure 43 shows a boxplot of the change in OMAS scores between 10-16 weeks plotted against the GIC score responses at 16 weeks. From the diagram you can observe that as the GIC score response improves, the median score of the OMAS increases. Despite this, all of the interquartile range boxes for the very much improved to the minimally worse groups overlap. The numbers are low for the much worse and very much worse, with only one participant in each of these groups, therefore these results are inconclusive due to the low numbers in these groups.

6.3.3.f Interpretability

Table 105 - Frequency of highest and lowest possible OMAS scores

Time point	Total responses at time point (n)	Frequency of 0 scores (n)	Percent of 0 scores (%)	Frequency of 100 scores (n)	Percent of 100 scores (%)
Baseline	545	57	10.5	1	0.2
6 weeks	423	5	1.2	2	0.5
10 weeks	389	0	0	7	1.8
16 weeks	410	0	0	32	7.8
Total	1767	62	3.5	42	2.4

Table 105 shows the frequency of scores at the highest and lowest level per time point and overall to enable assessment of edge effects within the OMAS scores. At baseline, 57 of 545 OMAS scores (10.5%) were at the lowest score (0). At 16 weeks, 32 of 410 OMAS scores (7.8%) were at the highest score (100). Of the scores across all time points, 62 of 1767 (3.5%) OMAS scores were at the lowest level and 42 of 1767 (2.4%) OMAS scores were at the highest score. None of these figures reach the pre-defined level (15%) to be classed as floor or ceiling effects in the sample studied here.

Table 106 - Mean changes in OMAS scores from 10-16 weeks in each GIC response group

GIC response	Sample (n=)	Mean change in OMAS score 10-16 weeks	Standard deviation
Very much improved	62	20.7	14.8
Much improved	156	15.6	16.9

Minimally improved	83	9.7	14.7
No change	21	6.4	12.3
Minimally worse	12	0.4	14.4
Much worse	1	-20	-
Very much worse	1	-25	-

Table 106 shows the change in OMAS score by GIC response group. The mean change in OMAS was 6.4 in the no change group, 9.7 in the minimally improved group, 15.6 in the much improved group and 20.7 in the very much improved group. The estimated MIC of the OMAS between 10 – 16 weeks is 9.7 points in this patient population, in this context of use.

6.3.4 Measurement properties of OMAS by subgroup

6.3.4.a Convergent validity

Table 107 - Correlations of scores of OMAS with scores of comparator instruments and domains by subgroup

Comparator instrument/domain	Subgroup	Sample size (n=)	Pearson's correlation (r=)	Significance (2-tailed)
DRI	≤47yrs group	179	-0.78	<0.01
	≥48yrs group	232	-0.77	<0.01
MOXFQ	≤47yrs group	101	-0.86	<0.01
	≥48yrs group	122	-0.85	<0.01
MOXFQ-pain domain	≤47yrs group	103	-0.76	<0.01
	≥48yrs group	122	-0.72	<0.01
MOXFQ-walking standing domain	≤47yrs group	101	-0.86	<0.01
	≥48yrs group	122	-0.85	<0.01
	≤47yrs group	102	-0.71	<0.01

MOXFQ-social interaction domain	≥48yrs group	122	-0.78	<0.01
EQ-5D value set	≤47yrs group	195	0.71	<0.01
	≥48yrs group	247	0.75	<0.01
EQ-5D-mobility domain	≤47yrs group	195	-0.76	<0.01
	≥48yrs group	247	-0.76	<0.01
EQ-5D-pain & discomfort domain	≤47yrs group	195	-0.62	<0.01
	≥48yrs group	247	-0.66	<0.01
EQ-5D-self-care domain	≤47yrs group	195	-0.61	<0.01
	≥48yrs group	247	-0.57	<0.01
EQ-5D-anxiety & depression domain	≤47yrs group	195	-0.28	<0.01
	≥48yrs group	247	-0.39	<0.01
EQ-5D-usual activities domain	≤47yrs group	195	-0.74	<0.01
	≥48yrs group	247	-0.75	<0.01
EQ-5D-overall health state	≤47yrs group	195	0.55	<0.01
	≥48yrs group	246	0.57	<0.01

The correlations of OMAS scores and scores of comparator instruments and domains by subgroup are found in table 107. These are comparable to the results found in the convergent validity for the total sample. The association between OMAS and EQ-5D anxiety & depression domain is higher in the ≥48yrs group at -0.39 compared to the ≤47yrs group which is -0.28. Because the correlations found in this subgroup analysis were very similar to those found in the total sample analysis, I have not included scatterplots here, as they do not show any new information to those presented previously, found in section 6.3.3.a.

6.3.4.b Discriminative validity

Table 108 - Descriptive statistics for participants with operatively and non-operatively fractures by subgroup

Time point	Age group	Fracture management	Sample (n=)	Mean OMAS score	Standard deviation	Standard error of the mean
OMAS baseline	≤47yrs group	Operative	165	19.61	16.51	1.29
		Non-operative	118	20.76	17.30	1.59
	≥48yrs group	Operative	133	24.40	19.54	1.69
		Non-operative	139	20.29	15.62	1.33
OMAS 6 weeks	≤47yrs group	Operative	114	35.66	21.02	1.97
		Non-operative	83	42.05	22.89	2.51
	≥48yrs group	Operative	111	35.77	20.56	1.95
		Non-operative	115	43.00	20.80	1.94
OMAS 10 weeks	≤47yrs group	Operative	101	44.65	22.86	2.28
		Non-operative	78	53.72	23.04	2.61

	≥48yrs group	Operative	102	49.41	23.48	2.32
		Non-operative	109	54.22	20.12	1.93
OMAS 16 weeks	≤47yrs group	Operative	108	63.52	23.39	2.25
		Non-operative	81	67.47	22.25	2.47
	≥48yrs group	Operative	101	63.07	23.91	2.38
		Non-operative	120	65.46	23.43	2.14

Table 109 - Results of t-test for comparison of mean scores of operative and non-operatively managed patients by subgroup

Time point	Age group	t-test for equality of means						
		t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% confidence interval of the difference	
							Lower	Upper
Baseline	≤47yrs group	-0.57	281	0.569	-1.16	2.03	-5.15	2.84
	≥48yrs group	1.91†	252.51	0.057	4.11	2.15	-0.13	8.35
6 weeks	≤47yrs group	-2.03	195	0.044	-6.39	3.15	-12.60	-0.18
	≥48yrs group	-2.63	224	0.009	-7.23	2.75	-12.66	-1.81
10 weeks	≤47yrs group	-2.62	177	0.010	-9.06	3.46	-15.89	-2.24
	≥48yrs group	-1.60	209	0.111	-4.81	3.00	-10.73	1.11
16 weeks	≤47yrs group	-1.17	187	0.242	-3.95	3.37	-10.59	2.69
	≥48yrs group	-0.75	219	0.455	-2.39	3.19	-8.68	3.91

† = t-test not assuming equal variance due to result of Levene's test

The descriptive statistics for operatively and non-operatively managed patients in each subgroup per time point are found in table 108. Prior to running the t-test, Levene's test for equality of variances was completed. One test was significant (baseline in ≥48yrs group f=5.15 p=0.024). The remainder of tests were not significant (baseline ≤47yrs group f=0.11 p=0.736, six weeks ≤47yrs group f=0.58 p=0.449, six weeks ≥48yrs group f=0.03 p=0.867, 10 weeks ≤47yrs group f=0.15 p=0.695, 10 weeks ≥48yrs group f=2.53 p=0.113, 16 weeks ≤47yrs group f=0.71 p=0.399, 16 weeks ≥48yrs group f=0.52 p=0.473). Therefore, t-test assuming equal variances was used for all apart from the baseline ≥48yrs group, for which the t-test not assuming equal variances was used.

Results of these t-tests are shown in table 109. Results show that the mean OMAS scores are not statistically significantly different in the groups, expect for in three instances. The ≤47yrs group shows a statistically significant difference at six weeks (t= -2.03 p=0.044) between the operatively and non-operatively managed individuals. The mean difference was -6.39. At the same time point, a statistically significant difference was found between operatively and non-operatively managed individuals in the ≥48yrs group (t= -2.63 p=0.009), with a mean difference of -7.23. At the 10 week time point, there was a statistically significant difference in the mean OMAS scores in the ≤47yrs group (t= -2.62 p=0.010) with a mean difference of -9.06.

The analyses for discriminative validity for Weber classification and number of malleoli involved in the fracture has not been completed for the subgroups because the numbers included in the groups were too low to run the subgroup analyses.

6.3.4.c Internal consistency

Table 110 shows the Cronbach's alpha of the OMAS by subgroup. In the ≤47yrs group there were a total of 848 questionnaires, however 39 contained missing data so this analysis was completed on a sample of 809 fully completed OMAS scores. For the ≥48yrs group, there were 1025 scores in total, however 67 contained missing data so this meant the analysis was based upon a sample of 958 fully completed OMAS questionnaires

Table 110 - Cronbach's alpha of OMAS scores of the total questionnaire by subgroup

Age group	Cronbach's alpha	Cronbach's alpha based on standardized items
≤47yrs group	0.79	0.87
≥48yrs group	0.73	0.83

Table 111 - Cronbach's alpha of OMAS scores if item deleted by subgroup

Age group	Questionnaire item (total points available)	Mean	Standard deviation	Cronbach's if item deleted
≤47yrs group (n=809)	Pain (25)	13.08	9.41	0.80
	Stiffness (10)	1.88	3.91	0.78
	Swelling (10)	3.83	3.89	0.77
	Climbing stairs (10)	5.66	3.20	0.75
	Running (5)	0.67	1.71	0.78
	Jumping (5)	1.01	2.01	0.77
	Squatting (5)	1.74	2.38	0.76
	Supports (10)	4.04	4.76	0.74
	Work, activities of daily life (20)	8.76	8.19	0.75
≥48yrs group (n=958)	Pain (25)	14.44	9.20	0.74
	Stiffness (10)	2.78	4.48	0.73
	Swelling (10)	3.63	3.72	0.72
	Climbing stairs (10)	5.49	3.17	0.67
	Running (5)	0.63	1.66	0.72
	Jumping (5)	0.76	1.80	0.72
	Squatting (5)	1.57	2.32	0.71
	Supports (10)	4.20	4.80	0.69
	Work, activities of daily life (20)	9.60	8.26	0.68

The results of Cronbach's alpha by subgroup (table 111) shows that the internal consistency remains within acceptable ranges in both the ≤47yrs group and the ≥48yrs group (α=0.79 and α=0.73 respectively). This shows that the OMAS is slightly more internally consistent in the younger age group, but both subgroups are within the acceptable limits of 0.70 - 0.95. Pain remains to be a redundant item in both subgroups, increasing in both subgroups to α=0.80 in the ≤47yrs group and 0.74 in the ≥48yrs group. Additionally, in the ≥48yrs group, the Cronbach's alpha remains the same when the item for stiffness is deleted, indicating a possible redundancy of this item in the OMAS in this particular subgroup.

Table 112 - Cronbach's alpha of ankle function subscale by subgroup

Age group	Cronbach's alpha	Cronbach's alpha based on standardized items
≤47yrs group	0.79	0.89
≥48yrs group	0.74	0.86

Table 112 shows the Cronbach's alpha analysis ankle function subscale of the OMAS by subgroup. There was a total of 848 responses in the ≤47yrs group. Seventeen of these, however, contained missing data within the specific subscale, therefore a sample of 831 was used for this analysis in the ≤47yrs group subgroup. In the ≥48yrs group. There was a total of 1025 responses, however 42 of these contained missing data within the specific subscale items, therefore a total of 983 responses were included in this analysis.

Results of this shows α=0.79 in the ≤47yrs group, with α=0.89 for standardized items. In the ≥48yrs group α=0.74 and α=0.86 for standardized items. There are similar to results found in the total sample analysis.

Table 113 - Cronbach's alpha of ankle symptoms subscale by subgroup

Age group	Cronbach's alpha	Cronbach's alpha based on standardized items
≤47yrs group	0.48	0.59
≥48yrs group	0.44	0.54

Table 113 shows the Cronbach's alpha of the ankle symptoms subscale of the OMAS by subgroup. In the ≤47yrs group there were a total of 848 scores, however 25 were

excluded as they contained at least one item of missing data in that subscale, therefore the Cronbach's analysis on one this subscale in this subgroup was 823 OMAS scores. For the ≥48yrs group, there were a total of 1025 questionnaires, however 29 of these were excluded as they contained at least one item of missing data within the subscale, so the sample size for this calculation in this subgroup was 996.

Results of this shows α=0.48 in the ≤47yrs group, with α=0.59 for standardized items. In the ≥48yrs group α=0.44 and α=0.54 for standardized items. There are similar to results found in the total sample analysis found in section 6.3.3.d.

6.3.4.d Responsiveness

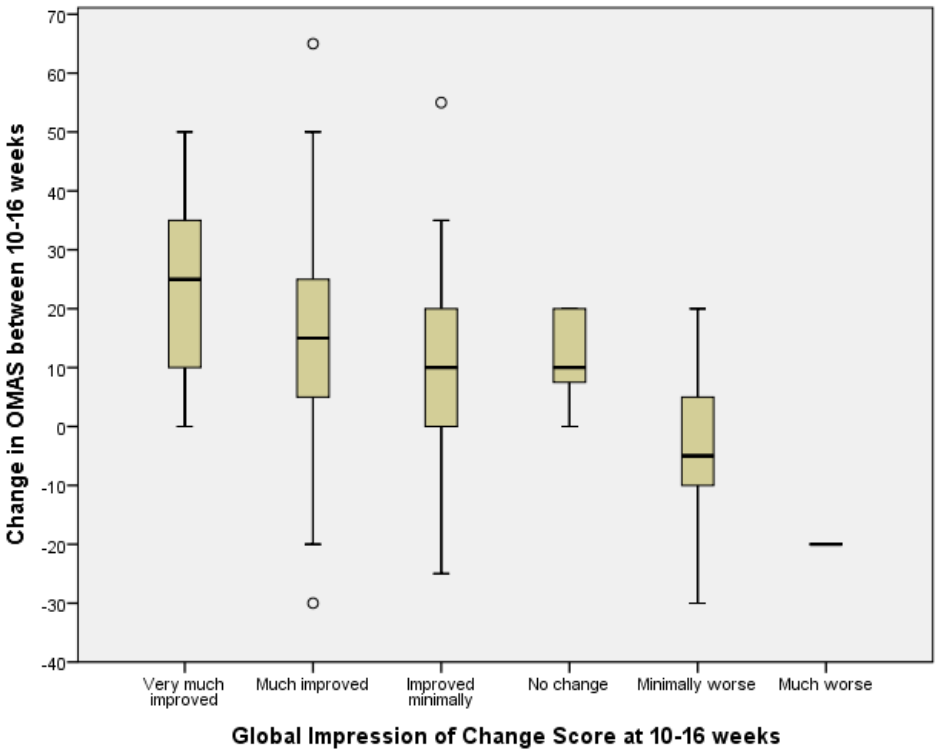


Figure 44 - Boxplot of GIC scores and change in OMAS scores between 10-16 weeks in ≤47years group

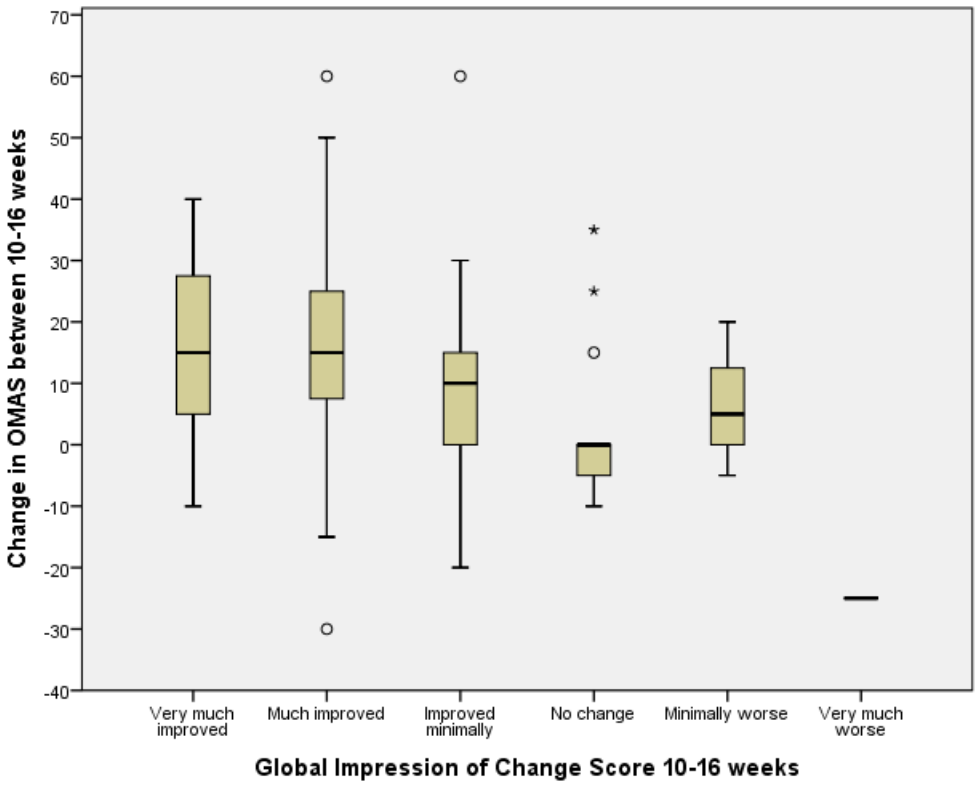


Figure 45 - Boxplot of GIC scores and change in OMAS scores between 10-16 weeks in ≥48years group

The boxplots shown in figures 44 and 45 show the responsiveness of the OMAS in the two subgroups. From these diagrams, we can see that the in the ≤47yrs group, the median of the OMAS scores increase as the GIC score improves. This is not seen so distinctly in the ≥48yrs group, where the lines for the median are relatively flat. This indicates that the OMAS may not be as responsive the older age group than the younger age group. Similar to the boxplot presented for responsiveness of the whole sample (section 6.3.3.e), the interquartile range boxes overlap in each GIC group.

6.3.4.e Interpretability

Table 114 - Frequency of highest and lowest possible OMAS scores by subgroup

Age group	Time point	Total responses at time point (n=)	Frequency of 0 scores (n=)	Percent of 0 Scores (%)	Frequency of 100 scores (n=)	Percent of 100 Scores (%)
≤47years	Baseline	183	31	10.2	0	0
	6 weeks	169	3	1.6	1	0.5

	10 weeks	192	0	0	4	2.4
	16 weeks	304	0	0	13	7.1
	Total	848	34	4	18	2.1
≥48years	Baseline	238	26	8.7	1	0.3
	6 weeks	238	2	0.8	1	0.4
	10 weeks	250	0	0	3	1.3
	16 weeks	299	0	0	19	8
	Total	1025	28	2.7	24	2.3

Table 114 shows the frequency of highest and lowest scores of OMAS by subgroup. As the results for the whole sample of participants did not demonstrate any edge effects beyond the pre-specified 15% threshold, the subgroup analyses also do not demonstrate edge effects in this sample. There were slightly more scores of 0 at baseline in the ≤47years group than the ≥48years group (10.2% to 8.7%) and slightly more scores of 100 at 16 weeks in the ≥48years group compared to the ≤47years group (8% to 7.1%).

Table 115 - Mean changes in OMAS scores from 10-16 weeks in each GIC response group by subgroup

Subgroup	**GIC response**	**Sample (n=)**	**Mean change in OMAS 10-16 weeks**	**Standard deviation**
≤47years	Very much improved	23	19.78	14.26
	Much improved	62	19.19	15.92
	Minimally improved	31	9.35	17.02
	No change	9	8.33	10.31

	Minimally worse	9	<0.0001	16.58
	Much worse	1	-20	-
	Very much worse	0		
≥48years	Very much improved	38	21.18	15.49
	Much improved	94	13.19	17.23
	Minimally improved	52	9.9	13.23
	No change	12	5.0	13.82
	Minimally worse	3	1.67	5.77
	Much worse	0		
	Very much worse	1	-25	-

Table 115 shows the mean change in OMAS scores for each GIC response group by subgroup. The mean change of OMAS in the minimally improved groups for the ≤47yrs and ≥48yrs groups were 9.35 and 9.9 respectively. The result is similar for the whole sample overall which was 9.7 points.

6.4 Discussion

Results presented here show that OMAS demonstrates sufficient convergent validity relative to the hypotheses set for expected magnitudes of correlations with scores of other instruments and domains. The only hypothesis which was not met was for the association of scores of OMAS with scores of EQ-5D anxiety and depression domain. As outlined prior to analysis, however, nine of the 12 hypothesis were required to be met in order to score sufficiently for construct validity. The strongest associations of OMAS scores were with scores of MOXFQ, MOXFQ walking-standing domain, EQ-5D mobility domain and DRI. The lowest correlations apart

from the EQ-5D anxiety and depression domain were the EQ-5D overall health state domain, the EQ-5D self-care domain and the EQ-5D pain and discomfort domain.

The low correlation of scores of OMAS with scores of the EQ-5D anxiety and depression supports findings of the qualitative work package, where some individuals felt that psychological wellbeing and mental health was an item missing from OMAS. Research completed by Pan, et al. [267] found similar results when correlating scores of the LEFS with scores of the SF-36 MCS and PCS in a population of individuals with traumatic lower limb injuries. Authors found high correlation between the physical component of SF-36 and LEFS but low correlations between the mental component score and LEFS in the patient population. This indicates that scores of PROMs measuring function are not likely to be correlated with scores of measures of psychological and emotional wellbeing following traumatic injury such as ankle fractures.

Overall, OMAS did not meet a sufficient number of hypotheses to be considered discriminative valid, with three of the 20 hypotheses outlined prior to analysis being met. The tests assessing the discriminative validity of the OMAS showed that there were statistically significant differences in the mean OMAS scores of operatively and non-operatively managed patients, with higher scores in the non-operatively managed patients by 6.89 at six weeks and 6.97 points at 10 weeks post injury. Whilst these differences were significant, they likely don't reach the level of clinical importance, considering the MIC here was estimated to be 9.7 points. The differences in mean OMAS scores were not statistically significant between operative and non-operative groups at baseline and 16 weeks. Results for the subgroups in this analysis showed similar results, apart from the difference between mean scores in the ≥48yrs group at 10 weeks, which did not show statistically significant difference between these patient groups.

The results of the analysis regarding the number of malleoli involved in the fracture, shows that the only fracture type groups which show statistically significant differences were the unimalleolar and bimalleolar fracture types at six weeks, where unimalleolar fractures had a mean OMAS score of 7.14 points higher than the bimalleolar group. This was the only of the four hypotheses to be met in this particular analysis. Again, this difference is likely not clinically important.

Results of the comparison of Weber classification groups showed an unexpected result, whereby the only significant result was the Weber B fracture group had a

higher average OMAS score than Weber A fractures at the baseline time point by 8.7 points (i.e. the Weber B fracture group scored better on OMAS than the Weber A fracture group). This is inconsistent with the hypothesis set prior to analysis and is probably explained by the RCT data which was used to complete this validation work. The trial inclusion criteria (as outlined in section 1.5.2) shows that for inclusion in AIR, individuals should have a fracture for which the treating clinician feels that plaster cast is a reasonable management option. Therefore, minor Weber A fractures, such as avulsion fractures of the lateral malleolus are likely to be ineligible for the trial in which this validation was embedded. Therefore, the Weber A fractures included in the AIR trial are likely to be more severe than the Weber A fractures found in the actual population of individuals with ankle fracture.

This means that the results found here for assessment of the discriminative validity of OMAS in adults with an ankle fracture are not generalizable to the population of interest due to the fact that RCT data were used to complete this analysis. This is a further limitation of using data from a clinical trial and, for this reason, the results for discriminative validity based upon Weber classification are inconclusive as they are likely not representative of the population.

A further issue with the Weber classification includes the ability of the classification system to accurately interpret the extent of soft tissue involvement in a fracture. A study by Hermans, et al. [268] assessed the correlation between radiological assessment and MRI for assessing syndesmotic injury in a sample of ankle fractures. The authors found that the Lauge-Hansen fracture classification system demonstrated more sensitivity in detecting a syndesmotic injury compared to a combination of Weber and AO-Muller classification. It is not possible to assess the extent of ligamentous damage using X-Ray and this is the imagining that the patients included in AIR would have likely had. Therefore, we cannot be sure that the Weber classification acts as a reliable surrogate for extent of soft tissue damage around the ankle following fracture.

Interestingly, this study also found that in 33% of the included cases, there was a posterior malleolar fracture which was not visible on plain X-Ray but was visible on the MRI [268]. Considering that the majority of the patients included in AIR will have only had a plain X-ray taken, this means the assessment of the number of malleoli involved in the fracture may not have been accurate with the imaging available.

The analysis to assess the discriminative validity involved testing 20 hypotheses. Testing these hypotheses using a significance level of 0.05 means that one of these results will statistically show a significant result by chance, rather than because of a true difference. Whilst there were several hypotheses tested, this felt appropriate to complete as part of an exploratory project into the functioning of a PROM. However, these results should be approached with caution due to the number of hypotheses tested as part of this analysis.

The results of Cronbach's alpha of the score demonstrate that OMAS is likely internally consistent, with an overall result of α=0.76 overall, α= 0.79 and α= 0.73 in the ≤47years and ≥48years groups respectively. Results of the Cronbach's alpha if item deleted analyses show that the pain item of the questionnaire is likely redundant, both in the overall sample and subgroups. Furthermore, the stiffness item demonstrates item redundancy in the ≥48years subgroup. The redundancy of the pain item and possibly the stiffness item in the ankle symptoms subscale, which comprises only three items is an indication that the overall scale of OMAS may be more likely to be weighted towards a measurement of the construct of patient reported ankle function, rather than symptomatic domains such as pain and swelling.

Principal component analysis of OMAS completed as part of this project shows two separate subscales present within the score, which I describe as the ankle function subscale and the ankle symptoms subscale. The ankle function subscale explains 45.32% of the variance observed in the score and consists of six items, whereas the ankle symptoms subscale consists of three items. The analysis has summarised the internal structure of the score and given an understanding of the subscales within the score. This is similar to the two subscales found within the OHS in individuals undergoing total hip replacement which were identified as pain and function [269].

As per COSMIN recommendations, the internal consistency was subsequently assessed in each of these two subscales found during the PCA. The results of this showed that the Cronbach's alpha was within acceptable range for the ankle function subscale, but not for the ankle symptoms subscale. These results were comparable in the age subgroups, also. These results, however, should be approached with caution; the results of Cronbach's alpha is affected by the number of items in a scale i.e. the higher the number of items within a (sub) scale, the higher the alpha [226 270]. Considering that the ankle function subscale has six items and the ankle symptoms subscale has three items, the results of Cronbach's alpha for this subscale is

inconclusive for this reason. Whilst this means that we cannot draw conclusions on the internal consistency of the ankle symptoms subscale, the results show that the ankle function subscale shows sufficient internal consistency in this population.

OMAS was shown to be generally responsive to self-reported changes in patient status, seen more evidently in the ≤47years group than the ≥48years group. It's also evident that the interquartile range boxes overlap between the GIC score response groups in all the boxplots presented here. This indicates that a proportion of individuals are stating that they've changed on the GIC score, but the OMAS isn't detecting these changes. This is possibly because, as outlined in chapter four, OMAS was not thought to be comprehensive for all individuals in terms of the whole construct of ankle fracture recovery, and therefore it may not be assessing all factors of importance which affect the individuals' sense of change in their status. In other words, the OMAS may not be detecting real changes in some individuals because it does not include all the factors which are considered part of recovery for them.

As we learnt in chapter four, there were other items participants felt were relevant in recovery, such as psychological wellbeing, wound issues and impact on sleep, all of which are items not explicitly considered in the OMAS questionnaire content. It's pertinent to note here that scores of EQ-5D domain of anxiety and depression correlated poorly with scores of OMAS. This indicates that these instruments are measuring two separate constructs, but individuals might be considering a combination of these factors when responding to the GIC question. It's possible that individuals are changing on the many facets of ankle fracture recovery, but OMAS is not capable of detecting all these improvements which contribute to an improved outcome for individuals, hence why the interquartile range boxes overlap.

The data presented here shows that there was no edge effects observed in the scores of OMAS, with the frequencies of highest and lowest scores being less than 15% in all time points included in this project. There is evidence to suggest that full recovery from an ankle fracture can take 12 months or more [271]. From the work presented here, it's not possible to conclude whether ceiling effects occur in OMAS scores at later time points than 16 weeks, and indeed studies completed by other researchers indicate this might be the case.

A study by Turhan, et al. [170] assessed the floor and ceiling effects of the score at a mean follow up time point of 4.3 years and found that 29% of responses were at the highest score of 100, which according to the criteria applied here and in many other

studies [108 175 177] would be classed as a ceiling effect. Whilst the primary end point for the AIR trial is at 16 weeks, we can be therefore confident that edge effects are unlikely to be present in the results of AIR. This should, however, be considered as a potential issue at later time points for studies which use a primary end point of later than this, and requires further investigation.

Based on results presented here, the minimally important change of OMAS is estimated at 9.7 points. This is consistent with the MIC used in several randomised controlled trials for adults with ankle fracture, the authors of whom used a MIC of 10 points [41 72 272]. Other studies have used slightly lower MIC in the calculation of their sample sizes, such as Kortekangas, et al. [273] who used a MIC of 8.8 in their study assessing different immobilisation methods for ankle fracture. Briet, et al. [274] used an MIC of 7 points for their trial of weight-bearing protocols following surgical fixation of ankle fractures.

Whilst this is based upon an anchor method for assessing the MIC, whereby the patients have responded to a GIC question to verify their subjective change in their ankle status, it's important to note that this is an estimate of the MIC. It's not known whether the level of "minimally improved" is acceptable for individuals who are undergoing possibly risky interventions. A study completed by Roos, et al. [275] used slightly different anchor questions when assessing meaningful and important changes in patient undergoing anterior cruciate ligament reconstruction. These anchor items may have been more useful in this instance, whereby answers include an indication of both the magnitude of the change they have experiences and also how important that change is to them.

If a clinical trial compares interventions with high risk profiles, then it might be that the improvement might need to be higher (e.g. much improved on the GIC score) to warrant the additional risks on the intervention. In this particular case within the context of the AIR trial, which evaluates two low risk and routinely used immobilisation methods for ankle fracture, using the mean change of the minimally improved group as an estimate of the MIC is likely sufficient for this purpose. This means that this estimate of the MIC is likely not appropriate for interventions with higher risk profiles, such as operative and non-operative management.

Results of the subgroup analyses in this project show that the measurement properties in both the younger and older participants are very similar. Convergent validity of the scores was adequate, with the main difference being in the EQ-5D

anxiety and depression domain, for which a slightly higher correlation was found between this domain and OMAS in the older age group compared to the younger age group. The internal consistency of the score was adequate in both subgroups and the pain item demonstrated redundancy, as well as the swelling item in the older age group. The ankle function subscale had an adequate internal consistency in both groups. For responsiveness, the score looks to be more responsive in the younger population included here upon inspection of the boxplots. The low numbers in this analysis when the sample was split into two subgroups means that these results should be interpreted with caution.

Results presented here are comparable to those reported by Garratt, et al. [169], who compared three different PROMs for assessing outcome following ankle fracture in translated Norwegian versions. The key differences to this research are that here I used a sample of both operatively and non-operatively managed patients, whereas the researchers studying the Norwegian version only assessed operatively managed individuals. The results obtained for construct validity were similar to those presented here, with similar low correlations of the OMAS scores with scores of the EQ-5D domains of anxiety and depression (r= -0.31) and self-care (r= -0.33). Score of the EQ-5D correlated highly overall (r= 0.79) with scores of OMAS, which again is similar to results presented here (r= 0.73). The results for internal consistency of the whole OMAS scale was 0.82 in this article, similar to results reported here.

Articles by Buker, et al. [168] and Turhan, et al. [170] translated the OMAS into Turkish and assessed measurement properties of the score in a sample of patients with ankle fracture. Buker, et al. [168] assessed operatively managed patients only and analysed the construct validity and reliability of the score. They found an internal consistency of α=0.76 which is similar to the results found here. The comparator instruments were not the same as those used here, however they showed high correlation of scores of OMAS with those of the subscales of the FAOS and the global self-rated function score.

Turhan, et al. [170] assessed a mixture of operatively and non-operatively managed patients. As discussed previously in this section, the authors of this study assessed patients at a mean of 4.3 years following injury, and found high levels of individuals scoring the highest possible score on OMAS. The researchers also correlated the scores of OMAS with scores of the domains of FAAM and the SF-12. They found moderate to high (r= 0.6 – 0.86) correlations within all these subscales. Interestingly,

the lowest level correlation of r= 0.6 was found with OMAS and scores of the SF-12 MCS, which is comparable to results found here with the EQ-5D anxiety and depression.

Nilsson, et al. [276] studied the measurement properties of the OMAS in a group of individuals recovering from operatively managed ankle fractures at six months and 12 months following injury. The authors of this study found a similar Cronbach's alpha of α=0.76, indicating acceptable internal consistency of the score overall. The study also assessed the convergent validity of scores of OMAS with scores of the FAOS and its subscales, all of which demonstrated high correlations (r >0.8). The authors also assessed the distributions of scores and at six months post injury, there were no edge effects present in the scores. At 12 months follow up, however, 15% of participants recorded 100 points on the questionnaire, which is a further indication that ceiling effects may be present in OMAS at later time points than 16 weeks, as previously discussed.

The strengths of this project include the large data set and sample size used. The setting of the validation research within a multi-centre RCT for individuals with ankle fracture is also a strength, as this thesis is concerned primarily with the measurement properties of these scores in this particular context. The multi-centre nature of the data we had available is also a strength because we include a range of individuals from different parts of the country, rather than being a single site analysis which may restrict the generalisability of results. This is also the first large evaluation of the OMAS score in a UK based setting, something which is warranted considering the number of studies using this PROM as an outcome measure, which was demonstrated at the start of this thesis during chapter two. Furthermore, no other researchers have completed PCA to understand the underlying structure of OMAS before, or estimated the MIC, which gives useful insight to the score.

Whilst the contextual basis of a multi-centre RCT is a strength, there are also aspects of this which could be considered limitations to this validation research. Inherently, RCT data such as is presented here is predisposed towards having a plentiful supply of baseline data and less follow up data available. As the outcome data presented in histograms by time point shows (sections 5.3.1b and 5.3.2b), there is comparatively more baseline data than follow up data. This is because to be entered into the trial, all participants completed the baseline questionnaire pack as a standard, which must have been completed before consent and randomisation. Not all follow up

questionnaires are collected, for several possible reasons such as participant withdrawal or participants not returning the questionnaires. This means that we naturally have more scores for baseline available than we do at the follow up time points.

The limitation of using clinical trial data is also reflected in the missing data present in the data set. In this project, I was limited to the data obtained in the clinical trial at the data of data extraction and had no control over the number of unit non-response here. The issues with unit non-response here is that it is highly likely that these data were not MCAR and that there is probably a reason why the individuals haven't returned these questionnaires, which subsequently could affect the results obtained for the measurement properties. For example, a large proportion of individuals might not have returned their questionnaires because they were doing well, having returned to pre-fracture function and therefore felt there was no point or value in returning their questionnaire. Conversely, others may have been struggling so much with their ankle that they did not want to complete a questionnaire about their ankle or want to walk to the post box to return it to the trial office. Missing these two possible subsets of individuals could have an effect on the measurement properties. For example the evaluation of floor and ceiling effects, results of which could be quite different if we had data for the individuals who fall into the categories above. Because we cannot be confident that the missing data is MCAR, this introduces bias.

Another limitation to the use of clinical trial data is the selection bias which may have been encountered, both in the trial and subsequently this secondary analysis. There is evidence to suggest that clinical trials often underrepresent individuals from minority ethnic background. A study by Rochon, et al. [277] studied the representation and reporting of data relating to ethnic background in trials published in a sample of medical journals and the researchers found that minority ethnic groups are frequently underrepresented in RCTs. Researchers have completed a review of the possible reasons for this underrepresentation of ethnic minorities in clinical trials in the UK [278]. The authors of this study reviewed the potential barriers to inclusion of ethnic minorities in RCTs, which included inappropriate exclusion criteria based on language restrictions, health provider attitudes and socio-cultural barriers [278].

Underrepresentation of ethnic groups is an issue in clinical trials because interventions may work differently for people from different ethnic backgrounds and therefore it's essential that the samples recruited to these gold-standard tests for

interventions are representative of the population which these interventions will ultimately treat [279].

The lack of representation of different cultural backgrounds is an issue because cultural interpretation can influence the performance of PROMs [280], therefore it's of paramount importance that samples being studied is inclusive of individuals from a range of ethnicities and backgrounds. We can see in table 65 in the descriptive statistics section that 84.8% of the sample identified as being from a white ethnic background. This issues was also discussed in chapter 4 which details the qualitative work included in this thesis, whereby we cannot be sure that the patient experiences of individuals from different cultural backgrounds would not be different.

The quality of the conclusions drawn for convergent and discriminative validity are dependent up on the accuracy of the hypotheses set a priori. There is little understanding of the comparator instruments' measurement properties in the population of adults with ankle fracture, as discussed in chapter three of this thesis, therefore it's difficult to be sure that the hypotheses used for assessing convergent and discriminative validity are wholly appropriate.

The hypotheses for discriminative validity were more difficult to set and has resulted in conclusions which have been more difficult to interpret, as previously discussed. The used of the Weber classification as surrogate for severity has caused some results which feel difficult to interpret, partly because of the sample of individuals studied. Upon reflection, the analysis using Weber classification and number of malleoli involved in the fracture was not an appropriate analysis to do and has resulted in a more exploratory analysis of the score in different fracture types, rather than offering information on the discriminative validity of OMAS.

It's likely that Weber classifications are not associated with differing OMAS scores. This is supported by a study by Verhage, et al. [281], who completed a retrospective cohort study on the outcome of operatively managed ankle fractures by the number of malleoli involved in the fracture. The authors of this study found that poorer functional outcomes were most strongly associated with involvement of the medial malleolus, regardless of how many other malleoli were involved in the fracture. They also noted that the presence of a posterior malleoli fracture was of less importance than the size of the posterior malleolar fragment. Results presented by this author team show that it's likely that the analysis I performed here was too simplistic and

that the involvement of different structures, both bony and soft tissue, is likely to be more complex than I initially hypothesised.

A further limitation is in the nature of the injury being studied. Assessing responsiveness for a chronic condition which fluctuates more readily is easier than a condition like an acute fracture because people with an injury like this generally tend to improve over time. Consequently, there aren't many people whose perceived ankle status remains constant or gets worse over time, which is reflected in the responses for the GIC score question shown in table 106. The nature of the recovery trajectories of injuries such as ankle fractures makes it very difficult to draw conclusions on the ability of PROMs to detect stability or negative changes in a patient's status.

In conclusion, results presented here show that OMAS has sufficient construct validity against the hypothesis formulated a priori within the sample analysed. There was a statistically significant difference in mean OMAS scores between operatively and non-operatively managed patients at six and 10 weeks post injury. The internal structure of the OMAS shows two underlying subscales; ankle function and ankle symptoms. The ankle function subscale is likely to be internally consistent based upon results of Cronbach's alpha, but no conclusions can be drawn on the internal consistency of the ankle symptoms subscale. OMAS is responsive to changes in patient status, but this responsiveness is less obvious in the ≥48 years subgroup. There were no edge effects present in the score at the time points assessed and the minimally important change was estimated at 9.7 points.

In the next chapter of this thesis, I will draw conclusions on this multi-methods project completed here, along with some recommendations and implications for clinical academics designing and planning for further clinical trials in the population of adults with ankle fracture. Recommendations for future work within this research area will also be discussed.

Throughout this chapter I have explored the validity, reliability, responsiveness and interpretability of OMAS using analysis of data from the AIR trial. I presented evidence of some sufficient measurement properties of the OMAS, such as convergent validity and internal consistency. I demonstrated through factor analysis of scores that there are two underlying subscales within the OMAS. I found there to be no floor and ceiling effects in the OMAS and made an estimate of the minimally important change of the score. The subsequent chapter will provide a discussion on the work presented here in this thesis and make recommendations for future work in this area. I will also reflect on the process of this PhD.

7. Discussion and reflections

7.1 Discussion

The overall aim of this thesis was to evaluate the way in which outcome is assessed in the population of adults with an ankle fracture in the context of RCTs and to determine whether it's appropriate. Overall, this has included understanding what outcomes are collected, a review of the quality of these outcomes, an exploratory study into the construct of interest that these PROMs assess and an evaluation of the measurement properties of a frequently used outcome measure in this particular type of study. This programme of research has shown that there is a wide range of concerns and issues that individuals recovering from this injury experience, most of which are not covered in the PROMs which are meant to assess their progress. Whilst OMAS, in some respects, does display sufficient measurement properties against commonly used criteria, it's likely not a holistic measure of the true construct of patient reported outcome.

The first review included in this thesis demonstrated that there is a wide range of outcomes collected in trials of interventions to manage adults with an ankle fracture. The increasing trend towards patient assessed outcomes from 1990's - 2010's is evident, as is the decrease in clinician assessed outcomes in this same time period. Patient reported outcomes are now the most frequently collected outcome measures in this type of study, with OMAS being the most commonly used out of these.

This work also highlighted variance in measurement methods used. For example, many different ways of measuring strength, swelling or radiographic outcome. If objective measures such as these are deemed to be important to collect in this type of study, then it's imperative that measurement methods are standardised to allow for accurate comparison between studies, and also for the purposes of pooling results in systematic reviews. Future research should also assess the value of these objective measures in relation to PROMs and the importance of clinical and objective measures in creating benefits for the patients. For example, researchers found no association between radiographic assessments and patient functional outcome measured through range of movement assessments and also PROM scores in individuals with distal radius fracture [282]. If these objective assessments are not deemed important to patients in terms of improvements to quality of life, the value of collecting these measures in RCTs must be questioned. Further work to identify

if these objective measures are associated with improved patient reported health related quality of life is warranted.

Chapter 3 sought to identify and critically appraise the evidence for the measurement properties of the region specific PROMs found to be used in the first review. Results of this showed that the PROMs routinely collected in studies to assess clinical effectiveness of interventions for ankle fractures were not able to be recommended for use due to a lack of evidence of sufficient measurement properties. The older and most researched PROM had some evidence for its measurement properties, but a significant lack of patient involvement in its development. In contrast, A-FORM was developed comparatively well, but will no information on the measurement properties of this PROM in its final format. Another gap identified by this review was the understanding of the experiences of adults who sustain an ankle fracture, the factors they find important to them and ultimately the construct of interest being assessed by these outcome measures. These two systematic reviews addressed the first objective outlined in section 1.7 of this thesis.

The subsequent work package outlined in chapters 4 focussed on the area of patient experience and the items of most important to individuals in their recovery, which addressed the second objective of this thesis. The findings of this study showed that individuals who have sustained an ankle fracture experience a range of difficulties which are far broader than the content of some PROMs used to assess this patient population. Participants described the difficulties with walking and weight-bearing restrictions, and this was especially evident in older participants of the study.

There was a high number of participants, in the older group most notably, who described difficulty with losing a sense of independence, whether that was in regard to their mobility, transport, personal care or household chores. Many described significant frustrations a sometimes feelings of depression, anxiety and low mood which they attributed to their injury and/or its effects on their life. Some individuals spoke of the frustrations of not being able to perform usual roles within their family, particularly with respect to dependents. Sleep disturbance was also described as important in the context of the ankle fracture recovery period for some individuals.

This research showed the broad nature of the construct of recovery, and the extensive and wide ranging effects that this injury, defined as a non-complex fracture [30], has on individuals' lives. The profound difficulty of weight-bearing restriction, particularly evident in the older population of those included in the study highlights

the need for further research into the need for these restrictions. This area of research has progressed significantly in the field of hip fracture surgery, whereby it is now common practice to enable patients to fully weight bear following fixation, in the understanding that the benefits outweigh the risks [283]. The post-operative management of ankle fractures may be found to require a similar approach. Indeed, there is research currently ongoing to address this question, which highlighted here as a significant and distressing issue for patients with this injury [54].

Throughout the interview study, I also explored the relevance and comprehensiveness OMAS score, results of which are outlined in chapter 5. This work contributed to answering the third objective outlined in section 1.7, regarding the measurement properties of OMAS as this was an assessment of the content validity of the score. This developed an understanding of patient opinions on the content of this widely used questionnaire. The results of this content validity study was mixed and difficult to interpret. Answers to the relevance of pain in their recovery was interesting; people usually explained that asking about pain is appropriate and important in a PROM for ankle fractures, but also explained that they didn't experience pain that was too debilitating or affected their daily activities. The relevance of swelling and stiffness was unclear and swelling was often discussed in relation to footwear rather than in terms of the swelling itself. Many people said supports such as walking aids were a nuisance, but acknowledged the necessity of them, especially in the stages of weight-bearing restrictions.

The OMAS was comprehensive for some individuals, with several people saying they felt it covered everything of importance to them. Others felt that the questionnaire omitted key items such as mental health, sleep, wound concerns and return to driving. Finally, there were issues with the response options available for some questions, with the most common issues being insufficient options, poorly worded response options and response options which were not reflective of their experience in relation to the item in question. People also described frustrations with the binary response options in some questions, where they felt their true status was not able to be reflected in the scores. For example, if their stiffness in their ankle had improved, but had not completely gone, it was difficult to feel this had been accurately captured in the available response of "stiffness" or "none". Some argued that, in some instances, binary response were appropriate because you can either do an activity or you can't. Others explained their feelings of having to go through a

period of being impaired at an activity to get to full function again, and that for some items in OMAS, the response options failed to account for this period of improvement.

In chapter 6 of this thesis, I sought to evaluate some of the remaining measurement properties of OMAS; the construct validity, internal structure, internal consistency, responsiveness and interpretability of the score. This aimed to gain a complete picture of the measurement properties of OMAS, which is the first evaluation of this commonly used PROM in a UK setting with the English version of the questionnaire.

OMAS demonstrated sufficient convergent validity against hypotheses formulated a priori with regards to expected associations with scores of other instruments. The only score which did not meet the hypothesis was the anxiety and depression domain of the EQ-5D. From this, it's clear that OMAS does not measure a similar construct to the domain of anxiety and depression in the EQ-5D. This has also been shown to be the case in other PROMs used to assess function in individuals with trauma and orthopaedic conditions. For example, Pan, et al. [267] completed correlations of scores of the LEFS and SF-36 subscales and found that scores of LEFS had poor correlations to the scores of the mental component of the SF-36. Scores of the physical component scale and scores of the LEFS were highly correlated, which is similar to findings presented here. This shows that OMAS, along with LEFS, focus predominantly on the physical and functional aspects of recovery and often do not include items which assess an individual's psychosocial health and wellbeing.

Participants of the qualitative study also raised difficulties individuals with sleep, which some specific as being an item missing from OMAS. A study by Shulman, et al. [284] assessed sleep issues for a 12 month period in a population of individuals recovering from common fractures, of which ankle fractures was one group studied. Authors assessed the prevalence, improvement and risk factors to poor sleep in this group of orthopaedic trauma patients. The authors found that 19% of individuals with an ankle fracture reported poor sleep at three months post injury. Authors assessed the associations between poor sleep and subscales of the SMFA and found that in the lower limb fracture groups, poor sleep was associated with the emotional functioning subscale of the SMFA, but none of the others such as body pain and body function. These associations were not mirrored in the patients with upper limb fractures, despite these groups of patients experiencing higher prevalence of sleep disturbance than lower limb fracture patients.

The discriminative validity results showed a statistically significant difference in mean OMAS scores between patients treated operatively and non-operatively at six and 10 weeks post-injury, however these differences were not observed at the baseline or 16 week time points. This difference could be interpreted in several ways; for example, it could be that this shows that operatively managed fractures are more severe injuries and this is reflected in the scores of OMAS between these two groups. Alternatively, the differences in these two groups could simply be due to effects form the surgery, rather than the severity of the injury. It's likely that individuals who have an operation to fix their ankle fracture have a more prolonged recovery within the first 10 weeks following their injury than individuals who don't. It would be interesting to ascertain data for OMAS between these two groups at later time points, beyond 16 weeks. The discriminative validity of the score in relation to fracture classification were difficult to interpret for several reasons previously outlined in chapter 6.

The internal structure of OMAS revealed two subscales; ankle function and ankle symptoms, with the ankle function subscale explaining 45% of the variance in the scores. The internal consistency of OMAS and its ankle function subscale was sufficient, although no conclusions were achievable for the internal consistency of the ankle symptoms subscale, due to the low number of items within this subscale. From results of the Cronbach's alpha if item deleted analysis, the pain item is likely to be redundant item in the score in both the sample overall and the subgroup analyses. The pain item is one of three items in the ankle symptoms subscale, and if this item is redundant, the ankle symptoms subscale becomes comparatively smaller in contribution compared to the ankle function subscale. For these reasons, it's likely that OMAS is predominantly a measure of patient reported function following ankle fracture.

The OMAS was assessed to be mildly responsive to changes in patient status, but more so in the younger age group than the older age group. Despite this, the interquartile range boxes overlapped between the GIC response groups, which indicates that the OMAS might not be sensitive to all aspects of change which patients may perceive as part of their recovery from ankle fracture. The estimate of the MIC of the OMAS was consistent with those used in completed and ongoing RCTs in this research area. No floor or ceiling effect were noted at the time points assessed in this study. As previously discussed in chapter 6, it's possible that ceiling

effects could be identified at time points later than 16 weeks following injury. Care should be taken when measuring the OMAS beyond 16 weeks following injury, and further studies are needed to explore this in more detail at later time points.

The limitations of this project include the use of clinical trial data and participants to answer the research questions posed. This was an advantage in some ways; it allowed the contextual focus to centre on clinical trial practice and also was a pragmatic way of completing the research. The population of individuals who are eligible and consent to participate in a clinical trial will always be, however, somewhat different to the general population of individuals found in the wider clinical setting, which introduces bias. Furthermore, the data collected in the format of collection was not designed for the specific purposes that I have used them for here, making some things difficult to complete or interpret. For example, completing a test-retest reliability analysis of these OMAS was not possible as participants were already burdened with completing trial questionnaires.

One of the main challenges of the qualitative work presented here is the lack of guidance on assessing the content validity of a PROM and interpreting the results of qualitative work in this area. Whilst I conclude from this work that OMAS is not fully relevant and comprehensive for all people, I question how a PROM can ever be fully relevant and comprehensive for all people with a particular condition? And if it can't, what level of relevance and comprehensiveness is sufficient *enough* to measure treatment effects at a population level in an RCT? Results from the qualitative work shows that OMAS is sufficient and acceptable for some patients included in this study. This was highlighted by Murray, et al. [285], when the suggestion was made to include some additional items in the Oxford Hip Score. The authors rejected this request, explaining that no questionnaire will ever be complete for every individual in the population it's intended for [285]. It may be that the levels of comprehensiveness and relevance of OMAS, is good enough for population level assessment in clinical trials.

From these results presented here, I conclude that the OMAS is unlikely to be measuring the true, multifaceted and holistic construct of patient important outcome for the majority of people recovering from this injury. Due to the comparatively small subscale of ankle symptoms and the possible item redundancy of the pain item, I suggest OMAS is likely measuring a construct which is predominantly concerned with patient reported ankle function following ankle fracture. This likely does not

cover the many facets of recovery which patients draw on to perceive their sense of progress throughout their rehabilitation. Particularly notable is the lack of consideration for psychological wellbeing in the OMAS questionnaire content, which many participants discussed as an issue in their recovery, but is not included in the questionnaire.

A study completed by Baumhauer, et al. [183] surveyed patients with foot and ankle conditions and asked which were the most important factors to them in their recovery. These items were then compared to the content of commonly used outcomes in foot and ankle research (AOFAS and FFI) to look for similarities and differences. The author team noted gender and age differences in patient response, also. The results of this study, although containing a broader range of foot and ankle conditions, agrees with findings of this thesis. The results showed that items of importance to patients included walking ability and pain. They concluded that, whilst the outcome measures overlapped with the patient important factors in some areas, up to 50% of the items on the outcome measures were not relevant to patients. This is similar again to results, here, with several individuals stating that items in OMAS are not relevant or important to them. Similar results have been found in other patient populations, such as arthritis and cancer [286 287].

Quite clearly, from results presented in this thesis and research by other authors in this area, that not all items collected in PROMs are relevant or important to patients with the condition they intend to assess. It's imperative that patients are collaboratively involved in the development of PROMs from the outset [288]. Here, I have attempted to retrofit patient views and opinions on a PROM which was very unlikely to have included input from patients during its development, which is a fundamental issue with the PROM [77 109]. It's possible, however, that this retrospective evaluation of a PROM will never be sufficient and that newer measures are required to replace these older instruments, to be sure that they are fully representative of the concerns affecting the specific population of individuals.

Future research should develop the knowledge on the measurement properties and overall quality of ankle fracture PROMs which have been developed with some patient involvement, such as the A-FORM. An understanding of the measurement properties of A-FORM, such as structural validity, internal consistency, construct validity and responsiveness to changes would enable greater confidence in the use of this PROM moving forwards. A-FORM is the only alternative ankle fracture

specific PROM and could be considered to be measuring a more complete representation of patient important outcome and therefore understanding its quality of functioning in people with ankle fracture is essential. This should also include a thorough qualitative enquiry with individuals with an ankle fracture to determine the content validity of the A-FORM. Preferably, this would be compared to the OMAS measure directly, to determine the most appropriate measure for this patient population in terms of content and relevance of the items.

The extent to which OMAS is appropriate to use in a trial examining interventions for adults with an ankle fracture is dependent upon the nature of the question that the trial seeks to answer. For example, if a study was to assess the impact of protein supplementation on the ankle strength of individuals recovering from an ankle fracture, ankle strength would be the construct of interest which would be most likely affected, due to the nature of the intervention. The resultant impact of any improved strength on the patient's ankle function and health related quality of life would not be understood from these results, however. OMAS has shown good construct validity against other measures of function and disability, such as the DRI and the MOXFQ. There is evidence to show that scores of OMAS does not show high association with measures of psychological functioning. This research shows that there is no one universal outcome measure for all research questions; consideration should be given to the nature of the construct that the interventions being assessed are likely to affect. Defining a range of outcome measures to be used across research projects is important to maintain consistency, whilst also allowing the flexibility of including additional measure related to the specific question being answered.

When deciding whether OMAS should be recommended for use following a review of this evidence presented here, there would likely be differences on opinion dependent upon who was reviewing the evidence. From the perspective of a measurement purist, grounded in the theoretical underpinnings of COSMIN methodology, OMAS would likely not be recommended for continued use. This is because it has insufficient content validity on the whole construct of patient outcome in ankle fracture, it was not based upon a foundation of concept elicitation qualitative interviews and was likely entirely clinician derived. From outside of the narrow field of PROMs methodology, however, for individuals who require using a pragmatic approach to answering important research questions to guide health care practice now, OMAS may be considered appropriate.

This is especially when considering some specific interventions which seek to ask specific research questions. For example, it would be highly unlikely that a comparison of two different methods of operative fracture fixation would have many differences in the psychosocial factors which the OMAS fails to cover. Therefore, in this case, measuring patient reported ankle function (i.e. using the OMAS) is probably good enough for this purpose. If this example was compared to interventions of a holistic rehabilitative and multi-disciplinary nature, it would be more appropriate to consider the ways in which psychosocial outcome could be affected by these interventions. This is because the target construct would be more multifaceted than pure ankle function and would be more reasonable to assume that interventions such as this could be expected to effect the psychological functioning of individuals with this injury. Therefore, in this instance, it would be more appropriate to consider the complete biopsychosocial functioning of the individuals who receive interventions of this nature. We know that implementation of methodological advancements in clinical research takes time, and often this time is not available when addressing urgent research questions, therefore a pragmatic approach is sometimes required.

Following further validation work on the available PROMs for ankle fracture, development of a core outcome set (COS) for use in clinical trials for individuals with ankle fracture is the next research priority. A COS is a standard set of outcome measures which are agreed upon being collected as a minimum for clinical trials for individuals with a particular disease, injury or condition [289]. Inconsistency in the outcomes being collected in similar types of study has been described as a form of research waste and can prevent outcomes being pooled in systematic reviews, which are the pinnacle of research methodology in evidence-based medicine [62 113]. Development of a COS would facilitate the comparison of results of research trials through homogeneity of outcomes, allowing for high quality systematic reviews and evidence syntheses to be conducted [112].

Core outcome sets have been implemented to varying degrees within areas of clinical research but perhaps the most successful implementation of them is in the field of rheumatology [290]. Whilst there are several clinical areas in trauma and orthopaedics which now have an associated COS, for example, low back pain [291], chronic post-surgical pain after knee replacement [292], hip fracture [293] and there is one in development for ankle fracture [294]. It is not known, however, how well these have

been implemented into the practice of performing randomised controlled trials. The benefits of core outcome sets can only be realised if they are implemented by the research community.

7.2 Implications and recommendations

The implications and recommendation following this work are:

1. There is a wide range of outcome measures used to assess endpoints in randomised controlled trials in the population of adults with an ankle fracture. PROMs are the most commonly used type of measure, with the most frequently used being the OMAS.
2. None of the PROMs collected in this type of study have sufficient evidence in order to recommend their use under COSMIN criteria. It was identified that further research was needed in this area.
3. The construct of interest for this patient population is broad and extends beyond the commonly understood symptomatic concerns of pain and swelling of the ankle. Individuals experience significant difficulties with weight-bearing restrictions, symptoms of anxiety and low mood, issues with sleep and concerns regarding their loss of independence. The older individuals in the study described these issues with a greater intensity than the younger participants.
4. The OMAS is not relevant and comprehensive for all individuals interviewed and it's response options are not sufficient in all cases for individuals to feel able to convey the issues they have in relation to a particular item. Items which were most commonly cited as being irrelevant were running and jumping. Items suggested as being missing from the measure were mental health, sleep, wound concerns and return to driving. Some individuals, however, felt that the OMAS content was appropriate.
5. OMAS is probably appropriate *enough* in some situations, for the assessment of the specific construct of interest of patient reported ankle function. For other research questions, where the aim is to assess the multifaceted nature of outcome in the patient population, OMAS is likely not appropriate. Further validation work into alternative PROMs which may be more appropriate in this population of patients, such as A-FORM and MOXFQ, is warranted. This should include qualitative and quantitative research methods.

6. The development of a consensus based core outcome set for this particular trial type would be advantageous, in order to standardise the outcomes collected in this area and ultimately reduce research waste.

7.3 Reflection

Here, I will reflect on the chapters contained within this thesis as well as the process overall.

The first systematic review outlined in chapter 2 was instrumental in regaining familiarity and confidence with research, following a break from this type of work. I found that, with the help of some training, such as the UReCA module detailed in the researcher training section, that I was able to gain my confidence with conducting systematic reviews quite quickly. I familiarised myself with Cochrane methodology and the PRISMA guidance, in order to complete this review. I found it especially difficult to manage the volume of returned records in this review. Use of an independent reviewer is also something I hadn't had experience with before and I devised ways of managing this process. The use of cloud storage technology was helpful in allowing effective and remote collaboration on this project.

The fact that this was not a traditional interventional review caused some difficulties, such as not being able to register the protocol due to ineligibility with PROSPERO. The development of the coding framework also caused some difficulty. Ideally, this would have been completed using consensus based methodology. However, I also felt that because it was simply a method of displaying my results, that this could potentially be a lot of work for comparatively not much gain in terms of research impact and future work in this area. I feel that, for the purposes of this specific project, the review and feedback consultations was sufficient to develop this basic framework used to displaying the results of this review. Since completing this review, I have seen several other authors use a specific framework, as used by two author teams completing similar reviews [295 296]. I think, if I were to complete this work again, I would follow this framework for consistency with other literature.

If I were to complete this review again, I would just focus on a specific type of trial, for example, only trials of rehabilitation. This would ensure that intervention types were more similar and therefore the construct of interest would be similar for each trial. It's understandable that so many outcome measures were collected when lots of different types of interventions were studied. It may have better focussed my research aims if I were to focus on trials similar to AIR, so rehabilitation studies only. It may be that if I did this, that the outcome measures collected may have been more homogenous.

I was really pleased when I had this review accepted for publication in BMC Musculoskeletal Disorders, which was my first experience of the publication process. I found the peer-review process much less daunting that originally expected and the comments and advice given by the reviewers were helpful in the advancement of my ideas and work. I also presented this work at the BOA congress and found this process helpful in guiding future work.

During the second systematic review outlined in chapter 3, I further developed my skills in performing systematic reviews. I found the review process much more manageable than the previous one, likely due to the experience I gained in the previous chapter's work. Furthermore, the lower number of article returns made it easier to have a more comprehensive second review system, whereby all articles were screened by two reviewers, rather than just the full texts. There were more cases in this review whereby we did not agree and this provoked thoughtful and insightful discussion to decide up on eligibility of the studies.

I found the COSMIN guidance challenging to use at times. This was largely because the guidance is relatively new in comparison to some of the older studies, making it very difficult to apply the methodology to some of the articles. Furthermore, the COSMIN risk of bias checklist was sometimes problematic in that it was a worst score counts method, which sometimes felt like I was unfairly downgrading articles of better quality than others, just because they had scored poorly on one of the criteria.

Secondly, the difficulty in reviewing multiple articles with different versions of PROMs in different languages, with no cross-cultural validation, meant that results have been pooled without knowing whether results are generalisable across populations with cultural differences. For example, the OMAS was evaluated in the Turkish and Norwegian languages. Overall, this review has highlighted the need for further validation work in this particular area. However, it felt difficult to apply this strict criteria to research which hadn't followed their guidance in the first place. There is still merit in translating PROMs so they can be used in different cultural contexts, but this wasn't done in accordance to COSMIN and therefore doesn't not add much value in this review.

I think this review was a turning point in my understanding of the measurement properties of PROMs, something which this process has really helped me to get my head around. On the surface, these concepts can be confusing and I used to

constantly have to look definitions up. However, this work has helped to cement these definitions and concepts and their importance. Furthermore, this review was a turning point in my understanding of the gaps in the literature in this areas. This work helped me focus my research in the following work packages and highlighted the significant gaps in the evidence base regarding the suitability of OMAS for use in the context of clinical trials for adults with ankle fracture.

I was thrilled when the work of this project was published in the Journal of Patient Reported Outcomes and felt increasing confidence with the peer-review process on this second time publishing systematic review work. I also presented this work at the BOA congress, which was useful for getting valuable feedback on this work and the future planned work packages.

The work outlined in chapters 4 and 5 was my first experience of qualitative research and it was something that I found quite daunting. One thing I wasn't prepared for was the length of time and emotional energy that an interview took in total. This was not only in conducting the interview itself, but reflecting on it afterwards. Initially I scheduled more than one interview a week but quickly realised that this was not giving myself enough time to reflect on the previous interview and sufficiently prepare for the next one.

During the initial interviews I found that I would have a tendency to talk too much and was reluctant to allow for silence or pauses. Through continual reflection, I improved on this, giving participants time think and respond, without fear of silences. I also made a conscious effort to not react too much to what people told me. I found this difficult as I did not want to appear unsympathetic, but equally did not want to appear as though I was encouraging certain answers and reinforcing certain responses through positive feedback. Achieving this balance took a few interviews but this is something I improved considerably throughout the process.

I think if I had completed this research again, I would approach it slightly differently. I found it awkward to go from the first part of the interview, which was semi-structured and more open-ended questions, to the second part of the interview on OMAS, which felt more structured, with the use of more closed-ended type questions. It took me a few interviews to get used to this abrupt change in interview style and I wonder if there would have been a better way of approaching this. For example, if it were possible, it might have been better to conduct the OMAS section in a questionnaire. However, I was conscious of placing too much questionnaire burden

on individuals already participating in a trial. Another option would have been to complete them with an entirely different sample of individuals and treat them as two separate interviews and data collection points.

For cost and feasibility purposes, I transcribed my own interviews. At first this seemed like an arduous task but in retrospect I found this to be of great value in understanding and immersing myself in the data I had collected. It also helped me to critically analyse my interviewing technique and to make the improvements in my technique which I discussed previously. Throughout the analysis, I initially struggled to get my head around the concept of generating themes from codes. I took an alternative approach at the initial stage, organising codes using post-it notes and large paper to try and organise my thoughts. The process became considerably easier and more enjoyable the more I practiced.

I was really pleased when I got part of the work presented here accepted for publication in BMJ Open. This was my first time conducting and publishing qualitative research, so to have this published was a real highlight in this process. I presented this work to an audience at the British Society of Rehabilitation Medicine and Society for Research in Rehabilitation joint meeting which was a great opportunity to showcase my work and gain valuable feedback.

The completion of work outlined in chapter 6 was a significant learning curve, particularly with re-familiarising myself with the statistical software I used. I have previous experience of SPSS, however this was in my undergraduate dissertation which I did ten years ago now, so a refresher was required! Therefore I attended specific software training for this purpose. I enjoyed the challenge and managed to re-familiarised myself with the programme. I was really surprised at how long the data preparation took in comparison with the actual data analysis.

Finding a way to complete the random time point sampling as described in the methods section was tricky at first. I tried to use the functions on SPSS to do this initially, which did not work. I then found a way to use excel with the function of the random number generator to assign the random time points and then import these into SPSS. This worked, then I just needed to use the function on SPSS to pull the data for each time point for each patient. The more I worked around issues like this, the easier SPSS became to navigate. I also learnt to always keep back-up copies of data for when my experimenting with data didn't go to plan!

I also learnt the importance of version control of data files acutely in this process, whereby I had a small issue with finding the correct and up to date version I was working on. This was helped by guidance and advice from a colleague on how to correctly store and label data files to avoid issues like this. This advice made life much easier from then onwards!

I think if I was to complete this project again, I would look in more depth about the different surrogates of injury severity I could have used for the completion of discriminative validity. This part of the analysis didn't work for several reasons explained in the discussion section, but there could be better ways of looking at injury severity in the population. Ultimately, this is probably not an assessment discriminative validity, as we don't know whether these groups are appropriate measures of injury severity or not. The results presented here would indicate that they are not. Upon reflection and in retrospect, this analysis is more exploratory in nature than other assessments of discriminative validity would seem.

Unfortunately the conference I had this work accepted to for a poster presentation was cancelled due to the coronavirus pandemic. The work outlined in this chapter has been written up and submitted for peer review with Bone and Joint Research.

Throughout the process of completing this thesis and the work contained in it, I have evaluated the ways in which outcome is measured in adults with an ankle fracture and determine its appropriateness, within the context of clinical trials. The unique and sometimes challenging nature of the split role detailed in the declaration section has taught me a great deal. Firstly I've learnt about the wider research process of clinical trial management, HRA and REC applications, the functions of trial oversight committees, quality assurance processes, risk assessment and the multi-disciplinary team in delivering high quality research.

Secondly, I learnt a lot about completing my own research, formulating questions and making decisions on the best methods to answer those questions. I have developed my systematic review skills and also skills here as I completed two as part of this thesis. I developed my quantitative research skills having previously done some in my undergraduate degree. Furthermore, during this project I completed my first piece of qualitative research, something which I initially found daunting but quickly grew to enjoy the process. My academic writing has also improved throughout this project and it something which I now feel more confident with compared to the beginning of the process.

I have also learnt a great deal about myself through learning to cope with the many challenges and uncertainty that comes with the process of academic research. Throughout this I have developed my problem-solving skills, patience, self-compassion and resilience. I've also learnt the very important factor of having a work-life balance. I think I have always been pretty good at maintaining a balance in this respect, but the unique nature of doing a PhD sometimes became a little overwhelming and I sometimes felt like I had to make real effort to ensure I was giving myself time away from this work. Ultimately, by taking the time I needed for non-work activities, the time I spent doing work was all the more easier, productive and enjoyable.

This difficulty in maintaining a balance was heightened when I was writing up and the country went into lockdown due to the coronavirus pandemic. Whilst I was used to working in some degree of isolation due to the nature of the work, I realised how much value I got from regular visits to the office, hockey training and running club. Having these opportunities taken away was difficult to begin with. Whilst this did present some silver linings, as I was able to focus on my work entirely, knowing I wasn't missing out on any fun social activities, it did mean that the opportunities for finding non-work activities to relax became narrower. It took more effort than usual to ensure I got this balance right in light of these circumstances.

One aspect of this process which I have enjoyed more as I progressed is writing for articles for publication and abstracts for conferences. Initially, I would have described my feelings towards this process as apprehensive and probably something I actively tried to avoid. The majority of my worry came from fearing failure. Initially I took comments quite personally and found it difficult to detach emotionally. Once I initially submitted work, I became more confident and I soon realised the peer-review process was much less daunting than I initially feared it to be. I really started to enjoy the process of publication writing. Gaining valuable feedback from peer-reviewers became something I could use to strengthen my work, rather than something to dread.

Moving forwards, I would like to build on the skills I have learnt throughout this PhD and I would like to develop a clinical academic career. Applying for research funding and development of a clinical academic profile within the field of trauma and orthopaedic rehabilitation is my ultimate goal. I know this PhD has been instrumental developing some of the skills I require in order to achieve this goal.

8. References

1. Drake RL, Vogl AW, Mitchell AWM. Lower Limb. In: Drake RL, Vogl AW, Mitchell AWM, eds. Gray's Anatomy for Students. 41st ed. Philadelphia: Elsevier 2016:512-638.

2. Thompson J. Foot and Ankle. Netter's Concise Orthopaedic Anatomy. 2nd ed. Philadelphia: Saunders Elsevier 2010:338-83.

3. Brockett CL, Chapman GJ. Biomechanics of the ankle. *Orthopaedics and Trauma* 2016;30(3):232-38.

4. Griend RAV, Savoie FH, Hughes JL. Fractures of the Ankle. In: Jr CAR, Green DP, Bucholz RW, eds. Rockwood and Green's fractures in adults. Philadelphia: PA Lippincott 1991:1983-2030.

5. Duckworth T. Orthopaedics and Fractures. 3rd ed. Oxford: Blackwell Science Ltd. 1995.

6. Deivaraju C, Vlasak R, Sadasivan K. Staged treatment of pilon fractures. *J Orthop* 2015;12(Suppl 1):S1-6.

7. Hansen J. Lower Limb. In: Hansen J, ed. In: Netter's Clinical Anatomy E-Book. Philadelphia: Elsevier 2018:291-366.

8. Kawalec JS. Mechanical testing of foot and ankle implants. In: Friis E, ed. Mechanical Testing of Orthopaedic Implants. Duxford: Woodhead Publishing 2017:231-53.

9. Court-Brown C, McQueen MM, Tornetta P, et al. Ankle Fractures. In: Court-Brown C, McQueen MM, Tornetta P, et al., eds. Trauma. Philadelphia: Wolters Kluwer Health 2005:366-82.

10. Willmott H. Trauma. In: Willmott H, ed. Trauma and Orthopaedics at a Glance. Oxford: Wiley-Blackwell 2015:77-110.

11. Donaldson LJ, Reckless IP, Scholes S, et al. The epidemiology of fractures in England. *Journal of Epidemiology and Community Health* 2008;62(2):174.

12. Ferrari SL. Pathophysiology of Osteoporosis. In: Ferrari SL, Roux C, eds. Pocket Reference to Osteoporosis. Cham: Springer International Publishing 2019:1-10.

13. Holroyd C, Cooper C, Dennison E. Epidemiology of osteoporosis. *Best Practice & Research Clinical Endocrinology & Metabolism* 2008;22(5):671-85.

14. National Health Service. Broken leg: NHS; 2018. Available from: https://www.nhs.uk/conditions/broken-leg/ [accessed on 13/04/2020].

15. Pasco JA, Sanders KM, Hoekstra FM, et al. The human cost of fracture. *Osteoporosis International* 2005;16(12):2046-52.

16. Diamantopoulos AP, Rohde G, Johnsrud I, et al. The epidemiology of low- and high-energy distal radius fracture in middle-aged and elderly men and women in Southern Norway. *PloS one* 2012;7(8):e43367-e67.

17. Court-Brown C, McQueen MM, Tornetta P, et al. Epidemiology. In: Court-Brown C, McQueen MM, Tornetta P, et al., eds. Trauma. Philadelphia: Wolters Kluwer Health 2005:1-6.

18. Daghino W, Massè A, Marcolli D. Malleolar Fractures. In: Daghino W, Massè A, Marcolli D, eds. Foot and Ankle Trauma Injuries 2018:21-32.

19. Daghino W, Massè A, Marcolli D. Tibial Pilon Fractures. In: Daghino W, Massè A, Marcolli D, eds. In: Foot and Ankle Trauma Injuries 2018:1-19.

20. Statistics N. Hospital Admitted Patient Care Activity, 2017-18: Hospital admitted patient care activity, 2017-18: Diagnosis: NHS Digital; 2018 [updated 19 September 2018.] Available from: https://digital.nhs.uk/data-and-information/publications/statistical/hospital-admitted-patient-care-activity/2017-18 [accessed on 07/04/2020].

21. Court-Brown C, Caeser B. Epidemiology of adult fractures: A Review. *Injury* 2006;37

22. Kannus P, Palvanen M, Niemi S, et al. Increasing Number and Incidence of Low-Trauma Ankle Fractures in Ederly People: Finnish Statistics During 1970-2000 and Projections for the Future. *Bone* 2002;31(3)

23. Daly PJ, Fitzgerald RH, Melton LJ, et al. Epidemiology of ankle fractures in Rochester, Minnesota. *Acta Orthopaedica Scandinavica* 1987;58(5):539-44.

24. Bengnér U, Johnell O, Redlund-Johnell I. Epidemiology of ankle fracture 1950 and 1980: Increasing incidence in elderly women. *Acta Orthopaedica Scandinavica* 1986;57(1):35-37.

25. Thur CK, Edgren G, Jansson K-Å, et al. Epidemiology of adult ankle fractures in Sweden between 1987 and 2004. *Acta Orthopaedica* 2012;83(3):276-81.

26. Court-Brown CM, McBirnie J, Wilson G. Adult Ankle Fractures - an increasing problem? *Acta Orthop Scand* 1998;69(1):43-47.

27. Reagan JT, Giangarra CE, Jasko JJ. Foot and Ankle Fractures. In: Giangarra CE, Manske RC, eds. Clinical orthopaedic rehabilitation: a team approach. Philadelphia: Elsevier 2018:246-54.

28. Stiell I, Wells G, Laupacis A, et al. Multicentre trial to introduce the Ottawa ankle rules for use of radiography in acute ankle injuries. Multicentre Ankle Rule Study Group. *BMJ* 1995;311(7005):594-7.

29. Pena FA. Ankle Injuries. In: Swiontkowski MF, ed. Manual of Orthopaedics. Philadelphia, USA: Wolters Kluwer Health 2012:486-500.

30. National Institute for Health and Care Excellence. Fractures (non-complex): assessment and management. [NICE Guideline]; 2016 Available from:

https://www.nice.org.uk/guidance/ng38/evidence/full-guideline-2358460765 [accessed on 07/04/2020].

31. Radiopaedia. Ankle fracture - Weber B. Available from: https://radiopaedia.org/cases/8787 [accessed on 07/04/2020].

32. Gueorguiev B, Moriarty FT, Stoddart M, et al. Principles of fractures. In: Blom A, Warwick D, Whitehouse M, eds. In: Apley & Solomon's System of Orthopaedics and Trauma. 10th ed. Boca Raton: CRC Press 2017:711-54.

33. Lampridis V, Gougoulias N, Sakellariou A. Stability in ankle fractures: Diagnosis and treatment. *EFORT Open Reviews* 2018;3(5):294-303.

34. Lasanianos NG, Kanakaris NK. Ankle Fractures. In: Lasanianos NG, Kanakaris NK, Giannoudis PV, eds. In: Trauma and Orthopaedic Classifications 2015:367-70.

35. Radiopaedia. Weber fracture classification (illustration). Available from: https://radiopaedia.org/cases/9642 [accessed on 06/04/2020].

36. Osman K, Gabr A, Haddad FS. Bone Healing. In: General Orthopaedics and Basic Science 2019:111-19.

37. Thompson J. Basic Science. Netter's Concise Orthopaedic Anatomy. 2nd ed. Philadelphia: Saunders Elsevier 2010:1-24.

38. Charnley J. The Pott's Fracture. In: Charnley J, ed. The Closed Treatment of Common Fractures. 4 ed. Cambridge: Cambridge University Press 2003:250-69.

39. Chartered Society of Physiotherapy. Physiotherapy works: Rehabilitation 2012 Available from: https://www.csp.org.uk/publications/physiotherapy-works-rehabilitation [accessed on 13/04/2020].

40. Bowyer G. Injuries of the ankle and foot. In: Blom A, Warwick D, Whitehouse M, eds. In: Apley & Solomon's System of Orthopaedics and Trauma. 10th ed. Boca Raton: CRC Press 2017:937-63.

41. Willett K, Keene DJ, Mistry D, et al. Close Contact Casting vs Surgery for Initial Treatment of Unstable Ankle Fractures in Older Adults: A Randomized Clinical Trial. *JAMA* 2016;316(14):1455-63.

42. Keene DJ, Willett K. Implications of the Ankle Injury Management (AIM) trial. *The Bone & Joint Journal* 2019;101-B(12):1472-75.

43. SooHoo NF, Krenek L, Eagan MJ, et al. Complication Rates Following Open Reduction and Internal Fixation of Ankle Fractures. 2009;91(5):1042-49.

44. ISRCTN Registry. In younger adults with unstable ankle fractures treated with close contact casting, is ankle function not worse than those treated with surgical intervention? 2020 [updated 20/04/2020.] Available from: https://www.isrctn.com/ISRCTN67007305 [accessed on 21/04/2020].

45. Charnley J. The mechanics of conservative treatment. In: Charnley J, ed. The Closed Treatment of Common Fractures. 4 ed. Cambridge: Cambridge University Press 2003:43-59.

46. Lehtonen H, Jarvinen TL, Honkonen S, et al. Use of a cast compared with a functional ankle brace after operative treatment of an ankle fracture. A prospective, randomized study. *Journal of Bone & Joint Surgery American* 2003;85(2):205-11.

47. Egol KA, Dolan R, Koval KJ. Functional outcome of surgery for fractures of the ankle. *The Journal of Bone and Joint Surgery British volume* 2000;82-B(2):246-49.

48. Lin CC, Donkers NA, Refshauge KM, et al. Rehabilitation for ankle fractures in adults. *Cochrane Database of Systematic Reviews* 2012(11)

49. Jester R. Rehabilitation and the Orthopaedic and Musculoskeletal Trauma Patient. In: Clarke S, Santy-Tomlinson J, eds. Orthopaedic and trauma nursing. Hoboken, NJ: Wiley Blackwell 2014:59-66.

50. Shaffer MA, Okereke E, Esterhai JL, Jr., et al. Effects of Immobilization on Plantar-Flexion Torque, Fatigue Resistance, and Functional Ability Following an Ankle Fracture. *Physical Therapy* 2000;80(8):769-80.

51. Ulku TK, Kocaoglu B, Caglar MM, et al. Proprioception After Ankle Injury, Surgery, and Rehabilitation. In: Kaya D, Yosmaoglu B, Doral MN, eds. Proprioception in Orthopaedics, Sports Medicine and Rehabilitation. Cham: Springer International Publishing 2018:143-48.

52. Ploeger MM, Burger C, Wimmer MD. Acute Fractures (Lower Leg, Ankle, Hindfoot, Midfoot, Forefoot). In: Valderrabano V, Easley M, eds. Foot and Ankle Sports Orthopaedics. Cham: Springer International Publishing 2016:147-55.

53. Gill JR, Lostis E, Bleibleh SS, et al. Weight-bearing in ankle fractures: An audit of UK practice. *The Foot* 2019;39:28-36.

54. ISRCTN Registry. Weight-bearing in ankle fractures 2019 [updated 16/04/2020] Available from: https://www.isrctn.com/ISRCTN12883981 [accessed on 21/04/2020].

55. Smeeing DPJ, Houwert RM, Briet JP, et al. Weight-bearing or non-weight-bearing after surgical treatment of ankle fractures: a multicenter randomized controlled trial. *European Journal of Trauma and Emergency Surgery* 2020;46(1):121-30.

56. Ganesh SP, Pietrobon R, Cecílio WAC, et al. The Impact of Diabetes on Patient Outcomes After Ankle Fracture. *JBJS* 2005;87(8)

57. SooHoo NF, Eagan M, Krenek L, et al. Incidence and factors predicting pulmonary embolism and deep venous thrombosis following surgical treatment of ankle fractures. *Foot and Ankle Surgery* 2011;17(4):259-62.

58. Patil S, Gandhi J, Curzon I, et al. Incidence of deep-vein thrombosis in patients with fractures of the ankle treated in a plaster cast. *The Journal of Bone and Joint Surgery British volume* 2007;89-B(10):1340-43.

59. Uebbing CM, Walsh M, Miller JB, et al. Fracture blisters. *West J Emerg Med* 2011;12(1):131-33.

60. Weatherall J, Mroczek K, McLaurin T, et al. Post-Traumatic Ankle Arthritis. *Bulletin for the Hospital of Joint Diseases* 2013;71(1)

61. Eastwood D. Neuromuscular disorders. In: Blom A, Warwick D, Whitehouse M, eds. In: Apley & Solomon's System of Orthopaedics and Trauma. 10th ed. Boca Raton: CRC Press 2017:229-78.

62. Greenhalgh T. Why read papers at all? How to read a paper The bascis of evidence-based medicine. 4th ed. West Sussex: Wiley-Blackwell 2010:1-9.

63. Kabisch M, Ruckes C, Seibert-Grafe M, et al. Randomized controlled trials: part 17 of a series on evaluation of scientific publications. *Dtsch Arztebl Int* 2011;108(39):663-68.

64. Greenhalgh T. Searching the literature. How to read a paper The basics of evidence-based medicine 4th ed. West Sussex: Wiley-Blackwell 2010:15-28.

65. Murad MH, Asi N, Alsawas M, et al. New evidence pyramid. *Evidence Based Medicine* 2016;21(4):125.

66. Greenhalgh T. Getting your bearings - what is thie paper about? How to read a paper The bascis of evidence-based medicine. 4th ed. West Sussex: Wiley-Blackwell 2010:31-46.

67. Newell R, Burnard P. Randomised controlled trials. Research for Evidence-Based Practice in Healthcare. 2nd ed. Oxford, UK: John Wiley & Sons 2010:192-202.

68. Higgins JPT, Altman DG, Gøtzsche PC, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ* 2011;343:d5928.

69. Altman D. Clinical Trials. In: Altman D, ed. Practical statistics for medical research. London: Chapman & Hall 1991:440-74.

70. Armour M, Ee C, Steiner GZ. Randomized Controlled Trials. In: Liamputtong P, ed. Handbook of Research Methods in Health Social Sciences. Singapore: Springer 2019:645-62.

71. Schulz KF, Altman DG, Moher D, et al. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *BMC Medicine* 2010;8(1):18.

72. Kearney RS, McKeown R, Stevens S, et al. Cast versus functional brace in the rehabilitation of patients treated for an ankle fracture: protocol for the UK study of ankle injury rehabilitation (AIR) multicentre randomised trial. *BMJ Open* 2018;8(12):e027242.

73. World Health Organisation. Constitution: World Health Organisation. Available from: https://www.who.int/about/who-we-are/constitution [accessed on 16/04/2020].

74. Landgraf JM. Practical considerations in the measurement of HRQoL in child/adolescent clinical trials. In: Hays PFaR, ed. Assessing quality of life in clinical trials. 2nd ed. Oxford: Oxford University Press 2005:339-67.

75. Hathcoat JD, Meixner C, Nicholas MC. Ontology and Epistemology. In: Liamputtong P, ed. Handbook of Research Methods in Health Social Sciences. Singapore: Springer Singapore 2019:99-116.

76. Wilson IB, Cleary PD. Linking Clinical Variables With Health-Related Quality of Life: A Conceptual Model of Patient Outcomes. *JAMA* 1995;273(1):59-65.

77. Beard DJ, Knezevic K, Al-Ali S, et al. The use of outcome measures relating to the knee. *Orthopaedics and Trauma* 2010;24(4):309-16.

78. Röhrig B, du Prel J-B, Wachtlin D, et al. Sample size calculation in clinical trials: part 13 of a series on evaluation of scientific publications. *Dtsch Arztebl Int* 2010;107(31-32):552-56.

79. van Laarhoven CJ, Meeuwis JD, van der Werken C. Postoperative treatment of internally fixed ankle fractures: a prospective randomised study. *Journal of Bone & Joint Surgery British* 1996;78(3):395-9.

80. Bent NP, Wright CC, Rushton AB, et al. Selecting outcome measures in sports medicine: a guide for practitioners using the example of anterior cruciate ligament rehabilitation. *British Journal of Sports Medicine* 2009;43(13):1006.

81. Devlin NJ, Appleby J. Getting the most out of PROMs. Putting health outcomes at the heart of NHS decision-mkaing. London: The King's Fund; 2010 Available from: https://www.kingsfund.org.uk/sites/default/files/Getting-the-most-out-of-PROMs-Nancy-Devlin-John-Appleby-Kings-Fund-March-2010.pdf18/04/2020 [accessed on 16/04/2020].

82. Fitzpatrick R, Davey C, Buxton M, et al. Evaluating patient-based outcome measures for use in clinical trials. *Health Technol Assessment* 1998;2(14)

83. De Vet H, Terwee, C., Mokkink, L., & Knol, D. Concepts, theories and models, and types of measurements. In: Terwee CB, Knol DL, de Vet HCW, et al., eds. Measurement in Medicine: A Practical Guide. Cambridge: Cambridge University Press 2011:7-29.

84. Ware JE. The Status of Health Assessment 1994. *Annual Review of Public Health* 1995;16(1):327-54.

85. Department of Health. High Quality Care For All: NHS Next Stage Review Final Report 2008 Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/228836/7432.pdf [accessed on 18/04/2020].

86. Deaprtment of Health. Equity and excellence: liberating the NHS UK: The Stationery Office Limited; 2010 Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/213823/dh_117794.pdf [accessed on 04/04/2020].

87. Roos MKLSLEM. WOMAC Osteoarthritis Index: Reliability, validity, and responsiveness in patients with arthroscopically assessed osteoarthritis. *Scandinavian Journal of Rheumatology* 1999;28(4):210-15.

88. Ganestam A, Barfod K, Klit J, et al. Validity and Reliability of the Achilles Tendon Total Rupture Score. *The Journal of Foot and Ankle Surgery* 2013;52(6):736-39.

89. Gummesson C, Ward MM, Atroshi I. The shortened disabilities of the arm, shoulder and hand questionnaire (Quick DASH): validity and reliability based on responses within the full-length DASH. *BMC Musculoskeletal Disorders* 2006;7(1):44. doi: 10.1186/1471-2474-7-44

90. Jenkinson C, Layte R, Jenkinson D, et al. A shorter form health survey: can the SF-12 replicate results from the SF-36 in longitudinal studies? *Journal of Public Health* 1997;19(2):179-86.

91. Devlin N, Shah K, Feng Y, et al. Valuing Health -Related Quality of Life: An EQ-5D-5L Value Set for England. *Health Economics* 2018;27:7-22

92. Weldring T, Smith SMS. Patient-Reported Outcomes (PROs) and Patient-Reported Outcome Measures (PROMs). *Health Serv Insights* 2013;6:61-68.

93. Mokkink LB, Terwee CB, Patrick Sr LJ, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *Journal of Clinical Epidemiology* 2010;63(7):737-45.

94. de Vet HCW TC, Mokkink LB, Knol DL. Validity. In: Terwee CB, Knol DL, de Vet HCW, et al., eds. Measurement in Medicine: A Practical Guide. Cambridge: Cambridge University Press 2011:150-201.

95. de Vet HCW TC, Mokkink LB, Knol DL. Reliability. In: Terwee CB, Knol DL, de Vet HCW, et al., eds. Measurement in Medicine: A Practical Guide. Cambridge: Cambridge University Press 2011:96-149.

96. de Vet HCW TC, Mokkink LB, Knol DL. Responsiveness. In: Terwee CB, Knol DL, de Vet HCW, et al., eds. Measurement in Medicine: A Practical Guide. Cambridge: Cambridge University Press 2011:202-26.

97. de Vet HCW TC, Mokkink LB, Knol DL. Interpretability. In: Terwee CB, Knol DL, de Vet HCW, et al., eds. Measurement in Medicine: A Practical Guide. Cambridge: Cambridge University Press 2011:227-74.

98. Williamson K. Research in Constructivist Frameworks using Ethnographic Techniques: Johns Hopkins University Press; 2006 Available from: https://muse.jhu.edu/article/202362/pdf [accessed on 05/04/2020].

99. Goldkuhl G. Pragmatism vs interpretivism in qualitative information systems research. *European Journal of Information Systems* 2012;21(2):135-46.

100. Bryman A. Social research strategies: quantitative research and qualitative research. In: Social Research Methods. 5th ed. Oxford: Oxford University Press 2016:16-38.

101. Bryman A. The nature of quantitative research. In: Social Research Methods. 5th ed. Oxford: Oxford University Press 2016:148-69.

102. Khanna P. Positivism and Realism. In: Liamputtong P, ed. Handbook of Research Methods in Health Social Sciences. Singapore: Springer Singapore 2019:151-68.

103. Johnson RB, Onwuegbuzie AJ. Mixed Methods Research: A Research Paradigm Whose Time Has Come. *Educational Researcher* 2004;33(7):14-26.

104. Peter RG, Artur P, Peter H. A Pragmatic Research Philosophy for Sport and Exercise Psychology. *The Sport Psychologist* 2005;19(1):18-31.

105. Morgan DL. Paradigms Lost and Pragmatism Regained: Methodological Implications of Combining Qualitative and Quantitative Methods. *Journal of Mixed Methods Research* 2007;1(1):48-76.

106. Fielding NG. Triangulation and Mixed Methods Designs: Data Integration With New Research Technologies. *Journal of Mixed Methods Research* 2012;6(2):124-36.

107. Fetters MD, Curry LA, Creswell JW. Achieving Integration in Mixed Methods Designs - Principles and Practices. *Health Services Research* 2013;48:2134-56.

108. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Quality of Life Research* 2010;19(4):539-49.

109. Prinsen CAC, Mokkink LB, Bouter LM, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Quality of Life Research* 2018;27:1147-57.

110. Raftery J, Young A, Stanton L, et al. Clinical trial metadata: defining and extracting metadata on the design, conduct, results and costs of 125 randomised clinical trials funded by the National Institute for Health Research Health Technology Assessment programme. *Health Technol Assess* 2015;19(11)

111. National Institute for Health Research (NIHR). About us: NIHR; 2020 Available from: https://www.nihr.ac.uk/about-us/ [accessed on 03/05/2020].

112. Kirkham JJ, Gargon E, Clarke M, et al. Can a core outcome set improve the quality of systematic reviews? – a survey of the Co-ordinating Editors of Cochrane review groups. *Trials* 2013;14(1):21.

113. Clarke M, Williamson P. Core outcome sets and trial registries. *Trials* 2015;16(1):216.

114. Riddle DL, Stratford PW, Bowman DH. Findings of extensive variation in the types of outcome measures used in hip and knee replacement clinical trials: A systematic review. *Arthritis Care & Research* 2008;59(6):876-83.

115. Wylde V, Bruce J, Beswick A, et al. Assessment of Chronic Postsurgical Pain After Knee Replacement: A Systematic Review. *Arthritis Care & Research* 2013;65(11):1795-803.

116. Naal FD, Impellizzeri FM, Rippstein PF. Which Are the Most Frequently Used Outcome Instruments in Studies on Total Ankle Arthroplasty? *Clinical Orthopaedics and Related Research* 2010;468(3):815-26.

117. Whittle JH, Peters SE, Manzanero S, et al. A systematic review of patient-reported outcome measures used in shoulder instability research. *Journal of Shoulder and Elbow Surgery* 2020;29(2):381-91.

118. Froud R, Patel S, Rajendran D, et al. A Systematic Review of Outcome Measures Use, Analytical Approaches, Reporting Methods, and Publication Volume by Year in Low Back Pain Trials Published between 1980 and 2012: Respice, adspice, et prospice. *PloS one* 2016;11(10):e0164573-e73.

119. Hunt KJ, Hurwit D. Use of Patient-Reported Outcome Measures in Foot and Ankle Research. *The Journal of Bone and Joint Surgery* 2013;95(e118)

120. Moher D, Liberati A, Tetzlaff J, et al. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLOS Medicine* 2009

121. Cochrane. Cochrane Handbook for Systematic Reviews of Interventions version 6.0 (updated July 2019) 2019 Available from: www.training.cochrane.org/handbook [accessed on 05/04/2019].

122. Blackburn EW, Aronsson DD, Rubright JH, et al. Ankle Fractures in Children. *The Journal of Bone and Joint Surgery* 2012;94-A(13)

123. Olerud C, Molander H. A Scoring Scale for Symptom Evaluation After Ankle Fracture. *Archives of Orthopaedic and Traumatic Surgery* 1984;103

124. Dawson J, Boller I, Doll H, et al. The MOXFQ patient-reported questionnaire: Assessment of data quality, reliability and validity in relation to foot and ankle surgery. *Foot* 2011;21(2):92-102.

125. Johanson NA, Liang MH, Daltroy L, et al. American Academy of Orthopaedic Surgeons Lower Limb Outcomes Assessment Instruments: Reliability, Validity, and Sensitivity to Change. *Journal of Bone and Joint Surgery - Series A* 2004;86(5):902-09.

126. McPhail SM, Williams CM, Schuetz M, et al. Development and validation of the ankle fracture outcome of rehabilitation measure (A-FORM). *Journal of Orthopaedic & Sports Physical Therapy* 2014;44(7):488-99, B1-2.

127. Karlsson J, Peterson L. Evaluation of ankle joint function: the use of a scoring scale. *The Foot* 1991;1:15-19.

128. Martin RL, Irrgang JJ, Burdett R, G. , et al. Evidence of Validity for the Foot and Ankle Ability Measure (FAAM). *Foot and Ankle International* 2005;26(11)

129. Sierevelt IN, Zwiers R, Schats W, et al. Measurement properties of the most commonly used Foot-and Ankle-Specific Questionnaires: the FFI, FAOS and FAAM. A systematic review. *Knee Surg Sports Traumatol Arthrosc* 2018;26:2059-73.

130. Kitaoka HB, Alexander IJ, Adelaar RS, et al. Clinical Rating Systems for the ankle-Hindfoot, Midfoot, Hallux, and Lesser Toes. *Foot and Ankle International* 1994;15(7)

131. Mazur J, Schwartz E, Simon S. Ankle Arthrodesis. Long-Term Follow Up with Gait Analysis. *The Journal of Bone and Joint Surgery* 1979;61-A(7)

132. Kaikkonen A, Kannus P, Jarvinen M. A Performance Test Protocol and Scoring Scale for the Evaluation fo Ankle Injuries. *The American Journal of Sports Medicine* 1994;22(4)

133. Niki H, Aoki H, Inokuchi S, et al. Development and reliability of a standard rating system for outcome measurement of foot and ankle disorders I: development of standard rating system. *Journal of Orthopaedic Science* 2005;10(5):457.

134. Hughes JL, Weber H, Willenegger H, et al. Evaluation of Ankle Fractures: non-operative and operative Treatment. *Clinical Orthopaedics & Related Research* 1978;138

135. Baird RA, Jackson ST. Fractures of the Distal Part of the Fibula with Associated Disruption of the Deltoid Ligament. *The Journal of Bone and Joint Surgery* 1987;69-A(9)

136. Phillips WA, Schwartz HS, Keller CS. A prospective, randomized study of the management of severe ankle fractures. *Journal of Bone and Joint Surgery - Series A* 1985;67(1):67-78.

137. Linde F, Hvass I, Jurgensen U, et al. Compression bandage in the treatment of anke sprains. A comparitive prospective study. *Scandinavian Journal of Rehabilitation Medicine* 1984;16(4)

138. Begg C, Cho M, Eastwood S, et al. Improving the Quality of Reporting of Randomized Controlled Trials: The CONSORT Statement. *JAMA* 1996;276(8):637-39.

139. Moher D, Schulz KF, Altman D, et al. The CONSORT Statement: Revised Recommendations for Improving the Quality of Reports of Parallel-Group Randomized Trials. *JAMA* 2001;285(15):1987-91.

140. Ng R, Broughton N, Williams C. Measuring Recovery After Ankle Fractures: A Systematic Review of the Psychometric Properties of Scoring Systems. *Journal of Foot & Ankle Surgery* 2018;57(1):149-54.

141. Kearney R, Achten J, Lamb S, et al. A Systematic review of patient reported outcome measures to assess Achilles tendon rupture management: What's being used and should we be using it? *British Journal of Sports Medicine* 2011;4(6):1102-09.

142. Bucholz RW, Henry S, Henley MB. Fixation with bioabsorbable screws for the treatment of fractures of the ankle. *Journal of Bone & Joint Surgery American* 1994;76(3):319-24.

143. Park YH, Ahn JH, Choi GW, et al. Comparison of Clamp Reduction and Manual Reduction of Syndesmosis in Rotational Ankle Fractures: A Prospective Randomized Trial. *Journal of Foot and Ankle Surgery* 2018;57(1):19-22.

144. Günther KP, Sun Y. Reliability of radiographic assessment in hip and knee osteoarthritis. *Osteoarthritis and Cartilage* 1999;7(2):239-46.

145. Costa M. Do Patients Care What Their X-rays Look Like? *JAMA Network Open* 2020;3(1):e1919570-e70.

146. Brauer CA, Rosen AB, Olchanski NV. Cost-Utility Analyses in Orthopaedic Surgery. *Journal of Bone and Joint Surgery* 2005;87-A(6)

147. Hijji FY, Schneider AD, Pyper M, et al. The Popularity of Outcome Measures Used in the Foot and Ankle Literature. *Foot & Ankle Specialist* 2019;13(1):58-68.

148. Guyton GP. Theoretical Limitations of the AOFAS Scoring Systems: An Analysis Using Monte Carlo Modeling. *Foot & Ankle International* 2001;22(10):779-87.

149. Pinsker E, Daniels TR. AOFAS Position Statement Regarding the Future of the AOFAS Clinical Rating Systems. *Foot & Ankle International* 2011;32(9):841-42.

150. Jia Y, Huang H, Gagnier JJ. A systematic review of measurement properties of patient-reported outcome measures for use in patients with foot or ankle diseases. *Quality of Life Research* 2017;26(8):1969-2010.

151. Button G, Pinney S. A Meta-analysis of Outcome Rating Scales in Foot and Ankle Surgery: Is There a Valid, Reliable, and Responsive System? *Foot & Ankle International* 2004;25(8):521-25.

152. Gagnier JJ. Patient reported outcomes in orthopaedics. *Journal of Orthopaedic Research* 2017;35(10):2098-108.

153. Budiman-Mak E, Conrad KJ, Roach KE. The foot function index: A measure of foot pain and disability. *The Journal of Clinical Epidemiology* 1991;44(6):561-70.

154. Braunstein M, Baumbach SF, Regauer M, et al. The value of arthroscopy in the treatment of complex ankle fractures -A protocol of a randomised controlled trial. *BMC Musculoskeletal Disorders* 2016;17 (1)(210)

155. Roos EM, Brandsson S, Karlsson J. Validation of the Foot and Ankle Outcome Score for Ankle Ligament Reconstruction. *Foot and Ankle International* 2001;22(10):788-94.

156. Fayers P, Machin D. Principles of measurement scales. In: Fayers P, Hays RD, eds. Quality of Life The assessment, analysis and reporting of patient-reported outcomes. 3rd ed. Oxford: Wiley-Blackwell 2016:35-56.

157. Roos EM, Roos HP, Lohmander LS, et al. Knee Injury and Osteoarthritis Outcome Score (KOOS)—Development of a Self-Administered Outcome Measure. *Journal of Orthopaedic & Sports Physical Therapy* 1998;28(2):88-96.

158. McKeown R, Kearney R, Ellard D. A systematic review of the quality of ankle specific patient reported outcome measures in the assessment of patients recovering from ankle fracture 2018 Available from: https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=103112 [Accessed on 06/04/2020]

159. Mokkink LB, Prinsen CA, Patrick DL, et al. COSMIN methodology for systematic reviews of patient reported outcome measures (PROMs) - user manual 2018 Available from: https://cosmin.nl/wp-content/uploads/COSMIN_manual_syst-review-PROMs_V1.0.pdf [accessed on 03/07/2018].

160. Conijn AP, Jens S, Terwee CB, et al. Assessing the Quality of Available Patient Reported Outcome Measures for Intermittent Claudication: A Systematic Review Using the COSMIN Checklist. *European Journal of Vascular and Endovascular Surgery* 2015;49(3):316-34.

161. McKenna SP, McEwen J, Hunt SM, et al. Changes in the perceived health of patientsrecovering from fractures. *Public Health* 1984;98(2):97-102.

162. Abram SGF, Middleton R, Beard DJ, et al. Patient-reported outcome measures for patients with meniscal tears: a systematic review of measurement properties and evaluation with the COSMIN checklist. *BMJ Open* 2017;7(10):e017247.

163. Terwee CB, Jansma EP, Riphagen II, et al. Development of a Methodological PubMed Search Filter for Finding Studies on Measurement Properties of Measurement Instruments. *Quality of Life Research* 2009;18(8):1115-23.

164. Mokkink LB, de Vet HCW, Prinsen CA, et al. COSMIN Risk of Bias checklist for systematic reviews of Patient Reported Outcome Measures. *Quality of life Research* 2017;27:1171-79.

165. Wu SH, Liang HW, Hou WH. Reliability and validity of the Taiwan Chinese version of the Foot Function Index. *Journal of the Formosan Medical Association* 2008;107(2):111-8.

166. Terwee CB, Prinsen CAC, Chiarotto A, et al. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. *Quality of Life Research* 2018;27(5):1159-70.

167. Abma I, L. , Rovers M, Van der Wees P. Appraising convergent validity of patient-reported outcome measures in systematic reviews: constructing and interpreting outcomes. *BMC Research Notes* 2016;9(226)

168. Buker N, Savkin R, Gokalp O, et al. Validity and Reliability of Turkish Version of Olerud-Molander Ankle Score in Patients With Malleolar Fracture. *Journal of Foot and Ankle Surgery* 2017;56(6):1209-12.

169. Garratt AM, Naumann MG, Sigurdsen U, et al. Evaluation of three patient reported outcome measures following operative fixation of closed ankle fractures. *BMC Musculoskeletal Disorders* 2018;19(1)(134)

170. Turhan E, Demirel M, Daylak A, et al. Translation, cross-cultural adaptation, reliability and validity of the Turkish version of the Olerud-Molander Ankle Score (OMAS). *Acta Orthopaedica et Traumatologica Turcica* 2017;51(1):60-64.

171. Zelle BA, Francisco BS, Bossmann JP, et al. Spanish Translation, Cross-Cultural Adaptation, and Validation of the American Academy of Orthopaedic Surgeons Foot and Ankle Outcomes Questionnaire in Mexican-Americans with Traumatic Foot and Ankle Injuries. *Journal of Orthopaedic Trauma* 2017;31(5):e158-e62.

172. Keene DJ, Costa ML, Tutton E, et al. Progressive functional exercise versus best practice advice for adults aged 50 years or over after ankle fracture: protocol for a pilot randomised controlled trial in the UK - the Ankle Fracture Treatment: Enhancing Rehabilitation (AFTER) study. *BMJ Open* 2019;9(11):e030877.

173. McPhail SM, Dunstan J, Canning J, et al. Life impact of ankle fractures: qualitative analysis of patient and clinician experiences. *BMC Musculoskeletal Disorders* 2012;13(224)

174. Terwee C, Bot S, Boer Md, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology* 2007;60

175. Lim CR, Harris K, Dawson J, et al. Floor and Ceiling effects in the OHS: an analysis of the NHS PROMs data set. *BMJ Open* 2015;5(7)

176. McHorney CA, Tarlov AR. Individual-patient monitoring in clinical practice: are available health status surveys adequate? *Quality of Life Research* 1995;4(4)

177. Wamper KE, Sierevelt IN, Poolman RW, et al. The Harris hip score: Do ceiling effects limit its usefulness in orthopaedics? *Acta Orthopaedica* 2010;81(6)

178. Laflamme M, Belzile EL, Bédard L, et al. A Prospective Randomized Multicenter Trial Comparing Clinical Outcomes of Patients Treated Surgically With a Static or Dynamic Implant for Acute Ankle Syndesmosis Rupture. *Journal of Orthopaedic Trauma* 2015;29(5)

179. Siddique A, Prasad CVR, O'Connor D. Early Active Mobilization Versus Cast Immobilization in Operatively Treated Ankle Fractures. *European Journal of Trauma* 2005;31(4):398-400.

180. White TO, Bugler KE, Appleton P, et al. A prospective randomised controlled trial of the fibular nail versus standard open reduction and internal fixation for fixation of ankle fractures in elderly patients. *The Bone & Joint Journal* 2016;98-B(9):1248-52.

181. Arinzon Z, Shabat S, Peisakh A, et al. Gender differences influence the outcome of geriatric rehabilitation following hip fracture. *Archives of Gerontology and Geriatrics* 2010;50(1):86-91.

182. Hall JA, Dornan MC. Patient sociodemographic characteristics as predictors of satisfaction with medical care: A meta-analysis. *Social Science & Medicine* 1990;30(7):811-18.

183. Baumhauer JF, McIntosh S, Rechtine G. Age and Sex Differences Between Patient and Physician-Derived Outcome Measures in the Foot and Ankle. *JBJS* 2013;95(3)

184. Court-Brown CM, Duckworth AD, Clement ND, et al. Fractures in older adults. A view of the future? *Injury* 2018;49(12):2161-66.

185. Inglis M, McClelland B, Sutherland LM, et al. Synthetic versus plaster of Paris casts in the treatment of fractures of the forearm in children. *The Bone & Joint Journal* 2013;95-B(9):1285-89.

186. Brod M, Tesler LE, Christensen TL. Qualitative research and content validity: developing best practices based on science and experience. *Quality of Life Research* 2009;18(9):1263.

187. Rothman M, Burke L, Erickson P, et al. Use of Existing Patient-Reported Outcome (PRO) Instruments and Their Modification: The ISPOR Good Research Practices for Evaluating and Documenting Content Validity for the Use of Existing Instruments and Their Modification PRO Task Force Report. *Value in Health* 2009;12(8):1075-83.

188. Braun V, Clarke V. Some very important starting information. In: Braun V, Clarke V, eds. Successful Qualitative Research A Practical Guide for Beginners. London: Sage Publications 2013:3-18.

189. Braun V, Clarke V. Planning and designing qualitative research. In: Braun V, Clarke V, eds. Successful Qualitative Research A Practical Guide for Beginners. London: Sage Publications 2013:42-73.

190. Sim J, Wright C. Designing an Exploratory Study. Research in Health Care. Cheltenham: Stanley Thornes 2000:45-67.

191. Nathan S, Newman C, Lancaster K. Qualitative Interviewing. In: Liamputtong P, ed. Handbook of Research Methods in Health Social Sciences. Singapore: Springer Singapore 2019:391-410.

192. Hewitt J. Ethical Components of Researcher—Researched Relationships in Qualitative Interviewing. *Qualitative Health Research* 2007;17(8):1149-59.

193. O'Brien BC, Harris IB, Beckman TJ, et al. Standards for Reporting Qualitative Research: A Synthesis of Recommendations. 2014;89(9):1245-51.

194. Griffiths F. Preparing to design research for health care practice. In: Research Methods for Health Care Practice. London: Sage Publications Ltd 2009:68-97.

195. Bryman A. Focus groups. In: Social Research Methods. 5th ed. Oxford: Oxford University Press 2016:500-24.

196. Braun V, Clarke V. Interactive data collection 1: interviews. In: Braun V, Clarke V, eds. Successful Qualitative Research A Practical Guide for Beginners. London: Sage Publications 2013:77-105.

197. Bryman A. Interviewing in qualitative research. In: Social Research Methods. 5th ed. Oxford: Oxford University Press 2016:466-98.

198. Rees S, Tutton E, Achten J, et al. Patient experience of long-term recovery after open fracture of the lower limb: a qualitative study using interviews in a community setting. *BMJ Open* 2019;9(10):e031261.

199. Griffiths F, Mason V, Boardman F, et al. Evaluating recovery following hip fracture: a qualitative interview study of what is important to patients. *BMJ Open* 2015;5(1):e005406.

200. Hallberg I, Ek A-C, Toss G, et al. A striving for independence: a qualitative study of women living with vertebral fracture. *BMC Nursing* 2010;9(1):7.

201. Richards H, Emslie C. The 'doctor' or the 'girl from the University'? Considering the influence of professional roles on qualitative interviewing. *Family Practice* 2000;17(1):71-75.

202. Gagnon M, Jacob JD, McCabe J. Locating the qualitative interview: reflecting on space and place in nursing research. *Journal of Research in Nursing* 2014;20(3):203-15.

203. Braun V, Clarke V. Reflecting on reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health* 2019;11(4):589-97.

204. Berger R. Now I see it, now I don't: researcher's position and reflexivity in qualitative research. *Qualitative Research* 2013;15(2):219-34.

205. Finlay L. "Outing" the Researcher: The Provenance, Process, and Practice of Reflexivity. *Qualitative Health Research* 2002;12(4):531-45.

206. Braun V, Clarke V. To saturate or not to saturate? Questioning data saturation as a useful concept for thematic analysis and sample-size rationales. *Qualitative Research in Sport, Exercise and Health* 2019:1-16.

207. Braun V, Clarke V, Hayfield N, et al. Thematic Analysis. In: Liamputtong P, ed. Handbook of Research Methods in Health Social Sciences. Singapore: Springer Singapore 2019:843-60.

208. Braun V, Clarke V. Using thematic analysis in psychology. *Qualitative Research in Psychology* 2006;3(2):77-101.

209. Nowell LS, Norris JM, White DE, et al. Thematic Analysis: Striving to Meet the Trustworthiness Criteria. *International Journal of Qualitative Methods* 2017;16(1)

210. Braun V, Clarke V. Thematic analysis. In: Cooper HM ed. APA handbook of research methods in psychology, Vol 2: Research designs: Quantitative, qualitative, neuropsychological, and biological. Washington, DC, US: American Psychological Association 2012:57-71.

211. McKeown R, Kearney RS, Liew ZH, et al. Patient experiences of an ankle fracture and the most important factors in their recovery: a qualitative interview study. *BMJ Open* 2020;10(2):e033539.

212. Ponsford J, Hill B, Karamitsios M, et al. Factors Influencing Outcome After Orthopedic Trauma. 2008;64(4):1001-09.

213. Bhandari M, Busse JW, Hanson BP, et al. Psychological distress and quality of life after orthopedic trauma: an observational study. *Canadian journal of surgery Journal canadien de chirurgie* 2008;51(1):15-22.

214. Gong HS, Lee JO, Huh JK, et al. Comparison of depressive symptoms during the early recovery period in patients with a distal radius fracture treated by volar plating and cast immobilisation. *Injury* 2011;42(11):1266-70.

215. Crichlow RJ, Andres PL, Morrison SM, et al. Depression in Orthopaedic Trauma Patients: Prevalence and Severity. *JBJS* 2006;88(9)

216. Keene DJ, Vadher K, Willett K, et al. Predicting patient-reported and objectively measured functional outcome 6 months after ankle fracture in people aged 60 years or over in the UK: prognostic model development and internal validation. *BMJ Open* 2019;9(7):e029813.

217. Steven JL, Nina B, Lars S, et al. Pain-related fear, catastrophizing and pain in the recovery from a fracture. *Scandinavian Journal of Pain* 2010;1(1):38-42.

218. Clay FJ, Newstead SV, McClure RJ. A systematic review of early prognostic factors for return to work following acute orthopaedic trauma. *Injury* 2010;41(8):787-803.

219. van der Sluis CK, Eisma WH, Groothoff JW, et al. Long-term physical, psychological and social consequences of a fracture of the ankle. *Injury* 1998;29(4):277-80.

220. Pinsker EB, Sale JEM, Gignac MAM, et al. “I don't have to think about watching the ground”: a qualitative study exploring the concept of vigilance as an important outcome for ankle reconstruction. *Arthritis Care & Research* 2019

221. Wylde V, Learmonth ID, Cavendish VJ. The Oxford hip score: the patient's perspective. *Health and Quality of Life Outcomes* 2005;3(1):66.

222. Todres L, Galvin K. Pursuing Both Breadth and Depth in Qualitative Research: Illustrated by a Study of the Experience of Intimate Caring for a Loved One with Alzheimer's Disease. *International Journal of Qualitative Methods* 2005;4(2):20-31.

223. Broom A, Hand K, Tovey P. The role of gender, environment and Individual biography in shaping qualitative interview data. *International Journal of Social Research Methodology* 2009;12(1):51-65.

224. COSMIN Group. Measurement Properties of Outcome Measurement Intruments Available from: https://www.cosmin.nl/wp-content/uploads/COSMIN_taxonomy.pdf [accessed on 15/04/2020].

225. Reeve BB, Fayers P. Applying item response theory modelling for evaluating questionnaire item and scale properties. In: Fayers P, Hays R, eds. Assessing quality of life in clinical trials. 2nd ed. Oxford: Oxford University Press 2005:55-73.

226. DeVellis RF. Classical Test Theory. *Medical Care* 2006;44(11):S50-S59.

227. Hambleton RK, Jones RW. An NCME Instructional Module on Comparison of Classical Test Theory and Item Response Theory and Their Applications to Test Development. *Educational Measurement: Issues and Practice* 1993;12(3):38-47.

228. Morales RA. Evaluation of Mathematics Achievement Test: A Comparison between CTT and IRT. *The International Journal of Educational and Psychological Assessment* 2009;1(1):19-26.

229. Abedlazeez N. Exploring DIF: Comparison of CTT and IRT methods. *OIDA International Journal of Sustainable Development* 2010;1(7):11-46.

230. Prieto L, Alonso J, Lamarca R. Classical test theory versus Rasch analysis for quality of life questionnaire reduction. *Health and Quality of Life Outcomes* 2003;1(1):27.

231. Hays RD, Morales LS, Reise SP. Item response theory and health outcomes measurement in the 21st century. *Medical care* 2000;38(9 Suppl):II28-II42.

232. Sébille V, Hardouin J-B, Le Néel T, et al. Methodological issues regarding power of classical test theory (CTT) and item response theory (IRT)-based approaches for the comparison of patient-reported outcomes in two groups of patients - a simulation study. *BMC Medical Research Methodology* 2010;10(1):24.

233. Anthoine E, Moret L, Regnault A, et al. Sample size used to validate a scale: a review of publications on newly-developed patient reported outcomes measures. *Health and Quality of Life Outcomes* 2014;12(1):2.

234. Nunnally JC. Psychometric Theory. New York: McGraw-Hill 1967.

235. Osborne JW, Costello AB. Sample size and subject to item ratio in principal componnts analysis. *Practical Assessment, Research, and Evaluation* 2004;9

236. Allison P. Missing Data. In: Millsap RE, Maydeu-Olivares A eds. The SAGE Handbook of Quantitative Methods in Psychology. London: SAGE Publications Ltd, 2009: 73-90.

237. Sterne JAC, White IR, Carlin JB, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ* 2009;338:b2393

238. Chavance M. Handling Missing Items in Quality of Life Studies. *Communications in Statistics - Theory and Methods* 2004;33(6):1371-83.

239. Excellence NIfHaC. Position statement on use of the EQ-5D-5L value set for England (updated October 2019) 2019 Available from: https://www.nice.org.uk/about/what-we-do/our-programmes/nice-guidance/technology-appraisal-guidance/eq-5d-5l [accessed on 06/04/2020].

240. van Hout B, Janssen MF, Feng Y-S, et al. Interim Scoring for the EQ-5D-5L: Mapping the EQ-5D-5L to EQ-5D-3L Value Sets. *Value in Health* 2012;15(5):708-15.

241. Parsons H, Bruce J, Achten J, et al. Measurement properties of the Disability Rating Index in patients undergoing hip replacement. *Rheumatology* 2014;54(1):64-71.

242. Denham RA. Internal fixation for unstable ankle fractures. *The Journal of Bone and Joint Surgery British volume* 1964;46-B(2):206-11.

243. Joy G, Patzakis MJ, Harvey JPJ. Precise Evaluation of the Reduction of Severe Ankle Fractures: Technique and Correlation with End Results. *JBJS* 1974;56(5)

244. Hancock MJ, Herbert RD, Stewart M. Prediction of Outcome After Ankle Fracture. *Journal of Orthopaedic & Sports Physical Therapy* 2005;35(12):786-92.

245. Porter DA, May BD, Berney T. Functional Outcome after Operative Treatment for Ankle Fractures in Young Athletes: A Retrospective Case Series. *Foot & Ankle International* 2008;29(9):887-94.

246. Kennedy JG, Johnson SM, Collins AL, et al. An evaluation of the Weber classification of ankle fractures. *Injury* 1998;29(8):577-80.

247. Egol KA, Pahk B, Walsh M, et al. Outcome after Unstable Ankle Fracture: Effect of Syndesmotic Stabilization. 2010;24(1):7-11.

248. Broos PLO, Bisschop APG. Operative treatment of ankle fractures in adults: correlation between types of fracture and final results. *Injury* 1991;22(5):403-06.

249. Segal G, Elbaz A, Parsi A, et al. Clinical outcomes following ankle fracture: a cross-sectional observational study. *Journal of Foot and Ankle Research* 2014;7(1):50.

250. Hong CC, Nashi N, Prosad Roy S, et al. Impact of trimalleolar ankle fractures: How do patients fare post-operatively? *Foot and Ankle Surgery* 2014;20(1):48-51.

251. Jaskulka R, Ittner G, Schedl R. Fractures of the Posterior Tibial Margin; Their Role in The Prognosis of Malleolar Fractures. *The Journal of Trauma* 1989;29(11)

252. Tejwani NC, Pahk B, Egol KA. Effect of posterior malleolus fracture on outcome after unstable ankle fracture. *Journal of Trauma - Injury, Infection and Critical Care* 2010;69(3):666-69.

253. Langenhuijsen JF, Heetveld MJ, Ultee JM, et al. Results of Ankle Fractures with Involvement of the Posterior Tibial Margin. 2002;53(1):55-60.

254. Floyd F, Widaman K. Factor Analysis in the Development and Refinement of Clinical Assessment Instruments. *Psychol Assess* 1995;7(3):286-99.

255. Kaiser HF. The Application of Electronic Computers to Factor Analysis. *Educational and Psychological Measurement* 1960;20(1):141-51.

256. Stallman HM, Hurst CP. The Factor Structure of the Frost Multidimensional Perfectionism Scale in University Students. *Australian Psychologist* 2011;46(4):229-36.

257. Wood DL, Sawicki GS, Miller MD, et al. The Transition Readiness Assessment Questionnaire (TRAQ): Its Factor Structure, Reliability, and Validity. *Academic Pediatrics* 2014;14(4):415-22.

258. John MT, Reissmann DR, Feuerstahler L, et al. Exploratory factor analysis of the Oral Health Impact Profile. *J Oral Rehabil* 2014;41(9):635-43.

259. Bland JM, Altman DG. Statistics notes: Cronbach's alpha. *BMJ* 1997;314(7080):572.

260. Lin C, Moseley A, Refshauge K, et al. The lower extremity functional scale has good clinimetric properties in people with ankle fracture. *Physical Therapy* 2009;89(6):580-8.

261. Harris K, Lim CR, Dawson J, et al. The Oxford knee score and its subscales do not exhibit a ceiling or a floor effect in knee arthroplasty patients: an analysis of the National Health Service PROMs data set. *Knee Surgery, Sports Traumatology, Arthroscopy* 2017;25(9):2736-42.

262. Elsoe R, Ostgaard SE, Larsen P. Population-based epidemiology of 9767 ankle fractures. *Foot and Ankle Surgery* 2018;24(1):34-39.

263. Looker AC, Melton LJ, Harris T, et al. Age, gender, and race/ethnic differences in total body and subregional bone density. *Osteoporosis International* 2009;20(7):1141-49.

264. Wishart JM, Need AO, Horowitz M, et al. Effect of age on bone density and bone turnover in men. *Clinical Endocrinology* 1995;42(2):141-46.

265. Dziuban C, Shirkey E. When is a correlation matrix appropriate for factor analysis? Some decision rules. *Psychol Bull* 1974;81(6):358-61.

266. Tobias S, Carlson JE. Brief report: Bartlett's test of sphericity and chance findings in factor analysis. *Multivariate Behavioral Research* 1969;4(3):375-77.

267. Pan S-L, Liang H-W, Hou W-H, et al. Responsiveness of SF-36 and Lower Extremity Functional Scale for assessing outcomes in traumatic injuries of lower extremities. *Injury* 2014;45(11):1759-63.

268. Hermans JJ, Wentink N, Beumer A, et al. Correlation between radiological assessment of acute ankle fractures and syndesmotic injury on MRI. *Skeletal Radiology* 2012;41(7):787-801.

269. Harris KK, Price AJ, Beard DJ, et al. Can pain and function be distinguished in the Oxford Hip Score in a meaningful way? *Bone & Joint Research* 2014;3(11):305-09.

270. Field A. Exploratory Factor Analysis. In: Discovering Statistics using IBM SPSS Statistics. London: Sage Pulications 2018.

271. Egol KA, Tejwani NC, Walsh MG, et al. Predictors of short-term functional outcome following ankle fracture surgery. *Journal of Bone & Joint Surgery, American Volume* 2006;88(5):974-79.

272. Schubert J, Lambers KTA, Kimber C, et al. Effect on Overall Health Status With Weightbearing at 2 Weeks vs 6 Weeks After Open Reduction and Internal Fixation of Ankle Fractures. *Foot & Ankle International* 2020:1071100720908853.

273. Kortekangas T, Haapasalo H, Flinkkilä T, et al. Three week versus six week immobilisation for stable Weber B type ankle fractures: randomised, multicentre, non-inferiority clinical trial. *BMJ* 2019;364:k5432.

274. Briet JP, Houwert RM, Smeeing DP, et al. Weight bearing or non-weight bearing after surgically fixed ankle fractures, the WOW! Study: study protocol for a randomized controlled trial. *Trials* 2015;16(175):18.

275. Roos EM, Boyle E, Frobell RB, et al. It is good to feel better, but better to feel good: whether a patient finds treatment ‘successful’ or not depends on the questions researchers ask. *British Journal of Sports Medicine* 2019;53(23):1474.

276. Nilsson GM, Eneroth M, Ekdahl CS. The Swedish version of OMAS is a reliable and valid outcome measure for patients with ankle fractures. *BMC Musculoskeletal Disorders* 2013;14 (109)

277. Rochon PA, Mashari A, Cohen A, et al. The Inclusion of Minority Groups in Clinical Trials: Problems of Under Representation and Under Reporting of Data. *Accountability in Research* 2004;11(3-4):215-23.

278. Hussain-Gambles M, Atkin K, Leese B. Why ethnic minority groups are under-represented in clinical trials: a review of the literature. *Health & Social Care in the Community* 2004;12(5):382-88.

279. Roberson NL. Clinical trial participation: Viewpoints from racial/ethnic groups. *Cancer* 1994;74(S9):2687-91.

280. Guillemin F, Bombardier C, Beaton D. Cross-cultural adaptation of health-related quality of life measures: Literature review and proposed guidelines. *Journal of Clinical Epidemiology* 1993;46(12):1417-32.

281. Verhage SM, Schipper IB, Hoogendoorn JM. Long-term functional and radiographic outcomes in 243 operated ankle fractures. *Journal of Foot and Ankle Research* 2015;8(1):45.

282. Plant CE, Parsons NR, Costa ML. Do radiological and functional outcomes correlate for fractures of the distal radius? *The Bone & Joint Journal* 2017;99-B(3):376-82.

283. Ariza-Vega P, Jiménez-Moleón JJ, Kristensen MT. Non-Weight-Bearing Status Compromises the Functional Level Up to 1 yr After Hip Fracture Surgery. *American Journal of Physical Medicine & Rehabilitation* 2014;93(8)

284. Shulman BS, Liporace FA, Davidovitch RI, et al. Sleep Disturbance After Fracture is Related to Emotional Well-Being Rather Than Functional Result. *Journal of Orthopaedic Trauma* 2015;29(3)

285. Murray DW, Fitzpatrick R, Rogers K, et al. The use of the Oxford hip and knee scores. *The Journal of Bone and Joint Surgery British volume* 2007;89-B(8):1010-14.

286. Hewlett SA. Patients and clinicians have different perspectives on outcomes in arthritis. *The Journal of Rheumatology* 2003;30(4):877.

287. Kessler L, Ramsey SD. The Outcomes of the Cancer Outcomes Research Symposium A Commentary. *Medical Care* 2002;40(6):III104-III08.

288. Staniszewska S, Haywood KL, Brett J, et al. Patient and Public Involvement in Patient-Reported Outcome Measures. *The Patient - Patient-Centered Outcomes Research* 2012;5(2):79-87.

289. Prinsen CAC, Vohra S, Rose MR, et al. How to select outcome measurement instruments for outcomes included in a "Core Outcome Set" – a practical guideline. *Trials* 2016;17(1):449.

290. Kirkham JJ, Boers M, Tugwell P, et al. Outcome measures in rheumatoid arthritis randomised trials over the last 50 years. *Trials* 2013;14(1):324.

291. Chiarotto A, Boers M, Deyo RA, et al. Core outcome measurement instruments for clinical trials in nonspecific low back pain. *Pain* 2018;159(3):481-95.

292. Wylde V, MacKichan F, Bruce J, et al. Assessment of chronic post-surgical pain after knee replacement: Development of a core outcome set. *European Journal of Pain* 2015;19(5):611-20.

293. Haywood KL, Griffin XL, Achten J, et al. Developing a core outcome set for hip fracture trials. *The Bone & Joint Journal* 2014;96-B(8):1016-23.

294. Haywood K, Pearson N, Tutton E, et al. Developing a core outcome set for distal tibia and ankle fractures (COSTA) 2019 Available from: http://www.comet-initiative.org/Studies/Details/1488 [accessed on 24/04/2020].

295. Dorman SL, Shelton JA, Stevenson RA, et al. Management of medial humeral epicondyle fractures in children: a structured review protocol for a systematic review of the literature and identification of a core outcome set using a Delphi survey. *Trials* 2018;19(1):119.

296. Crosby BT, Behbahani A, Olujohungbe O, et al. Developing a core outcome set for paediatric wrist fractures: a systematic review of prior outcomes. *Bone & Joint Open* 2020;1(5):121-30.

9. Appendices

9.1 Appendix 1 – chapter 2: search strategy

Table 116 - Search strategy: Medline, chapter 2

Number	Term	Returns
1	exp ANKLE/	8510
2	exp Ankle Joint/	13849
3	exp Ankle Injuries/	8941
4	exp Fractures, Bone/	167059
5	fracture*.mp.	242117
6	1 or 2 or 3	26960
7	4 or 5	244160
8	6 and 7	5848
9	exp TIBIA/	31611
10	exp Tibial Fractures/	14058
11	exp FIBULA/	8068
12	7 and 11	3434
13	7 and 9	5669
14	8 or 10 or 12 or 13	22739
15	exp Randomized Controlled Trial/	454845
16	exp Controlled Clinical Trial/	542188

17	randomized.mp. or exp RANDOMIZED CONTROLLED TRIALS AS A TOPIC/	672656
18	placebo.mp.	176058
19	randomly.mp.	246332
20	trial.mp.	987780
21	15 or 16 or 17 or 18 or 19 or 20	1303244
22	14 and 21	1146

Table 117 - Search Strategy: Embase, chapter 2

Number	Term	Returns
1	exp Ankle/	31462
2	exp ankle injury/	12120
3	exp ankle fracture/	5654
4	exp fracture/	283825
5	1 and 4	2889
6	exp distal tibia/	457
7	4 and 6	146
8	exp distal tibia fracture/	738
9	exp distal fibula/	166
10	4 and 9	38
11	exp distal fibula fracture/	86
12	2 or 3 or 5 or 7 or 8 or 10 or 11	14718
13	randomized controlled trial/	490547

14	exp controlled clinical trial/	668429
15	randomized.mp. or exp “randomized controlled trial (topic)” or exp placebo/	1098532
16	randomly.mp.	374772
17	trial.mp. or exp “clinical trial (topic)”/	1870869
18	13 or 14 or 15 or 16 or 17	2335340
19	12 and 18	1148

Table 118 - Search strategy: CINAHL, chapter 2

Number	Term	Returns
1	(MH “Ankle”)	3495
2	(MH “Ankle Joint”)	2764
3	(MH “Ankle Injuries+”)	3878
4	(MH “Fractures+”)	27923
5	(MH “Tibia”)	2879
6	(MH “Fibula”)	716
7	S1 OR S2 OR S3	9215
8	S4 AND S7	1503
9	S5 OR S6	3392
10	S4 AND S9	595
11	S8 OR S10	1954
12	(MH “Tibial Fractures+”)	1579
13	(MH “Fibula Fractures”)	146

14	S11 OR S12 OR S13	3232
15	(MH “Randomized Controlled Trials”) OR (MH “Clinical Trials”)	130098
16	“randomly”	41835
17	“trial”	99705
18	“placebo” OR (MH “Placebos”)	33546
19	S15 OR S16 OR S17 OR S18	214422
20	S14 AND S19	187

Table 119 - Search strategy: AMED, chapter 2

Number	Term	Returns
1	exp Ankle/	1263
2	exp ankle injuries/	1321
3	exp fractures bone/	2301
4	1 and 3	34
5	2 or 4	1355
6	exp Fibula/	231
7	exp Tibia/	894
8	6 or 7	1057
9	3 and 8	111
10	5 or 9	1432
11	exp Randomized controlled trials/	1908
12	controlled clinical trial.mp.	428

13	exp Clinical trials/ or randomized.mp.	11375
14	randomly.mp.	5360
15	trial.mp.	9887
16	11 or 12 or 13 or 14 or 15	18769
17	10 and 16	126

Table 120 - Search strategy: Cochrane CENTRAL Trials database, chapter 2

Number	Term	Returns
1	MeSH descriptor: [Ankle Joint] explode all trees	646
2	MeSH descriptor: [Ankle Injuries] explode all trees	608
3	MeSH descriptor: [Ankle] explode all trees	471
4	MeSH descriptor: [Tibia] explode all trees	560
5	MeSH descriptor: [Fibula] explode all trees	77
6	MeSH descriptor: [Ankle Fractures] explode all trees	44
7	#1 or #2 or #3 or #4 or #5 or #6	2140
8	MeSH descriptor: [Fractures, Bone] explode all trees	5193
9	#8 and #7	238
10	Limit to “Trials”	210

Table 121 - Search strategy: ISRCTN registry, chapter 2

Number	Term	Returns
1	Ankle Fracture	57

Table 122 - Search strategy: ClinicalTrials.gov registry, chapter 2

Number	Term	Returns
1	Ankle Fracture	104
2	Limit to “Interventional (Clinical Trial)”	84

9.2 Appendix 2 – chapter 2: data extraction form

Table 123 - Blank data extraction form for systematic review, chapter 2

Year of publication	Authors	Title (and link to registry record)	Was the primary outcome measure specified? (Y/N)	Sample size (planned if protocol or registry record)	Types of intervention studied in the trial	Primary outcome measure	Secondary outcome measure/s

9.3 Appendix 3 – chapter 3: search strategy

Table 124 - Search strategy: MEDLINE, chapter 3

Number	Term	Returns
1	exp ANKLE JOINT/ or exp ANKLE/ or exp ANKLE INJURIES/ or ankle.mp. or exp Fibula/ or exp TIBIA/	86638
2	exp Ankle Fractures/ or exp Tibial Fractures/ or distal tibia fracture.mp. or distal fibula fracture.mp.	15192
3	exp Fractures, Bone/ or fracture*.mp.	248872
4	(Olerud Molander Ankle Score or OMAS or "Olerud and Molander Ankle Score" or MOXFQ or Manchester-Oxford Foot Questionnaire or M-OXFQ or AAOS or "AAOS Foot and Ankle Outcome*" or "AAOS Foot and Ankle Outcomes Questionnaire" or American Academy of Orthopaedic Surgeons).mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	1019
5	(Foot Function Index or FFI or “Foot and Ankle Measure” or FAAM or Karlsson Score or A-FORM or “Ankle Fracture Outcome of Rehabilitation Measure” or “Foot and Ankle Outcome Survey” or “KOOS Foot and Ankle Outcome Survey”).mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	26329
6	1 and 3	15841
7	6 or 2	25148
8	4 or 5	27336
9	7 and 8	147
10	(Validation Studies or Comparative Study).pt.	1875546
11	exp Psychometrics/	67784
12	psychometr*.ti,ab.	33609

13	(clinimetr* or clinometr*).tw.	768
14	exp "Outcome Assessment (Health Care)"/	945982
15	outcome assessment.ti,ab.	2851
16	outcome measure*.tw.	177528
17	exp Observer Variation/	39389
18	observer variation.ti,ab.	928
19	exp Health Status Indicators/	273102
20	exp "Reproducibility of Results"/	360511
21	reproducib*.ti,ab.	124740
22	exp Discriminant Analysis/	9471
23	(reliab* or unreliab* or valid* or coefficient or homogeneity or homogeneous or internal consistency).mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	1033477
24	(cronbach* and (alpha or alphas)).ti,ab.	15075
25	(item and (correlation* or selection* or reduction*)).ti,ab.	15682
26	(agreement or precision or imprecision or precise values or test-retest).mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	298087
27	(test and retest).ti,ab.	20045
28	(reliab* and (test or retest)).ti,ab.	67280
29	(stability or interrater or inter-rater or intrarater or intra-rater or intertester or inter-tester or intratester or intra-tester or interobserver or inter-observer or intraobserver or intraobserver or intertechnician or inter-technician or intratechnician or intra-technician or interexaminer or inter-examiner or intraexaminer or intra-examiner or interassay or inter-assay or intraassay or intra-assay or interindividual or inter-individual or intraindividual or	522008

	intra-individual or interparticipant or inter-participant or intraparticipant or intra-participant or kappa or kappas or repeatab*).mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	
30	((replicab* or repeated) and (measure or measures or findings or result or results or test or tests)).ti,ab.	154761
31	(generaliza* or generalisa* or concordance).ti,ab.	63167
32	(intraclass and correlation*).ti,ab.	17343
33	(discriminative or known group or factor analysis or factor analyses or dimension* or subscale*).ti,ab.	431177
34	(multitrait and scaling and (analysis or analyses)).ti,ab.	124
35	(item discriminant or interscale correlation* or error or errors or individual variability).ti,ab.	215556
36	(variability and (analysis or values)).ti,ab.	75359
37	(uncertainty and(measurement or measuring)).ti,ab.	4318
38	(standard error of measurement or sensitive* or responsive*).mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	1536738
39	((minimal or minimally or clinical or clinically) and (important or significant or detectable) and (change or difference)).ti,ab.	173713
40	(small* and (real or detectable) and (change or difference)).ti,ab.	5447
41	(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).mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	9588

42	8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 1or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27 or 28 or 29 or 30 or 31 or 32 or 33 or 34 or 35 or 36 or 37 or 38 or 39	6030304
43	9 and 42	124

Search Strategy EMBASE

Number	Term	Returns
1	exp ankle/ or exp ankle injury/ or exp tibia/ or exp distal tibia/ or exp fibula/ or exp distal fibula/	89795
2	ankle fracture.mp. or exp ankle fracture/ or exp distal tibial fracture or exp distal fibula fracture/	5750
3	(distal tibia fracture or distal fibula fracture).mp. [mp=title, abstract, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword, floating subheading word, candidate term word]	698
4	fracture*.mp.	366187
5	1 and 4	19820
6	2 or 3 or 5	20384
7	(OMAS or Olerud Molander Ankle Score or "Olerud and Molander ankle Score" or Manchester-Oxford Foot Questionnaire or M-OXFQ or MOXFQ or AAOS or "AAOS Foot and Ankle Outcome Questionnaire" or " American Academy of Orthopaedic Surgeons Foot and Ankle*").mp. [mp=title, abstract, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword, floating subheading word, candidate term word]	1024
8	(Foot Function Index or FFI or "Foot and Ankle Measure" or FAAM or Karlsson Score or A-FORM or "Ankle Fracture Outcome of Rehabilitation Measure" or "Foot and Ankle Outcome Survey" or KOOS Foot and Ankle Outcome Survey").mp. [mp=title, abstract, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword, floating subheading word, candidate term word]	39585
9	7 or 8	40594

10	6 and 9	140
11	exp intermethod comparison/	218718
12	exp data collection method/	867010
13	exp validation study/	63825
14	exp feasibility study/	89537
15	exp psychometry/	79993
16	exp reproducibility/	189310
17	(reproducib* or audit or psychometr* or clinimetr* or cliniometr*).abti.	286861
18	exp observer variation/	18799
19	observer variation.ab,ti.	1445
20	exp discriminant analysis/	15885
21	exp validity/	77283
22	(reliab* or valid*or coefficient or internal consistency).ab,ti.	1389795
23	((Cronbach* and (alpha or alphas)) or item correlations or item selection or item selections or item reduction or item reductions or agreement or precision or imprecision or precise values or test-retest or (test and retest)).ab,ti.	439738
24	((reliab* and (test or retest)) or stability or interrater or inter-rater or intrarater or intra-rater or intertester or inter-tester or intratester or intra-tester or interobserver or inter-observer or intraobserver or intra-observer or intertechnician or inter-technician or intratechnician or intra-technician or interexaminer or inter-examiner or interassay or inter-assay or intraassay or intra-assay or interindividual or inter-individual or intra-individual or intraindividual or inter-participant or interparticipant or intraparticipant or intra-participant).ab,ti.	564805
25	(kappa or kappas or coefficient of variation or repeatab* or ((replica* or repeated) and (measure or measures or findings or result or results or test or tests))).ab,ti.	393093
26	(generaliza* or generalisa* or concordance or (intraclass and correlation*)).ab,ti.	120139

27	(discriminative or known group or factor analysis or facture analyses or facture structure or facture structures or dimensionality or subscale* or multitrait scaling analysis or multitrait scaling analyses).ab,ti.	118544
28	(item discriminant or interscale correlation or interscale correlations or ((error or errors) and (measure* or correlat* or evaluat* or accuracy or accurate or precision or mean)) or individual variability or interval variability or rate variability or variability analysis or (uncertainty and (measurement or measuring))).ab,ti.	232870
29	(standard error or measurement or sensitive* or responsive* or (limit and detection) or minimal detectable concentration or interpretab* or (small* and (real or detectable) and (change or difference)) or meaningful change or minimal important change or minimal important difference or minimally important change or minimally important difference or minimal detectable change or minimal detectable difference or minimally detectable change or minimally detectable difference or minimal real change or minimal real difference or minimally real change or minimally real difference 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).ab,ti.	1819366
30	11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27 or 28 or 29	5191379
31	10 and 30	68

Table 125 - Search strategy: CINAHL, chapter 3

Number	Term	Returns
1	(MH “Ankle”) OR (MH “Ankle Injuries+”) OR (MH “Ankle Joint”)	9717
2	(MH “Tibia”)	2964
3	(MH “Fibula”)	737
4	(MH “Fractures+”)	28805
5	“fracture*”	40620
6	1 OR 2 OR 3	12768
7	4 OR 5	40748

8	6 AND 7	2526
9	(MH “Tibial Fractures+”) OR (MH “Ankle Fractures”) OR (MH “Fibula Fractures”)	2619
10	8 OR 9	3864
11	“(Olerud Molander Ankle Score or OMAS or “Olerud and Molander Ankle Score”) OR (MOXFQ or Manchester-Oxford Foot Questionnaire or M-OXFQ) OR (“AAOS” or “AAOS Foot and Ankle Outcome Questionnaire” or “American Academy of Orthopaedic Surgeons Foot and Ankle*”)”	317
12	“Foot Function Index or FFI or “Foot and Ankle Measure” or FAAM or Karlsson Score or A-FORM or “Ankle Fracture Outcome of Rehabilitation Measure” or “Foot and ankle Outcome Survey” or “KOOS Foot and Ankle Outcome Survey””	0
13	MH “Psychometrics”) or (TI psychometr* or AB psychometr*) or (TI clinimetr* or AB clinimetr*) or (TI clinometr* OR AB clinometr*) or (MH “Outcome Assessment”) or (TI outcome assessment or AB outcome assessment) or (TI outcome measure* or AB outcome measure*) or (MH “Health Status Indicators”) or (MH “Reproducibility of Results”) or (MH “Discriminant Analysis”) or ((TI reproducib* or AB reproducib*) or (TI reliab* or AB reliab*) or (TI unreliab* or AB unreliab*)) or ((TI valid* or AB valid*) or (TI coefficient or AB coefficient) or (TI homogeneity or AB homogeneity)) or (TI homogeneous or AB homogeneous) or (TI “coefficient of variation” or AB “coefficient of variation”) or (TI “internal consistency” or AB “internal consistency”) or (MH “Internal Consistency+”) or (MH “Reliability+”) or (MH “Measurement Error+”) or (MH “Content Validity+”) or “hypothesis testing” or “structural validity” or “cross-cultural validity” or (MH “Criterion-Related Validity+”) or “responsiveness” or “interpretability” or (TI reliab* or AB reliab*) and ((TI test or AB test) OR (TI retest or AB retest)) or (TI stability or AB stability) or (TI interrater or AB interrater) or (TI	349401

	inter-rater or AB inter-rater) or (TI intrarater or AB intrarater) or (TI intra-rater or AB intrarater) or (TI intertester or AB intertester) or (TI inter-tester or AB inter-tester) or (TI intratester or AB intratester) or (TI intra-tester or AB intra-tester) or (TI interobserver or AB interobserver) or (TI inter-observer or AB inter-observer) or (TI intraobserver or AB intraobserver) or (TI intra-observer or AB intra-observer) or (TI intertechnician or AB intertechnician) or (TI inter-technician or AB inter-technician) or (TI intratechnician or AB intratechnician) or (TI intra-technician or AB intra-technician) or (TI interexaminer or AB interexaminer) or (TI inter-examiner or AB inter-examiner) or (TI intraexaminer or AB intraexaminer) OR (TI intra-examiner or AB intra-examiner) or (TI intra-examiner or AB intraexaminer) or (TI interassay or AB interassay) or (TI inter-assay or AB inter-assay) or (TI intraassay or AB intraassay) or (TI intra-assay or AB intra-assay) or (TI interindividual or AB interindividual) or (TI inter-individual or AB inter-individual) OR (TI intraindividual or AB intraindividual) or (TI intra-individual or AB intra-individual) or (TI interparticipant or AB interparticipant) or (TI inter-participant or AB inter-participant) or (TI intraparticipant or AB intraparticipant) or (TI intra-participant or AB intra-participant) or (TI kappa or AB kappa) or (TI kappa’s or AB kappa’s) or (TI kappas or AB kappas) or (TI repeatab* or AB repeatab*) or (TI responsive* or AB responsive*) or (TI interpretab* or AB interpretab*)	
14	11 AND 13	108

9.4 Appendix 4 – chapter 3: reviewer form for eligibility assessment

Table 126 - Reviewer form for assessment of eligibility, chapter 3

Authors	Date	Journal	Title	Abstract	Potentially eligible based on title and abstract review? (Y/N)	Eligible based on full text review? (Y/N)

9.5 Appendix 5 – chapter 3: completed COSMIN risk of bias checklist

Table 127 - COSMIN risk of bias checklist - box 1: PROM design, chapter 3

Ratings: V= very good; A = adequate; D = doubtful; I = inadequate; N/A= not applicable *Key: 1st Rev= First reviewer; 2nd Rev= Second Reviewer; C=Consensus*							
PROM and language:		A-FORM English			OMAS		
Article:		McPhail et al. (2014)			Olerud and Molander (1984)		
1st and 2nd Reviewer Results and Consensus:		1st Rev	2nd Rev	C	1st Rev	2nd Rev	C
1a. PROM Design							
General design requirements							
1	Is a clear description provided of the construct to be measured?	V	V	V	I		
2	Is the origin of the construct clear: was a theory, conceptual framework or disease model used or clear rationale provided to define the construct to be measured?	V	V	V	D		

3	Is a clear description provided of the target population for which the PROM was developed?	V	V	V	V		
4	Is a clear description provided of the context of use (i.e. discriminative, evaluative purpose, and/or predictive)	V	V	V	D		
5	Was the PROM development study performed in a sample representing the target population for which the PROM was developed?	V	V	V	D		
Concept elicitation (relevance and comprehensiveness)							
6	Was an appropriate qualitative data collection method used to identify relevant items for a new PROM?	V	V	V	I		
7	Were skilled group moderators/interviewers used?	D	D	D	N/A		
8	Were the group meetings or interviews based on an appropriate topic or interview guide?	A	A	A	N/A		
9	Were the group meetings or interviews recorded and transcribed verbatim?	V	V	V	N/A		
10	Was an appropriate approach used to analyse the data?	V	V	V	N/A		
11	Was at least part of the data coded independently?	V	V	V	N/A		

12	Was data collection continued until saturation was reached?	V	V	V	N/A		
13	For quantitative studies: was the sample size appropriate?	N/A	N/A	N/A	N/A		
	SUBTOTAL QUALITY CONCEPT ELICITATION STUDY *Lowest score of items 6-13*	A	A	A	I		
	TOTAL QUALITY OF THE PROM DESIGN *Lowest score of items 1-13*	D	D	D	I		
1b. Cognitive interview study or other pilot test							
14	Was a cognitive interview study or other pilot test performed? *If NO skip items 15-35*	No	No	No	No		
General Design Requirements							
15	Was the cognitive interview study or other pilot test performed in a sample representing the target population?						
Comprehensibility							
16	Were patients asked about the comprehensibility of the PROM? *If NO or not clear, skip items 17-25*						

17	Were all items tested in their final form?						
18	Was an appropriate qualitative method used to assess the comprehensibility of the PROM instructions, items, response options, and recall period?						
19	Was each item tested in an appropriate number of patients?						
20	Were skilled interviewers used?						
21	Were the interviews based on an appropriate interview guide?						
22	Were the interviews recorded and transcribed verbatim?						
23	Was an appropriate approach used to analyse the data?						
24	Were at least two researchers involved in the analysis?						
25	Were problems regarding the comprehensibility of the PROM instructions, items, response options, and recall period appropriately addressed by adapting the PROM?						
SUBTOTAL QUALITY OF COMPREHENSIBILITY STUDY *Lowest score of items 15-25*							
Comprehensiveness							

26	Were patients asked about the comprehensiveness of the PROM? *If NO or not clear, skip items 27-35*	No	No	No	No		
27	Was the final set of items tested?						
28	Was an appropriate method used for assessing the comprehensiveness of the PROM?						
29	Was each item tested in an appropriate number of patients?						
30	Were skilled interviewers used?						
31	Were the interviews based on an appropriate interview guide?						
32	Were the interviews recorded and transcribed verbatim?						
33	Was an appropriate approach used to analyse the data?						
34	Were at least two researchers involved in the analysis?						
35	Were problems regarding the comprehensiveness of the PROM appropriately addressed by adapting the PROM?						
SUBTOTAL QUALITY OF COMPREHENSIVENESS STUDY *Lowest score of items 15, 26-35*		N/A	N/A	N/A	N/A		

TOTAL QUALITY OF THE PILOT STUDY *Lowest score of items 14-35*	N/A	N/A	N/A	N/A		
TOTAL QUALITY OF THE PROM DEVELOPMENT STUDY *Lowest score of items 1-35*	D	D	D	I		

Table 128 - COSMIN risk of bias checklist boxes 3-10: A-FORM, AAOS and FFI articles, chapter 3

Score: V= very good; A = adequate; D = doubtful; I = inadequate; N/A= not applicable

Key: 1st Rev= First reviewer; 2nd Rev= Second Reviewer; C=Consensus

PROM and language:	A-FORM English			AAOS Spanish			FFI Taiwan Chinese		
Article:	McPhail et al. (2014)			Zelle et al. (2017)			Wu et al. (2008)		
1st and 2nd Reviewer Results and Consensus:	1st Rev	2nd Rev	C	1st Rev	2nd Rev	C	1st Rev	2nd Rev	C
3. *Structural validity*									
Does the scale consist of effect indicators, i.e. is it based on a reflective model? Unidimensionality or structural validity?									

1	For CTT: Was exploratory or confirmatory factor analysis performed?									
2	For IRT/Rasch: does the chosen model fit to the research question?									
3	Was the sample size included in the analysis adequate?									
4	Were there any other important flaws?									
TOTAL Lowest score of items 1-4										
4. Internal consistency										
Does the scale consist of effect indicators, i.e. is it based on a reflective model?										
1	Was an internal consistency statistic calculated for each unidimensional (sub)scale separately?									
2	For continuous scores: Was Cronbach's alpha or omega calculated?									

3	For dichotomous scores: Was Cronbach's alpha or KR-20 calculated?									
4	For IRT-based scores: Was standard error of the theta (SE (θ)) or reliability coefficient of estimated latent trait value (index of (subject or item) separation) calculated?									
5	Were there any other important flaws?									
TOTAL Lowest score of items 1-5										
5. Cross-cultural validity\measurement invariance										
1	Were the samples similar for relevant characteristics except for the group variable?									
2	Was an adequate approach used to analyse the data?									
3	Was the sample size included in the analysis adequate?									
4	Were there any other important flaws?									

TOTAL Lowest score of items 1-4										
6. Reliability										
1	Were patients stable in the interim period on the construct to be measured?				D	D	D			
2	Was the time interval appropriate?				V	I	D			
3	Were the test conditions similar for the measurements? e.g. type of administration, environment, instructions				D	D	D			
4	For continuous scores: Was an intraclass correlation coefficient (ICC) calculated?				D	I	D			
5	For dichotomous/nominal/ordinal scores: Was kappa calculated?				N/A	N/A	N/A			
6	For ordinal scores: Was a weighted kappa calculated?				N/A	N/A	N/A			
7	For ordinal scores: Was the weighting scheme described? e.g. linear, quadratic				N/A	N/A	N/A			

8	Were there any other important flaws?				D	D	D			
TOTAL Lowest score of items 1-8					D	D	D			
7. Measurement error										
1	Were patients stable in the interim period on the construct to be measured?									
2	Was the time interval appropriate?									
3	Were the test conditions similar for the measurements? e.g. type of administration, environment, instructions									
4	For continuous scores: Was the Standard Error of Measurement (SEM), Smallest Detectable Change (SDC) or Limits of Agreement (LoA) calculated?									
5	For dichotomous/nominal/ordinal scores: Was the percentage (positive and negative) agreement calculated?									
6	Were there any other important flaws?									

TOTAL *Lowest score of items 1-6*										
8. Criterion validity (NOT APPLICABLE)										
1	For continuous scores: Were correlations, or the area under the receiver operating curve calculated?									
2	For dichotomous scores: Were sensitivity and specificity determined?									
3	Were there any other important flaws?									
TOTAL *Lowest score of items 1-3*										
9. Hypotheses testing for construct validity										
9a. Comparison with other outcome measurement instruments (convergent validity)										
1	Is it clear what the comparator instrument(s) measure(s)?				V	V	V	V		
2	Were the measurement properties of the comparator instrument(s) adequate?				D	D	D	D		

3	Was the statistical method appropriate for the hypotheses to be tested?				A	A	A	A		
4	Were there any other important flaws?				D	V	V	A		
TOTAL *Lowest score of items 1-4*					D	D	D	D		
9b. Comparison between subgroups (discriminative or known-groups validity)										
5	Was an adequate description provided of important characteristics of the subgroups?									
6	Was the statistical method appropriate for the hypotheses to be tested?									
7	Were there any other important flaws?									
TOTAL *Lowest score of items 5-7*										
10. Responsiveness										
10a. Criterion approach (i.e. comparison to a gold standard) (NOT APPLICABLE)										
1	For continuous scores: Were correlations between change scores, or the area under the									

	Receiver Operator Curve (ROC) curve calculated?									
2	For dichotomous scales: Were sensitivity and specificity (changed versus not changed) determined?									
3	Were there any other important flaws?									
TOTAL *Lowest score of items 1-3*										
10b. Construct approach (i.e. hypotheses testing; comparison with other outcome measurement instruments)										
4	Is it clear what the comparator instrument(s) measure(s)?									
5	Were the measurement properties of the comparator instrument(s) adequate?									
6	Was the statistical method appropriate for the hypotheses to be tested?									
7	Were there any other important flaws?									
TOTAL *Lowest score of items 4-7*										

10c. Construct approach: (i.e. hypotheses testing: comparison between subgroups)										
8	Was an adequate description provided of important characteristics of the subgroups?									
9	Was the statistical method appropriate for the hypotheses to be tested?									
10	Were there any other important flaws?									
TOTAL *Lowest score of items 8-10*										
10d. Construct approach: (i.e. hypotheses testing: before and after intervention)										
11	Was an adequate description provided of the intervention given?									
12	Was the statistical method appropriate for the hypotheses to be tested?									
13	Were there any other important flaws?									
TOTAL *Lowest score of items 11-13*										

Table 129 - COSMIN risk of bias checklist boxes 3-10: OMAS articles, chapter 3

	PROM and language:	OMAS Turkish			OMAS Norwegian			OMAS Turkish		
	Article:	Buker et al. (2017)			Garratt et al. (2018)			Turhan et al. (2017)		
	1st and 2nd Reviewer Results and Consensus (C)	1st	2nd	(C)	1st	2nd	(C)	1st	2nd	(C)
3. Structural validity										
	Does the scale consist of effect indicators, i.e. is it based on a reflective model? Unidimensionality or structural validity?				Yes Structural validity					
1	For CTT: Was exploratory or confirmatory factor analysis performed?				V					
2	For IRT/Rasch: does the chosen model fit to the research question?				N/A					

3	Was the sample size included in the analysis adequate?				V					
4	Were there any other important flaws?				D					
TOTAL *Lowest score of items 1-4*					D					
4. Internal consistency										
Does the scale consist of effect indicators, i.e. is it based on a reflective model?		Yes	Yes	Yes	Yes			Yes	Yes	Yes
1	Was an internal consistency statistic calculated for each unidimensional (sub)scale separately?	I	D	D	D			I	D	D
2	For continuous scores: Was Cronbach's alpha or omega calculated?	V	V	V	V			D	V	V
3	For dichotomous scores: Was Cronbach's alpha or KR-20 calculated?	N/A	N/A	N/A	N/A			N/A	N/A	N/A

4	For IRT-based scores: Was standard error of the theta (SE (θ)) or reliability coefficient of estimated latent trait value (index of (subject or item) separation) calculated?	I	N/A	N/A	N/A			I	N/A	N/A
5	Were there any other important flaws?	V	N/A	V	V			V	N/A	V
TOTAL *Lowest score of items 1-5*		I	V	D	D			I	D	D
5. Cross-cultural validity\measurement invariance										
1	Were the samples similar for relevant characteristics except for the group variable?									
2	Was an adequate approach used to analyse the data?									
3	Was the sample size included in the analysis adequate?									

4	Were there any other important flaws?									
TOTAL *Lowest score of items 1-4*										
6. Reliability										
1	Were patients stable in the interim period on the construct to be measured?	D	V	D	D			V	D	D
2	Was the time interval appropriate?	V	D	D	D			V	D	D
3	Were the test conditions similar for the measurements? e.g. type of administration, environment, instructions	I	I	I	V			A	D	D
4	For continuous scores: Was an intraclass correlation coefficient (ICC) calculated?	A	V	A	V			A	V	A
5	For dichotomous/nominal/ordinal scores: Was kappa calculated?	N/A	N/A	N/A	N/A			N/A	N/A	N/A

6	For ordinal scores: Was a weighted kappa calculated?	N/A	N/A	N/A	N/A			N/A	N/A	N/A
7	For ordinal scores: Was the weighting scheme described? e.g. linear, quadratic	N/A	N/A	N/A	N/A			N/A	N/A	N/A
8	Were there any other important flaws?	D	I	I	V			D	D	D
TOTAL *Lowest score of items 1-8*		I	I	I	D			A	D	D
7. Measurement error										
1	Were patients stable in the interim period on the construct to be measured?				D			V	D	V
2	Was the time interval appropriate?				D			V	D	D
3	Were the test conditions similar for the measurements? e.g. type of administration, environment, instructions				V			A	D	D

4	For continuous scores: Was the Standard Error of Measurement (SEM), Smallest Detectable Change (SDC) or Limits of Agreement (LoA) calculated?				V			V	A	V
5	For dichotomous/ nominal/ ordinal scores: Was the percentage (positive and negative) agreement calculated?				N/A			I	N/A	N/A
6	Were there any other important flaws?				V			D	V	V
TOTAL *Lowest score of items 1-6*					D			I	D	D
8. Criterion validity (NOT APPLICABLE)										
1	For continuous scores: Were correlations, or the area under the receiver operating curve calculated?									

2	For dichotomous scores: Were sensitivity and specificity determined?									
3	Were there any other important flaws?									
TOTAL *Lowest score of items 1-3*										
9. Hypotheses testing for construct validity										
9a. Comparison with other outcome measurement instruments (convergent validity)										
1	Is it clear what the comparator instrument(s) measure(s)?	V	V	V	V			V	V	V
2	Were the measurement properties of the comparator instrument(s) adequate?	D	I	D	A			A	D	D
3	Was the statistical method appropriate for the hypotheses to be tested?	A	A	A	A			A	A	A

4	Were there any other important flaws?	V	V	V	V			V	V	V
TOTAL *Lowest score of items 1-4*		D	I	D	A			A	D	D
9b. Comparison between subgroups (discriminative or known-groups validity)										
5	Was an adequate description provided of important characteristics of the subgroups?				A					
6	Was the statistical method appropriate for the hypotheses to be tested?				A					
7	Were there any other important flaws?				V					
TOTAL *Lowest score of items 5-7*					A					
10. Responsiveness										
10a. Criterion approach (i.e. comparison to a gold standard) (NOT APPLICABLE)										
1	For continuous scores: Were correlations between change									

	scores, or the area under the Receiver Operator Curve (ROC) curve calculated?									
2	For dichotomous scales: Were sensitivity and specificity (changed versus not changed) determined?									
3	Were there any other important flaws?									
TOTAL Lowest score of items 1-3										
10b. Construct approach (i.e. hypotheses testing; comparison with other outcome measurement instruments)										
4	Is it clear what the comparator instrument(s) measure(s)?									
5	Were the measurement properties of the comparator instrument(s) adequate?									

6	Was the statistical method appropriate for the hypotheses to be tested?									
7	Were there any other important flaws?									
TOTAL Lowest score of items 4-7										
10c. Construct approach: (i.e. hypotheses testing: comparison between subgroups)										
8	Was an adequate description provided of important characteristics of the subgroups?									
9	Was the statistical method appropriate for the hypotheses to be tested?									
10	Were there any other important flaws?									
TOTAL Lowest score of items 8-10										
10d. Construct approach: (i.e. hypotheses testing: before and after intervention)										

11	Was an adequate description provided of the intervention given?									
12	Was the statistical method appropriate for the hypotheses to be tested?									
13	Were there any other important flaws?									
TOTAL *Lowest score of items 11-13*										

9.6 Appendix 6 – chapters 4&5: interview study documents

Participant invite letter

WARWICK
CLINICAL TRIALS UNIT

AIR
ANKLE INJURY REHABILITATION

AIR Qualitative Sub-study Invitation Letter

Dear

We are contacting you because you are participating in the AIR Study. This study is run by the University of Warwick and is investigating the best way of treating ankle fractures in adults. You indicated on your consent form that you would be happy to be contacted about future research in this area and we would now like to invite you to contribute to another part of the study. This study is an interview study to determine what it's like to live with an ankle fracture and determine what is most important to individuals with this injury.
Before you decide if you want to take part, please take the time to carefully read the enclosed participant information sheet. This information will inform you of what will be involved.

A member of the research team will contact you in about a one week to discuss the study and find out if you are interested in taking part. You can directly contact the study team during office hours on: [redacted] or email air@warwick.ac.uk

If you decide not to take part, the usual care you receive from your GP and other healthcare providers will not be affected in any way.
If you have any questions and/or would like to talk to someone please call the above number.

Yours sincerely,

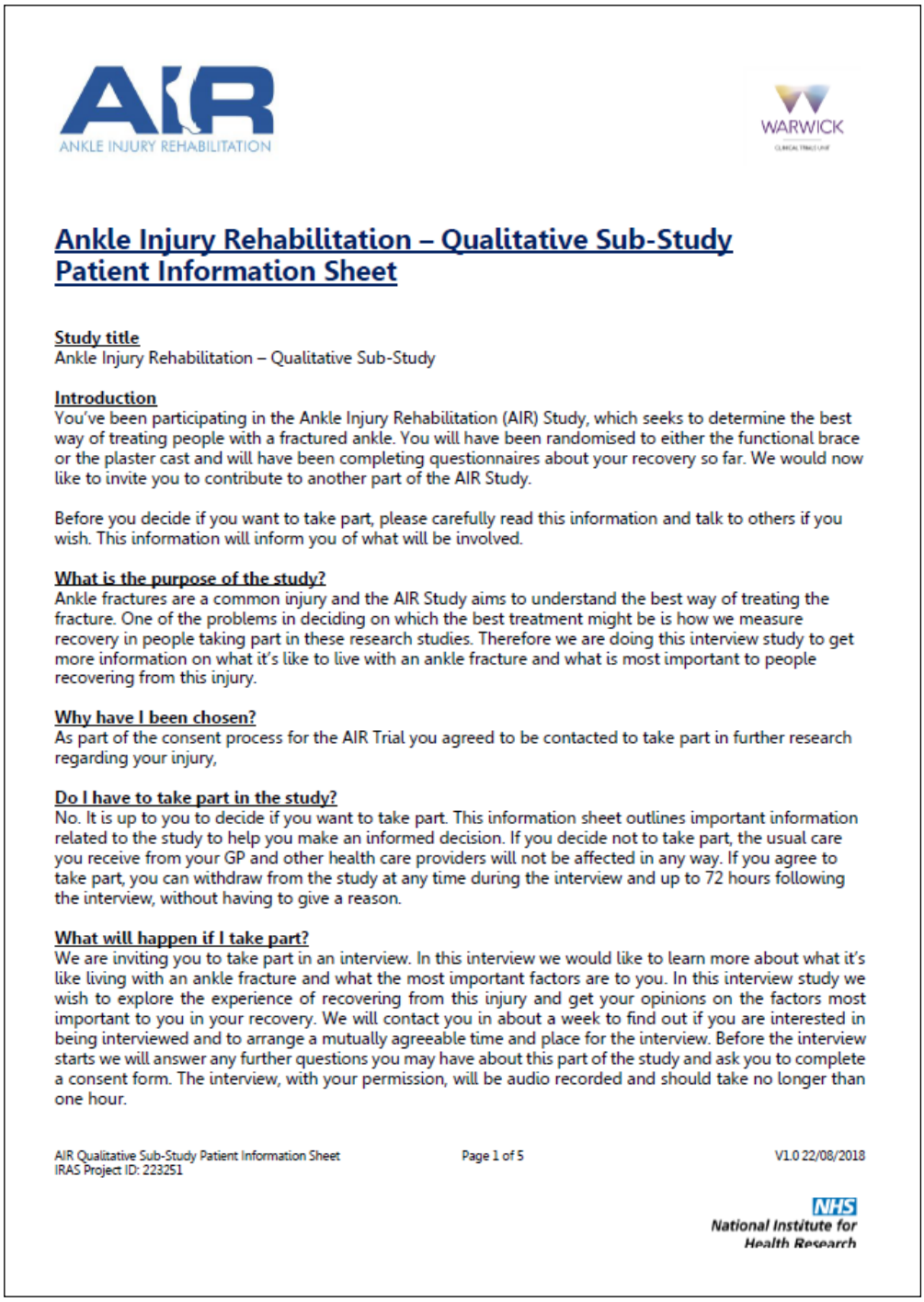
Dr Rebecca Kearney (Chief Investigator)

Enclosed:
Participant Information Sheet

AIR Qualitative Sub-Study Invitation Letter V1.0 23/07/2018

NHS
National Institute for Health Research

Participant information sheet



ANKLE INJURY REHABILITATION

WARWICK

Ankle Injury Rehabilitation – Qualitative Sub-Study Patient Information Sheet

Study title
Ankle Injury Rehabilitation – Qualitative Sub-Study

Introduction
You've been participating in the Ankle Injury Rehabilitation (AIR) Study, which seeks to determine the best way of treating people with a fractured ankle. You will have been randomised to either the functional brace or the plaster cast and will have been completing questionnaires about your recovery so far. We would now like to invite you to contribute to another part of the AIR Study.

Before you decide if you want to take part, please carefully read this information and talk to others if you wish. This information will inform you of what will be involved.

What is the purpose of the study?
Ankle fractures are a common injury and the AIR Study aims to understand the best way of treating the fracture. One of the problems in deciding on which the best treatment might be is how we measure recovery in people taking part in these research studies. Therefore we are doing this interview study to get more information on what it's like to live with an ankle fracture and what is most important to people recovering from this injury.

Why have I been chosen?
As part of the consent process for the AIR Trial you agreed to be contacted to take part in further research regarding your injury,

Do I have to take part in the study?
No. It is up to you to decide if you want to take part. This information sheet outlines important information related to the study to help you make an informed decision. If you decide not to take part, the usual care you receive from your GP and other health care providers will not be affected in any way. If you agree to take part, you can withdraw from the study at any time during the interview and up to 72 hours following the interview, without having to give a reason.

What will happen if I take part?
We are inviting you to take part in an interview. In this interview we would like to learn more about what it's like living with an ankle fracture and what the most important factors are to you. In this interview study we wish to explore the experience of recovering from this injury and get your opinions on the factors most important to you in your recovery. We will contact you in about a week to find out if you are interested in being interviewed and to arrange a mutually agreeable time and place for the interview. Before the interview starts we will answer any further questions you may have about this part of the study and ask you to complete a consent form. The interview, with your permission, will be audio recorded and should take no longer than one hour.

AIR Qualitative Sub-Study Patient Information Sheet
IRAS Project ID: 223251
Page 1 of 5
V1.0 22/08/2018

NHS
National Institute for Health Research

What are the possible risks of taking part?
We do not think there are any risks with taking part in this interview study. The interview will last no longer than one hour and can be ended at any time you like.

What are the possible benefits of taking part?
Whilst there may not be any direct benefits to you, the information we gather will help to inform research on measuring recovery in people who have fractured their ankle.

Will my details be kept confidential?
Yes. All information that is collected during the study will be kept confidential at all times and held in compliance with the Data Protection Act 2018.
The interviews will be transcribed and made anonymous (all your identifiable information will be deleted) by a member of the research team or by a third party transcription service. These will be contracted to work on the research project and compliant with University data protection procedures. Anonymised transcripts may be shared with other carefully selected researchers for further analysis. Transcripts will be held securely and only accessed by authorised study personnel. In the future we may use the recordings, and transcripts, of the interviews for other research. We might agree to share these with other, carefully selected, researchers; any such sharing will be closely monitored by the University. We may use anonymised written quotations from your interviews to illustrate academic presentations or publications based on this research

What if there is a problem?
This study is covered by the University of Warwick's insurance and indemnity cover. If you have any concerns about this study, please contact the Chief Investigator of the study:

Dr Rebecca Kearney
Associate Professor, Warwick Clinical Trials Unit
Warwick Medical School
The University of Warwick
Coventry, CV4 7AL
[redacted]

Who should I contact if I wish to make a complaint?
Any complaint about the way you have been dealt with during the study or any possible harm you might have suffered will be addressed. Please address your complaint to the person below, who is a senior University of Warwick official entirely independent of this study:

Deputy Director/Head of Research Governance
Research and Impact Services
University House
University of Warwick
Coventry, CV4 8UW
researchgovernance@warwick.ac.uk
[redacted]

Who can I contact for further information?
If you have any questions about the study, or your involvement in it, either now or in the future, do please contact the study team using the details below:

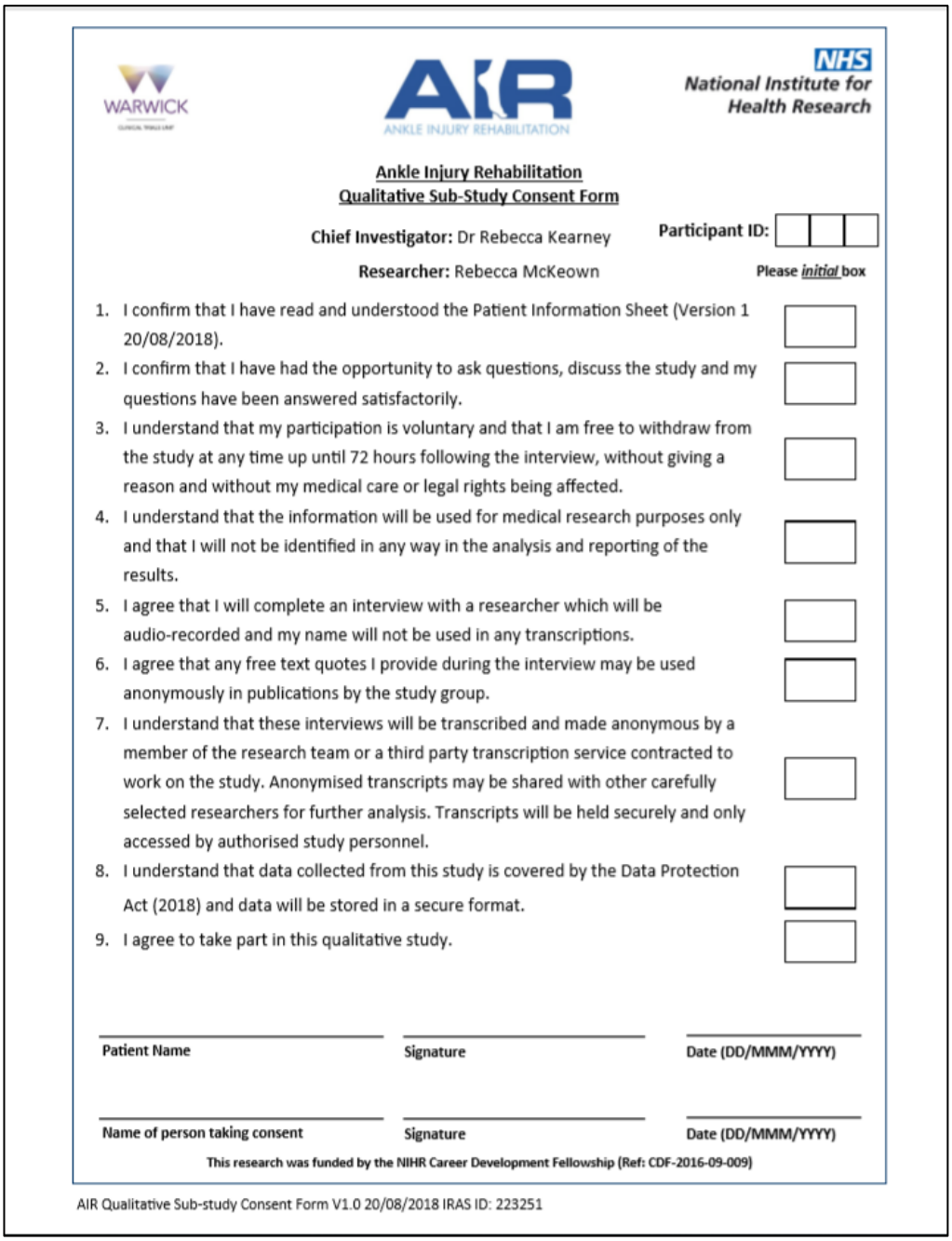
Rebecca McKeown, Trial Manager
Warwick Clinical Trials Unit, Warwick Medical School
Clinical Sciences Research Laboratories
Clinical Sciences Building
University Hospitals Coventry and Warwickshire
Coventry
CV2 2DX
air@warwick.ac.uk
[redacted]

Who is organising and paying for the study?
The study is being co-ordinated by the University of Warwick, and is being led by Dr Rebecca Kearney. The study is funded by the National Institute for Health Research – Career Development Fellowship Reference CDF-2016-09-009

Who has reviewed the study?
Any research that involves the NHS and patients is subject to review by an independent group of people, called a Research Ethics Committee. This committee is there to protect your interests. This study has been reviewed and given favourable opinion by West Midlands – Edgbaston Research Ethics Committee, Reference 17/WM/0239.

Thank you for taking time to read this information leaflet and for considering the study.

Participant consent form



Ankle Injury Rehabilitation
Qualitative Sub-Study Consent Form

Chief Investigator: Dr Rebecca Kearney

Researcher: Rebecca McKeown

Participant ID: ☐☐☐

Please *initial* box

1. I confirm that I have read and understood the Patient Information Sheet (Version 1 20/08/2018). ☐
2. I confirm that I have had the opportunity to ask questions, discuss the study and my questions have been answered satisfactorily. ☐
3. I understand that my participation is voluntary and that I am free to withdraw from the study at any time up until 72 hours following the interview, without giving a reason and without my medical care or legal rights being affected. ☐
4. I understand that the information will be used for medical research purposes only and that I will not be identified in any way in the analysis and reporting of the results. ☐
5. I agree that I will complete an interview with a researcher which will be audio-recorded and my name will not be used in any transcriptions. ☐
6. I agree that any free text quotes I provide during the interview may be used anonymously in publications by the study group. ☐
7. I understand that these interviews will be transcribed and made anonymous by a member of the research team or a third party transcription service contracted to work on the study. Anonymised transcripts may be shared with other carefully selected researchers for further analysis. Transcripts will be held securely and only accessed by authorised study personnel. ☐
8. I understand that data collected from this study is covered by the Data Protection Act (2018) and data will be stored in a secure format. ☐
9. I agree to take part in this qualitative study. ☐

Patient Name	Signature	Date (DD/MMM/YYYY)
Name of person taking consent	Signature	Date (DD/MMM/YYYY)

This research was funded by the NIHR Career Development Fellowship (Ref: CDF-2016-09-009)

Interview topic guide

Topic Guide – AIR Qualitative Sub-Study

"My name's Rebecca and I'm a researcher from the University of Warwick. We're conducting this research study linked to the AIR trial, which you're already participating in. You've been given an information sheet, the opportunity to ask questions and sign a consent form related to this project. I would like to remind you that you can withdraw at any time from this study and it will not affect you in any way. Today, during this interview I would like to ask you about what it's like to live with an ankle fracture in order to explore what factors are most important in your recovery. After that, I'd like to discuss one of the questionnaires we collect in the AIR trial. There are no right or wrong answers here, we would just like get understand your experiences. Can you please confirm for the recording that you understand this and are willing to continue?"

Part 1 – patient experiences

Could you start by telling me all about your ankle fracture?

Could you explain how your ankle fracture has impacted your day-today life?

How has your ankle fracture affected your walking?

Could you talk to me about the impact of your ankle fracture on your family life?

Could you explain what was most important to you when recovering from your ankle fracture?

What bothered you most throughout your recovery from your injury?

Did your ankle fracture affect your mood in any way?

How did your ankle fracture affect your work?

How did your ankle fracture affect your use of free time?

Probing questions for further discussion:

You mentioned that...Could you tell me more about that?

Can you explain that idea a bit more?

And how did that experience make you feel?

Could you explain a little bit more about your experience with…?

Part 2 – Relevance & comprehensiveness of OMAS content:

At this point, we would give the participant a copy of the constructs covered in the OMAS to review and discuss each item on the score:

Intro: "This is one of the questionnaires included in the AIR trial questionnaires. It's important to note that this questionnaire was not derived by myself or any member of the AIR Team. We just wanted to get your thoughts on the content of the questionnaire. What I would like to do is go through each item with you and discuss each item. Does that make sense?"

1. Pain

2. Stiffness

3. Swelling

4. Climbing Stairs

5. Running

6. Jumping

7. Squatting

8. Supports and walking aids

9. Work and ADL's

For each item, ask the following: Could you explain how relevant this aspect was in your recovery?

Do you think it's an important thing to ask you about whilst recovering from an ankle fracture?

Could you comment on the available response options?

Is there anything you don't understand in this question?

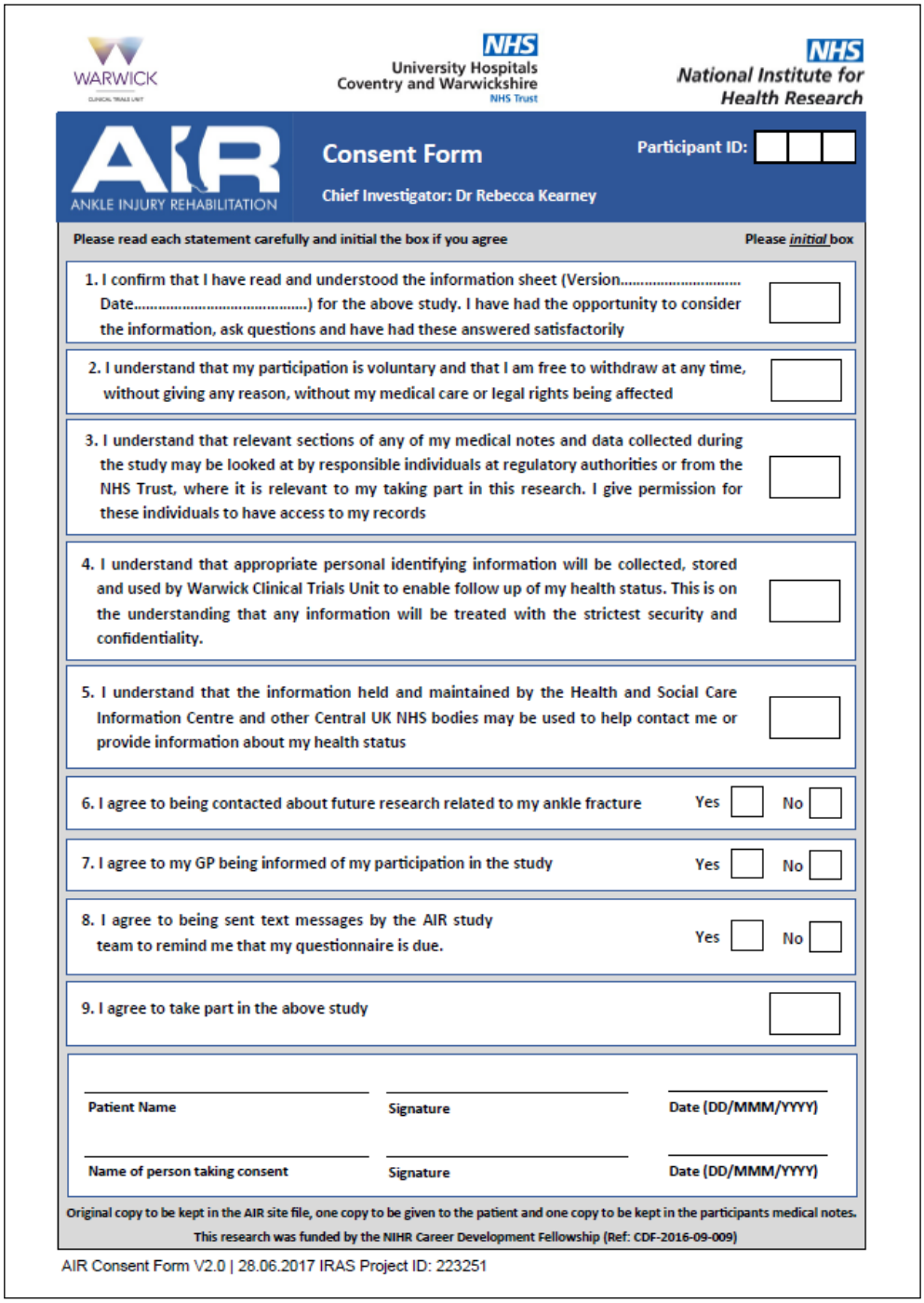
Comprehensiveness – Is there anything which you feel was important in your recovery which has been covered in this questionnaire?

"Is there anything else you would like to add or discuss?"

"Thank you very much for your time, I really appreciate it."

9.7 Appendix 7 – chapter 6: AIR trial documents

AIR trial participant consent form



WARWICK CLINICAL TRIALS UNIT

NHS University Hospitals Coventry and Warwickshire NHS Trust

NHS National Institute for Health Research

AIR ANKLE INJURY REHABILITATION

Consent Form

Chief Investigator: Dr Rebecca Kearney

Participant ID: ☐☐☐

Please read each statement carefully and initial the box if you agree	Please *initial* box
1. I confirm that I have read and understood the information sheet (Version..... Date.....) for the above study. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily	☐
2. I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason, without my medical care or legal rights being affected	☐
3. I understand that relevant sections of any of my medical notes and data collected during the study may be looked at by responsible individuals at regulatory authorities or from the NHS Trust, where it is relevant to my taking part in this research. I give permission for these individuals to have access to my records	☐
4. I understand that appropriate personal identifying information will be collected, stored and used by Warwick Clinical Trials Unit to enable follow up of my health status. This is on the understanding that any information will be treated with the strictest security and confidentiality.	☐
5. I understand that the information held and maintained by the Health and Social Care Information Centre and other Central UK NHS bodies may be used to help contact me or provide information about my health status	☐
6. I agree to being contacted about future research related to my ankle fracture	Yes ☐ No ☐
7. I agree to my GP being informed of my participation in the study	Yes ☐ No ☐
8. I agree to being sent text messages by the AIR study team to remind me that my questionnaire is due.	Yes ☐ No ☐
9. I agree to take part in the above study	☐

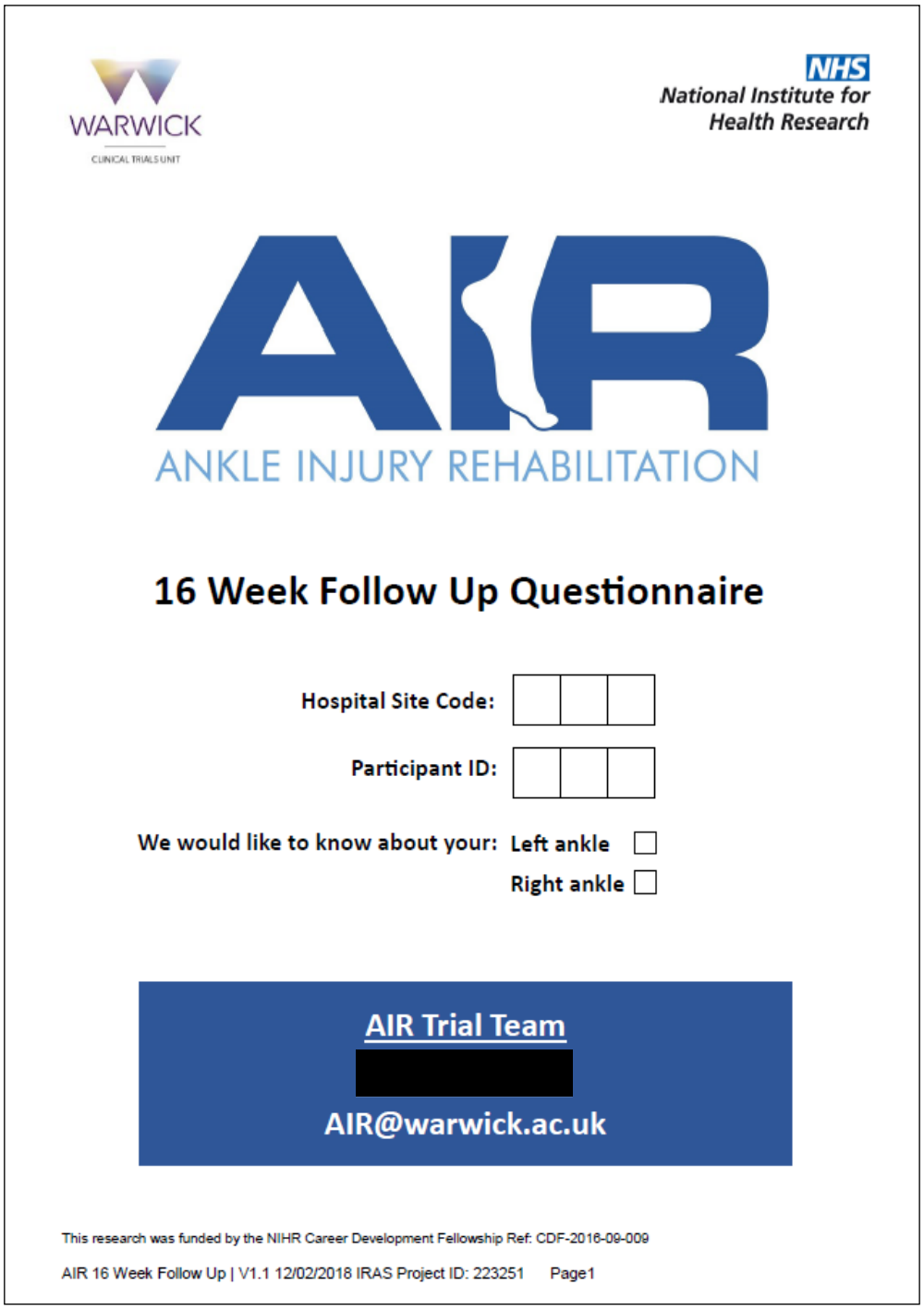
Patient Name	Signature	Date (DD/MMM/YYYY)
Name of person taking consent	Signature	Date (DD/MMM/YYYY)

Original copy to be kept in the AIR site file, one copy to be given to the patient and one copy to be kept in the participants medical notes.

This research was funded by the NIHR Career Development Fellowship (Ref: CDF-2016-09-009)

AIR Consent Form V2.0 | 28.06.2017 IRAS Project ID: 223251

AIR trial 16 week case report form showing relevant questionnaires



WARWICK
CLINICAL TRIALS UNIT

NHS
National Institute for Health Research

AIR
ANKLE INJURY REHABILITATION

16 Week Follow Up Questionnaire

Hospital Site Code: | | | |

Participant ID: | | | |

We would like to know about your: Left ankle ☐
Right ankle ☐

AIR Trial Team
[redacted]
AIR@warwick.ac.uk

This research was funded by the NIHR Career Development Fellowship Ref: CDF-2016-09-009

AIR 16 Week Follow Up | V1.1 12/02/2018 IRAS Project ID: 223251 Page1

ANKLE INJURY REHABILITATION

Instructions for Participants

The following questions will help us understand about your pain, function and how easily you are able to complete your day to day activities.

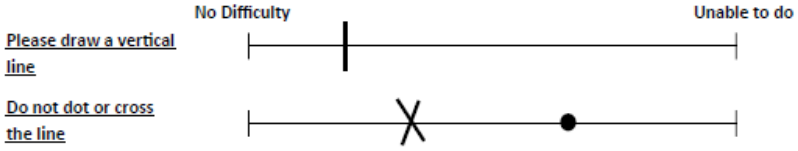
Please answer all questions. Although it may seem that some questions are asked more than once, it's still important that you answer all of them.

Please use a **BLUE** or **BLACK** pen to complete. Please **do not** use a pencil.

Please **do not** sign this form or add your name to it. If you make a mistake, please cross it through with a single line, initial and date it. Add the correct answer next to this.

If you have injured both of your ankles, please answer the questions based upon the ankle which has been included in the trial, indicated on the front of this booklet.

Some questions will require you to indicate how easy or difficult it is for you to carry out certain activities using a visual analogue scale, like the example below.



Please draw a vertical line on the scale, to indicate how difficult you find the specific activity. Do not dot or cross the line. If you don't do a task or didn't in the required time frame, then please guess how well you could do it.

Once complete, please return to the AIR Trial Office in the Freepost envelope provided.
If there are any questions which you don't understand or need help with, please speak call the AIR office on [redacted]

AIR
ANKLE INJURY REHABILITATION

Section 1 - Background

1. Please indicate the date on which you are completing this questionnaire:

Date of completion: D D / M M M / Y Y Y Y

2. Are your contact details likely to change within the next three months? Yes ☐ No ☐

If yes, please complete the "Change of Participant Details Form" located on page 22.

3. Did you have surgery for your ankle fracture? Yes ☐ No ☐

If YES, please complete Section 2 "Wound Complications"

If NO, please skip to Section 3, "Length of Stay"

Section 2 - Wound Complications (if you have had an operation)

1. In the last 6 weeks, have you had any issues with your wound healing? Yes ☐ No ☐

If YES, please select the types of issues from the list below

If NO, please skip to Section 3, "Length of Stay"

Please select all that apply from the list below:

	Yes	No
a) Has there been any discharge or fluid leaking from any part of the wound?	☐	☐
If yes, was it: Clear or blood stained?	☐	☐
Yellow/green pus?	☐	☐
b) Has there been increasing pain or discomfort in the area around the wound?	☐	☐
c) Has the area around the wound become increasingly swollen?	☐	☐
d) Have the edges of any part of the wound separated or gaped open?	☐	☐

2. Please document the date that you noticed these symptoms: (if you are unsure of the exact date, please give the month and year) D D - M M M - Y Y Y Y

3. Did any health care worker take a sample from the wound to send it to the laboratory? Yes ☐ No ☐

ANKLE INJURY REHABILITATION

Section 7 - Olerud Molander Ankle Score

By crossing ☒ one box in each group below, please indicate which statements best describe you health state **TODAY**.

1. Pain
- ☐ None
- ☐ While walking on uneven surfaces
- ☐ While walking on even surface outdoors
- ☐ While walking indoors
- ☐ Constant and severe

2. Stiffness
- ☐ None
- ☐ Stiffness

3. Swelling
- ☐ None
- ☐ Only evenings
- ☐ Constant

4. Stair climbing
- ☐ No problems
- ☐ Impaired
- ☐ Impossible

5. Running
- ☐ Possible
- ☐ Impossible

6. Jumping
- ☐ Possible
- ☐ Impossible

7. Squatting
- ☐ No problem
- ☐ Impossible

8. Supports
- ☐ None
- ☐ Taping, wrapping
- ☐ Stick or crutch

9. Work, activities of daily life
- ☐ Same as before injury
- ☐ Loss of tempo
- ☐ Changed to a simpler job/part-time work
- ☐ Severely impaired work capacity

ANKLE INJURY REHABILITATION

Section 8 - Global Impression of Change

1. Please indicate how your ankle is compared to how it was previously.

Over the past 6 weeks, since my last questionnaire, my ankle is.....

- ☐ Very much improved
- ☐ Much improved
- ☐ Improved minimally
- ☐ No change
- ☐ Minimally worse
- ☐ Much worse
- ☐ Very much worse

ANKLE INJURY REHABILITATION

Section 9 - Manchester-Oxford Foot and Ankle Questionnaire

Please tick one box for each statement.

1. During the past 4 weeks, this has applied to me:

I have pain in my foot/ankle

None of the time	Rarely	Some of the time	Most of the time	All of the time
☐	☐	☐	☐	☐

2. During the past 4 weeks, this has applied to me:

I avoid walking long distances because of the pain in my foot/ankle

None of the time	Rarely	Some of the time	Most of the time	All of the time
☐	☐	☐	☐	☐

3. During the past 4 weeks, this has applied to me:

I change the way I walk due to pain in my foot/ankle

None of the time	Rarely	Some of the time	Most of the time	All of the time
☐	☐	☐	☐	☐

4. During the past 4 weeks, this has applied to me:

I walk slowly because of the pain in my foot/ankle

None of the time	Rarely	Some of the time	Most of the time	All of the time
☐	☐	☐	☐	☐

5. During the past 4 weeks, this has applied to me:

I have to stop and rest my foot/ankle because of the pain

None of the time	Rarely	Some of the time	Most of the time	All of the time
☐	☐	☐	☐	☐

AIR
ANKLE INJURY REHABILITATION

6. During the past 4 weeks, this has applied to me:

I avoid some hard or rough surfaces because of pain my foot/ankle

None of the time	Rarely	Some of the time	Most of the time	All of the time
☐	☐	☐	☐	☐

7. During the past 4 weeks, this has applied to me:

I avoid standing for a long time because of the pain in my foot/ankle

None of the time	Rarely	Some of the time	Most of the time	All of the time
☐	☐	☐	☐	☐

8. During the past 4 weeks, this has applied to me:

I catch the bus or use the car instead of walking because of the pain in my foot/ankle

None of the time	Rarely	Some of the time	Most of the time	All of the time
☐	☐	☐	☐	☐

9. During the past 4 weeks, this has applied to me:

I feel self-conscious about my foot/ankle

None of the time	Rarely	Some of the time	Most of the time	All of the time
☐	☐	☐	☐	☐

10. During the past 4 weeks, this has applied to me:

I feel self-conscious about the shoes I have to wear

None of the time	Rarely	Some of the time	Most of the time	All of the time
☐	☐	☐	☐	☐

11. During the past 4 weeks, this has applied to me:

The pain in my foot/ankle is more painful in the evening

None of the time	Rarely	Some of the time	Most of the time	All of the time
☐	☐	☐	☐	☐

ANKLE INJURY REHABILITATION

12. During the past 4 weeks, this has applied to me:

I get shooting pains in my foot/ankle

None of the time	Rarely	Some of the time	Most of the time	All of the time
☐	☐	☐	☐	☐

13. During the past 4 weeks, this has applied to me:

The pain in my foot/ankle prevents me from carrying out my work/everyday activities

None of the time	Rarely	Some of the time	Most of the time	All of the time
☐	☐	☐	☐	☐

14. During the past 4 weeks, this has applied to me:

I am unable to do all my social or recreational activities because of pain in my foot/ankle

None of the time	Rarely	Some of the time	Most of the time	All of the time
☐	☐	☐	☐	☐

15. During the past 4 weeks...

How would you describe the pain you usually have in your foot/ankle?

None	Very mild	Mild	Moderate	Severe
☐	☐	☐	☐	☐

16. During the past 4 weeks...

Have you been troubled by pain from your foot/ankle in bed at night?

No nights	Only 1 or 2 nights	Some nights	Most nights	Every night
☐	☐	☐	☐	☐

ANKLE INJURY REHABILITATION

Section 10 - Quality of Life

Under each heading please tick the ONE box that best describes your health **TODAY**.

1. MOBILITY

- I have no problems in walking about ☐
- I have slight problems in walking about ☐
- I have moderate problems in walking about ☐
- I have severe problems in walking about ☐
- I am unable to walk about ☐

2. SELF-CARE

- I have no problems with washing or dressing myself ☐
- I have slight problems washing or dressing myself ☐
- I have moderate problems washing or dressing myself ☐
- I have severe problems washing and dressing myself ☐
- I am unable to wash or dress myself ☐

3. USUAL ACTIVITIES (e.g. work, study, housework, family or leisure activities)

- I have no problems with performing my usual activities ☐
- I have slight problems with performing my usual activities ☐
- I have moderate problems with performing my usual activities ☐
- I have severe problems with performing my usual activities ☐
- I am unable to perform my usual activities ☐

4. PAIN/DISCOMFORT

- I have no pain or discomfort ☐
- I have slight pain or discomfort ☐
- I have moderate pain or discomfort ☐
- I have severe pain or discomfort ☐
- I have extreme pain or discomfort ☐

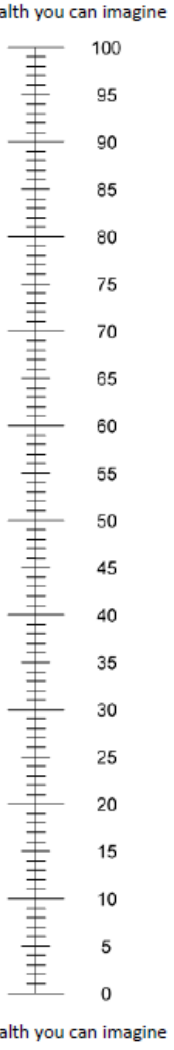
5. ANXIETY/DEPRESSION

- I am not anxious or depressed ☐
- I am slightly anxious or depressed ☐
- I am moderately anxious or depressed ☐
- I am severely anxious or depressed ☐
- I am extremely anxious or depressed ☐

ANKLE INJURY REHABILITATION

- We would like to know how good or bad your health is **TODAY**.
- This scale is numbered from 0 to 100
- 100 means the best health you can imagine.
- 0 means the worst health you can imagine.
- Mark an X on the scale to indicate how your

YOUR HEALTH **TODAY**= ☐



ANKLE INJURY REHABILITATION

Section 11 - Disability Rating Index

How do you manage the following activities TODAY?

After each question, please mark ONE POINT on the line

Without difficulty — Not at all

	Office use:
Dressing (without help)	
Out-door walks	
Climbing Stairs	
Sitting for a long time	
Standing bent over a sink	
Carrying a bag	
Making a bed	
Running	
Light work	
Heavy work	
Lifting heavy objects	
Participating in exercise/sports	

ANKLE INJURY REHABILITATION

AIR Trial Team
Warwick Clinical Trials Unit
Warwick Medical School
The University of Warwick
Coventry
CV4 7AL
AIR@warwick.ac.uk
Tel: [redacted]