



An evaluation of the measurement properties of the Olerud Molander Ankle Score in adults with an ankle fracture

Rebecca McKeown^{b,*}, Helen Parsons^b, David R. Ellard^a, Rebecca S. Kearney^a

^a Warwick Clinical Trials Unit, Warwick Medical School, University of Warwick and University Hospitals Coventry and Warwickshire, United Kingdom

^b Warwick Clinical Trials Unit, Warwick Medical School, University of Warwick, United Kingdom

Abstract

Objectives The aim of this study is to evaluate the measurement properties of the Olerud Molander Ankle Score in adults with an ankle fracture.

Methods Patients completed outcome measure questionnaires at baseline, six, 10- and 16-weeks postinjury as part of an ongoing clinical trial on ankle fracture rehabilitation. The internal consistency, convergent validity, structural validity and interpretability of the Olerud Molander Ankle Score was assessed. This was achieved through using the respective analysis methods of Cronbach's alpha, correlation coefficients, principal component analysis, evaluation of floor and ceiling scores and estimation of the minimally important change using anchor-based methods.

Results The Olerud Molander Ankle Score showed adequate convergent validity against hypotheses set in relation to scores of comparator instruments. Principal component analysis demonstrated that the measure has two subscales: ankle function and ankle symptoms. The internal consistency of the measure and the ankle function subscale was sufficient, but inconclusive for the ankle symptoms subscale. There were no floor and ceiling effects present within the scores and the estimated minimally important change was 9.7 points.

Conclusion The Olerud Molander Ankle Score demonstrates sufficient measurement properties and is likely to be primarily measuring the construct of patient reported function following ankle fracture. Further research should evaluate the relevance of other domains to individuals recovering from an ankle fracture, such as social participation and psychological wellbeing. The development of a core outcome set would be advantageous to standardise outcome measurement collection in this area.

© 2021 The Author(s). Published by Elsevier Ltd on behalf of Chartered Society of Physiotherapy. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Keywords: Ankle fracture; Rehabilitation; Patient reported outcome measures; Quality of life; Measurement properties; Validity

Contribution of the paper

This study evaluates the measurement properties of the Olerud Molander Ankle Score in adults with an ankle fracture within the context of a clinical trial for rehabilitation strategies in this patient population

- The score demonstrated sufficient construct validity, internal consistency, no floor or ceiling effects and the minimally important change was estimated at 9.7 points.

- Factor analysis showed two subscales present; ankle function and ankle symptoms, with the latter subscale only containing three items. The ankle function subscale is likely to be internally consistent, but results for this are inconclusive for the ankle symptoms subscale as it contains too few items.
- The score is likely measuring the construct of patient reported ankle function following ankle fracture and should therefore be used when this specific construct of interest is under evaluation.
- Other measures which may be more holistic in capturing the multifaceted, patient important recovery from ankle

* Corresponding author.

E-mail addresses: r.mckeown.1@warwick.ac.uk (R. McKeown), h.parsons@warwick.ac.uk (H. Parsons), d.r.ellard@warwick.ac.uk (D.R. Ellard), r.s.kearney@warwick.ac.uk (R.S. Kearney).

fracture may be required to measure outcome in studies whereby function is not the sole construct of interest.

Introduction

Ankle fracture is the fifth most common fracture affecting adults and contributes to the increasing socioeconomic burden of lower limb fractures in the UK [1]. The injury typically affects younger males and older females and the incidence is increasing, likely secondary to the increasing proportion of older individuals in the population [2]. As recommended by authors of a Cochrane review into rehabilitation for adults with ankle fracture, more high quality randomised controlled trials (RCTs) are required to answer important questions in this research area [3]. However, high quality results of RCTs are dependent on high quality methods of outcome assessments for the population of interest.

Outcome measures can be categorised as clinician reported or patient reported. Over recent decades, there has been an increasing trend towards the use of patient reported outcome measures (PROMs) in clinical research studies [4]. However, the outcomes must demonstrate adequate measurement properties and the use of outcome measures with insufficient validity and reliability in clinical trials is unethical and a waste of resource [5]. The Olerud Molander Ankle Score (OMAS) was developed in 1984 and is a popular PROM used in clinical trials for this population [6,7], however results of some systematic reviews have highlighted concerns regarding the lack of evidence to demonstrate sufficient measurement properties of this instrument [8,9]. There are several studies which assess the measurement properties of OMAS in different language versions [10–13] but none which do so in the English language version.

The aim of this study is to assess the structural and convergent validity, internal consistency and interpretability of the OMAS in a population of adults recovering from an ankle fracture. The construct of interest is patient reported outcome following ankle fracture and the context of use is outcome measurement in RCTs.

Ethical approval

Ethical approval for this validation project was sought and gained from the West Midlands Edgbaston NHS Research Ethics Committee (Reference 17/WM/0239) on 30/04/2019.

Patients and methods

This project uses secondary analysis of pre-existing data collected as part of the AIR trial; a UK based multicentre RCT across 20 NHS trusts. The AIR trial compared plaster cast to functional brace in the management of adults with a closed ankle fracture, managed with or without surgery (ISRCTN15537280) [14]. Participants were aged 18 years or

over with a closed ankle fracture for which the treating clinician felt plaster cast was a reasonable management option. Detailed eligibility criteria can be found in the supplementary files.

Data collection

Data were collected at baseline, 6, 10 and 16 weeks postinjury [14]. Baseline data and questionnaires were collected in fracture clinic prior to randomisation. Follow up questionnaires were posted to participants and returned to the trial office using freepost envelopes. In the instances where postal questionnaires were not returned, the trial team telephoned the participants to obtain the data over the phone within the relevant time frame. The questionnaires collected were OMAS [6], EuroQol EQ-5D-5L (EQ-5D) [15], the Manchester Oxford Foot and Ankle Questionnaire (MOXFQ) [16], the Disability Rating Index (DRI) [17] and a Global Impression of Change (GIC) score [4]. The MOXFQ was collected at baseline and 16 weeks only and GIC scores were collected at 16 weeks only. This set of outcome measures were collected to measure outcome within the trial and support a comparative evaluation of OMAS.

The OMAS is a nine item questionnaire to assess patient reported outcome following a fracture of the ankle [6]. It contains single response, multiple choice questions and scores from 0 to 100, with higher scores indicating better outcomes. It contains no subscales and is reported as a single index score with no indicated recall period.

The MOXFQ is a 16-item questionnaire to assess recovery in individuals with foot and ankle conditions [16]. It contains three subscales of walking-standing, pain and social interaction. The scores are 0–100 with lower scores indicating better outcomes.

The DRI is a 12-item questionnaire which assesses levels of disability resulting from musculoskeletal conditions of the lower limb [17]. It comprises Likert scale questions for individuals to rate how easy or difficult functional tasks of daily living are. Scores are 0–100 with lower scores indicating higher levels of disability.

The EQ-5D is a preference-based measure of health-related quality of life which supports cost utility analysis. It contains domains of pain, usual activities, anxiety and depression, self-care, mobility all which contribute to the calculation of a index score. The final question is an overall assessment of quality-of-life (visual analogue scale) on the day of response, ranging from 0 to 100, with 100 being perfect health. The UK tariff crosswalk value set was used to calculate the index score, ranging from –0.594 to 1 [15]. Individual item scores were used when using the domains within the EQ-5D-5L for convergent validity assessments.

The GIC score is a seven-point Likert scale relating to the degree to which an individual feels they have improved in a specified time frame, ranging from very much worse to very much improved [4] and is shown in Table 2. This question was included in the 16-week questionnaires, asking

individuals to specify how much their ankle had improved since the previous questionnaire (10 weeks).

Sample size

The sample size for exploratory factor analysis should be 10 times the number of items in the questionnaire or greater than 100 [5]. There are nine items in the OMAS, therefore data for a minimum of 100 patients are required to complete exploratory factor analysis. To ensure the effects of repeated measures were avoided for both convergent validity and factor analysis, each participant was only entered into these analyses once. A single time point was selected for each participant using a random number generator and the questionnaire for the random time point was included in the analysis. If a questionnaire did not exist for the randomly allocated time point, then this participant was not entered into the analysis.

Data management

Data were inputted into a secure online database accessible only to authorised trial personnel. Quality assurance processes were followed to ensure minimisation of error during data entry. All data were retrieved from the online database and pseudonymised by allocating a trial ID number by the trial statistician (HP). Dates of birth were converted to age at randomisation and no information on randomising site was retrieved to protect participant confidentiality. No information on allocated interventions was retrieved. Data were stored on secure university servers using password protected files only available to authorised trial personnel. In the instances where an item of data was missing from a questionnaire, we used complete case analysis, thus excluding questionnaires which contained one or more items of missing data. Mean imputation would have been the preferred method of dealing with missing data, however individual items within OMAS are not worth an equal number of points, therefore mean imputation is not possible with this PROM. The Data Protection Act (2018) was adhered to in relation to the processing of all data. All data analysis was completed using IBM SPSS Statistics (Version 24).

Data analysis

Descriptive statistics will be used to present the participant demographics and injury information of the individuals included in this study. Data presented will be age, sex, side of injury, mechanism of injury, fracture classification and fracture management. The fracture classification used to present injury type will be the Weber classification which is widely used clinically [18].

Convergent validity refers to how the scores of a PROM perform in relation to scores of another PROM [19]. Convergent validity was assessed using hypotheses testing for expected associations of OMAS scores with those of com-

Table 1
Hypotheses set a priori for assessment of convergent validity.

Hypothesis	Expected correlation
Scores of OMAS will be highly negatively associated with scores of MOXFQ and the domains of MOXFQ of walking-standing, pain and social interaction ($\times 4$ hypotheses)	$r \leq -0.7$
Scores of OMAS will be moderately negatively associated with scores of DRI ($\times 1$ hypothesis)	$r \leq -0.5$
Scores of OMAS will be moderately positively associated with EQ-5D overall scores and EQ-5D visual analogue scale domain ($\times 2$ hypotheses)	$r \geq 0.5$
Scores of OMAS will be moderately negatively associated with EQ-5D domains of mobility, pain & discomfort, anxiety & depression, self-care and usual activities ($\times 5$ hypotheses)	$r \leq -0.5$

Key: r = Pearson's correlation.

parator instruments. Scores of OMAS will be correlated with scores of the comparator instruments and their subscales collected in the trial. As the data were normally distributed, Pearson's correlation coefficient was used to correlate the scores. There are 12 hypotheses set a priori for this analysis, outlined in Table 1. A total of 75% hypotheses should be met to achieve sufficient convergent validity of OMAS [20], therefore in this case, nine of these hypotheses should be met for OMAS to demonstrate sufficient convergent validity.

Structural validity refers to whether the PROM is an adequate reflection of the dimensionality of the construct it intends to measure with regard to underlying subscales and components of the score [19]. To our knowledge, there is no evidence regarding the internal structure or underlying subscales of the OMAS [6]. We explored the dimensionality using principal component analysis (PCA). Orthogonal (varimax) rotation was used and components which had eigenvalues greater than 1 were extracted. Item loadings of a magnitude of 0.45 and above were considered sufficient for an item to be included within a component.

Internal consistency is the degree to which items contained within a PROM are interrelated with one another [19]. The internal consistency of the measure was assessed using Cronbach's alpha. This was calculated for the total score and also each subscale found within the score following findings of PCA described above, as recommended by the COSMIN group [21]. Acceptable scores were $\alpha = 0.70$ – 0.95 inclusive. Cronbach's alpha if item deleted was also calculated, whereby the result of the analysis is presented in the cases of each item being deleted. This can indicate where items might be redundant and therefore not contributing to the overall score. Item redundancy was defined as Cronbach's alpha remaining constant or increasing when an item is removed from the analysis.

Interpretability is the clinical or qualitative meaning that one can derive from scores, or changes in scores, of a PROM [19]. Interpretability was assessed by examining floor and ceiling (edge) effects within the score and the minimally important change (MIC) of the score. Edge effects are impor-

tant because their presence can limit the scores ability to detect changes in individuals who are either at the highest or lowest points on the scale. An edge effect was defined as present if 15% or more of the scores were at the lowest (0) and highest (100) level and this was calculated overall and by time point. This threshold has been recommended by the COSMIN research group [22] and used in other studies of a similar nature [23,24]. An assessment of the MIC of OMAS was made using an anchor-based method. This was assessed by calculating the mean change in OMAS of respondents who reported to be “minimally improved” on the GIC score response between the 10 and 16-week follow up questionnaires.

Results

Participant demographics and injury information

Data for 620 participants were obtained from the trial database on 26th July 2019. The mean age was 46 years (standard deviation of 16.7) with a minimum of 18 years and maximum of 94 years. Table 2 shows the injury and demographic information of the sample. Follow up rates for the trial by time point can be found in supplementary file C.

Convergent validity

Table 3 shows correlations of scores of OMAS with scores of the comparator instruments and domains of instruments. The lower sample sizes for the MOXFQ outcomes are because this outcome was only collected at baseline at 16 weeks.

Structural validity

The PCA was completed using a sample of 438 study participants, comprised of 132 baseline scores, 110 six-week scores, 100 10-week scores and 96 16-week scores. The mean OMAS score was 43 with a standard deviation of 26.36. Fig. 2 (in supplementary files) shows the scree plot with a reference line at eigenvalue = 1. Two components can be seen to have eigenvalues greater than 1, hence two components were extracted. Fig. 1 shows the component plot in the rotated space for these two components. Table 4 shows results for the extracted components.

Component 1 comprised six items; squatting, jumping, climbing stairs, running, work/activities of daily life and supports. This subscale accounted for 45% of the variance within scores. Due to the type of items contained within it, this was identified as the ankle function subscale. Component 2 comprised three items: pain, stiffness and swelling. This accounted for 12% of the total variance within scores and was identified as the ankle symptoms subscale.

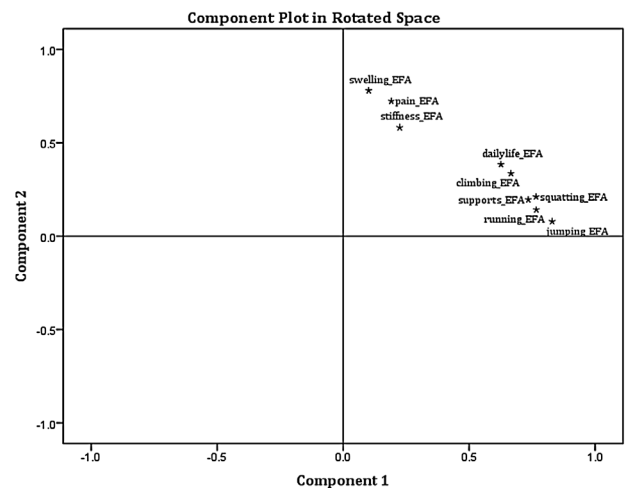


Fig. 1. Component plot in the rotated space.

Key: swelling_EFA – swelling item, pain_EFA – pain item, stiffness_EFA – stiffness item, dailylife_EFA – work, activities of daily life item, climbing_EFA – climbing stairs item, supports_EFA – supports item, squatting_EFA – squatting item, running_EFA – running item, jumping_EFA – jumping item.

Internal consistency

Results of Cronbach’s alpha for OMAS was $\alpha = 0.76$. Table 4 shows results of Cronbach’s alpha if item deleted. Possible item redundancy is shown as the scale internal consistency increases when the pain item is deleted. Results of Cronbach’s alpha for the ankle function subscale was $\alpha = 0.76$ and $\alpha = 0.46$ for the ankle symptoms subscale, however, this was completed on a subscale containing only 3-items, so should be interpreted with caution.

Floor and ceiling effects

Table 5 shows the frequency and percentage of highest and lowest possible scores at each time point and overall. At baseline, 11% of responses received were at the lowest level and at 16 weeks 8% of responses were at the highest level. Overall, this was 4% at the lowest score and 2% at the highest score overall.

Minimally important change

The mean change in OMAS scores for each GIC group are shown in Table 6. The mean change in OMAS scores in the minimally improved group, and therefore the MIC in this context, is 9.7 points.

Discussion

The aim of this research was to explore the measurement properties of OMAS in a sample of individuals with ankle fracture, in the context of a multicentre RCT. Results of RCTs can be combined to perform systematic reviews, the strongest

Table 2
Injury and demographic information of sample studied.

Descriptor		Frequency (n)	Percentage (%)
Side of injury	Right	298	48
	Left	303	49
	Missing	19	3
Sex	Male	275	44
	Female	345	56
	Low energy fall	381	62
	High energy fall	93	15
Mechanism of injury	Other	52	8
	Contact sport injury	46	7
	Road traffic accident	24	4
	Crush injury	2	0
	Missing	22	4
Fracture management	Operative	331	53
	Non-operative	289	47
	Weber A	38	6
	Weber B	409	66
Weber classification	Weber C	104	17
	N/A (not classifiable; no lateral malleolus fracture)	25	4
	Missing	44	7

Key: N/A = not applicable.

Table 3
Results of correlations for the assessment of convergent validity using hypotheses testing.

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	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-visual analogue scale	441	0.56	<0.01

Table 4
Results of the principal component analysis of OMAS.

Item on questionnaire (total points available)	Component 1 loadings of >0.45: ankle function subscale	Component 2 loadings of >0.45: ankle symptoms subscale	Item com- munity	Corrected item-total correlation	Cronbach's α if item deleted (whole scale $\alpha = 0.76$)
Pain (25)		0.72	0.56	0.47	0.77
Stiffness (10)		0.58	0.39	0.33	0.75
Swelling (10)		0.78	0.62	0.42	0.74
Climbing stairs (10)	0.67		0.56	0.62	0.72
Running (5)	0.77		0.61	0.58	0.75
Jumping (5)	0.83		0.69	0.60	0.74
Squatting (5)	0.77		0.63	0.65	0.73
Supports (10)	0.74		0.58	0.61	0.71
Work, activities of daily life (20)	0.63		0.54	0.61	0.71

level of evidence, to enable high-quality, evidence-based rehabilitation protocols to be developed. Results presented here demonstrate there are two subscales within the score: ankle function and ankle symptoms. The Cronbach's alpha analysis for the overall PROM show that it is likely internally

consistent overall, along with the ankle function subscale. However, the result for Cronbach's alpha of the ankle symptoms subscale is inconclusive, as it contains only three items and alpha is dependent upon the number of items within a score [25]. Furthermore, the pain item contained within

Table 5
Results of the assessment of floor and ceiling effects of OMAS.

Time point	Total responses at time point (n)	Frequency and percent of 0 scores n (%)	Frequency and percent of 100 scores n (%)
Baseline	545	57 (11)	1 (0)
6 weeks	423	5 (1)	2 (1)
10 weeks	389	0 (0)	7 (2)
16 weeks	410	0 (0)	32 (8)
Total	1767	62 (4)	42 (2)

Table 6
Mean change in OMAS scores for each GIC response group.

GIC response at 16 weeks	Frequency (n=)	Mean change in OMAS (standard deviation where applicable)
Very much improved	62	20.7 (14.8)
Much improved	156	15.6 (16.9)
Improved minimally	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

the ankle symptom subscale shows a degree of item redundancy. Results of the component analysis and the possible item redundancy of the pain item indicates that OMAS may predominantly be measuring the construct of patient reported function following ankle fracture.

Scores of OMAS demonstrate adequate convergent validity in relation to scores of comparator PROMs. The only association which did not meet the *a priori* hypothesis was for EQ-5D anxiety and depression domain, which is interesting because it might be expected that scores of OMAS may be associated with this domain. For example, if a person has poor ankle function, this could cause reduction in activity and social participation which could result in symptoms of anxiety and depression. Indeed, research shows that symptoms of anxiety and depression can affect individuals who have sustained an ankle fracture [26,27]. However, these results show a low association between these two scores, indicating that OMAS is not capturing this domain which can affect this population. It is likely that OMAS does not capture the holistic concept of recovery from ankle fracture for this reason.

We found no evidence of edge effects in the scores of the sample included here. The MIC is consistent with those used in RCTs in this research area [14,28,29] and will be helpful in enabling accurate sample size calculations to be made in trials with this patient population. Other authors have assessed the measurement properties of OMAS in different languages. One study translated OMAS into Turkish and evaluated the measurement properties, finding that OMAS showed strong correlations with the comparator scores used and adequate internal consistency [10].

Another author group which also assessed a Turkish version of OMAS found adequate internal consistency of scores

and positive associations with scores of the Foot and Ankle Ability Measure and SF-12 [13]. However, they also found a higher ceiling effect than results shown here, possibly because of the longer follow up period studied (4.3 years). This shows that, whilst ceiling effects are likely not an issue up to 16 weeks postinjury, they may become an issue at later follow up stages. Another study assessed the MIC of the Swedish version of OMAS, findings results of 12 points, which is similar to those found here [11], along with acceptable internal consistency.

The strengths of this study include the large sample size which met the criteria outlined for the principal component analysis. Furthermore, the sample was inclusive of both operatively and non-operatively managed patients, which helps ensure results are generalisable. Furthermore, this is overdue validation evidence for a very commonly used PROM, with comparatively little evidence for its functioning in the patient population it is intended for, particularly in the English language version of the instrument.

The main limitation of this study is that it was embedded in an existing RCT, meaning we were restricted to the sample of participants within the trial, who may not be representative of the entire population. To be eligible for the trial, individuals had to be suitable for plaster cast treatment. This means, for example, the majority of minor, avulsion type fractures will have been excluded from the sample because they are not suitable for a plaster cast. Therefore, we cannot be sure that these results would apply to individuals with this type of ankle fracture. Another limitation is that of missing data and missing questionnaires; we cannot be sure that data is missing completely at random, which will have introduced bias.

We were also limited to the specific data which was collected in the clinical trial. For example, the assessment of interpretability was based upon the 10–16-week time frame, which is the period for which this question was asked during the trial. This meant we were limited to assessing the interpretability of the score during this particular time frame and were unable to assess this at earlier to later stages in recovery. Furthermore, we were unable to add in additional questionnaires to assess test-retest reliability and interpretability at different time-points in an individual's recovery, because the participants were already completing a significant number of questionnaires as part of the trial. A further limitation is the lack of evidence for the comparator outcome measures in the specific population of adults with an ankle fracture.

Future research in this area should focus on ascertaining the measurement properties of OMAS for individuals with fractures which may not have been included in this sample, such as simple avulsion fractures, which would not be eligible for this study. The measurement properties of the score at later time points would also be useful, to ensure that the measure is has sufficient measurement properties for studies which use a longer follow up period.

Researchers designing studies to evaluate interventions and rehabilitation protocols for individuals with an ankle fracture should be mindful that the OMAS is likely a measure of patient reported ankle function and symptoms and probably does not capture a complete biopsychosocial construct of patient outcome following this injury. The development of a core outcome set for this population would be advantageous and findings presented here can aid with the decision on whether OMAS should be included in this [30]. A core outcome set would ensure that high quality meta-analyses can be performed on RCT results to facilitate the development of effective and evidence-based rehabilitation protocols for individuals with an ankle fracture.

Ethical approval

Ethical approval for this validation project was sought and gained from the West Midlands Edgbaston NHS Research Ethics Committee (Reference 17/WM/0239) on 30/04/2019.

Funding statement

The lead researcher is funded by a National Institute for Health Research (NIHR) Career Development Fellowship (Reference CDF-2016-09-009) for this research project.

This publication presents independent research funded by the NIHR. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care.

Conflicts of interest

RM, HP and DRE confirm they have no conflicts of interest. RSK is a member of the UK NIHR HTA CET board, NIHR ICA Doctoral panel and previous member of the NIHR RfPB board. RSK has been awarded current and previous NIHR research grants.

Acknowledgements

We would like to thank the participants of the trial for taking part in this study. We would also like to thank the AIR trial management group for their help with this work.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at <https://doi.org/10.1016/j.physio.2021.03.015>.

References

- [1] Court-Brown C, McQueen MM, Tornetta P, Einhorn TA. *Epidemiology*. In: Court-Brown C, McQueen MM, Tornetta P, Einhorn TA, editors. *Trauma*. Philadelphia: Wolters Kluwer Health; 2005. p. 6.
- [2] Court-Brown CM, Duckworth AD, Clement ND, McQueen MM. Fractures in older adults. A view of the future? *Injury* 2018;49(12):2161–6.
- [3] Lin CC, Donkers NA, Refshauge KM, Beckenkamp PR, Khera K, Moseley AM. Rehabilitation for ankle fractures in adults. *Cochr Datab Syst Rev* 2012;(11). N.PAG-N.PAG.
- [4] Fitzpatrick R, Davey C, Buxton M, Jones D. Evaluating patient-based outcome measures for use in clinical trials. *Health Technol Assess* 1998;2(14).
- [5] Prinsen CAC, Mokkink LB, Bouter LM, Alonso J, Patrick DL, de Vet HCW, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res* 2018;27:1147–57.
- [6] Olerud C, Molander H. A scoring scale for symptom evaluation after ankle fracture. *Arch Orthopaed Traum Surg* 1984;103.
- [7] McKeown R, Rabiou A-R, Ellard DR, Kearney RS. Primary outcome measures used in interventional trials for ankle fractures: a systematic review. *BMC Musculosk Disord* 2019;20(1):388.
- [8] McKeown R, Ellard DR, Rabiou A-R, Karasouli E, Kearney RS. A systematic review of the measurement properties of patient reported outcome measures used for adults with an ankle fracture. *J Pat-Rep Outcomes* 2019;3(1):70.
- [9] Ng R, Broughton N, Williams C. Measuring recovery after ankle fractures: a systematic review of the psychometric properties of scoring systems. *J Foot Ankle Surg* 2018;57(1):149–54.
- [10] Buker N, Savkin R, Gokalp O, Ok N. Validity and reliability of Turkish version of Olerud-Molander ankle score in patients with malleolar fracture. *J Foot Ankle Surg* 2017;56(6):1209–12.
- [11] Nilsson GM, Eneroth M, Ekdahl CS. The Swedish version of OMAS is a reliable and valid outcome measure for patients with ankle fractures. *BMC Musculosk Disord* 2013;14:109, no pagination.
- [12] Garratt AM, Naumann MG, Sigurdson U, Utvåg SE, Stavem K. Evaluation of three patient reported outcome measures following operative fixation of closed ankle fractures. *BMC Musculosk Disord* 2018;19(1):134.
- [13] Turhan E, Demirel M, Daylak A, Huri G, Doral MN, Celik D. Translation, cross-cultural adaptation, reliability and validity of the Turkish version of the Olerud-Molander Ankle Score (OMAS). *Acta Orthop Traumatol Turc* 2017;51(1):60–4.
- [14] Kearney RS, McKeown R, Stevens S, Parsons N, Parsons H, Wells P, 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.
- [15] Devlin N, Shah K, Feng Y, Mulhern B, Hout BV. In: Research OoHE, editor. *Valuing health-related quality of life: an EQ-5D-5L value set for England*. London: Office for Health Economics; 2016.
- [16] Dawson J, Boller I, Doll H, Lavis G, Sharp R, Cooke P, 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.
- [17] Parsons H, Bruce J, Achten J, Costa ML, Parsons NR. Measurement properties of the Disability Rating Index in patients undergoing hip replacement. *Rheumatology* 2014;54(1):64–71.

- [18] Lampridis V, Gougoulis N, Sakellariou A. Stability in ankle fractures. *EFORT Open Rev* 2018;3(5):294–303.
- [19] Mokkink LB, Terwee CB, Patrick Sr LJ, Alonso A, Stratford PW, Knol DL, *et al.* The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol* 2010;63(7):737–45.
- [20] Abma IL, Rovers M, Van der Wees P. Appraising convergent validity of patient-reported outcome measures in systematic reviews: constructing and interpreting outcomes. *BMC Res Notes* 2016; 9(226).
- [21] Mokkink LB, de Vet HCW, Prinsen CA, Patrick DL, Alonso J, Bouter L, *et al.* COSMIN Risk of Bias checklist for systematic reviews of Patient Reported Outcome Measures. *Qual Life Res* 2017;27:1171–9.
- [22] Terwee C, Bot S, Boer MD, Windt DVD, Knol D, Dekker J, *et al.* Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 2007;60.
- [23] Lin C, Moseley A, Refshauge K, Bundy A. The lower extremity functional scale has good clinimetric properties in people with ankle fracture. *Phys Therapy* 2009;89(6):580–8.
- [24] Lim CR, Harris K, Dawson J, Beard DJ, Fitzpatrick R, Price AJ. Floor and Ceiling effects in the OHS: an analysis of the NHS PROMs data set. *BMJ Open* 2015;5(7).
- [25] Vaske JJ, Beaman J, Sponarski CC. Rethinking internal consistency in Cronbach's alpha. *Leisure Sciences* 2017;39(2):163–73.
- [26] McPhail SM, Dunstan J, Canning J, Haines TP. Life impact of ankle fractures: qualitative analysis of patient and clinician experiences. *BMC Musculosk Disord* 2012;13(224).
- [27] McKeown R, Kearney RS, Liew ZH, Ellard DR. Patient experiences of an ankle fracture and the most important factors in their recovery: a qualitative interview study. *BMJ Open* 2020;10(2):e033539.
- [28] Willett K, Keene DJ, Mistry D, Nam J, Tutton E, Handley R, *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.
- [29] Kortekangas T, Haapasalo H, Flinkkilä T, Ohtonen P, Nortunen S, Laine H-J, *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.
- [30] Clarke M, Williamson P. Core outcome sets and trial registries. *Trials* 2015;16(1):216.

Available online at www.sciencedirect.com

ScienceDirect