SEFAS The Self-Reported Foot and Ankle Score

Cöster, Maria

2015

Link to publication

Citation for published version (APA):
Cöster, M. (2015). SEFAS The Self-Reported Foot and Ankle Score Department of Orthopaedics, Lund University

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.
SEFAS
The Self-Reported Foot and Ankle Score

Maria Cöster, MD

DOCTORAL DISSERTATION
by due permission of the Faculty of Medicine, Lund University, Sweden.

To be defended at The Lecture Hall, Level 5, Department of Orthopaedics,
Skåne University Hospital, Inga Marie Nilssons gata 28, Malmö, Sweden
21 May 2015, 13.00

Faculty opponent
Associate Professor Ola Rolfson, Department of Orthopaedics,
Sahlgrenska University Hospital,
Gothenburg, Sweden
Abstract

**Background:** Patient-reported outcome (PRO) instruments are used in all medical disciplines to evaluate patients with different diseases and also to follow results after treatments such as surgery. Currently there is no generally accepted and fully validated PRO instrument for the foot and ankle.

**Aims:** To translate the New Zealand Ankle Questionnaire into a Swedish version, [the Self-Reported Foot and Ankle Score (SEFAS)] and to evaluate the validity, reliability and responsiveness of SEFAS in patients with disabilities in the foot and/or ankle. The second aim was to compare SEFAS with four other PRO instruments: EQ-5D, SF-36, Foot and Ankle Outcome Score (FAOS) and American Orthopaedic Foot and Ankle Society (AOFAS) score.

**Methods:** In study I, we translated SEFAS into Swedish. We then included 135 patients registered in the Swedish Ankle Registry with osteoarthritis or inflammatory arthritis in the ankle joint. These patients completed SEFAS, FAOS, EQ-5D and SF-36. In study II we included 224 patients scheduled for surgery with a variety of foot and ankle disabilities who completed the same PRO instruments as in study I. In study III we included 206 patients scheduled for surgery with a variety of foot and ankle disabilities who completed SEFAS and AOFAS. Validity, reliability and responsiveness in addition to time to complete the instruments were then evaluated in studies I–III. In study IV we included 21 patients scheduled for surgery due to a flatfoot deformity. These patients completed SEFAS, EQ-5D and SF-36 before surgery and 6 and 24 months after surgery.

**Results:** In studies I-III we found good validity, reliability and responsiveness for SEFAS in patients with disorders in the forefoot, hindfoot and ankle. The results for SEFAS were better than or comparable to EQ-5D, SF-36, FAOS and AOFAS. In study IV we found that SEFAS was able to capture an improvement by surgery in patients with flatfoot deformity and that the improvement continued up to 24 months after surgery.

**Conclusion:** SEFAS is a PRO instrument with good validity, reliability and responsiveness. We recommend SEFAS as a PRO instrument when evaluating surgery in the foot and ankle, also in national registries.

**Key words**
ankle, foot, measurement properties, patient-reported outcome measures, reliability, responsiveness, SEFAS, validity

**Classification system and/or index terms (if any)**

**Supplementary bibliographical information**

- ISSN and key title: 1652-8220
- ISBN: 978-91-7619-133-0

**Recipient’s notes**

- Number of pages: 167
- Price

**Security classification**

I, the undersigned, being the copyright owner of the abstract of the above-mentioned dissertation, hereby grant to all reference sources permission to publish and disseminate the abstract of the above-mentioned dissertation.

Signature: Maria Cöster  Date: 2015-04-15
SEFAS
The Self-Reported Foot and Ankle Score

Maria Cöster

Lund University
Financial support for this study was received from:
The Swedish Association of Local Authorities and Regions
FoU Kalmar County Council
The Research Council in Southeast Sweden (FORSS)
FoU Region Skåne
ALF
Skåne University Hospital Foundation
Maggie Stephens Foundation
Österlund Foundation
Kock Foundation
Hand & Foot Surgery Center
Stiftelsen Skobranschens utvecklingsfond
Svenska Reumakirurgiska föreningen
Stiftelsen BGS Kvinnliga Ortopeder i Sverige
Swedish Foot and Ankle Society

“When you can measure what you are speaking about, and express it in numbers, you know something about it; but when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind: it may be the beginning of knowledge, but you have scarcely, in your thoughts, advanced to the stage of science, whatever the matter may be.”

[William Thomson, Lord Kelvin (1824–1907)]
# Contents

<table>
<thead>
<tr>
<th>Section</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contents</td>
</tr>
<tr>
<td>List of abbreviations</td>
</tr>
<tr>
<td>List of definitions</td>
</tr>
<tr>
<td>Original Papers</td>
</tr>
<tr>
<td>Introduction</td>
</tr>
<tr>
<td>Background</td>
</tr>
<tr>
<td>Function and disability</td>
</tr>
<tr>
<td>Outcome</td>
</tr>
<tr>
<td>Outcome measures</td>
</tr>
<tr>
<td>Outcome measure instruments</td>
</tr>
<tr>
<td>Clinician-based or patient-reported outcome measure instruments</td>
</tr>
<tr>
<td>Generic health-related outcome measure instruments</td>
</tr>
<tr>
<td>EuroQol 5 Dimensions (EQ-5D)</td>
</tr>
<tr>
<td>Short form 36 (SF-36)</td>
</tr>
<tr>
<td>Specific outcome measure instruments</td>
</tr>
<tr>
<td>Region-specific outcome measure instruments</td>
</tr>
<tr>
<td>American Orthopedic Foot and Ankle Society (AOFAS) scores</td>
</tr>
<tr>
<td>Foot and Ankle Outcome Score (FAOS)</td>
</tr>
<tr>
<td>Manchester-Oxford Foot Questionnaire (MOXFQ)</td>
</tr>
<tr>
<td>Foot and Ankle Ability Measure (FAAM)</td>
</tr>
<tr>
<td>Foot Function Index (FFI)</td>
</tr>
<tr>
<td>Ankle Osteoarthritis Score (AOS)</td>
</tr>
<tr>
<td>Visual Analogue Scale Foot and Ankle (VAS FA)</td>
</tr>
<tr>
<td>Self-Reported Foot and Ankle Score (SEFAS)</td>
</tr>
<tr>
<td>How do we select the most appropriate outcome measure?</td>
</tr>
<tr>
<td>Translation and adaptation of an outcome measure</td>
</tr>
<tr>
<td>Quality assessment of a PRO instrument</td>
</tr>
</tbody>
</table>
Aims of the thesis 37
  General aim: 37
Hypothesis 39
Summary of methods 41
  Assessment of quality and measurement properties 41
  Validity 42
  Reliability 44
  Responsiveness 45
  Interpretability 45
  Patient-friendliness 46
Summary of study subjects 47
  1. The Swedish National Ankle Registry 48
  2. The Kalmar and Eksjö cohort 48
Summary of statistical analysis 53
Summary of ethical considerations 55
Summary of papers 57
  Paper I 57
    Validity, reliability and responsiveness of a self-reported foot and ankle score (SEFAS) 57
  Paper II 60
    Validity, reliability and responsiveness of the Self-reported Foot and Ankle Score (SEFAS) in forefoot, hindfoot and ankle disorders 60
  Paper III 64
    Comparison of the Self-reported Foot and Ankle Score (SEFAS) and the American Orthopedic Foot and Ankle Society Score (AOFAS) 64
  Paper IV 66
    Surgery of Adult Acquired Flatfoot Deformity due to Posterior Tibial Tendon Dysfunction Reduces Pain and Improve Function and Health related Quality of Life 66
General discussion 71
  Outcome assessment and outcome instruments 71
  Should both generic and specific outcome instruments be used? 72
  Which generic outcome measure should be used? 73
Which foot- and ankle-specific instrument should be used? 73
Advantages and disadvantages of SEFAS 75
Are there new foot- and ankle-specific outcome measures under development? 76
Strengths and limitations 79
Conclusions 81
Future perspectives 83
Normative data 83
Rasch analysis 83
Minimal clinically important difference (MCID) 83
Spreading SEFAS to the medical community 84
Summary in English 87
Introduction 87
Methods 87
Results 88
Conclusion 88
Svensk sammanfattning 89
Acknowledgments 93
References 95
Appendices
Appendix I
Self-reported foot and ankle Score (SEFAS)
Appendix II
Patientrapporterade SEFAS: Frågeformulär
bra utvärderingsmetod vid fot- och fotledsbesvär
Appendix III
The Swedish National Ankle Registry
Original papers I-IV
List of abbreviations

AAFD  Adult acquired flatfoot deformity
AAOS  American Association of Orthopaedic Surgeons
ADL   Activities of daily living
A-HF  Ankle-hindfoot
AOFAS American Orthopaedic Foot and Ankle Society
AOS   Ankle Osteoarthritis Score
BMI   Body mass index
BP    Bodily pain
CAT   Computerized adaptive testing
CI    Confidence interval
COSMIN Consensus-based standards for the selection of health measurement instruments
CV    Coefficient of variation
EFAS  European Foot and Ankle Society
EQ-5D EuroQol 5 dimensions
ES    Effect size
FAAM  Foot and Ankle Ability Measure
FAOS  Foot and Ankle Outcome Score
FAS   Foot and Ankle Surgery
FDL   Flexor digitorum longus
FFI   Foot Function Index
GH    General health
HMTP  Hallux metatarsophalangeal-interphalangeal
HrQoL Health-related quality of life
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICC</td>
<td>Intraclass correlation coefficient</td>
</tr>
<tr>
<td>ICF</td>
<td>International Classification of Functioning, Disability and Health</td>
</tr>
<tr>
<td>KOOS</td>
<td>Knee injury and Osteoarthritis Score</td>
</tr>
<tr>
<td>LCL</td>
<td>Lateral column lengthening</td>
</tr>
<tr>
<td>LMTP</td>
<td>Lesser metatarsophalangeal-interphalangeal</td>
</tr>
<tr>
<td>MCID</td>
<td>Minimal clinically important difference</td>
</tr>
<tr>
<td>MCS</td>
<td>Mental component summary</td>
</tr>
<tr>
<td>MDCO</td>
<td>Medial displacement calcaneal osteotomy</td>
</tr>
<tr>
<td>MF</td>
<td>Midfoot</td>
</tr>
<tr>
<td>MFPDI</td>
<td>Manchester Foot Pain and Disability Index</td>
</tr>
<tr>
<td>MH</td>
<td>Mental health</td>
</tr>
<tr>
<td>MID</td>
<td>Minimal important difference</td>
</tr>
<tr>
<td>MOXFQ</td>
<td>Manchester Oxford Foot Questionnaire</td>
</tr>
<tr>
<td>NHP</td>
<td>Nottingham Health Profile</td>
</tr>
<tr>
<td>NZAQ</td>
<td>New Zealand total Ankle Questionnaire</td>
</tr>
<tr>
<td>OA</td>
<td>Osteoarthritis</td>
</tr>
<tr>
<td>PCS</td>
<td>Physical component summary</td>
</tr>
<tr>
<td>PF</td>
<td>Physical function</td>
</tr>
<tr>
<td>PF CAT</td>
<td>Physical function computerized adaptive testing</td>
</tr>
<tr>
<td>PRO</td>
<td>Patient-reported outcome</td>
</tr>
<tr>
<td>PROM</td>
<td>Patient-reported outcome measure</td>
</tr>
<tr>
<td>PTTD</td>
<td>Posterior tibial tendon dysfunction</td>
</tr>
<tr>
<td>QoL</td>
<td>Quality of life</td>
</tr>
<tr>
<td>QUALY</td>
<td>Quality-assessed life years</td>
</tr>
<tr>
<td>RE</td>
<td>Role limitations due to emotional problems</td>
</tr>
<tr>
<td>RP</td>
<td>Role limitations due to physical problems</td>
</tr>
<tr>
<td>SD</td>
<td>Standard deviation</td>
</tr>
<tr>
<td>SF</td>
<td>Social functioning</td>
</tr>
<tr>
<td>SF-36</td>
<td>Short form 36</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>------------------------------------</td>
</tr>
<tr>
<td>SIP</td>
<td>Sickness Impact Profile</td>
</tr>
<tr>
<td>SMO</td>
<td>Supramalleolar osteotomy</td>
</tr>
<tr>
<td>TAR</td>
<td>Total ankle replacement</td>
</tr>
<tr>
<td>TMT</td>
<td>Tarsometatarsal</td>
</tr>
<tr>
<td>VAS</td>
<td>Visual analogue scale</td>
</tr>
<tr>
<td>VAS FA</td>
<td>Visual analogue scale foot and ankle</td>
</tr>
<tr>
<td>VT</td>
<td>Vitality</td>
</tr>
<tr>
<td>WHO</td>
<td>World Health Organization</td>
</tr>
</tbody>
</table>
# List of definitions

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agreement</td>
<td>An estimate of the measurement error of an outcome measure.</td>
</tr>
<tr>
<td>Construct</td>
<td>The organization of a set of ideas or words used to define, understand and assess a given phenomenon like pain or disability. Several items in a PRO instrument can be parts of a construct.</td>
</tr>
<tr>
<td>Construct validity</td>
<td>An estimate of whether an outcome instrument relates to other instruments consistent with theoretically derived hypotheses.</td>
</tr>
<tr>
<td>Content validity</td>
<td>A measure of the extent to which questions in an outcome measure instrument are relevant to the target population.</td>
</tr>
<tr>
<td>Convergent validity</td>
<td>Exists when two outcome instruments that measure similar concepts correlate highly with each other.</td>
</tr>
<tr>
<td>Dimension</td>
<td>Dimensions in a PRO instrument refer to different areas of health and a PRO instrument may be composed of several dimensions, such as pain or mental health.</td>
</tr>
<tr>
<td>Discriminant validity</td>
<td>Exists when two outcome instruments that measure different concepts do not correlate highly with each other.</td>
</tr>
<tr>
<td>Floor and ceiling effects</td>
<td>Show the proportion of individuals who achieve the highest or lowest possible numeric value of an outcome measure.</td>
</tr>
<tr>
<td>Term</td>
<td>Definition</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Internal consistency</td>
<td>An estimate of the homogeneity of an outcome instrument, i.e. the degree to which the items are related to each other.</td>
</tr>
<tr>
<td>Interobserver reliability</td>
<td>Measures the variation occurring between several observers.</td>
</tr>
<tr>
<td>Interpretability</td>
<td>Means the comprehensibility or understandability of an outcome measure.</td>
</tr>
<tr>
<td>Intraobserver reliability</td>
<td>Measures the variation occurring from a single observer as a result of more than one exposure.</td>
</tr>
<tr>
<td>Item</td>
<td>Defines as an individual question or statement in a PRO instrument.</td>
</tr>
<tr>
<td>Measurement error</td>
<td>The systemic and random error of a patient’s score that is not attributed to true changes due to, for example, improvements after surgery.</td>
</tr>
<tr>
<td>Measurement properties</td>
<td>Different concepts of validity, reliability and responsiveness of an outcome measure instrument.</td>
</tr>
<tr>
<td>Normative data</td>
<td>Population reference data</td>
</tr>
<tr>
<td>Psychometric properties</td>
<td>Different concepts of validity, reliability and responsiveness of an outcome measure instrument.</td>
</tr>
<tr>
<td>Reliability</td>
<td>Defines as the degree to which an instrument is free from measurement errors</td>
</tr>
<tr>
<td>Response categories</td>
<td>Predetermined response options in an outcome measure instrument such as “occasionally” or “often”.</td>
</tr>
<tr>
<td>Responsiveness</td>
<td>Measures whether an outcome instrument is sensitive to changes over time or changes due to an intervention.</td>
</tr>
<tr>
<td>Target population</td>
<td>The entire group of individuals a researcher is interested in; the group</td>
</tr>
</tbody>
</table>
about which the researcher wishes to draw conclusions.

<table>
<thead>
<tr>
<th>Test-retest reliability</th>
<th>Identifies the stability of an instrument over time, i.e. the reproducibility</th>
</tr>
</thead>
<tbody>
<tr>
<td>Validity</td>
<td>The degree to which an outcome measure instrument, questionnaire or test measures what it is intended to measure.</td>
</tr>
</tbody>
</table>
Original Papers

This thesis reviews the following papers, which are referred to in the text by their Roman numerals:

I. Validity, reliability, and responsiveness of a self-reported foot and ankle score (SEFAS)
   Cöster M, Karlsson M, Nilsson J-Å, Carlsson Å
   *Acta Orthopedica* 2012; 83(2):197–203

II. Validity, reliability, and responsiveness of the Self-reported Foot and Ankle Score (SEFAS) in forefoot, hindfoot, and ankle disorders
    Cöster MC, Bremander A, Rosengren B, Magnusson H, Carlsson C, Karlsson M

III. Comparison of the Self-Reported Foot and Ankle Score (SEFAS) and the American Orthopedic Foot and Ankle Society Score (AOFAS)
    Cöster MC, Rosengren B, Bremander A, Brudin L, Karlsson M
    *Foot & Ankle International* 2014; 35(10):1031–6

IV. Surgery for Adult Acquired Flatfoot Deformity due to Posterior Tibial Tendon Dysfunction Reduces Pain and Improve Function and Health-Related Quality of Life
   Cöster MC, Rosengren B, Bremander A, Karlsson M
   Accepted to *Foot and Ankle Surgery* April 2015
Introduction

Musculoskeletal diseases and disorders impose a large burden on the healthcare system, since these disabilities affect hundreds of millions of people around the world (147). In Sweden, 20–30% of patients in the primary health care system describe these disabilities (122). Furthermore, musculoskeletal disorders include a variety of diagnoses, of which traumatic injuries, osteoarthritis (OA), inflammatory arthritis, osteoporosis and low back pain are the most frequently discussed conditions. All these conditions could affect and create symptoms from the feet or ankles, often including pain and impaired function (147). Foot or ankle pain and disability may also be the result of a variety of other disorders and deformities that occur in the forefoot, the hindfoot and/or the ankle. These problems can be found in both young and old individuals. Foot disabilities are reported to occur in 10–20% of individuals in the general population (56, 65) with an even higher proportion found in the old population (51, 94, 136). Foot and ankle disabilities are also especially widespread in specific subgroups such as athletes, most often reported among long-distance and marathon runners (4). Foot and ankle disabilities are associated in all ages with impaired health-related quality of life (HRQoL) (56, 65, 94) and in the old population also impaired gait and balance, with increased risk of sustaining falls and fractures (94, 96).

Currently we do not know how many of the individuals with foot and ankle disabilities who require surgical treatment. In an analysis of the Australian Medicare Benefits Schedule database from 1997–2006 (95), a database where most of the Australian surgical procedures in foot and ankle are registered, between 94 217 and 104 538 procedures (in the population of 21.6 million inhabitants) were reported per year during this period. All ages and both genders (52% men, 48% women) were represented in this database, with an increased incidence of surgery performed on individuals over 55 years. With the future increasing proportion of old people in community, these data speak for an even greater demand for foot and ankle surgery in the future (95). The number of surgical procedures for foot and ankle is also high in Sweden, with currently 20 000–25 000 elective surgical procedures performed each year according to the Swedish National Board of Health and Welfare (Socialstyrelsen) (110). However, the future demands are unknown.

Foot and ankle surgery includes a great variety of procedures such as surgery for arthritis, sports injuries, fractures, tendon disorders, post-traumatic foot disabilities, diabetic foot problems, heel pain, high-arched and flat feet, neurological foot
disorders, great and lesser toe disabilities. Since the foot is one of the most complex musculoskeletal systems in the body, several treatment strategies and surgical procedures are used for the different disorders and disabilities in order to decrease pain, improve function and health-related quality of life (HRQoL). However, our knowledge is limited regarding the utility and the clinical and patient-reported outcome of the different surgical procedures in the foot or ankle.

To improve our knowledge several national registries have been developed. In Sweden, for example, there are national registries that follow patients after hip, knee, shoulder and ankle arthroplasties, but also after degenerative spine surgery, fractures, ankle fusions and cruciate ligament injuries (www.kvalitetsregister.se). Most of the registries use patient-reported outcome (PRO) instruments for outcome reports and research evaluations. The Swedish Ankle Registry (www.swedankle.se) has reported the outcome of ankle arthroplasties since 1993, and of ankle fusions since 2008, but this registry only captures a small fraction of all performed foot and ankle surgery. There is therefore a need to create an extended registry that includes all types of surgical procedures in the foot and ankle. Such a registry is under construction (Riksfot) with the goal of being operational in 2015.

For this new registry it is imperative to define a suitable PRO instrument for the foot and the ankle since there is an international recommendation that PRO instruments should be included when following the outcome of surgical procedures (114). The Swedish Association of Local Authorities and Regions (Sveriges Kommuner och Landsting) and the National Board of Health and Welfare (Socialstyrelsen) also require PRO instruments in national registries to provide financial support. Until now there has been no generally accepted foot- and ankle-specific PRO instrument that is fully validated. Such a PRO instrument must be easy and fast to complete and ought to be able to evaluate all foot and ankle surgery. With this background, we stated the overall aim of this thesis – to find and validate a foot- and ankle-specific PRO instrument that could be used to evaluate surgical procedures in the foot and ankle in national registries, and also the surgical outcome in the general clinical practice and in other kinds of research studies.
Background

Function and disability

Physical disability and pain may follow a wide range of musculoskeletal disorders. These disabilities are often described in a variety of terms, making it difficult to attempt any comparisons across studies, populations and disorders. The World Health Organization (WHO) has therefore established an International Classification of Functioning, Disability and Health (ICF) to serve as a framework for measuring health and disability at both individual and population levels worldwide and to provide a standardized language when describing health and health-related states (139). ICF uses a model with an integration of biological, individual and social perspectives of health. According to the model there are three levels of human functioning: (i) body structure and function, (ii) activity and (iii) participation. In this model disability is defined as dysfunction in one or more of the levels impairments, activity limitations and participation restrictions (Figure 1). Disability is also regarded as an interaction between health and contextual factors, i.e. both environmental factors such as social attitudes and structures, climate and terrain, and personal factors such as age, gender, past and current experience, social background and education (139) (Figure 1).

Figure 1.
International Classification of Functioning, Disability and Health (ICF)
Outcome

Outcome is an expression that covers the patients’ perspective on the results of medical interventions such as surgery (129). In 1998 Clancy and Eisenberg defined outcomes research in Science (24) as “the study of the end results of health services that takes patients’ experiences, preferences and values into account”. This definition was instantly accepted as part of an aim to provide scientific evidence-based decisions by all those who participate in the health care process. WHO supports this view in the ICF model when they declare, regarding health and disability evaluations, that the patient’s perspective is the most important (139).

Orthopedic surgeons treat patients with musculoskeletal disorders in order to achieve improvements in pain, physical function and health-related quality of life (HrQoL). The outcome after surgery should be captured by validated patient-reported outcome (PRO) instruments in order to cover the patients’ perspective. The researcher must then take into account whether a numerically registered change in the summarized score of an instrument also means a meaningful clinical change in health (129).

Outcome measures

Clinicians, including surgeons, have in the past relied on clinical measurements when evaluating the outcome and efficacy of treatments. Examples of such outcome measures in orthopedics are observed joint range of movements, observed deformities, measures of muscle strength, measured laxity and functional performance tests. Radiographic parameters before and after surgery are other measures that often has defined the outcome in orthopedic literature (74, 76). However, these measures do not take the patients’ perspective into account. Current research has also shown that these measures include subjectivity in the registrations, even if they are often presented as objective outcome evaluations. The surgeons, in the hope of achieving a specific result, often perform the assessments by themselves, creating a risk of bias when estimating the results. Many measures or tests also have low intra- or interobserver reliability. Both intra- and interobserver reliability can be low when determining these alleged objective measures, such as range of motion and muscle strength estimated manually or with the use of a dynamometer (3, 15, 59, 67, 101, 149). Another fact is that clinical measurements often also have a low degree of correlation with patient-reported HrQoL, physical function or disability (14, 64, 111). Thus, individuals may have different perceptions of exactly the same outcome of the measurement dependent on their experiences, conditions and needs. The same measured muscle strength, for example, can be rated as poor in a younger individual but good in an older individual. This is another flaw when using these alleged objective measures.
Outcome measure instruments

New and other types of outcome instruments have therefore been developed, instruments that focus on the patients’ view of the disability. These instruments, often referred to as questionnaires or scores, consist of items or questions with different categories of responses. The simplest ones are the binary response categories including only “yes” or “no” as answers. However, more instruments use the Likert set of response categories, including three or more standardized response options such as “none”, “mild”, “moderate” and “severe”. Similar instruments are usually referred to as Likert scales. Visual analogue scales (VAS) or graphic rating scales (GRS) are other response categories or scales that are used in outcome measure instruments (129). In VAS for example, pain can be rated on a ruler numbered from 0 (worst possible pain) to 100 (no pain). A summary of different types of response categories or scales are presented in Figure 2.

![Likert Scale](image)

![Graphic Rating Scale](image)

![Visual Analogue Scale](image)

Figure 2.
Different types of scales and response categories

Outcome measure instruments may further consist of different subscores, where each subscore includes items within the same construct. Pain and physical function are two important constructs, often measured in instruments evaluating musculoskeletal disorders. The subscores are created when the different responses for the group of items within the same construct are transformed into numerical values and then summarized. There are several ways to calculate the summarized subscore with more or less complicated weighting systems (108, 129). If all items in the instrument are summarized in a similar way to one numerical value, this is referred to as the total score. Scores and scales are often used as the same concept in the literature. In this thesis we most often use the word score.
Clinician-based or patient-reported outcome measure instruments

The outcome measure instruments are often divided into clinician-based and patient-reported and labeled as either generic or specific.

Clinician-based outcome measures are based on assessments by a professional observer, usually a clinician (surgeon). Joint range of motion, deformities, gait abnormalities, response to provocative maneuvers and muscle strength are examples of such tests and measures that are included in these outcome measures (129). In some of these instruments there is a part with patient-reported data included. Clinician-based outcome measure instruments have previously been considered as objective, but nowadays, with more thorough evaluations of the reliability, these measures must be considered as partially subjective (129).

Patient-reported outcome (PRO) instruments include any outcome evaluation that is based on data specifically provided by the patients (Figure 3). These outcome measure instruments have become increasingly popular, and nowadays there is a variety of such instruments available to assess musculoskeletal disorders. If the PRO instrument is generic, it provides a general estimation of health regardless of disease or affected anatomical region, and if it is specific, it can be disease-specific, region-specific or dimension-specific (129).

Figure 3.
Completing the PRO instrument SEFAS.
Generic health-related outcome measure instruments

Generic health-related outcome measures are relevant to almost all populations, and these instruments measure multiple aspects of health. Generic instruments can be used to compare one patient population with another, regardless of disease or disability. However, these outcome measures are generally less responsive to changes and, due to the inclusion of more items, often more complicated and time-consuming to complete than the specific measures. Many of the generic outcome measures have been used for years and are thoroughly evaluated. Examples of such measures are the EuroQol 5 Dimensions (EQ-5D) (49, 53, 55, 132), the Short form 36 (SF-36) (107, 130, 131) the Sickness Impact Profile (SIP) (11), and the Nottingham Health Profile (NHP) (144, 145) (Table 1a)

Table 1a.
Generic health-related outcome measure instruments

<table>
<thead>
<tr>
<th>Generic outcome measure instruments</th>
</tr>
</thead>
<tbody>
<tr>
<td>EQ-5D (EuroQol 5 dimensions)</td>
</tr>
<tr>
<td>SF-36 (Short-form 36)</td>
</tr>
<tr>
<td>SIP (The Sickness Impact Profile)</td>
</tr>
<tr>
<td>NHP (Nottingham Health Profile)</td>
</tr>
</tbody>
</table>

**EuroQol 5 Dimensions (EQ-5D)**

EQ-5D ([www.euroqol.org](http://www.euroqol.org)) (18, 49, 53, 132) is a generic PRO instrument that evaluates Health-Related Quality of Life (HrQoL). The instrument has been developed by the EuroQol group in order to provide a simple outcome measure for clinical and economic appraisal. EQ-5D can be used when evaluating a variety of health-related conditions and treatments. EQ-5D consists of a descriptive profile, a single index and a self-reported VAS scale for the patient’s self-estimated health status. The descriptive profile includes 5 dimensions: (i) mobility, (ii) self-care, (iii) usual activities, (iv) pain/discomfort and (v) anxiety/depression. In each dimension there are three identical response options: (a) no problem, (b) some problem and (c) severe problems/unable to. The combinations of answers generate different EQ-5D states, and altogether there are 243 possible states in addition to the states dead and unconscious. The single index value, also referred to as the EQ-5D index, is a weighted total value of the score that is calculated after being adjusted for cultural differences in the response pattern. Different national or regional value sets (tariffs) are used for these calculations. The tariff from the United Kingdom is used in
Sweden. The highest EQ-5D index 1 represents full health and the lowest EQ-5D index 0 represents dead. There are also negative values, but for health-economic calculations and evaluations all negative values are set at 0. The EQ-VAS measures the patient’s self-rated health on a vertical visual analogue scale, with best imaginable health and worst imaginable health being the two endpoints of the scale. The EuroQol group has also collected population reference or normative data in general population surveys from 24 countries. With the aid of these data, there are now country-specific, gender-specific and age-specific normative EQ-5D index and EQ-VAS values available. Such data can be used to compare patients with specific disorders or diseases with healthy individuals, and also to compare populations or patient cohorts with different disorders or disabilities. The normative data also make it possible to compare the benefits of a specific treatment and calculate health economy and the cost-benefit of these treatments. EQ-5D is fully validated with confirmed good measurement properties (53, 132).

Short form 36 (SF-36)

SF-36 (www.sf-36.org) (107, 130, 131, 141) is a generic measure of HrQoL widely used for patients with musculoskeletal disorders. It is also used to compare treatment between different surgical disciplines. SF-36 consists of 36 items that measure functional ability, well-being and overall health. The 36 items have a Likert scale rating system and these 36 items are collected within eight defined dimensions (or subscores) of health: (i) physical functioning (PF), (ii) role limitations due to physical problems (RP), (iii) bodily pain (BP), (iv) general health perception (GH), (v) vitality (VT), (vi) social functioning (SF), (vii) role limitations due to emotional problems (RE) and (viii) mental health (MH). The results from responses of the items are summarized, weighted and then, for each subscore, transformed to a value from 0 to 100, with 0 representing worst possible health, and 100 best possible health. The first four subscores in SF-36 are related to physical function, summarized and reported as the physical component summary scale (PCS) and the latter four to mental health, summarized and reported as the mental component summary scale (MCS). The SF-36 is thoroughly evaluated with confirmed good measurement properties and country-specific, gender-specific and age-specific normative data exist in more than 40 countries. SF-36 is often used as comparison when other outcome measures are validated (85, 89, 124).

Specific outcome measure instruments

There are different types of specific outcome measures. Disease- or condition-specific outcome measures are designed to reflect symptoms and functional limitations
experienced by individuals with a specific disease or condition (55, 108). Region-specific outcome measures are designed to reflect symptoms and limitations that arise from specific anatomical parts of the body, for example the foot or the ankle. Dimension-specific outcome measures assess one specific aspect of health status, for example pain (55). Specific outcome measures are usually better at capturing changes in clinical status, i.e. are more responsive, and usually include fewer items than the generic ones, thus being easier and less time-consuming to complete. However, they cannot be used for comparisons between different diseases or to compare general health in different populations.

Region-specific outcome measure instruments

There are a variety of PRO instruments that evaluate musculoskeletal disorders and measure outcome after surgery. Most of these are region-specific and designed to evaluate different anatomical regions of the body, for example conditions in the shoulder, elbow, hip or knee (10, 23, 40, 41, 45, 61, 82, 84, 117, 120, 121, 123, 129) (Table 1b).

Table 1b.
Region-specific outcome measure instruments for both upper and lower extremities

<table>
<thead>
<tr>
<th>Region-specific outcome measure instruments</th>
<th>Shoulder and Elbow</th>
<th>Hip and Knee</th>
</tr>
</thead>
<tbody>
<tr>
<td>DASH (Disability of Arm, Shoulder and Hand)</td>
<td></td>
<td>WOMAC (Western Ontario and McMaster Universities Osteoarthritis index)</td>
</tr>
<tr>
<td>UEFI (Upper Extremity Functional Index)</td>
<td></td>
<td>HHS (Harris Hip Score)</td>
</tr>
<tr>
<td>OSS (Oxford Shoulder Score)</td>
<td></td>
<td>HOOS (Hips disability and Osteoarthritis Outcome Score)</td>
</tr>
<tr>
<td>SPADI (Shoulder Pain and Disability Index)</td>
<td></td>
<td>Nonarthritic hip score</td>
</tr>
<tr>
<td>SDQ (Shoulder Disability Questionnaire)</td>
<td></td>
<td>OHS (Oxford Hip Score)</td>
</tr>
<tr>
<td>OES (Oxford elbow score)</td>
<td></td>
<td>KOOS (Knee Injury and Osteoarthritis Outcome Score)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Lysholm score</td>
</tr>
<tr>
<td></td>
<td></td>
<td>OKS (Oxford Knee Score)</td>
</tr>
</tbody>
</table>

For the foot and ankle there is also a wide range of specific outcome measures (52, 54, 63, 80, 90, 92, 105, 113) (Table 2). Hunt et al. (72) reported recently, in a systematic review, that 139 different outcome measures specifically for the foot and ankle were used in research during the period 2002 to 2011. However, only 30 of these outcome measures
measures were used in 5 or more publications, indicating that many of these are only locally used. The American Orthopedic Foot and Ankle Society (AOFAS) scores were by far the most frequently used instrument in this review. However, the AOFAS score is neither a strict patient-reported nor a sufficiently validated outcome measure. To cite the authors of the review, “Our data underscore the need for a paradigm shift toward the use of consistent, valid, reliable outcome measure instrument for studies of foot and ankle procedures and disorders and the ultimate goal is to progress toward a consensus in the use of outcome measure instruments for various foot and ankle disorders”. This view is supported by a variety of other foot and ankle surgeons and researchers worldwide (21, 93).

Table 2.
Foot- and ankle-specific outcome measure instruments

<table>
<thead>
<tr>
<th>Foot-and ankle-specific outcome measure instruments</th>
</tr>
</thead>
<tbody>
<tr>
<td>AOFAS (American Orthopedic Foot and Ankle Society)*</td>
</tr>
<tr>
<td>AOS (Ankle Osteoarthritis Score)</td>
</tr>
<tr>
<td>FAAM (Foot and Ankle Ability Measure)</td>
</tr>
<tr>
<td>FAOS (Foot and Ankle Outcome Score)</td>
</tr>
<tr>
<td>FFI (Foot Function Index)</td>
</tr>
<tr>
<td>MOXFQ (Manchester-Oxford Foot Questionnaire)</td>
</tr>
<tr>
<td>SEFAS (SELF-reported Foot and Ankle Score)</td>
</tr>
<tr>
<td>VAS FA (Visual-Analogue Scale Foot and Ankle)</td>
</tr>
</tbody>
</table>

* AOFAS is partial clinician-based, all other instruments patient-reported

American Orthopedic Foot and Ankle Society (AOFAS) scores

The AOFAS scores (www.eorif.com/ankle-foot-outcome-measures) (78) have been developed with the purpose of evaluating all foot and ankle disorders and all surgical procedures in this region. The AOFAS scores consist of 4 scores: (i) the ankle-hindfoot score (A-HF), (ii) the midfoot score (MF), (iii) the hallux metatarsophalangeal-interphalangeal score (HMTP) and (iv) the lesser metatarsophalangeal-interphalangeal score (LMTP), where each score is related to a specified anatomic region in the foot or the ankle. The scores contain both clinician-
based items such as range of motion, alignment, gait and stability, and patient-reported items such as pain, function, walking distance, walking ability on different surfaces and shoe wear. The scores contain 8 to 9 items with 3 to 4 defined response options as a Likert scale. The response to each item is rated from 0 to 5 up to 40 depending on the specific item. AOFAS is scored from 0 to 100, with 0 points representing the most severe disability and 100 representing normal functions. The AOFAS scores are extensively used in foot- and ankle-related literature, but they are only partially validated, partial clinician-based and not available in Swedish (30, 85, 87, 124).

Foot and Ankle Outcome Score (FAOS)

The FAOS (www.koos.nu/FAOSEng.pdf) (116) is a patient-reported outcome measure that evaluates symptoms and functional limitations related to the foot and ankle, being adapted and developed from the Knee Injury and Osteoarthritis Score (KOOS) (117). FAOS was developed in Sweden and consists of 42 items in 5 subscores, with each item scored on a 5-point Likert scale ranging from 0 to 4. The subscores evaluate (i) pain, (ii) other symptoms, (iii) activities of daily living, (iv) function in sport and recreation and (v) ankle-related quality of life (QoL). A raw score is calculated from the sum of the items in each subscore and transformed to a value ranging from 0 to 100, with 0 representing the worst and 100 the best score. The different subscores are presented graphically as FAOS profiles, but there is no summarized total score that includes all subscores. Furthermore, the FAOS has not been validated for all foot and ankle disabilities. The results of published validation studies also differ depending on evaluated disability, with superior measurement properties found for patients with ankle instability compared to hallux valgus or adult acquired flatfoot deformity (AAFD) (22, 88, 116). However, Golightly et al. (57), in a large community-based study of 1670 adults with and without osteoarthritis (OA), found acceptable reliability and validity for the FAOS score. The FAOS has been translated from Swedish into eight different languages, but has not been extensively used in published literature.

Manchester-Oxford Foot Questionnaire (MOXFQ)

The MOXFQ (35–38) is developed from the Manchester Foot Pain and Disability Index (MFPDI), a patient-reported outcome measure designed to measure a variety of foot and ankle disorders. MOXFQ is a 16-item instrument with each item answered on a 5-point Likert scale. Scores for each item are summed to form three separate subscores representing the three constructs (i) walking/standing problems, (ii) foot pain and (iii) issues related to social interaction including feelings of self-consciousness about foot and footwear appearance. The different items within one
subscore are then calculated and converted to a score from 0 to 100, with 0 representing the least and 100 the most severe disability. The three MOXFQ subscores can also be summed, converted to a metric from 0 to 100 to create a summary index score called the **MOXFQ-Index** (102). MOXFQ has been thoroughly validated with good measurement properties for a wide range of foot and ankle disorders and the PRO instrument has been translated from English into Italian and has also been validated. It has not been translated into Swedish.

**Foot and Ankle Ability Measure (FAAM)**

The FAAM (91) is another patient-reported outcome measure that evaluates the health status of patients with a variety of foot and ankle disorders. FAAM consists of 21 items that evaluate activities of daily living (ADL) in one subscore and 8 items that evaluate sport ability in another subscore. Each item is scored on a 5-point Likert scale and the items in each subscore are transformed to a score from 0 to 100, with 0 representing the greatest disability and 100 normal function. FAAM has been validated with good measurement properties in a variety of patients with foot and ankle disorders, and the original English version of FAAM has been translated into several other languages and also validated. The FAAM is not translated into Swedish and has not been used extensively in the literature.

**Foot Function Index (FFI)**

The FFI (19) is a patient-reported outcome measure designed to measure foot- and ankle-related pain, disability and restrictions of activity in patients with rheumatoid arthritis. FFI contains 23 items in three subscores scored from 0 to 100 points on a reversed VAS scale, with 0 representing no difficulty and 100 points the most severe disability. Several modified versions have been developed as to make FFI possible to use in populations with foot and ankle disorders beyond those with rheumatoid arthritis (FFI-RL, FFI-RS) (79). FFI has been validated, but several problems in some of the measurement properties and with the VAS measure were then detected (2, 138, 148). The FFI has been translated from English into several other languages and also validated (104). FFI is not translated into Swedish and has not been used extensively in the literature.

**Ankle Osteoarthritis Score (AOS)**

The AOS (50) is a patient-reported, disease-specific and region-specific outcome measure, which is developed from the FFI and used to measure pain and disability related to OA in the ankle. The score is composed of two subscores, (i) pain and (ii)
disability, with 18 items scored on a VAS scale running from 0 to 100 points. Like the FFI, 0 represents normal health and 100 the greatest disability. This outcome measure is evaluated for patients with OA in the ankle, but not applicable for other foot and ankle disorders. AOS is not translated into Swedish and has not been used extensively in the literature.

**Visual Analogue Scale Foot and Ankle (VAS FA)**

VAS FA (112) is a patient-reported outcome measure developed in Germany to measure a variety of foot and ankle disorders. VAS FA contains 20 items and is scored on a VAS scale in the three different categories (i) pain, (ii) function and (iii) other complaints. With a computerized evaluation system the total score runs from 0 to 100 points, with 0 representing the worst and 100 the best status. There are normative data for this score (128), but VAS FA has only been partly validated. VAS FA is translated from German into several other languages, but not into Swedish. VAS FA has not been used extensively in the literature.

**Self-Reported Foot and Ankle Score (SEFAS)**

SEFAS (Appendix I) (www.swedankle.se) is a foot- and ankle-specific patient-reported outcome measure based on the New Zealand total ankle questionnaire (NZAQ) (66). NZAQ was originally derived from the validated Oxford Hip Score (OHS) (40). Before the New Zealand national joint registry started to use the NZAQ in 2000 they made some changes that adapted the original instrument OHS to patients with ankle disorders. Four of the items in the original instrument were replaced by foot- and ankle-specific questions, while the other 8 items were identical to the hip score, except that the word “hip” was replaced by the word “ankle”. NZAQ is still used in the New Zealand national joint registry but it has never been validated. For this reason, in papers I–III (27, 28, 30) we evaluated the Swedish version of NZAQ from a variety of aspects. When the Swedish version of this PRO instrument was created it was called the Self-Reported Foot and Ankle Score (SEFAS). The translation and adaptation to Swedish conditions in paper I was done according to a standardized cross-cultural adaptation procedure described in the literature (58).

SEFAS contains 12 items with 5 response options on a Likert scale scored from 0 to 4, where a sum of 0 points represents the most severe disability and 48 represents normal function. The instrument includes no subscores, but covers different important constructs, such as pain, function and activity limitations. When SEFAS is not correctly completed or when there are missing items we use the following approach: (1) when results from 2 or more items are missing, the instrument is disregarded; (2) when the result from 1 item is missing, the mean result of the remaining 11 items is used; (3) when the patients give 2 answers for 1 item, the worse
outcome is recorded; and (4) when the patients have put a mark between 2 answers, the worse outcome is recorded.

How do we select the most appropriate outcome measure?

Outcome measure instruments vary in content and quality and are developed for a variety of purposes. The use of an inappropriate outcome measure can lead to results that are not comparable with other studies, to incorrect conclusions and as a result non-evidence-based practice. When we select an outcome measure, it is important to first define the patient population to be evaluated and the condition or disability to be assessed. The next step is to identify available instruments that may address this target population and the constructs we want to evaluate. Finally, the qualities of the measurement instrument, i.e. the content, the methodological quality and the utility of the instrument have to be assessed. The recommendations in musculoskeletal research are often to select one region-specific and one generic instrument (13, 16, 39). If no suitable outcome measure exists, a new one has to be created. The process of creating a new outcome measure is an extended procedure that not will be described in this thesis. An easier approach is to translate, adapt and validate an existing appropriate outcome measure instrument that has been developed in another language for use in the new language and country.

Translation and adaptation of an outcome measure

There is need for specific guidelines when patient-reported outcome measures are translated and cross-culturally adapted for use in a new country, a new culture, and/or a new language. These guidelines are required to achieve equivalence between the original and the translated and adapted instrument. In 1993 Guillemin et al. (58) proposed guidelines that were later on refined and used by the American Association of Orthopaedic Surgeons (AAOS) Outcomes Committee (8), among others. These guidelines consist of 6 stages presented in Figure 4. The first stage (stage I) is the forward translation by two independent translators from the original language to the new language. One of the translators should have a medical background and be aware of the concepts examined and the other one should reflect the language used by the target population (the population who are going to be evaluated) without having any medical knowledge. In stage II the two translations are synthesized to one common version, which in stage III is translated back to the original language by two translators with the original language as their mother language. In stage IV an expert committee reviews the translated and the back-translated versions and produces a new version of the instrument in the new language. This pre-final version (stage V) should
then be completed by 30 to 40 patients from the target population, to evaluate whether they understand the items. The developers of the translated outcome measure instrument and an expert committee then assess all written reports from the 5 stages and submit the final translated and adapted version (stage VI). As the final step, the measurement properties of the instrument must be evaluated before this new version can be used. Modified versions of these guidelines have been synthesized in recent years, with modifications that focus more on the cultural adaptation (32).

Figure 4.
The six stages in the translation and cultural adaptation process when creating a new PRO instrument

Quality assessment of a PRO instrument

The measurement properties of a PRO instrument must be evaluated before the outcome instrument can be used in research studies (1, 47, 109). This process is easier when using a standardized assessment. The COSMIN (Consensus-based standards for the selection of health measurement instruments) group has developed a checklist that contains both consensus on what measurement properties to use, how these properties should be defined and how they should be evaluated in terms of study design and statistical analysis (97-100). This checklist is nowadays internationally used when PRO instruments are created and assessed. The checklist requires that, when the quality of a PRO instrument is assessed, four main categories should be distinguished and evaluated: (i) validity, (ii) reliability, (iii) responsiveness and (iv) interpretability. These four categories will be further explained and discussed in the section on methods below.

The evaluation of the measurement properties of a PRO instrument is often referred to as the validation of a PRO instrument.
Aims of the thesis

General aim:

The overall aim of this thesis was

- to translate and adapt the Self-Reported Foot and Ankle Score (SEFAS) to Swedish conditions
- to validate the instrument according to international guidelines
- to evaluate whether SEFAS is a better PRO instrument for evaluating foot and ankle surgery than other commonly used specific and generic instruments.

Specifically we wanted to answer the following research questions:

I. Is SEFAS easy to use and understand after the translation and cultural adaptation?

II. Are the measurement properties of SEFAS acceptable in patients with inflammatory arthritis or osteoarthritis in the ankle?

III. Is the outcome of the validation better for SEFAS than EQ-5D, SF-36 and FAOS in patients with inflammatory arthritis or osteoarthritis in the ankle?

IV. Are the measurement properties of SEFAS acceptable in patients with forefoot, hindfoot and ankle disorders?

V. Is the outcome of the validation better for SEFAS than for EQ-5D, SF-36 and FAOS in patients with forefoot, hindfoot and ankle disorders?

VI. Are the measurement properties of SEFAS and AOFAS similar or better for SEFAS in patients with great toe, hindfoot and ankle disorders?

VII. Is SEFAS an appropriate PRO instrument to use for evaluating the outcome after surgery for acquired flatfoot deformity?

VIII. Is SEFAS better than EQ-5D and SF-36 to use for evaluating outcome in patients with acquired flatfoot deformity?
Hypothesis

Based on the experiences from the use of the New Zealand total ankle questionnaire (NZAQ) in the New Zealand national ankle registry, we hypothesized that SEFAS would become a foot- and ankle-specific PRO instrument with good measurement properties that makes it suitable for use in national foot and ankle registries, in other research and in clinical evaluations.
Summary of methods

Assessment of quality and measurement properties

In this thesis we used the COSMIN guidelines (99) when we assessed the quality of the Self-reported Foot and Ankle Score. Four main categories should be distinguished and evaluated: (i) validity, (ii) reliability, (iii) responsiveness and (iv) interpretability (Figure 5).

Figure 5.
Quality including measurement properties of a PRO instrument according to the COSMIN checklist
Validity identifies the degree to which a PRO instrument measures the constructs or the dimensions (for example physical function) that it is intended to measure. Reliability is defined as the degree to which an instrument is free from measurement errors and the concept include both the stability of an instrument over time and the homogeneity of the items in the instrument at one point in time. Responsiveness is the ability of a PRO instrument to detect changes over time, such as changes occurring after surgery. The guidelines when evaluating a PRO instrument with recommended statistical methods, sample sizes, evaluations and accepted standards (17, 25, 43, 127, 129, 134, 135) are summarized in Table 3.

### Table 3.
Guidelines evaluating measurement properties of PRO instruments

<table>
<thead>
<tr>
<th>Measurement property</th>
<th>Time interval between administrations</th>
<th>Statistical method</th>
<th>Accepted standard</th>
<th>Sample size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Construct validity</td>
<td>1 administration</td>
<td>Spearman’s correlation coefficient</td>
<td>&gt;0.60 strong correlation, 0.30-0.60 moderate correlation, &lt;0.30 weak correlation</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hypothesis testing</td>
<td>&gt;75% of the results in correspondence with the hypotheses</td>
<td></td>
</tr>
<tr>
<td>Test-retest reliability</td>
<td>1-2 weeks</td>
<td>Intraclass correlation coefficient (ICC)</td>
<td>&gt;0.70</td>
<td>50</td>
</tr>
<tr>
<td>Internal consistency</td>
<td>1 administration</td>
<td>Cronbach’s alpha (CA)</td>
<td>0.70-0.95</td>
<td>100</td>
</tr>
<tr>
<td>Measurement error</td>
<td>1-2 weeks</td>
<td>Bland Altman plots Smethod (Dahlberg)</td>
<td>data are described graphically</td>
<td>50</td>
</tr>
<tr>
<td>Responsiveness</td>
<td>Before and after surgery</td>
<td>Effect size (ES)</td>
<td>&gt;0.8 large, 0.5-0.8 moderate, &gt;0.2-0.5 small, &lt;0.2 trivial</td>
<td>50</td>
</tr>
<tr>
<td>Floor and ceiling effects</td>
<td>1 administration</td>
<td>Standardized response means (SRM)</td>
<td>(SRM slightly lower values)</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Descriptive statistics</td>
<td>&gt;15% achieve the lowest or highest possible numerical value of a score</td>
<td></td>
</tr>
</tbody>
</table>

Validity

Validity consists of three measurement properties: (i) content validity, (ii) construct validity with hypotheses testing and (iii) criterion validity.

Content validity is a measurement property, mainly not based on statistical calculations, which describes how well the different items in an instrument relate to the patients’ perceptions of disability. The property also assesses the patients’ perceptions of the comprehensiveness and relevance of the different items. Most important is that the content of the items is suitable for the specific population that is
to be evaluated (the target population). It is necessary, for example, to have several items related to pain for patients with pain as a predominant symptom, such as in individuals with musculoskeletal disorders. Evaluation of the content validity should be done before all other analyses of validity are made. The evaluation is performed in different steps, including interviews with patients and professionals and discussions in focus groups. The understandability of the items are discussed in the focus groups, and these discussions may lead to changes in the text or even removal of some items. One recommended way to examine the relevance of the different items is to let patients grade each item from 1 to 3 in respect of importance, where 1 point represents an unimportant, 2 an important and 3 a most important item. The items with an average score of ≥ 2 are then considered as being relevant. When all these evaluations have been done and adjustments have been made, the content validity of the instrument can be regarded as good (99, 134, 135).

**Construct validity** identifies the extent to which the PRO instrument relates to other instruments according to hypotheses defined in advance. **Convergent (construct) validity** is present when two instruments (scores) or subscores that theoretically should be related, that is, measuring similar constructs, show a strong correlation. **Discriminant (construct) validity** is present when two instruments (scores) or subscores that theoretically should not be related, that is, measuring different constructs, show a weak correlation. It is important to demonstrate that there is both convergent and divergent validity according to the predefined hypotheses. The general recommendations are that already existing outcome measures should be used for this comparison (129). When evaluating musculoskeletal disorders, SF-36 is often the PRO instrument that is selected for this comparison. Correlations between region-specific PRO instruments and subscores in SF-36 that are related to physical function are often strong, but correlations between region-specific PRO instruments and subscores in SF-36 that are related to mental health are often weak. To evaluate the correlations between scores it is recommended to use the Spearman’s correlation coefficient (r) and to have at least 100 participants in these analyses. An r-value of >0.60 is considered a strong, 0.30–0.60 a moderate and <0.30 a weak correlation. The overall construct validity should be rated as good if ≥75% of the predefined hypotheses can be confirmed (43).

**Criterion validity** identifies the degree to which the scores of a new PRO instrument are an adequate reflection of a gold standard score. According to a consensus from the COSMIN group, there is no gold standard for PRO instruments (97).
Reliability

Reliability consists of three measurement properties: (i) test-retest reliability, (ii) internal consistency and (iii) measurement error.

Test-retest reliability identifies the extent to which the same results are obtained on repeated administrations of the instrument to the same patient when no change has occurred in clinical status. Intraobserver reliability and interobserver reliability are also included in the concept of reliability, used in clinician-based outcome measures, but not in patient-reported ones. Intraobserver reliability, like test-retest reliability, is a form of reproducibility where the same observer completes the same instrument on two or more occasions without changes in clinical status, whereafter the different ratings are compared. Interobserver reliability is defined as the extent to which the same results are obtained by two or more different observers/clinicians using the same instrument for the same patient on the same occasion. General recommendations when estimating reliability are to calculate the intraclass correlation coefficient (ICC) with 95% confidence intervals. An ICC value >0.70 is considered good in a sample size of at least 50 patients (48, 127, 134).

Internal consistency identifies the extent to which the items in an instrument are correlated to each other. This evaluation is thus a measure of the homogeneity of the items. Cronbach’s alpha (CA) is an adequate measure, where values between 0.70 and 0.95 are considered good with a sample size of at least 100 individuals. Too high value of CA indicates that the items are too highly correlated, which implies difficulties in the interpretation of results and provides the patient with several items with no additional information since the items capture the same disability (127).

Measurement error is the systematic or random error of a patient’s score. There are no clear guidelines stating what method to use when evaluating the measurement error (127, 134). One of the recommended methods is calculations and constructions of Bland Altman plots (12, 127). The plots are constructed from two sets of completed instruments. The Bland Altman plots show the mean values of the two measurements for each patient (x-axis) in relation to the differences of the same paired values (y-axis) and the data are described graphically. The sample size when conducting Bland Altman plots should be at least 50 patients. Another method that evaluates the measurement error is to calculate the intra-individual variability of the functional measures expressed as standard error of a single determination ($S_{method}$), together with the coefficient of variation (CV in %) for the score (33). The formula for the calculation of $S_{method}$ is: $S_{method} = \sqrt{\frac{\sum d_i^2}{2n}}$ where $d_i$ is the difference between the $i_{th}$ paired measurement and $n$ is the number of differences and the CV% is calculated as the $S_{method}$ divided by the overall mean (33).
Responsiveness

Responsiveness, in contrast to validity and reliability, includes only one measurement property (Figure 5). Responsiveness identifies the ability of a PRO instrument to detect changes over time. For this property the COSMIN group recommends also comparing the responsiveness with other validated PRO instruments and defining hypotheses in advance. Different kinds of statistical tests are proposed when evaluating responsiveness, but most studies recommend either Effect size (ES) or Standardized Response Means (SRM) with a sample size of at least 50 individuals. ES is calculated as the difference between the means before and after treatment, divided by the pre-treatment standard deviation (SD) of that measure. SRM is calculated as the difference between the means before and after treatment divided by the SD of the change. ES values >0.80 are considered large, 0.50 to 0.80 moderate, 0.20 to 0.49 small and <0.2 trivial (25, 48, 73, 77). SRM values are usually slightly lower than the corresponding ES values (81).

Interpretability

The interpretability of a PRO instrument is not a true measurement property, but important for the quality of the instrument. According to the COSMIN group (99), interpretability comprises three different concepts: (i) floor or ceiling effects, (ii) missing items and (iii) minimal clinically important difference (MCID) or minimal important difference (MID).

Floor or ceiling effects are considered to be present when more than 15% of the individuals reach the highest or lowest possible numeric value of a score. With existing floor effects it is impossible to demonstrate deterioration in the score when the patient is getting worse, and with a ceiling effect it is impossible to demonstrate an improvement in the score when the patient is improving in clinical status. The ideal PRO instrument has no floor or ceiling effects (99, 134, 140).

Missing items will occur in all PRO instruments, and how this should be handled for the specific instrument must be clarified in easy instructions for everyone who will use the instrument (75). Instruments can be incorrectly completed in a variety of ways. Patients may not fill in every item or mark two alternatives in the same item, and there must be rules to handle all types of mistakes. Another reflection is that, with too many missing items, the instrument could be an inappropriate instrument for that population in that many items could be too difficult to answer.

The third concept of interpretability is MCID, which identifies the extent to which score differences or changes are meaningful in respect of clinical benefits. MCID and normative data should therefore be determined for a PRO instrument in a
representative population. The *MCID* is the smallest difference in the instrument score that is perceived as being of clinical significance. With an adequately calculated *MCID*, it is possible to discuss and define a threshold when a treatment should be regarded as being an improvement of clinical significance (in contrast to statistical significance that has no relevance in respect of clinical significance). Available normative data can also be helpful when discussing whether an improvement by treatment is relevant and cost-beneficial (7, 26, 60, 126).

**Patient-friendliness**

Besides these quality assessments, there are other aspects that have to be considered. The PRO instrument should be easy to complete within a short time frame, something that makes the instrument patient-friendly (129). The instrument should also be easy to administer at the clinic or hospital where it is used. PRO instruments that require complicated weighted calculations lead to more work and time consumption for the professionals and the hospitals that administer the instruments. It can therefore be better in some situations for both the patients and the clinicians to use an instrument that is simpler, shorter and less complicated than a more time-consuming instrument, even if the former instrument is less sensitive. For use in national registries, it is important to choose an instrument which is patient-friendly and suitable, to attract as many individuals as possible to complete the instrument.

*Figure 6.*
Preparations made for hindfoot surgery
Summary of study subjects

In this thesis we used two different populations:

1. **In paper I** we included patients who were registered in the Swedish National Ankle Registry due to planned or performed surgery during the period 1 February 2008 to 31 January 2010. The included patients had also all completed the PRO instruments used in the registry.

2. **In papers II–IV** we included patients who were scheduled for surgery in the foot or ankle at the orthopedic departments of the two Swedish county hospitals in Kalmar and Eksjö during the period 1 January 2011 to 30 September 2013.

**Table 4.**
Summary of patients in the four studies

<table>
<thead>
<tr>
<th>Study</th>
<th>Patients (n)</th>
<th>Female/male (n/n)</th>
<th>Age (median)</th>
<th>Diagnoses</th>
<th>PRO instruments</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>135</td>
<td>74/61</td>
<td>63</td>
<td>Inflammatory arthritis or osteoarthritis in the ankle</td>
<td>SEFAS, FAOS, EQ-5D and SF-36</td>
</tr>
<tr>
<td>II</td>
<td>224</td>
<td>156/68</td>
<td>57</td>
<td>Arthritis in foot and ankle joints, tendon and ligament disorders, deformities in forefoot and hindfoot</td>
<td>SEFAS, FAOS, EQ-5D and SF-36</td>
</tr>
<tr>
<td>III</td>
<td>206</td>
<td>142/64</td>
<td>55</td>
<td>Arthritis in foot and ankle joints, tendon and ligament disorders, deformities in great toe and hindfoot</td>
<td>SEFAS and AOFAS</td>
</tr>
<tr>
<td>IV</td>
<td>21</td>
<td>15/6</td>
<td>60</td>
<td>Adult acquired flatfoot deformity (AAFD)</td>
<td>SEFAS, EQ-5D and SF-36</td>
</tr>
</tbody>
</table>
1. The Swedish National Ankle Registry

The Swedish Ankle Registry (www.swedankle.se) was started in 1997 to register patients nationwide with total ankle replacements (TARs). When the registry started, patients who had undergone TARs during the period 1993 to 1996 were included retrospectively. From the beginning, the registry included information about gender, diagnoses, date of surgery, type of prosthesis, per- and postoperative complications, simultaneous surgery, previously performed surgery in the same foot or ankle and the name of operating and assisting surgeons.

In 2008 the registry started to register also ankle fusions in Sweden, and in 2009 supramalleolar tibial osteotomies (SMO). Concerning ankle fusions, the registry included information about diagnoses, date of operations, per- and postoperative complications, simultaneous surgery, previously performed surgery in the same foot or ankle, the name of the operating surgeon, and also preoperative alignment, type of arthrodesis and fixation technique. For the SMO the same information as for the ankle fusions was included except that the type of wedge in the osteotomy was registered instead of the type of arthrodesis. In 2008 the registry also started to use PRO instruments in the evaluation of the patients. From this year the generic EQ-5D and SF-36 and the ankle-specific SEFAS and FAOS were provided before surgery and 6, 12 and 24 months after surgery. In 2010 the FAOS was excluded from the registry.

In recent years new data has been added to the registry. The height and weight of the patients, the ASA (American Society of Anesthesiologists) classification, the Charnley classification of arthropathy and comorbidity, smoking habits and – if used – the type of bone transplant, are now also included in the registry.

When the registry was evaluated on 31 December 2103, there were 1064 primary TARs, 297 revisions of TARs, 1345 ankle fusions and 112 revision ankle fusions reported. During the period when we included patients in study I, 142 primary TARs, 60 revisions of TARs, 217 ankle fusions and 29 revision ankle fusions were reported (62) (Appendix III) (www.swedankle.se).

2. The Kalmar and Eksjö cohort

The Kalmar and Eksjö cohort consists of included patients at the two orthopedic departments at these hospitals in Sweden. The patients were scheduled for and underwent surgery for different types of disorders in the foot or ankle by the foot and ankle surgeons at these hospitals and they were all included during the period 1 January 2011 to 30 September 2013. During this period, 609 surgical procedures were performed in the foot or ankle at the two hospitals. Of these, a total of 258 were included in papers II–IV.
Table 5a. Flowcharts Paper I-II

Paper I

135 patients
(74 women / 61 men)
with inflammatory arthritis or osteoarthritis in the ankle

SEFAS FAOS
EQ-5D SF-36

Construct validity
Floor or ceiling effects
Internal consistency

62 patients
40 women / 22 men
Test-retest reliability
Measurement error

37 patients
22 women / 15 men
Responsiveness

Paper II

224 patients
(156 women / 68 men)
with disorders in the forefoot, midfoot, hindfoot or ankle

SEFAS FAOS
EQ-5D SF-36

Construct validity
Floor or ceiling effects
Internal consistency

89 patients
62 women / 27 men
Test-retest reliability
Measurement error

142 patients
102 women / 40 men
Responsiveness
Table 5b. Flowcharts Paper III-IV

**Paper III**

- 206 patients (142 women / 64 men) with disorders in great toe, hindfoot or ankle

**SEFAS and AOFAS**

- Correlation between scores
- Floor or ceiling effects
- Internal consistency

- 106 patients (SEFAS)
  - 73 women / 33 men
  - Test-retest reliability
  - Measurement error

- 72 patients (AOFAS)
  - 50 women / 22 men
  - Intra-observer reliability
  - Inter-observer reliability
  - Measurement error

- 127 patients
  - 90 women / 37 men
  - Responsiveness

**Paper IV**

- 21 patients (15 women / 6 men) with acquired flatfoot deformity due to posterior tibial tendon dysfunction

**SEFAS EQ-5D SF-36**

- Absolute values in scores
- Responsiveness

- 21 patients
  - 15 women / 6 men
  - Absolute values in scores
  - Changes in scores
  - Responsiveness
Summary of statistical analysis

Statistical calculations were performed with Statistical Package for the Social Sciences (SPSS) software version 17.0 (IBM Software StatisticsR, Armonk, NY) and STATISTICA version 10.0 (Statsoft Inc., Tulsa, OK).

Standard procedures were used for descriptive statistics. Data are presented as means with standard deviations (SD), means with 95% confidence intervals (CI), median with range and/or as frequencies and proportions. The 95% confidence intervals (CI) for correlations were calculated according to Fisher’s z-transformation. We calculated the confidence intervals for ES and SRM according to the method described by Becker (9).

The more specific statistical analyses used in papers I–IV are described separately in the summary of methods and in the method sections in the summary of the papers.
Summary of ethical considerations

All the studies presented in this thesis were approved by the ethics committee of Lund University, Sweden (2009/698), and the studies were performed according to the Declaration of Helsinki. The patients registered in the national ankle registry were informed in writing or orally before registration (paper I). All the subjects in papers II–IV received oral and written information about the purpose and procedure of the study and written informed consent was obtained from these patients.
Summary of papers

Paper I

Validity, reliability and responsiveness of a self-reported foot and ankle score (SEFAS)

Aim
The first aim of this study was to translate, cross-culturally adapt and modify the New Zealand Ankle Questionnaire to a Swedish version. The second aim was to evaluate the measurement properties of the translated instrument, called the Self-reported Foot and Ankle score (SEFAS), in terms of validity, reliability, and responsiveness in patients with inflammatory arthritis or osteoarthritis in the ankle, and to compare the results with one region-specific and two generic PRO instruments.

Methods
The New Zealand national joint registry derived the New Zealand Ankle Questionnaire (NZAQ) from the validated Oxford hip score (OHS) and started to use this PRO instrument without validation in their registry in 2000. We translated and adapted SEFAS to Swedish conditions according to a standardized cross-cultural adaptation procedure, and then evaluated the measurement properties of SEFAS together with the Foot and Ankle Outcome Score (FAOS), the EuroQol 5 dimensions (EQ-5D) with EuroQol visual analogue scale (EQ-VAS) and the Short form 36 (SF-36).

Convergent and divergent construct validity were evaluated when SEFAS was compared with FAOS, EQ-5D and SF-36. Hypotheses were specified in advance concerning correlations between both SEFAS and subscores in the foot- and ankle-specific and the generic outcome measures. As pain and function seem to be the constructs of major interest among patients with foot and ankle disorders, we also defined pain-specific and function-specific items in SEFAS and related these separately to the specific subscores within the same constructs in the other instruments. Test-retest reliability was evaluated in a subgroup of the patients who completed all four PRO instruments twice 6 months after surgery with a short period of time between the two occasions. The measurement error was calculated from the
paired PRO instruments as Bland Altman plots and as calculations of the intra-individual variability expressed as standard error of a single determination (\(S_{\text{method}}\)) and as calculations of the coefficient of variation \((CV)\) (%) for the scores.

Internal consistency and floor or ceiling effects were calculated for all four PRO instruments.

Responsiveness was evaluated for all four PRO instruments using completed instruments before and 6 months after surgery. We used effect size (ES) and standardized response mean (SRM) for the calculations. We compared the results from the different instruments without defining any hypotheses in advance.

**Subjects**

We included 135 patients (74 women) with a median age of 63 (range 26–85) years, who were reported to the Swedish Ankle Registry. All patients had completed the PRO instruments SEFAS, FAOS, EQ5-D and SF-36, 69 patients before and 66 after the surgery with ankle fusions or ankle arthroplasties. The patients were included during the period 1 February 2008 to 31 January 2010.

First 78 patients were asked to complete the PRO instruments twice 6 months after the surgery. Twelve of them were excluded due to non-completed instruments, and the remaining four patients completed the second set of instruments too late to be usable. Therefore, only 62 patients (40 women) with a median age of 64 (range 26–85) years were included for the evaluation of reliability.

Of the 69 patients who had completed the instruments before surgery, 37 patients (22 women) with a median age of 65 (range 24–80) years had also completed the PRO instruments 6 (range 5–7) months after surgery. These 37 patients with pre- and postoperative data were included for the evaluation of responsiveness. Thirty-two patients had a follow up less than 6 months at the time of data analysis and they were therefore excluded from evaluation of responsiveness. Since the FAOS was excluded from the registry on 1 January 2010, we only achieved completed pre-and postoperative FAOS from 20 of the 37 patients.

Drop-out analyses revealed no statistically significant differences in age and gender distribution (data not shown) when comparing the baseline data of the subcohort that had completed the PRO instruments once and the subcohort that had completed the PRO instruments twice 6 months after surgery, and the subcohort that had completed the PRO instruments only before and the subcohort that had completed the PRO instruments both before and after surgery.

**Statistical analysis**

We calculated Spearman’s correlation coefficient when evaluating the convergent and divergent construct validity for SEFAS in relation to the other scores. Data were
presented as r-values with 95% confidence interval (95% CI) according to Fisher’s z-transformation.

We evaluated test-retest reliability by intraclass correlation coefficient (ICC). Bland Altman plots were constructed from two sets of instruments. The plots consist of the mean values of the two measurements for each patient (x-axis) and the differences of the same paired values (y-axis). The data are described graphically. Standard error of a single determination ($S_{\text{method}}$) expresses the measurement error in scoring points. $S_{\text{method}}$ was calculated together with the relative measurement error by the coefficient of variation (CV). For the floor or ceiling effects, descriptive statistics were used. We evaluated the internal consistency by Cronbach’s alpha (CA). We evaluated the responsiveness by effect size (ES) and standardized response mean (SRM).

**Results**

After SEFAS was translated and culturally adapted, the score could be used as an ankle-specific PRO instrument.

More than 70% of our predefined hypotheses could be confirmed when convergent and divergent construct validity were evaluated for SEFAS in comparison with the other scores or subscores. Strong correlations were confirmed between SEFAS (total score) and the subscores that dealt with the constructs pain and function. Weak correlations were confirmed between SEFAS (total score) and the subscores that dealt with different constructs such as mental health or general health. We also found strong correlations when the pain- and function-specific items in SEFAS were related separately to the specific subscores within the same constructs in the other instruments.

ICC for SEFAS was 0.92 (95% CI 0.87 to 0.95), indicating good test-retest reliability. SEFAS had an acceptable measurement error with $S_{\text{method}}$ of 2.7 and coefficient of variation (CV) of 15%. Cronbach’s alpha for SEFAS was 0.96 when the calculations were done in the group of patients (n=62) used for evaluation of test-retest reliability and the measurement error. SEFAS was the only PRO instrument with no floor or ceiling effects.

ES for SEFAS was 1.44, indicating good responsiveness (Table 7).

**Conclusion**

SEFAS has good measurement properties, with properties comparable to or better than FAOS, SF-36 and EQ-5D, in patients with inflammatory arthritis and osteoarthritis in the ankle.

**Erratum**

The FAOS was removed from the Foot and Ankle Registry in 2010, but in the original paper I it was incorrectly stated that the FAOS was removed in 2011.
Additional analyses

Internal consistency was reanalyzed for SEFAS after the publication when we achieved data from 135 patients. The Cronbach’s alpha was then 0.92, indicating good internal consistency.

Figure 8.
Ankle osteoarthritis before surgery and after surgery with ankle fusion

Paper II

Validity, reliability and responsiveness of the Self-reported Foot and Ankle Score (SEFAS) in forefoot, hindfoot and ankle disorders

Aim

The first aim was to assess the quality and evaluate the measurement properties of SEFAS in patients with other disorders in the ankle than arthritis, but also for disorders in the forefoot and hindfoot. The second aim was to compare the results with one region-specific and two generic PRO instruments.

Methods

We evaluated the measurement properties of SEFAS and compared them with the same three PRO instruments (FAOS, EQ-5D and SF-36) as in paper I.
Content validity was thoroughly evaluated for SEFAS before we started to use and validate it in the target population, i.e. in patients with both foot and ankle disorders and disabilities. We replaced “ankle” with “foot and ankle” in all questions, after which patients with disorders in the forefoot, hindfoot and ankle discussed the different questions or items together with research nurses. Orthopedic surgeons, physiotherapists and nurses from the orthopedic department provided oral and written comments on the questions. After the discussions in focus groups were completed, additional changes were made to the instrument, but none of the items were excluded. We finally let patients with disorders in forefoot, hindfoot or ankle assess the understandability and relevance of the questions.

Convergent and the divergent construct validity were evaluated as described in paper I, also with hypotheses specified in advance. The pain- and function-specific items in SEFAS were related separately to the specific subscores within the same constructs in the other instruments.

Test-retest reliability was calculated by use of two sets of instruments that were completed a week apart before surgery.

Measurement error, internal consistency, floor or ceiling effects and responsiveness were evaluated and calculated for all PRO instruments as described in paper I.

Time to complete the instruments was measured by our research nurse.

Subjects

We included 224 patients (156 women) with a median age of 57 (range 16–87) years in the Kalmar-Eksjö cohort with disorders in the forefoot or the hindfoot and/or ankle. The patients were included during the period 1 January 2011 to 31 January 2013. All patients had completed the PRO instruments SEFAS, FAOS EQ-5D and SF-36, before surgery and 142 (102 women) with a median age of 58 (range 18–81) years also 6 (range 5–7) months after surgery. Eighty-nine patients (62 women), with a median age of 57 (range 22–81) years had completed the PRO instruments twice before surgery. These patients were included for the evaluation of reliability. The 142 patients with pre- and postoperative data were included for the evaluation of responsiveness.

Drop-out analyses revealed no statistically significant differences in age, height, weight, BMI and gender distribution (data not shown) when comparing the baseline data of the subcohort that had completed the PRO instruments once and the subcohort that had completed the PRO instruments twice before surgery, and the subcohort that had completed the PRO instruments only before and the subcohort that had completed the PRO instruments both before and after surgery.
Statistical analysis

In paper II we used the same statistics as in paper I, with the exception of the correlation analyses, where we presented both Pearson’s correlation coefficient and Spearman’s correlation coefficient, following the specific demand from the reviewers and the editor. We also calculated the confidence intervals for ES and SRM.

Results

SEFAS was completed twice as fast as SF-36 and FAOS but more slowly than EQ-5D.

The items in SEFAS were regarded as easy to understand and relevant for the target population. More than 80% of our predefined hypotheses could be confirmed when convergent and divergent construct validity were evaluated for SEFAS (total score) in comparison with the other scores or subscores. We also found strong correlations when the pain- and function-specific items in SEFAS were related separately to the specific subscores within the same constructs in the other instruments.

ICC for SEFAS was 0.92 (95% CI 0.85 to 0.96) in forefoot patients and 0.93 (95% CI 0.88 to 0.96) in hindfoot/ankle patients. SEFAS had an acceptable measurement error with $S_{method}$ of 2.3 and coefficient of variation (CV) of 8% in forefoot patients and 2.4 and 13% in hindfoot/ankle patients. CA for SEFAS was 0.84 in forefoot and 0.86 in hindfoot/ankle patients. SEFAS was the only PRO instrument without floor or ceiling effects.

The responsiveness for SEFAS was good, with an ES of 1.29 (95% CI 0.95 to 1.63) in patients with forefoot disorders and 1.05 (95% CI 0.77 to 1.33) for patients with hindfoot or ankle disorders (Table 7).

Conclusion

SEFAS has good measurement properties, with properties comparable to or better than FAOS, SF-36 or EQ-5D, in patients with disorders in the forefoot, hindfoot or/and ankle.

Errata

Unfortunately there are errors in descriptive Table 1 and the methods section in paper II where one female patient with an Achilles tendon disorder (in the hindfoot/ankle group) is missing. The correct data are presented in the new Table 6 below. It should be noted that all statistical calculations and all other tables in paper II included the correct number of patients in each group.
Figure 9.
Patients with different hindfoot deformities before surgery

Table 6.
Patient data paper II

<table>
<thead>
<tr>
<th></th>
<th>Patients with forefoot disorders</th>
<th>Patients with hindfoot or ankle disorders</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>118</td>
<td>106</td>
</tr>
<tr>
<td>Age median (range)</td>
<td>57 (16-87)</td>
<td>55 (18-81)</td>
</tr>
<tr>
<td>Sex male (%)</td>
<td>22 (19%)</td>
<td>46 (43%)</td>
</tr>
<tr>
<td></td>
<td>female (%)</td>
<td>96 (81%)</td>
</tr>
<tr>
<td></td>
<td>60 (57%)</td>
<td></td>
</tr>
<tr>
<td>Height (cm) mean ± SD</td>
<td>169 ± 8.6</td>
<td>172 ± 10.8</td>
</tr>
<tr>
<td>Weight (kg) mean ± SD</td>
<td>74 ± 13.4</td>
<td>84 ± 15.6</td>
</tr>
<tr>
<td>BMI (kg/ m²) mean ± SD</td>
<td>26 ± 4.3</td>
<td>29 ± 4.7</td>
</tr>
<tr>
<td>Diagnosis Arthritis</td>
<td>4</td>
<td>33</td>
</tr>
<tr>
<td>Achilles tendon disorders</td>
<td>0</td>
<td>13</td>
</tr>
<tr>
<td>Flatfoot</td>
<td>0</td>
<td>26</td>
</tr>
<tr>
<td>Cavovarus/neurological</td>
<td>2</td>
<td>21</td>
</tr>
<tr>
<td>Big toe disorders</td>
<td>91</td>
<td>0</td>
</tr>
<tr>
<td>Lesser toe disorders</td>
<td>17</td>
<td>0</td>
</tr>
<tr>
<td>Others</td>
<td>4</td>
<td>13</td>
</tr>
<tr>
<td>Surgery Arthrodesis</td>
<td>10</td>
<td>38</td>
</tr>
<tr>
<td>Calcaneal osteotomy</td>
<td>0</td>
<td>32</td>
</tr>
<tr>
<td>Tendon surgery</td>
<td>0</td>
<td>18</td>
</tr>
<tr>
<td>Osteotomy first metatarsal</td>
<td>77</td>
<td>1</td>
</tr>
<tr>
<td>Surgery in lesser toes</td>
<td>18</td>
<td>0</td>
</tr>
<tr>
<td>Tendon transfers</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>Others</td>
<td>11</td>
<td>8</td>
</tr>
</tbody>
</table>
Paper III

Comparison of the Self-reported Foot and Ankle Score (SEFAS) and the American Orthopedic Foot and Ankle Society Score (AOFAS)

Aim
The aim of the study was to compare the two foot- and ankle-specific outcome measure instruments SEFAS and AOFAS using measurement properties and other quality assessments in patients with great toe disorders and hindfoot/ankle disorders.

Methods
We used the following approach when items were missing in AOFAS; (1) when results from more than 2 items are missing, the instrument is disregarded; (2) when results from 1 or 2 items are missing the mean result of the remaining items is used.

The patients completed AOFAS with assistance from specially trained physiotherapists and the physiotherapists also conducted the clinician-dependent part of the instrument. The patients completed SEFAS without help from the physiotherapists.

Time to complete the instruments was registered by our research nurse for SEFAS and by the physiotherapists for AOFAS.

We then compared the two instruments with correlation analyses. Test-retest reliability (SEFAS), intraobserver reliability (AOFAS) and measurement errors (both instruments) were calculated by the use of two sets of instruments that were completed a week apart before surgery. Interobserver reliability was calculated by use of AOFAS scores registered by two physiotherapists on the same day. Floor and ceiling effects and internal consistency were calculated for both instruments. Responsiveness was calculated by the use of two sets of instruments that were completed before and 6 months after surgery.

Subjects
We included 206 patients (142 women) with a median age of 55 (range 18–81) years who were reported in the Kalmar/Eksjö cohort with disorders in the great toe or patients with disorders in the hindfoot or/and ankle. All patients completed SEFAS and AOFAS before surgery and 127 (90 women) with a median age of 55 (range 18–81) years also 6 (range 5–7) months after surgery. Our primary intention was to finish the inclusion 31 January 2013, but at that time we had too few patients for the evaluation of reliability. We therefore extended the inclusion with 34 new participants (24 women) with a median age of 60 (range 26–76) years during the period 1 February 2013 to 30 September 2013. These patients completed the two
instruments twice one week apart. All individuals with two completed sets of instruments were used for the evaluation of reliability. By these inclusion criteria we reached 106 patients (73 women), with a median age of 56 (range 22–81) years, who completed SEFAS twice within a week and 72 patients (50 women), with a median age of 56 (range 22–80) years, who completed AOFAS twice within a week. Two physiotherapists conducted in 39 (25 women) patients separate AOFAS scores during the same day for the evaluation of interobserver reliability. The 127 patients with pre- and postoperative data were included for the evaluation of responsiveness.

Drop-out analyses revealed no statistically significant differences in age, height, weight, BMI and gender distribution (data not shown) when comparing the baseline data of the subcohort that had completed the PRO instruments once and the subcohort that had completed the PRO-instruments twice before surgery and the subcohort that had completed the PRO instruments only before and the subcohort that had completed the PRO instruments both before and after surgery.

**Statistical analysis**

We calculated Spearman’s correlation coefficient to compare the two scores SEFAS and AOFAS. Data were presented as r-values with 95% CI according to Fisher’s z-transformation. Intraobserver and interobserver reliability for AOFAS were calculated by ICC. All other analyses were done using the same statistics as in papers I–II.

**Results**

Patients completed SEFAS a mean 3 times faster than AOFAS.

SEFAS and AOFAS correlated with an r-value of 0.49 (95% CI 0.31 to 0.67) in patients with great toe disorders and 0.67 (95% CI 0.53 to 0.81) in patients with hindfoot/ankle disorders.

SEFAS test-retest ICC values were 0.89 (95% CI 0.81 to 0.94) for great toe and 0.92 (95% CI 0.87 to 0.95) in hindfoot/ankle, while the intraobserver ICC values for AOFAS were 0.57 (95% CI 0.29 to 0.77) for great toe and 0.75 (95% CI 0.58 to 0.86) for hindfoot/ankle. AOFAS interobserver ICC values were 0.82 (95% CI 0.50 to 0.94) for great toe and 0.71 (95% CI 0.46 to 0.86) for hindfoot/ankle. The method for SEFAS was 2.5 and for AOFAS 11.2 scoring points in patients with great toe disorders and in hindfoot/ankle patients 2.4 and 9.4, respectively. The corresponding CV values were 8.3% and 18.8% in great toe patients and 11.7% and 19.2% in hindfoot/ankle patients. The Cronbach’s alpha (CA) value for SEFAS was 0.86 in patients with great toe disorders and 0.85 in patients with hindfoot/ankle disorders and 0.15 and 0.42, respectively, for AOFAS. None of the instruments had any floor or ceiling effects.

ES was 1.39 (95% CI 1.01 to 1.77) in SEFAS in patients with great toe disorders and 1.15 (95% CI 0.85 to 1.45) in patients with hindfoot/ankle disorders, and 1.73 (95% CI 1.26 to 2.20) and 1.05 (95% CI 0.76 to 1.35) for AOFAS, respectively (Table 7).
Conclusion

SEFAS is easier to use than AOFAS for both patients and clinicians. The measurement properties for SEFAS were comparable to or better than for AOFAS in patients with great toe and hindfoot/ankle disorders.

Figure 10.
Before and after forefoot surgery

Paper IV

Surgery of Adult Acquired Flatfoot Deformity due to Posterior Tibial Tendon Dysfunction Reduces Pain and Improve Function and Health related Quality of Life

Aim

The first aim was to prospectively evaluate the patient-reported outcome using SEFAS as an instrument in patients who had underwent surgery for stage II adult acquired flatfoot deformity (AAFD) due to posterior tibial tendon dysfunction (PTTD). The second aim was to compare the outcome evaluated with SEFAS and
the generic instruments EQ-5D and SF-36. The third aim was to find out if any improvement proceeded from 6 to 24 months after surgery.

Methods
We used stage II AFFD due to PTTD as a model for testing the usability of SEFAS in patients who were scheduled for surgery with one specific disease at two county hospitals. We specifically focused on the ability to capture changes in pain, function and health-related quality of life with surgery. Body weight and height were measured by standard equipment for the calculations of body mass index (BMI). The medical charts and the local complication registries were used to identify complications, length of hospital stays and sick leave. SEFAS, EQ-5D, EQ-VAS and SF-36 were completed before, 6 (range 5–9) months and 24 (range 20–32) months after surgery. Before surgery we also asked if the patients were smokers or not, and at the 24-month follow-up visit, (i) whether they would with the same disability as preoperatively, once again choose to undergo the surgical procedure, and (ii) if they had been improved by surgery. The absolute mean values and changes in the summarized scores were calculated after 6 months and 24 months to estimate patient-reported outcome after surgery. Responsiveness was calculated from before to 24 months after surgery.

Subjects
We included 21 patients (15 women) with a median age of 60 (range 37–72) years who were reported in the Kalmar/Eksjö cohort scheduled for surgery due to AAFD caused by PTTD with no additional inflammatory disease or osteoarthritis in the hindfoot.

Drop-out analyses were not performed because only two patients did not complete the instruments adequately, making them unusable for the analyses.

Statistical evaluations
We used two-tailed paired t-test to calculate changes in absolute score values from preoperative to 6 months after surgery, from 6 to 24 months after surgery and from preoperative to 24 months after surgery. ES was calculated to evaluate responsiveness as in papers I–III.

Results
Twenty-four months after surgery, 5 patients had no complaints, 10 were much better, 3 were better, 2 were unchanged and 1 had deteriorated. There was a significant improvement between preoperative and 6 months postoperative status registered in SEFAS and SF-36 (BP, PF) and also between 6 and 24 months registered in SEFAS and SF-36 PF. Furthermore, all scores showed significant improvement registered from preoperative to 24 months postoperative status. Responsiveness was good for SEFAS, with an ES value of 1.76 (95% CI 1.04 to 2.48) (Table 7).
Conclusion

Surgery of stage II AFFD due to PTTD results in less pain and improved function and is well captured by SEFAS. The generic outcome instruments also captured improvements in HRQoL. We registered a further improvement between 6 and 24 months after surgery, indicating that a minimum follow-up of 2 years is needed to evaluate outcome.

Figure 11.
Before and two years after flatfoot surgery
Table 7.
Effect Size (ES) summarized for all studies

<table>
<thead>
<tr>
<th>Outcome measure instruments</th>
<th>Paper I Effect Size (95% CI)</th>
<th>Paper II Effect Size (95% CI)</th>
<th>Paper III Effect Size (95% CI)</th>
<th>Paper IV Effect Size (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ankle Fore foot Ankle / hindfoot Great toe Ankle / hindfoot AAFD</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SEFAS FAOS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pain</td>
<td>1.78 (0.93 - 2.63)</td>
<td>1.10 (0.78 - 1.42)</td>
<td>0.90 (0.66 - 1.14)</td>
<td></td>
</tr>
<tr>
<td>Symptom</td>
<td>0.47 (0.10 - 0.84)</td>
<td>0.38 (0.12 - 0.64)</td>
<td>0.40 (0.27 - 0.69)</td>
<td></td>
</tr>
<tr>
<td>ADL</td>
<td>1.36 (0.67 - 2.05)</td>
<td>0.76 (0.50 - 1.02)</td>
<td>0.85 (0.62 - 1.08)</td>
<td></td>
</tr>
<tr>
<td>Sport/Recreation</td>
<td>0.37 (-0.21 - 0.95)</td>
<td>0.82 (0.54 - 1.10)</td>
<td>0.70 (0.44 - 0.96)</td>
<td></td>
</tr>
<tr>
<td>Quality of life</td>
<td>1.38 (0.67 - 2.08)</td>
<td>1.48 (1.12 - 1.84)</td>
<td>1.57 (1.24 - 1.90)</td>
<td></td>
</tr>
<tr>
<td>EQ-5D</td>
<td>0.93 (0.48 - 1.38)</td>
<td>0.95 (0.60 - 1.30)</td>
<td>0.57 (0.28 - 0.86)</td>
<td>0.72 (0.34 - 1.10)</td>
</tr>
<tr>
<td>EQ-VAS</td>
<td>0.65 (0.14 - 1.16)</td>
<td>0.46 (0.21 - 0.71)</td>
<td>0.34 (0.06 - 0.62)</td>
<td>0.58 (0.04 - 1.12)</td>
</tr>
<tr>
<td>SF-36 Physical functioning (PF)</td>
<td>0.67 (0.30 - 1.04)</td>
<td>0.94 (0.63 - 1.25)</td>
<td>0.73 (0.51 - 0.95)</td>
<td>0.92 (0.41 - 1.43)</td>
</tr>
<tr>
<td>Role limitations - physical (RP)</td>
<td>0.23 (-0.23 - 0.69)</td>
<td>0.44 (0.16 - 0.72)</td>
<td>0.51 (0.24 - 0.78)</td>
<td></td>
</tr>
<tr>
<td>Bodily pain (BP)</td>
<td>1.25 (0.68 - 1.82)</td>
<td>1.57 (1.17 - 1.97)</td>
<td>0.73 (0.47 - 0.99)</td>
<td>1.76 (1.10 - 2.42)</td>
</tr>
<tr>
<td>General health (GH)</td>
<td>-0.04 (-0.24 - 0.16)</td>
<td>0.07 (-0.13 - 0.27)</td>
<td>0.02 (-0.15 - 0.19)</td>
<td></td>
</tr>
<tr>
<td>Vitality (VT)</td>
<td>0.14 (-0.22 - 0.50)</td>
<td>0.45 (0.19 - 0.71)</td>
<td>0.35 (0.15 - 0.55)</td>
<td></td>
</tr>
<tr>
<td>Social functioning (SF)</td>
<td>0.31 (0.08 - 0.70)</td>
<td>0.37 (0.14 - 0.60)</td>
<td>0.24 (0.02 - 0.46)</td>
<td></td>
</tr>
<tr>
<td>Role limitation - emotional (RE)</td>
<td>0.44 (-0.04 - 0.92)</td>
<td>0.27 (-0.01 - 0.55)</td>
<td>0.24 (0.02 - 0.46)</td>
<td></td>
</tr>
<tr>
<td>Mental health (MH)</td>
<td>0.07 (-0.22 - 0.36)</td>
<td>0.15 (-0.09 - 0.39)</td>
<td>0.20 (0.00 - 0.40)</td>
<td></td>
</tr>
<tr>
<td>AOFAS HMTP scale</td>
<td></td>
<td></td>
<td></td>
<td>1.73 (1.26 - 2.20)</td>
</tr>
<tr>
<td>A-HF scale</td>
<td></td>
<td></td>
<td></td>
<td>1.05 (0.65 - 1.23)</td>
</tr>
</tbody>
</table>
General discussion

Outcome assessment and outcome instruments

“When you can measure what you are speaking about, and express it in numbers, you know something about it; but when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind: it may be the beginning of knowledge, but you have scarcely, in your thoughts, advanced to the stage of science, whatever the matter may be.”

[William Thomson, Lord Kelvin (1824–1907)]

This view has been generally adopted in medical science and scores have been used for decades for clinical evaluation in different medical disciplines. Examples of such scores are the Apgar score to assess new-born babies, the Glasgow Coma Scale (GCS) to assess level of consciousness, the Injury Severity Score (ISS) to assess the overall severity of injury, the Systemic Inflammatory Response Syndrome (SIRS) Criteria to assess sepsis, CHADS2-VASc to assess risk of stroke and the APACHE II Score to assess the severity of disease for adult patients admitted to intensive care units. Many of these scores have been developed for clinicians in their assessments of patients and do not take the view of the patient into account. Since the patients’ perspective on health has gradually come to be regarded as being of major importance, PRO instruments, as complements to clinical measurements, have gained increased attention.

The information we receive from a completed PRO instrument provides information about a variety of aspects of the patient’s health and quality of life. However, in order to make comparisons before and after treatment, between different treatments and between sub-groups of patients, it is necessary to sum up the different answers the patients provide to a summarized score, either as several subscores or as a total score. The summarized score, like the scores developed for clinicians above, is used to capture different constructs such as pain or function, but actually also the total clinical appearance of the patient. Such scores can then be used in clinics, in hospitals or in national registries and databases in order to assess the actual disability of the patient and follow the efficacy of surgical or non-surgical treatments. With standardized scores collected in the same way, it is also possible to compare the outcome of treatments in different clinics and hospitals and between different
surgeons. If using generic PRO instruments, it is also possible to compare totally different diseases and treatments by using general health as the final outcome variable. This estimate is also of great value when calculating the cost-benefit of different treatments and estimating quality-assessed life years (QUALY).

The region- or disease-specific PRO instruments provide items or questions adapted for a specific patient population. These outcome instruments are focused on a defined aspect of health and from the beginning they were often developed to focus on changes with treatments. For this reason these instruments are often more sensitive in capturing changes in status than the generic scores. For example, De Vries et al. (44) reported that the disease-specific outcome measure VascuQol was better than EQ-5D and SF-36 at detecting changes after treatment for peripheral arterial diseases, and Brazier et al. (16) found that the osteoarthritis (OA) specific instrument WOMAC was more responsive than EQ-5D and SF-36 in capture changes with surgery after OA. We similarly found, as reported in papers I and II, better responsiveness for SEFAS than for EQ-5D and SF-36 when evaluating changes in several patient populations subjected to foot and ankle surgery and in paper IV better responsiveness for SEFAS (but also SF-36 PF and SF-36 BP) than for EQ-5D and EQ-VAS when following changes in status from before to 24 months after surgery in patient with AAFD due to PPTD.

Should both generic and specific outcome instruments be used?

Recommendations mostly suggest that when evaluating medical treatment such as surgery, these evaluations should include both generic and specific PRO instruments. Generic outcome instruments provide questions that capture general health and as such allow comparisons across populations, across diseases and across different treatments. Examples of generic PRO instruments are SF-36, EQ-5D and NHP, instruments that have also been shown to include the ability to capture changes after treatments (20, 44, 114, 115, 143, 144). Since there are also population reference data (normative data) for the generic instruments EQ-5D and SF-36, it is for these instruments possible to compare data from a specific patient population with age- and/or gender-specific data from the general population. It is thus possible to estimate whether a surgical procedure leads to normal health or not. Similar aspects must be taken into account today with the limited resources in the health care system. It could actually be better to choose a cheaper method with less complications but only a slightly inferior outcome than a method with higher costs and risk of complications, but with marginally superior outcome. With the help of the PRO instruments it is also possible to compare the efficacy of different treatments in different medical
disciplines in relation to costs, most relevant information when prioritizing resources between medical disciplines within the health care system (46, 132).

**Which generic outcome measure should be used?**

In the orthopedic foot and ankle discipline, we are not only interested in defining the disability of the patient. We are also most interested in defining the outcome of both non-surgical and surgical procedures. The responsiveness, measured with for example ES, is then an expression of major interest. This expression tells us whether the outcome instrument can capture the surgically or non-surgically induced improvement or not. In papers I and II, when evaluating patients with several different disorders in the foot and ankle, we found that SF-36 had favorable ES in physical function (PF) and bodily pain (BP), indicating good responsiveness in these subscores, but not in the other 6 subscores (Table 7). However, we also found that SF-36 had a high percentage of floor and ceiling effects, indicating that changes in the clinical status of a patient may be missed by the subscores in SF-36. Similar conclusions have been drawn in other studies evaluating patients with musculoskeletal disorders (20). This indicates that SF-36 may have limitations when evaluating the efficacy of treatment in these types of diseases. The appropriate PRO instrument should also be short and quick and easy for the patient to complete. This is another disadvantage of SF-36, since the many items (n=36) in the instrument may lead to the instrument being insufficient completed, with a high degree of missing answers.

The ES for EQ-5D (but not EQ-VAS) for the patients in papers I–II were large (>0.8) or moderate (>0.5), indicating good or acceptable responsiveness. For EQ-5D, in contrast to SF-36, we found a low proportion of individuals with floor and ceiling effects, similar to data presented by Maher et al. (86) when evaluating elective fore- and midfoot surgery. The EQ-5D instrument in paper II was also completed much faster than the SF-36 instrument. Thus, it seems as if EQ-5D is a generic PRO instrument that has many favorable properties compared to SF-36 when evaluating health status associated with elective foot and ankle surgery.

**Which foot- and ankle-specific instrument should be used?**

Several foot- and ankle-specific outcome instruments are available, but few have undergone complete evaluation of the measurement properties. At present there is no generally accepted specific outcome measure instrument for patients with musculoskeletal disorders in the foot and/or the ankle. The internationally recommended measurement properties for SEFAS were therefore evaluated in papers
I to IV for patients with a variety of musculoskeletal disorders and surgical procedures in the forefoot, the hindfoot and the ankle, with the intention of finding the appropriate outcome measure usable for both simple and complex disabilities. We then found good or acceptable convergent and divergent validity, test-retest reliability, measurement error and responsiveness. There were no floor or ceiling effects. We further found that SEFAS was patient-friendly and easy to administer in the orthopedic departments where it was tested. The simplicity of SEFAS is an advantage that makes SEFAS suitable to use in large national registries.

When we compared SEFAS with foot- and ankle-specific instrument FAOS in papers I and II, we found that SEFAS in these patient populations had measurement properties comparable to or better than FAOS. Comparing these two specific PRO instruments, then, SEFAS seems to be the instrument to recommend when evaluating foot and ankle disorders. In paper III, when we compared SEFAS with the foot- and ankle-specific instrument most used in the literature and research, AOFAS, we found that SEFAS was completed faster than AOFAS and that SEFAS was superior in reliability to AOFAS. We could also confirm the disadvantages of AOFAS as being a partly clinician-dependent instrument, making it more difficult to administer than SEFAS. The clinician-based part in AOFAS also makes this instrument less suitable in national registries. In conclusion, all these quality assessments spoke in favor of SEFAS in comparison with AOFAS.

Another strictly patient-reported foot- and ankle-specific instrument that is adequately evaluated with good outcome measures, and thus could also be used when assessing surgery in the foot and ankle, is the Manchester Oxford Foot Questionnaire (MOXFQ) (35-37). MOXFQ consists of 16 different items, which are summarized in three separate subscores. Each subscore is then converted to a nominal scale from 0 to 100, with 0 representing the least and 100 the most severe disability. The minimal clinically important difference (MIDC) and the minimal important change (MIC) for the 3 subscores have also been determined (34, 38). The disadvantages of MOXFQ include the fact that the score is reversed in relation to almost all other scores and that a license is needed to use this PRO instrument. However, in spite of being thoroughly evaluated, SEFAS and MOXFQ have not been extensively used in the literature. These instruments will probably be more reported in the future, as MOXFQ is endorsed by the British Foot and Ankle Society for use when measuring surgical outcome after foot and ankle surgery, and as SEFAS is supported by the Swedish Foot and Ankle Society, used in the Swedish Ankle Registry and planned to be used in the forthcoming Swedish National Foot and Ankle Registry, Riksfot.
Advantages and disadvantages of SEFAS

In this thesis we found that SEFAS is an appropriate outcome instrument when evaluating foot and ankle disorders.

Advantages that could be mentioned are: (i) SEFAS is validated with good results in patients with both simple disabilities and complex deformities in the forefoot, hindfoot or the ankle that significantly affect activities of daily living; (ii) SEFAS is simple to complete in a short time; (iii) SEFAS includes only 12 items with 5 response categories for each item, which makes it easy to sum up to a total score without any complicated conversions; (iv) SEFAS has an easy and clear algorithm for handling missing data; (v) SEFAS consists of one common score without subscores, which makes it possible to capture several aspects of the patient’s health status within one score; (vi) no copyright license is required for SEFAS; (vii) the use of SEFAS in the Swedish National Ankle Registry and in the planned Swedish Foot and Ankle Registry will increase the chances for the instrument to be used in large populations and to be published.

Disadvantages that could be mentioned are: (i) SEFAS is hitherto only validated in Swedish (although an English version exists); (ii) SEFAS is not extensively used in the literature; (iii) SEFAS has no subscores, although the items are related to the two major constructs pain and function; (iv) no population reference data (normative data) for SEFAS is available; (v) SEFAS has not been exposed to an adequate Rasch analysis; (vi) SEFAS has not been evaluated in respect of minimal clinically important difference (MICD), making it not possible to define clinical significant differences; (vii) SEFAS lacks items directed at prolonged standing and shoe fitting, which are items that patients with foot and ankle disabilities have regarded as important (5).

It can be discussed if more items should be added to the SEFAS, but also if the instrument should consist of subscores including different constructs important for patients with foot and ankle disabilities (5, 106). At this moment we have chosen to have fewer items making the instrument more easy to use in a registry, but of course this can make the instrument less sensitive. More studies have to be performed assessing if SEFAS is sufficient sensitive for our purpose, i.e. to evaluate all kinds of foot and ankle surgery also in registries.
Are there new foot- and ankle-specific outcome measures under development?

There is ongoing work worldwide to find the optimal PRO instrument for evaluating the outcome after foot and ankle surgery. In the United States researchers are currently developing computerized adaptive testing (CAT) health instruments (118). These instruments should be short and simple to use for the patient, yet capable of capturing the entire health status. CAT is based on the “item response theory”, a concept not used or evaluated in this thesis, and not explained or discussed in detail. In summary, this theory includes a theoretical background where the outcome instrument is dependent on a computer and a collected question bank for different health domains. For example, when a lower limb disability is to be tested, the patient is directed to a question bank with specific questions about disabilities and disorders in the lower limb. Using a CAT instrument each patient gets individually selected questions from the question bank, each new question being selected from the entire bank based on the patient’s responses to the previous question. This approach means that each patient will answer a specific combination of questions based on the answers provided. With this method, the total number of questions will often be less than 10 and the instrument is then quick to finish. Physical function (PF) CAT is an example of such an instrument, which includes a bank of 124 different questions (69). PF CAT has previously been evaluated in the orthopedic population, with good measurement properties, and is currently being evaluated for a variety of specific foot and ankle disorders with the support of the American Orthopedic Foot and Ankle Society. Future research will tell us whether this CAT-based instrument has a role in the evaluation of foot- and ankle-specific disorders (68, 70, 71).

There is also work within the European Foot and Ankle Society (EFAS) with the goal of identifying a foot- and ankle-specific PRO instrument that could be generally used when following the results of foot and ankle surgery. The “Score Committee” within the EFAS is developing a totally new PRO instrument that could be used in different countries and administered in different languages. The items in this instrument are developed using patients with a variety of foot and ankle disabilities in different countries in Europe (personal communication). However, currently there is no data to support the superiority of such a score in comparison with existing scores.

With all this work going on, it is understandable that there is a debate about which instrument to use as the foot- and ankle-specific outcome standard, often with each research group claiming the superiority of its own score. As for example, the Oxford group, which developed the MOXFQ, commented on our paper I (42).

There is also still low knowledge regarding the available validated foot- and ankle-specific PRO instruments, probably explained by the complexity of the development and evaluation of such instruments. For example, Witteveen et al. (146) recently
published a manuscript in *Foot and Ankle Surgery* (FAS) where they qualitatively evaluated the opinions of patients and surgeons about several topics concerning ankle OA. In this manuscript they compared and discussed some existing outcome measures, without including SEFAS. This is notable since SEFAS was first validated in our series of papers in patients with ankle OA (paper I). Witteveen et al. ended their paper with the recommendation that “the next step is to create, test and validate a new PROM for ankle OA patients”. This is the reason why we were trying to inform the foot and ankle society about SEFAS as an available and validated foot- and ankle-specific PRO instrument by sending a letter to the editor of FAS (29).
Strengths and limitations

The strengths of the thesis are the structural and thorough evaluations of the measurement properties of SEFAS in patients with a variety of disorders and disabilities and surgical procedures in the foot or ankle, with comparison of the results with established generic and foot- and ankle-specific instruments. The total samples size must be regarded as a study strength, but it would have been advantageous if we also had been able in papers II and III to achieve power for separate evaluations of, for example, hindfoot and ankle disorders and to conduct gender-specific evaluations and evaluations of one specific diagnosis.

The limitations include the small sample size in paper IV. This paper should therefore be regarded as a pilot study. It would also have been preferable to follow the patients in paper IV as long as 5 year after surgery, since we by our data were unable to state if improvement does continue beyond the first 2 postoperative years.

Another limitation is that no patients with acute disabilities were included in the different studies, and SEFAS should therefore be evaluated in such patients before the instrument can be recommended to use for these patients. It would also have been of interest to evaluate SEFAS in patients with non-operative treatments.

SEFAS is by now only validated in the Swedish language and this is a limitation of the instrument. SEFAS ought to be translated and validated in different languages and in other countries before the instrument can be used worldwide.

Other limitations are that we did not estimate the minimal clinically important difference (MCID), a concept used to define the smallest meaningful score change. The MICD of the SEFAS has to be estimated in the future.

Furthermore, no age- and gender-specific normative data were collected for SEFAS, something that also ought to be done in the future.
Conclusions

We translated and culturally adapted the items in the Self-reported Foot and Ankle score (SEFAS) (paper I) to Swedish, and then evaluated the measurement properties of the instrument using international guidelines. This work gave us the possibility to achieve our general aim of this thesis. We found that SEFAS is a patient-reported outcome (PRO) instrument as good as or better than the region-specific instruments Foot and Ankle Outcome Score (FAOS) and American Orthopedic Foot and Ankle Society (AOFAS) score, and the generic instruments EuroQol 5 Dimensions (EQ-5D) and Short form 36 (SF-36), when evaluating foot and ankle surgery in patients with inflammatory arthritis, osteoarthritis (paper I) and other disorders in the forefoot, hindfoot and/or ankle (papers II-IV).

In the thesis we were also able to answer our specific research questions, in that we can state that

I. After translation, cultural adaptation and evaluation of the content validity, SEFAS is proven to have questions relevant for the target population, and be easy to understand and to use (I-II).

II. The assessed quality including the measurement properties are good for SEFAS in patients with inflammatory arthritis or osteoarthritis in the ankle (I).
III. The assessed quality of SEFAS, including the measurement properties, is similar to or better than for EQ-5D, SF-36 and FAOS in patients with inflammatory arthritis or osteoarthritis in the ankle (I).

IV. The assessed quality including the measurement properties of SEFAS are good in patients with forefoot, hindfoot and ankle disorders (II).

V. The assessed quality including the measurement properties of SEFAS are similar to or better than for EQ-5D, SF-36 and FAOS in patients with forefoot, hindfoot and ankle disorders (II).

VI. The assessed quality including the measurement properties of SEFAS are good and similar to or better than for AOFAS in patients with great toe, hindfoot and ankle disorders (III).

VII. SEFAS is an appropriate PRO instrument, as good as or even better than EQ-5D and SF-36, when evaluating the outcome after surgery for acquired flatfoot deformity (IV).

Based on these findings, we infer that SEFAS is more useful when evaluating foot and ankle-specific disorders than the region-specific FAOS and AOFAS and the generic EQ-5D and SF-36. We therefore recommend SEFAS for use as a foot- and ankle-specific PRO instrument when conducting research and clinical evaluation of foot and ankle surgery also within a national registry.
Future perspectives

Normative data

A major limitation is that no age- and gender-specific normative data for SEFAS exist. In paper IV, where we used SEFAS to evaluate the disability before and after surgery in patients with adult acquired flatfoot deformity (AAFD) due to posterior tibial tendon dysfunction (PPTD), normative data would have helped us to find out whether patients improved and became as good as the general population. Due to this limitation, there are ongoing studies by our research group with the purpose of developing Swedish age- and gender-specific normative data for SEFAS.

Rasch analysis

A method that has attracted more and more interest when evaluating PRO instruments is the Rasch analysis. One purpose of this analysis is to ensure that the scores or subscores are unidimensional, i.e. contain items that correspond to the same construct (83, 133). Rasch analyses can be performed in already existing PRO instruments, often resulting in some items being removed and/or the numbers of the response alternatives in the items being changed, for example, from four to three (103). A Rasch analysis has been performed for SEFAS for the total score and work is now in progress to include Rasch analyses separately for the two constructs pain and function in the score. These analyses will be presented in future reports. Since Rasch analyses were not included in this thesis, we do not describe or discuss these analyses further.

Minimal clinically important difference (MCID)

When assessing changes in patients over time it is of great importance to assess whether statistically significant changes also correspond to changes of clinical significance. It is then essential to realize that there is no agreement between researchers regarding the best method for determining MCID or MID. Different
methods are currently in use for estimating the smallest change in an outcome instrument that could be considered as being of clinical importance (6, 7, 26, 125, 142). One of these methods is the anchor-based method that compares the changes in the PRO instruments with other measures of change, as an “anchor”. Most commonly the “anchor” is the patients’ answer to a global assessment rating where patients rate themselves to some extent as “better,” “unchanged,” or “worse” after the treatment. Currently there are no data to indicate what absolute score value of SEFAS could be regarded as being MCID, but data regarding patient global assessment of treatment effect for SEFAS exist and our research group has initiated a study that will estimate this expression for SEFAS.

Spreading SEFAS to the medical community

Unfortunately, extensive evidence of a PRO instrument’s good quality is not sufficient to get the instrument generally used. There are examples of a great many outcome measures that are validated as being good in well-conducted studies, yet are seldom used (72, 90, 93). For this reason it is important to in different ways spread knowledge of a new PRO instrument.

First the original researchers must spread the knowledge of the PRO instrument and the quality assessments performed. We have done this, as shown in appendix II, in a review article concerning outcome measures and the validation process including results of SEFAS in the Journal of the Swedish Medical Association (Läkartidningen) (31). SEFAS and the outcome of the different validation studies were also presented, as shown in appendix III, in the German Journal of Foot and Ankle Surgery (Fuss & Sprunggelenk) (62). Furthermore, our letters and responses to the editor of the journal Foot and Ankle Surgery (FAS) (29) and Acta Orthopédica (42) are other examples of work with the aim of informing foot and ankle surgeons of the availability of a new validated foot- and ankle-specific PRO instrument. Schrier et al. recently published a review article, in which the authors found that SEFAS was a suitable PRO instrument for assessment of hallux valgus treatment (119).

Another concern with PRO instruments is that they have to be available and validated in different languages. The English-speaking population is large and therefore is it important and urgent to spread knowledge about the SEFAS to researchers who are willing to validate the instrument in English. Obviously, is it also important to pursue the same process for other languages, such as German and French. This work is ongoing. Recently the German Journal of Foot and Ankle Surgery published data from the Swedish National Ankle Registry, shown in appendix III, including a presentation of SEFAS and outcome of the different validation studies. The NZAQ, the original English version of SEFAS, is still used in the New Zealand Joint Registry as a part of the post-operative evaluation of ankle replacements and they refer to
SEFAS and OHS when they present their patient-reported results (137). Foot and ankle surgeons from different countries have shown their interest in starting the translation and validation process of SEFAS.

Another research facility that we hope could help us to spread knowledge regarding SEFAS is the start of the National Swedish Foot and Ankle Registry (Riksfot) supported by the Swedish Foot and Ankle Society. All surgical procedures in the foot and ankle performed in Sweden are planned to be included in this registry and SEFAS will then be used as the region-specific PRO instrument. Since around 20 000–25 000 elective foot and ankle procedures are performed per year in Sweden, according to the Swedish National Board of Health and Welfare (Socialstyrelsen), the registry has the possibility, in just a few years, to be of great value and to identify the efficacy of different surgical procedures and then also to be of value in spreading the usability of SEFAS.

Figure 12.
How do we make this patient satisfied with the surgery?
Summary in English

Introduction

Clinicians in the past have relied on clinical measurements when assessing the outcome of treatments, but the patient’s perspective has gradually gained more focus. Patient-reported outcome (PRO) instruments have therefore been developed in all medical disciplines, with generic PRO instruments to evaluate general health and specific PRO instrument focusing on particular anatomical regions or specific diseases. The generic instruments are then often used to compare health in populations with different diseases or different affected regions, while the specific instruments often are better at capturing changes over time, such as after treatments. Currently there are several widespread region-specific PRO instruments that evaluate wrists, elbows, shoulders, knees, hips or the spine, but there is no generally accepted and fully validated specific PRO instrument for the foot and ankle. The aim of this thesis was to translate the New Zealand Ankle Questionnaire into a Swedish version \textit{[Self-Reported Foot and Ankle Score (SEFAS)]} and to evaluate measurement properties in respect of validity, reliability, measurement error and responsiveness of SEFAS in patients with disabilities in the foot and ankle. The aim was also to compare SEFAS with the \textit{Foot and Ankle Outcome Score (FAOS), EuroQol 5 Dimensions (EQ-5D), Short form 36 (SF-36), and the American Orthopedic Foot and Ankle Society (AOFAS) score.}

Methods

In \textit{study I}, 135 patients (median age 63, range 26–85) registered in the Swedish Ankle Registry as having had surgery in the ankle joint due to inflammatory arthritis or osteoarthritis (OA) completed SEFAS, FAOS, EQ-5D and SF-36. The instruments from the 62 patients who had completed the instruments twice within a short interval were used to evaluate reliability, and the instruments from the 37 who had provided both pre- and postoperative answers to evaluate responsiveness. In \textit{study II}, 224 patients (median age 57, range 16–87) with different diseases scheduled for different types of surgery in the foot or ankle completed SEFAS, FAOS, EQ-5D and SF-36. The instruments from the 89 patients who completed them twice within a week were
used to evaluate reliability, and the instruments from the 142 who had provided both pre- and postoperative answers were used to evaluate responsiveness. In study III, 206 patients (median age 55, range 18–81) with different diseases scheduled for different types of surgery in the great toe, hindfoot or ankle completed SEFAS and AOFAS. Instruments from the 106 patients who completed SEFAS and the 72 who completed AOFAS twice within a week were used to evaluate reliability, and the instruments from the 127 who had provided both pre- and postoperative answers were used to evaluate responsiveness. In study IV, 21 patients (median age 60, range 37–72) with stage II adult acquired flatfoot deformity (AAFD) due to posterior tibial tendon dysfunction (PPTD), scheduled for different types of surgery based on the flatfoot deformity, completed SEFAS, EQ-5D and SF-36 before surgery and 6 and 24 months after surgery.

Results

SEFAS and EQ-5D were completed faster than the other PRO instruments. The studies supported our predefined hypotheses that there would be strong correlations between SEFAS and the other scores or subscores that dealt with similar constructs such as pain and function, and weak correlations between SEFAS and the scores or subscores measuring different constructs such as mental health or general health. SEFAS, FAOS, EQ-5D and SF-36 all had good reliability in terms of test-retest reliability and internal consistency, while AOFAS was inferior to SEFAS as regards reliability. Visual inspection of the Bland Altman plots and calculation of the standard error of a single determination (S\text{method}) indicated that the measurement error was acceptable for SEFAS. SEFAS and AOFAS were the only instruments without any floor or ceiling effects. The foot- and ankle-specific instruments SEFAS, FAOS and AOFAS were better at detecting changes after surgery than the generic SF-36 and EQ-5D instruments. Surgery of stage II AAFD due to PPTD improved the patients markedly. SEFAS was able to capture the improvement by surgery in these patients and the improvement continued up to 24 months after surgery.

Conclusion

SEFAS is a PRO instrument, quick to complete, with good measuring properties in respect of validity, reliability and responsiveness. For SEFAS these properties are similar to or better than the properties for the foot- and ankle-specific instruments FAOS and AOFAS and the generic instruments EQ-5D and SF-36. We recommend SEFAS as a PRO instrument when evaluating surgery in the foot and ankle, not least within national registries.
Svensk sammanfattning

Varje år utförs i Sverige 20 000-25 000 planerade kirurgiska ingrepp i fot- och fotled. Resultatet av dessa bör värderas bättre än vad som sker i dag, gärna i ett nationellt register. Traditionellt har man inom ortopedin utvärderat kirurgiska ingrepp med hjälp av röntgenbilder och/eller genom att mäta eller bedöma t.ex. muskelstyrka och ledrörlighet. Om denna utvärdering genomförs av opererande läkare är det en risk att bedömningen inte blir objektiv då operatörens bedömning ofta styrs mot det resultat man önskar uppnå. Dessutom är det visat att uppmätt muskelstyrka och ledrörlighet korrelerar dåligt till hur patienterna själv upplever sin funktion och livskvalitet. En elitidrottsman, en yngre och en äldre individ upplever t.ex. samma muskelstyrka på olika sätt.

Som en följd av ovanstående har patientrapporterade utvärderingsinstrument (frågeformulär) vunnit gehör. Dessa instrument innehåller frågor som utvärderar individens smärta, funktion och hälsorelaterade livskvalitet. Frågorna är utformade så att de passar den patientgrupp och de kirurgiska ingrepp som ska värderas. Instrumentet kan vara patientrapporterat (innehåller bara frågor till patienten) eller undersökarberoende (innehåller även frågor som kräver en undersökning av medicinsk personal).

På engelska, men även på svenska, används begreppet PROM (Patient-Reported Outcome Measure) när man beskriver utvärderingsinstrument som åskådliggör patientens upplevda besvär. Ett strikt PROM innehåller ingen klinisk undersökning och är därför ett lämpligt utvärderingsinstrument i nationella patientregister, men också för att utvärdera verksamheten på en klinik.

Utvärderingsinstrument kan delas in i allmänna (eller generiska), som värderar det allmänna hälsotillståndet, och specifika, som värderar en specifik kroppsdel eller en specifik sjukdom. Specifika instrument är i regel känsligare än de generiska när det gäller att fånga förändringar. Detta är gynnsamt när man vill utvärdera resultatet av kirurgi eller annan typ av behandling. De generiska formulären är däremot att föredra om man skall jämföra olika sjukdomar eller åkommor.

Utvärderingsinstrumenten består oftast av ett antal frågor med olika svarsalternativ, där varje svar ger poäng. En totalsumma, en summascore, kan räknas ut genom att lägga ihop de olika frågornas poäng i ett ifyllt instrument.

![Diagram](image)

**Figur 1.**
Psykometriska egenskaper som bör testas i ett patient-rapporterat utvärderingsinstrument
För att kunna jämföra olika behandlingsmetoder vid fot- och fotledsåkommor mellan vårdenheter, sjukhus och länder finns behov av ett kvalitetsgranskat utvärderingsinstrument som är specifikt för fot och fotled. Något sådant allmänt accepterat utvärderingsinstrument finns inte. Därför startades detta avhandlingsprojekt med syfte att översätta och utvärdera ett fot- och fotledsspecifikt frågeformulär från Nya Zeeland som hos oss fick namnet SEFAS (Self-reported Foot and Ankle Score).

I delarbete I översattes och modifierades SEFAS till svenska förhållanden. Vi fann att SEFAS validitet, reliabilitet och responsivitet hos patienter med artros eller artrit i fotleden var jämförbara eller bättre än de jämförda instrumenten EuroQol 5 Dimensions (EQ-5D), Short form 36 (SF-36) och Foot and Ankle Outcome Score (FAOS).

I delarbete II utfördes samma utvärderingar av patienter med olika framfots-, bakfots och fotledsbesvär. Vi fann även här jämförbara eller bättre resultat för SEFAS än för EQ-5D, SF-36 och FAOS.

I delarbete III jämfördes SEFAS med American Orthopedic Foot and Ankle Society (AOFAS) score, ett instrument som används i hela världen för utvärdering av fot- och fotledskirurgi. Vi fann i detta delarbete, där patienter med olika stortå-, bakfots och fotledsbesvär utvärderades, att SEFAS var jämförbar eller bättre än AOFAS, samt dessutom enklare att använda.

I delarbete IV fann vi att SEFAS även fungerar i praktiken när vi testade instrumentet på en grupp patienter som opererades för symtomgivande plattfothet på två mindre sjukhus. Vi fann vid denna utvärdering en förbättring i SEFAS summascore både 6 och 24 månader efter operationen. Detta talar dels för att SEFAS verkligen registrerar en förbättring efter kirurgi på grund av plattfothet men också att förbättringen fortsätter upp till 24 månader efter operationen.

När vi sammanställer delarbetena är vår slutsats att SEFAS har goda psykometriska egenskaper, jämförbara eller bättre än de instrument som används i dag. Vi rekommenderar därför användningen av SEFAS som ett fot- och fotledsspecifikt patientrapporterat utvärderingsinstrument, både vid verksamhetsuppföljning, klinisk forskning och vid registrering i nationella register. Inför internationell användning av SEFAS måste instrumentet först översättas och utvärderas på det nya språket.
I would like to thank everyone who, in one way or another, helped and supported me to accomplish this thesis – all the patients participating in the studies, all my colleagues, friends and members of my family.

In particular I would like to express my warmest gratitude to:

First and foremost my supervisor and friend, Professor Magnus Karlsson, for your support at any time of the day, your professional guidance in scientific thinking and writing, your never-ending patience and also for your enthusiasm even though the research field and the part of the body, the foot and ankle, are not your specialty and interest. Even though we worked in different places have you been able to support me incredibly well with the aid of phone and email.

My co-supervisor Åke Carlson, who introduced the SEFAS and the validation process to me, for your support with great enthusiasm and energy during my PhD studies and research.

My co-supervisor Björn Rosengren, for your support and your invaluable precise comments and analyses with the purpose of refining the drafted work.

Ann Bremander, co-author, for your help, with great enthusiasm and knowledge, in getting me to better understand the complexity of the validation of PRO instruments.

Lars Brudin and Jan-Åke Nilsson, co-authors, for sharing your knowledge and experience of statistics and for invaluable help with the databases.

Anders Henricson, co-author, for your support and for your effort to find important publications also in my research field.

Håkan Magnusson, co-author, for your support and friendship

Alan Crozier, for revising my English.

Carin Holmberg, secretary, for all your help in contacts with the University of Lund.

Ingrid Djukanovic, my first research nurse in Kalmar. For helping me to start the Kalmar-Eksjö cohort and set up focus groups with patients and professionals to evaluate SEFAS.

Anna Pettersson, my second research nurse in Kalmar, for showing great patience in getting all our patients to complete the PRO instruments in the right time. Without
your structure and energy we would not have been able to obtain such a complete material.

*Emil Kilander*, my colleague in Eksjö, for including patients and for all the exciting discussions when we performed the surgical procedures on these patients.

*Rebecca Nobin, Markus Henmalm, Magnus Hellström* and *Johan Christiansson*, orthopedic foot and ankle surgeons in Kalmar, for helping to collect patients with foot and ankle disorders for the different studies, and for stimulating discussions during our surgical sessions.

*Carina Nilsson, Cecilia Madebrink, Charlotte Eriksson* and *Karin Henning*, physical therapists in Kalmar and Eksjö, for performing all the examinations for the AOFAS scores.

*Ann-Marie Oscarsson*, administrator in Kalmar, for assistance with research accounts and other practical matters even when I was no longer employed.

My colleagues and all staff at the operating theatre, outpatient clinic and patient wards in Kalmar, for your continuous support, good friendship and inspiring atmosphere.

All foot and ankle colleagues around Sweden and especially the ones involved in the Swedish Foot and Ankle Society, for inspiring me to get involved in the development of foot and ankle surgery as an important part of orthopedics in Sweden and to accomplish this thesis. Thanks for all the valuable discussions about surgical methods in the foot and ankle, and also for good friendship. We are a Swedish “foot and ankle family”.

*Fredrik Montgomery*, clinical mentor and friend, for teaching and getting me devoted to foot and ankle surgery.

*Ingrid Atteryd-Heiman*, my friend who lived in Lund until 2012, for letting me stay in her apartment when I completed the courses included in the PhD studies and for stimulating discussions about life.

My beloved children *Marcus, Fredrik* and *Angelica*, for just being there to support me with love and enthusiasm throughout the work on my thesis; for helping me with different practical matters related to the thesis: Marcus with statistics, Fredrik with computer-related matters and Angelica with the graphical design and illustration of the figures.

My beloved husband *Håkan*, for coaching my clinical career as an orthopedic surgeon, and also now throughout my thesis, for your endless loving support, patience and for always being there when I need you.


33. Dahlberg G. Statistical methods for Medical and Biological students. 2nd ed. London: George Allen and Unwin Ltd; 1940.


80. Lau JT, Mahomed NM, Schon LC. Results of an Internet survey determining the most frequently used ankle scores by AOFAS members. Foot Ankle Int 2005; 26(6):479–82.


113. Riskowski JL, Hagedorn TJ, Hannan MT. Measures of foot function, foot health, and foot pain: American Academy of Orthopedic Surgeons Lower Limb Outcomes Assessment: Foot and Ankle Module (AAOS-FAM), Bristol Foot Score (BFS), Revised Foot Function Index (FFI-R), Foot Health Status Questionnaire (FHSQ), Manchester Foot Pain and Disability Index (MFPDI), Podiatric Health Questionnaire (PHQ), and Rowan Foot Pain Assessment (ROFPAQ). Arthr Care Res 2011; 63 (Suppl 11):229–39.


126. Strand V, Boers M, Idzerda L, Kirwan JR, Kvien TK, Tugwell PS, et al. It’s good to feel better but it’s better to feel good and even better to feel good as soon as possible for as long as possible. Response criteria and the importance of change at OMERACT 10. J Rheumatol 2011; 38(8):1720–7.


129. Suk M. Musculoskeletal Outcomes measures and Instruments: Davos, Switzerland: AO Foundation; 2009.


