

QUESTIONNAIRE FOR USABILITY EVALUATION OF ORTHOPAEDIC SHOES: CONSTRUCTION AND RELIABILITY IN PATIENTS WITH DEGENERATIVE DISORDERS OF THE FOOT

Michiel J. A. Jannink,¹ Jaap de Vries,² Roy E. Stewart,³ Johan W. Groothoff³
and Gustaaf J. Lankhorst⁴

From the ¹Roessingh Research and Development, Enschede, ²Rehabilitation Centre "Het Roessingh", Enschede, ³Department of Health Sciences/Northern Centre for Health Care Research, University of Groningen and ⁴Department of Rehabilitation Medicine, VU Medical Centre Amsterdam, The Netherlands

Objective: To develop a self-report questionnaire for patients with degenerative disorders of the foot to evaluate the usability of their orthopaedic shoes, and to assess the reproducibility and responsiveness of the instrument.

Design: Development of the Questionnaire for Usability Evaluation of orthopaedic shoes was based on a literature search, structured expert interviews and a ranking procedure. A cross-sectional study was carried out to determine the reproducibility and internal consistency of the questionnaire.

Setting and subjects: The study population comprised 15 patients with degenerative disorders of the foot, who had worn their orthopaedic shoes for at least 3 years and 15 patients with degenerative disorders of the foot, who had never worn orthopaedic shoes, but would receive them within 1 month.

Results: Within the questionnaire 4 effectiveness items (pain, instability, callus, wounds), 1 efficiency item (putting on and taking off shoes) and 7 satisfaction items (pinch, slip, weight of shoes, cold feet, perspiration, maintenance, cosmetic appearance) were developed. All items in the questionnaire met the test-retest criteria. The smallest real difference ranged from 0.23 to 3.82 cm on a Visual Analogue Scale (10 cm). Cronbach's alpha's for the domains of pain and instability ranged from 0.70 to 0.92.

Conclusion: The Questionnaire for Usability Evaluation should provide a good rationale to assess the usability of orthopaedic shoes and can be considered reliable.

Key words: shoes, osteoarthritis, usability, questionnaire.

J Rehabil Med 2004; 36: 242–248

Correspondence address: Michiel J. A. Jannink, Roessingh Research and Development, Roessinghsbleekweg 33b, 7522 AH Enschede, The Netherlands. E-mail: m.jannink@rrd.nl

Submitted June 2, 2003; accepted February 11, 2004

INTRODUCTION

Degenerative disorders of the foot are very common in older individuals. Population surveys have reported a 10–24% prevalence of self-reported foot abnormalities in adults, with the

highest prevalence in women and in those of 65 years of age and older (1). These foot complaints and abnormalities (e.g. hallux valgus, claw toes, metatarsalgia) may restrict ambulation, limit activities and adversely affect participation in daily life. For the above-mentioned degenerative disorders of the foot, orthopaedic shoes can be prescribed, especially in serious cases. A key feature of orthopaedic shoes is their usability.

Usability is defined in ISO 9241-11 as "the extent to which a product can be used by specific users to achieve goals with effectiveness, efficiency and satisfaction in a specified context of use".

Within the definition of usability, effectiveness is the accuracy and completeness with which users achieve specified goals. For example, a patient is able to walk to a supermarket without foot pain. The resources that are expended in relation to the accuracy and completeness with which users achieve goals are assessed to determine efficiency. Relevant resources may include mental or physical effort (for example, independently putting on and taking off orthopaedic shoes), time or financial costs. Satisfaction is defined as the comfort and acceptability of use and can be assessed in terms of attitudes to using the product (for example, how do patients feel about the cosmetic appearance or level of perspiration in their orthopaedic shoes?). Finally, the context of use refers to the physical and social environments in which a product is used. Measurement of usability is particularly important in view of the complexity of the interactions between the patient and his or her goals and the elements of the context of use, which can result in significantly different levels of usability for the same product when used in different contexts (2).

Insight into the use and usability of rehabilitation technological aids can be obtained from the results of several evaluation studies described in the literature (3). In these evaluation studies, more and more questionnaires are being used to establish the usability of rehabilitation technological aids. This is also the case in studies evaluating orthopaedic shoes prescribed for patients with degenerative disorders of the foot. One such questionnaire is the Foot Function Index (FFI), which measures the impact of complaints on foot function (4, 5). Foot function is measured in terms of pain, disability and limitation of activities. All of the items are rated on a Visual Analogue Scale (VAS) and have satisfactory clinimetric properties. The FFI has been used

in several studies, both in selected patient groups with generalized diseases (rheumatoid arthritis, osteoarthritis) and in patients with more localized foot complaints (heel pain, forefoot pain) (4, 5). However, because the FFI focuses only on the foot function, no evaluation can be made of the usability of the orthopaedic shoes, as defined in ISO 9241-11.

Another questionnaire, the SERVQUAL (SERVice QUALity measurement scale), can be used to assess patient satisfaction with orthopaedic shoes (6). In this questionnaire, consumer interests and experiences are assessed on a 5-point Likert scale. The questionnaire contains 30 items, covering 5 domains: tangibles, reliability, responsiveness, assurance and empathy. The SERVQUAL was used in the National Health Service in the UK, in rehabilitation, in hospital services, in nursing services and other healthcare services (6). The subscales of the SERVQUAL were found to be internally consistent, and have a satisfying content validity and reliability.

Other researchers have used questionnaires that they have developed specifically for their study, instead of more generally applied questionnaires. Caravaggi et al. (7) measured patient acceptance of a therapeutic shoe using a VAS. Torkki et al. (8) used a self-designed questionnaire that measured the duration and intensity of foot pain, ability to work, cosmetic disturbance, footwear problems, health-related quality of life, satisfaction and costs related to foot care. Kelly & Winson (9) evaluated the use of ready-made insoles in the treatment of metatarsalgia. The assessment included a questionnaire consisting of VAS pain scores, estimated walking distance and VAS symptom relief scores. Franssen & Edmonds (10) evaluated the effectiveness of off-the-shelf orthopaedic footwear for people with rheumatoid arthritis using a questionnaire to assess chronic foot pain, in terms of self-reported pain and physical functioning. No further information about the properties and methodological quality of these questionnaires is available.

Although several instruments currently exist to measure pain and disability associated with foot problems or to measure patient satisfaction and acceptance of rehabilitation aids, none of the above-mentioned questionnaires quantifies all aspects of the usability (effectiveness, efficiency, satisfaction and context of use) of orthopaedic shoes. The purpose of this study was to develop a self-report instrument (questionnaire) to measure all aspects of the usability (effectiveness, efficiency, satisfaction, and context of use) of orthopaedic shoes in patients with degenerative disorders of the foot, which is reliable with regard to reproducibility and homogeneity.

METHODS

Development of the Questionnaire for Usability Evaluation of orthopaedic shoes

Collection of items. The development of the Questionnaire for Usability Evaluation (QUE) of orthopaedic shoes is based on the standard methodology for the development of questionnaires for research purposes, i.e. a literature search, structured expert interviews and a ranking procedure. In the literature search, articles and reference books were sought in MEDLINE (1970–2001), EMBASE (1970–2001) and the database of the Cochrane Collaboration Field “Rehabilitation and

Related Therapies”, using the following combinations of keywords: foot, ankle, osteoarthritis, claw toes, hammer toes, hallux valgus, metatarsalgia, plantar fasciitis, calcaneal spur, calcaneal bursitis, plantar fibromatosis, flat foot, cavus foot, shoes, orthopaedic shoes and orthopaedic footwear. In addition to this search, the reference lists of relevant publications were carefully checked. The initial selection of articles was based on the title and the content of the abstract. The following inclusion criteria were applied by 2 researchers (MJ and JdV): [1] studies concerning the evaluation of orthopaedic shoes and degenerative disorders of the foot; [2] published, full-length articles; [3] language: English, German or Dutch. The literature search resulted in the identification of 5 reference books (11–15) and 18 articles (16–33). Based on this literature search, it can be stated that the concept of usability with respect to orthopaedic shoes has been explored only superficially. Little formal knowledge is therefore available. In order to obtain additional information from clinical practice, structured expert interviews were held with a group of rehabilitation specialists ($n = 10$), orthopaedic surgeons ($n = 3$), orthopaedic shoe technicians ($n = 10$) and patients with degenerative disorders of the foot ($n = 10$). These experts (specialists and patients) were interviewed about (foot) problems and aspects regarding orthopaedic shoes at the effectiveness, efficiency, satisfaction and context of use level relevant for the specific experts.

Selection and ranking of items. The same experts ($n = 33$) were asked to rank the usability items, based on 2 criteria: subjective experienced incidence of these usability items in clinical practice, and measure of relevance. The literature search, the expert interviews and the ranking procedure resulted in a list of 12 usability items, which could be measured by means of a questionnaire. These 12 items are: pain during daily activities, stability during daily activities, callus, wounds, pinch, slip, and weight of shoes, cold feet, perspiration, putting on/taking off shoes, maintenance and cosmetic appearance.

Based on these 12 items the first version of the QUE of orthopaedic shoes was developed, consisting of 2 parts (QUE pre-test and QUE post-test). The QUE pre-test should be completed before patients receive their orthopaedic shoes. It measures the current state of subjective experienced foot problems and shoe problems while the patient is still wearing ready-made shoes, and measures the expectations patients have with regard to the orthopaedic shoes they will receive. The QUE pre-test consists of 67 questions distributed over the 12 usability items. Pain during daily activities (standing, walking, climbing stairs, riding a bicycle, activities of daily life and work) [18], stability during daily activities (standing, walking, climbing stairs, riding a bicycle, activities of daily life and work) [21], callus [3], wounds [3], pinch [3], slip [3], weight of shoes [3], cold feet [3], perspiration [3], putting on/taking off shoes [3], maintenance [2] and cosmetic appearance [2].

The QUE post-test measures the current state of subjective experienced foot and shoe problems of a patient who wears orthopaedic shoes, and has to be completed after the orthopaedic shoes have been worn for at least 3 months. The QUE post-test consists of 45 questions distributed over 12 items. Pain during daily activities [12], stability during daily activities [14], callus [2], wounds [2], pinch [2], slip [2], weight of shoes [2], cold feet [2], perspiration [2], putting on/taking off shoes [2], maintenance [2] and cosmetic appearance [1].

Face validity (whether the questions, on the face of it, appear to be measuring the variables they claim to measure) was reviewed by experts from various fields: rehabilitation medicine, rehabilitation research, human movement sciences and orthopaedic shoe technology.

Response format. The QUE pre-test and the QUE post-test consist of questions at a dichotomous level (yes/no) and questions at an interval level (VAS). Each VAS question consists of a 100-mm line bounded by 2 anchor phrases denoting the extremes of possible answers. Patients indicated their answers by making a mark across the line. Pilot-testing indicated that the respondents understood the direction of the choices and how to fill in the answers.

Reliability characteristics of the QUE for orthopaedic shoes

The QUE pre-test and QUE post-test for orthopaedic shoes were tested for reliability, in terms of reproducibility and internal consistency.

Reproducibility is defined as the ability to measure attributes in a reproducible and consistent manner when administered on several occasions to stable subjects (34). Internal consistency refers to the statistical coherence of the scale items.

Study population. Thirty patients were recruited from the outpatient clinic of a rehabilitation centre. Inclusion criteria were: (i) degenerative disorders of the foot; (ii) wearing orthopaedic shoes for at least 3 years ($n = 15$; experienced group, who will fill in the QUE post-test) or will be wearing them within 1 month ($n = 15$; inexperienced group, who will fill in the QUE pre-test); (iii) able to read Dutch; (iv) over 18 years of age; and (v) in a stable phase of the degenerative foot disorders.

Design of the test-retest reproducibility study. Patients completed the first version of the QUE pre-test and QUE post-test in the outpatient clinic of a rehabilitation centre twice, with an interval of 2 weeks. It was not expected that any clinically relevant changes in the patients' health status would be found during this 2-week interval. Because of the diversity of the questions, the age of the study population (elderly people) and the time required to complete the QUE (+/- 30 minutes), it was expected that at the second occasion patients would not remember their first responses.

Data analysis. Reproducibility. Reproducibility refers to the agreement in scores between 2 measurements. This is quantified with Cohen's kappa and the intraclass correlation coefficient (ICC). Cohen's kappa represents the proportion of agreement. In general, with a kappa value of less than 0.40, the agreement is considered to be poor to fair, 0.41–0.60 indicates moderate agreement, 0.61–0.80 good agreement, and when kappa exceeds 0.80 the agreement is very good. The ICC is often preferred over the Pearson's correlation as a measure of reproducibility, because it combines systematic and random errors into a single statistic. In this study the ICC (absolute agreement, two-way random) model was used, measuring the degree of absolute agreement among measurements (36). To detect longitudinal changes in time the standard error of measurement (SE_m) was calculated. The SE_m provides an interpretation of the magnitude of this within-subject variability, which is also known as the error variance (34). SE_m is calculated according to:

$$SE_m = \sqrt{MS_{\text{error}}} \quad (35).$$

Assuming that the 2 measurement errors are independent of each other, an interval or error band can be calculated, expressing the uncertainty of the difference between the 2 true scores. The difference between both measurements should be at least $1.96 \cdot \sqrt{2} \cdot SE_m$ to be 95% confident of a real difference between the true scores. The quantity $1.96 \cdot \sqrt{2} \cdot SE_m$ is called the "Smallest Real Difference" (SRD) and indicates the point where the difference between 2 consecutive assessments exceeds the measurement error or "noise".

Homogeneity. Homogeneity refers to the statistical coherence of scale items, and can be expressed in Cronbach's alpha correlation coefficients. This coefficient is based on the (weighted) average correlation of items within a scale, and indicates whether each item in the scale is contributing to the variance in the overall score. The internal consistency was only computed for the pain and instability items. The pain item consists of several sub-items (pain during standing, walking, climbing stairs, riding a bicycle, activities of daily life and work). Instability also consists of several sub-items (instability during standing, walking, climbing stairs, riding a bicycle, activities of daily life and work). The other items (callus, wounds, pinch, slip, weight, cold feet, perspiration, putting on/taking off, maintenance and cosmetic appearance) have no sub-items. Internal consistency is considered to be good if Cronbach's alpha is higher than 0.70. However, because of the small study population ($n = 15$) the computed Cronbach's alphas in this study will give only an indication of the internal consistency of the pain and instability items.

RESULTS

Study population

The characteristics of the study population are summarized in

Table I. There was no difference between the inexperienced group ($n = 15$) and the experienced ($n = 15$) group in age ($p = 0.289$) or gender ($p = 0.705$). The inexperienced group had a mean age of 61.5 years ($SD = 14.4$ years) and the experienced group had a mean age of 55.8 years ($SD = 14.3$). Both groups consisted predominately of women (9 females in the inexperienced group and 10 females in the experienced group) who were not working for various reasons. The most common reasons were that they had retired because of age or disability. The level of education was also comparable between the 2 groups.

Reproducibility

In this test–retest study, 3 aspects of reliability were examined. In Table II these reliability aspects are listed for the "inexperienced group", who filled in the QUE pre-test questionnaire, which had 20 questions at a nominal level that correlated significantly ($p < 0.05$). Questions, which did not correlate significantly or were not relevant for 75% or more of this study population, were removed from the questionnaire. The Cohen's kappa of 9 questions was between 0.60 and 0.80 ($p < 0.05$), which can be regarded as good, and for 11 questions the Cohen's kappa was above 0.80 ($p < 0.05$), which can be regarded as very good. In the "inexperienced group" 1 person did not fill in the questions at interval level. As a consequence this person is omitted from the analysis for the calculation of the ICC and the SRD. The ICC for the interval items regarding the effectiveness ranged between 0.726 and 0.996, and for items regarding satisfaction it ranged between 0.835 and 0.990. Both of these ranges were considerably high. However, the SRD showed ranges of 0.42–2.67 for items of effectiveness implying that differences in VAS scores over 0.42–2.67 cm should be found before it can be concluded that there is a detectable change in effectiveness beyond measurement error can be concluded. The SRD for the item of efficiency showed a range of 1.22–2.44, and for the item of satisfaction a range of 0.70–2.70.

Table III lists the reliability aspects for the "experienced group" of 15 patients who filled in the QUE post-test question-

Table I. *Characteristics of the study population (n = 30)*

| | "Inexperienced group" ($n = 15$) | "Experienced group" ($n = 15$) |
|-----------------------|------------------------------------|----------------------------------|
| Mean age (SD) (years) | 61.5 (14.4) | 55.8 (14.3) |
| Gender | | |
| Male | 6 (40%) | 5 (33%) |
| Marital status | | |
| Never married | | 3 (20%) |
| Married | 11 (73%) | 12 (80%) |
| Widowed or divorced | 4 (27) | |
| Living alone | 4 (27%) | 3 (20%) |
| Level of education | | |
| Primary school | 5 (33%) | 8 (53%) |
| Secondary school | 7 (47%) | 4 (27%) |
| High school | 1 (7%) | 1 (7%) |
| College | 2 (13%) | 2 (13%) |
| Employed | 2 (13%) | 4 (27%) |

Table II. Psychometric summary of the "inexperienced" Questionnaire for Usability Evaluation pre-test scales (n = 15)

| Domain | No. of Items "nominal"* | Cohen's kappa | Internal consistency | No. of Items "interval"* | ICC (n = 14) | SRD (n = 14) | Internal consistency |
|---------------------------|----------------------------|------------------|-------------------------|-----------------------------|--------------|--------------|-------------------------|
| <i>Effectiveness</i> | | | | | | | |
| Pain | 6 | 0.6–1.0 | 0.87 | 9 | 0.870–0.996 | 0.42–2.60 | 0.90 |
| Instability | 7 | 0.714–1.0 | 0.82 | 9 | 0.747–0.993 | 0.99–2.67 | 0.85 |
| Callus | 1 | 0.875 | | 2 | 0.935–0.948 | 1.99–2.15 | |
| Wounds | 1 | 0.733 | | 2 | 0.726–0.978 | 1.80–1.86 | |
| <i>Efficiency</i> | | | | | | | |
| Putting on and taking off | 1 | 0.765 | | 2 | 0.764–0.978 | 1.22–2.44 | |
| <i>Satisfaction</i> | | | | | | | |
| Pinch | 1 | 1.0 | | 1 | 0.989 | 1.09 | |
| Slip | 1 | 0.6 | | 2 | 0.948–0.990 | 0.70–1.60 | |
| Weight | 1 | 1.0 | | 2 | 0.892–0.978 | 1.46–2.08 | |
| Cold feet | | | | 1 | 0.975 | 1.36 | |
| Perspiration | 1 | 1.0 | | 2 | 0.905–0.930 | 2.11–2.42 | |
| Maintenance | | | | 2 | 0.781–0.812 | 2.47–2.70 | |
| Cosmetics | | | | 2 | 0.835–0.908 | 2.60–2.64 | |

* $\alpha = 0.05$.

ICC = intraclass correlation coefficient; SRD = smallest real difference.

naire. Twenty-one questions at a nominal level correlated significantly ($p < 0.05$). The Cohen's kappa of 14 questions was higher than 0.80 ($p < 0.05$), and can be considered as very good. For 6 questions the Cohen's kappa was between 0.61 and 0.80 ($p < 0.05$), which can be considered as good. One question, experienced instability during "climbing stairs" can be considered as moderate, with a Cohen's kappa between 0.41 and 0.60 ($p < 0.05$). In the "experienced group", also 1 person did not fill in the questions at interval level. As a consequence, this person is omitted from the analysis for the calculation of the ICC and the SRD. The ICC for interval items regarding the effectiveness of orthopaedic shoes ranged between 0.853 and 0.999, and for items regarding satisfaction it ranged between 0.839 and 0.994. Both ranges were considerably high. The SRD showed a range of 0.15–2.62 for items of effective-

ness, so differences in VAS scores over 0.15–2.62 cm should be found before it can be concluded that there is a detectable change in value of use beyond measurement error. For the items of satisfaction, a range of 0.66–2.62 was found, implying that differences in VAS scores over 0.66–2.62 cm should be found before it can be concluded that the changes were not caused by measurement error.

Homogeneity

In Table II the internal consistency of the pain and instability items are listed for the "inexperienced group", who filled in the QUE pre-test questionnaire. However, because of the small study population ($n = 15$) this can only give an indication of the internal consistency of the pain and instability items. The Cronbach's alpha for pain and instability items at a nominal

Table III. Psychometric summary of the "experienced" Questionnaire for Usability Evaluation post-test scales (n = 15)

| Domain | No. of Items "nominal"* | Cohen's kappa | Internal consistency | No. of Items "interval"* | ICC (n = 14) | SRD (n = 14) | Internal consistency |
|---------------------------|----------------------------|------------------|-------------------------|-----------------------------|--------------|--------------|-------------------------|
| <i>Effectiveness</i> | | | | | | | |
| Pain | 5 | 0.7–0.896 | 0.70 | 6 | 0.943–0.999 | 0.15–2.62 | 0.90 |
| Instability | 6 | 0.524–1.0 | 0.82 | 7 | 0.853–0.988 | 0.97–2.03 | 0.92 |
| Callus | 1 | 0.867 | | 1 | 0.950 | 2.54 | |
| Wounds | 1 | 0.867 | | 1 | 0.943 | 2.04 | |
| <i>Efficiency</i> | | | | | | | |
| Putting on and taking off | 1 | 0.634 | | 1 | 0.940 | 0.47 | |
| <i>Satisfaction</i> | | | | | | | |
| Pinch | 1 | 0.842 | | 1 | 0.958 | 1.85 | |
| Slip | 1 | 1.0 | | 1 | 0.994 | 0.66 | |
| Weight | 1 | 1.0 | | 1 | 0.986 | 1.06 | |
| Cold feet | 1 | 0.815 | | 1 | 0.839 | 2.40 | |
| Perspiration | 1 | 0.602 | | 1 | 0.954 | 2.09 | |
| Maintenance | | | | 2 | 0.966–0.975 | 1.00–1.81 | |
| Cosmetics | | | | 1 | 0.888 | 2.62 | |
| <i>Amount of use</i> | | | | | | | |
| Days a week | 1 | 1.0 | | | | | |
| Hours a day | 1 | 0.694 | | | | | |

* $\alpha = 0.05$.

ICC = intraclass correlation coefficient; SRD = smallest real difference.

level (yes/no) ranged from 0.82 to 0.87, and for pain and instability items at an interval level (VAS-scores), it ranged from 0.85 to 0.90.

In Table III the internal consistency of the pain and instability items is listed for the “experienced group”, who filled in the QUE post-test questionnaire. For pain and instability items at a nominal level (yes/no), Cronbach’s alpha ranged from 0.70 to 0.82, and for pain and instability items at an interval level (VAS-scores), it was between 0.90 and 0.92.

The Cronbach’s alphas for pain are based on 6 sub-items (pain during standing, walking, climbing stairs, riding a bicycle, activities of daily life and work). The Cronbach’s alphas for instability are based on 7 sub-items (instability during standing, walking, walking on a rough surface, climbing stairs, riding a bicycle, activities of daily life and work).

DISCUSSION

Assessment of person-perceived usability is essential when evaluating rehabilitation interventions such as orthopaedic shoes. In this study, usability was defined as “the extent to which a product can be used by specific users to achieve goals with effectiveness, efficiency and satisfaction in a specified context of use”. There is no available questionnaire that quantifies all aspects of the usability of orthopaedic shoes. The purpose of this study was to develop a questionnaire to measure the usability of orthopaedic shoes in patients with degenerative disorders of the foot, based on the ISO 9241-11 definition of usability, which was reliable with regard to reproducibility and homogeneity.

The development of the QUE for orthopaedic shoes was based on a literature search, structured expert interviews and a ranking procedure. Since the purpose of this literature search was to make an inventory of possible items regarding the usability of orthopaedic shoes, no assessment was made of the methodological quality of the studies. Based on the systematic review it can be stated that the concept of usability with respect to orthopaedic shoes has been explored only superficially. To overcome publication bias 33 “experts” in the field of orthopaedic footwear were asked to provide additional information gained from their own clinical practice. It should be mentioned that this additional information is valid for the Dutch situation, and needs to be further examined before extrapolation to other countries. Since the literature does not report any data on the incidence of the usability items, the ranking of these items was carried out by the same experts and based on their experiences in clinical practice. An epidemiological study is recommended to identify objective rates of incidence.

The literature search, the expert interviews and the ranking procedure resulted in a list of 12 items to assess the usability of orthopaedic shoes.

Based on these 12 items, the Questionnaire for Usability Evaluation of orthopaedic shoes was developed, consisting of

2 parts (QUE pre-test and QUE post-test). The QUE pre-test (final version) consists of 56 questions, and measures different aspects of foot complaints and the expectations inexperienced people have with regard to the effectiveness, efficiency, satisfaction and context of use of their orthopaedic shoes. The QUE post-test (final version) consists of 45 questions, and measures different aspects of foot complaints and the experience people have with regard to the effectiveness, efficiency, satisfaction and context of use of their orthopaedic shoes. Pilot-testing indicated that the patients understood the direction of the choices and how to fill in the answers. Within this pilot-study patients filled in the questionnaire and were later interviewed about the comprehensibility, direction of choices and how to fill in the answers stated in the questionnaire. However, this was not formally tested. Face validity was based on the experts’ judgement of the items. The experts came from various fields: rehabilitation medicine, rehabilitation research, human movement sciences and orthopaedic shoe technology. In future studies the currently available questionnaire needs to be examined by linking the items to the ICF reference framework. Thus it is possible to link the QUE to other already existing instruments.

The test–retest reliability of the QUE was also satisfactory, compared with the reliability of the Dutch version of the FFI (ICC = 0.70–0.83). However, the FFI focuses only on the foot function, and does not provide any information about the usability of orthopaedic shoes.

Reproducibility coefficients, expressed as a dimensionless number between 0 and 1, do not lend themselves to a straightforward interpretation. For this purpose the SRD is better suited. The SRDs are expressed in the same dimensions as the questions in the QUE pre-test and QUE post-test.

In this study some of the SRDs were found to be relatively large (up to 27% of the total VAS scale). However this is not a problem, because patients with degenerative disorders of the foot have severe pain before they are provided with orthopaedic shoes, which results in high VAS scores. The goal of prescribing orthopaedic shoes, however, is to reduce a lot of the pain they experience during their daily activities.

The other way to test the reliability of a questionnaire is to calculate Cronbach’s alpha. The internal consistency (based on Cronbach’s alpha) of the QUE was also satisfactory, compared with the FFI ($\alpha = 0.88–0.94$) and the SERVQUAL ($\alpha > 0.70$). However, it should be mentioned that, because of the small study population ($n = 30$), the Cronbach’s alphas calculated in this test-retest study give only an indication of the internal consistency of the pain and instability items. Further investigation in a larger study population will be necessary to draw firm conclusions with regard to the internal consistency. It is then also possible to analyse the results using other psychometric methods including factor or principle component analysis.

Based on this study, it can be concluded that the QUE assesses all aspects of the usability (effectiveness, efficiency, satisfaction and context of use) of orthopaedic shoes, which no other

questionnaire does. Four items were developed within the domain of effectiveness (pain, instability, callus and wounds), one item was developed within the domain of efficiency (putting on and taking off orthopaedic shoes) and 7 items were developed within the domain of satisfaction (pinch, slip, weight of shoes, cold feet, perspiration, maintenance and cosmetics). All the above-mentioned items relate to various different aspects of the context of use. Furthermore, the QUE can be considered as a reliable questionnaire with which to assess the usability of orthopaedic shoes, also compared with other, more generic questionnaires. The multidimensional structure of the QUE should provide a good rationale to evaluate the usability of orthopaedic shoes.

ACKNOWLEDGEMENTS

This study was supported by ZON/MW, The Netherlands. We are grateful to Ms M. J. G. Moll for searching the Database of the Cochrane Field "Rehabilitation and Related Therapies". The Questionnaire for Usability Evaluation of orthopaedic shoes is also available on request from the corresponding author.

REFERENCES

- Gorter KJ, Kuyvenhoven MM, De melker RA. Nontraumatic foot complaints in older people. A population-based survey of risk factors, mobility, and well-being. *J Am Podiatr Med Assoc* 2000; 90: 397-402.
- Stewart T. Ergonomic requirements for office work with visual display terminals (VDTs): Part 11: Guidance on usability. International Organization for Standardization ISO 9241, 1998.
- Jannink MJA, Van Dijk H, De Vries J, Groothoff JW, Lankhorst GJ. A systematic review on methodology and usability evaluation of orthopaedic shoes. *Clin Rehab* 2004; 18: 15-26.
- Kuyvenhoven MM, Gorter KJ, Zuthoff P, Budiman-Mak E, Conrad KJ, Post MW. The foot function index with verbal rating scales (FFI-5pt): a clinimetric evaluation and comparison with the original FFI. *J Rheumatol* 2002; 29: 1023-1028.
- Budiman-Mak E, Conrad KJ, Roach KE. The Foot Function Index: a measure of foot pain and disability. *J Clin Epidemiol* 1991; 44: 561-570.
- Geertzen JH, Gankema HG, Groothoff JW, Dijkstra PU. Consumer satisfaction in prosthetics and orthotics facilities. *Prosthet Orthot Int* 2002; 26: 64-71.
- Caravaggi C, Faglia E, De Giglio R, Mantero M, Quarantiello A, Sommariva E, et al. Effectiveness and safety of a nonremovable fiberglass off-bearing cast versus a therapeutic shoe in the treatment of neuropathic foot ulcers: a randomized study. *Diabetes Care* 2000; 23: 1746-1751.
- Torkki M, Malmivaara A, Seitsalo S, Hoikka V, Laippala P, Paavolainen P. Surgery vs orthosis vs watchful waiting for hallux valgus: a randomized controlled trial. *JAMA* 2001; 285: 2474-2480.
- Kelly A, Winson I. Use of ready-made insoles in the treatment of lesser metatarsalgia: a prospective randomized controlled trial. *Foot Ankle Int* 1998; 19: 217-220.
- Fransen M, Edmonds J. Off-the-shelf orthopedic footwear for people with rheumatoid arthritis. *Arthritis Care Res* 1997; 10: 250-256.
- Postema K, Toornend JLA, Zilvold G, Schaars AH. Orthopedisch maatschoeisel in de medische praktijk. Houten: Bohn Stafleu Van Loghum; 1991.
- Jahss MH. Disorders of the foot. Philadelphia WB Saunders Co.; 1982.
- Gould JS. The foot book. Baltimore: Williams and Wilkins; 1988.
- Rabl CRH, Nyga W. Orthopädie des Fusses. Stuttgart: Ferdinand Enke Verlag, 6e druk; 1982.
- Baumgartner R, Stinus H. Die orthopädische Versorgung des Fusses. Stuttgart/New York: Georg Thieme Verlag; 1995.
- Kudriavtsev VA, Rozhkov AV, Ogorodnikov VI. Insertable orthopedic devices in the treatment of flatness of the anterior section of the foot and valgus deviation of the great toe. *Ortop Travmatol Protez* 1990; 9-11.
- Silfverskiöld JP. Common foot problems. Relieving the pain of bunions, keratoses, corns, and calluses. *Postgrad Med* 1991; 89: 183-188.
- Bogy LT, Fladger CB. Historic perspective of lesser metatarsalgia. *Clin Podiatr Med Surg* 1990; 7: 569-572.
- Osterman HM, Stuck RM. The aging foot. *Orthop Nurs* 1990; 9: 43-47, 76.
- Gould JS. Metatarsalgia. *Orthop Clin North Am* 1989; 20: 553-562.
- Chandler TJ, Kibler WB. A biomechanical approach to the prevention, treatment and rehabilitation of plantar fasciitis. *Sports Med* 1993; 15: 344-352.
- Gill LH, Kiebzak GM. Outcome of nonsurgical treatment for plantar fasciitis. *Foot Ankle Int* 1996; 17: 527-532.
- Sadat-Ali M. Plantar fasciitis/calcanal spur among security forces personnel. *Mil Med* 1998; 163: 56-57.
- Coughlin MJ, Carlson RE. Treatment of hallux valgus with an increased distal metatarsal articular angle: evaluation of double and triple first ray osteotomies. *Foot Ankle Int* 1999; 20: 762-770.
- Coughlin MJ. Treating hallux valgus deformities: versatility is required. *Orthopedics* 1990; 13: 935.
- Coughlin MJ. Hallux valgus. Causes, evaluation, and treatment. *Postgrad Med* 1984; 75: 174-178, 183, 186-187.
- Chao W, Wapner KL, Lee TH, Adams J, Hecht PJ. Nonoperative management of posterior tibial tendon dysfunction. *Foot Ankle Int* 1996; 17: 736-741.
- Boer H, Seydel ER. Medical opinions, beliefs and prescription of orthopaedic footwear: a survey of Dutch orthopaedists and rehabilitation practitioners. *Clin Rehabil* 1998; 12: 245-253.
- Rao UB, Joseph B. The influence of footwear on the prevalence of flat foot. A survey of 2300 children. *J Bone Joint Surg Br* 1992; 74: 525-527.
- Noe DA, Voto SJ, Hoffmann MS, Askew MJ, Gradisar IA. Role of the calcaneal heel pad and polymeric shock absorbers in attenuation of heel strike impact. *J Biomed Eng* 1993; 15: 23-26.
- Leung AK, Mak AF, Evans JH. Biomedical gait evaluation of the immediate effect of orthotic treatment for flexible flat foot. *Prosthet Orthot Int* 1998; 22: 25-34.
- Donley BG, Tisdell CL, Sferra JJ, Hall JO. Diagnosing and treating hallux valgus: a conservative approach for a common problem. *Cleve Clin J Med* 1997; 64: 469-474.
- Postema K, Burm PE, Zande ME, Limbeek J. Primary metatarsalgia: the influence of a custom moulded insole and a rockerbar on plantar pressure. *Prosthet Orthot Int* 1998; 22: 35-44.
- Streiner DL, Norman GR. Health measurement scales. A practical guide to their development and use. Oxford: Oxford University Press; 1993.
- Stratford PW. Reliability: consistency or differentiating among subjects? *Phys Ther* 1989; 69: 299-300.
- McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. *Psychological Methods* 1996; 1: 30-46.

APPENDIX

| Domain | Items | Description of the items | |
|---------------|---|--|--|
| Effectiveness | Pain during daily activities | Pain during stance | |
| | | Pain during walking | |
| | | Pain during climbing stairs | |
| | | Pain during riding a bicycle | |
| Instability | Instability during activities of daily life | Pain during work | |
| | | Instability during stance | |
| | | Instability during walking | |
| | | Instability during walking on uneven ground | |
| | | Instability during climbing stairs | |
| | | Instability during riding a bicycle | |
| Callus | Instability during activities of daily life | Instability during work | |
| | | Corns are small hard conical hyperkeratosis due to friction and pressure | |
| Wounds | Callus are thickenings of keratin due to pressure | Abnormality/damage of skin texture (e.g. ulceration, colour) | |
| | | Efficiency | Putting on and taking off orthopaedic shoes |
| Satisfaction | Pinch | The sticking, squeezing of the shoe | |
| | | Slip | The occurrence of slipping of the heel in the shoe |
| | | Weight | The experienced (subjective) weight of the shoe |
| | | Cold feet | The occurrence of cold feet |
| | | Perspiration | The occurrence of perspiration |
| | | Maintenance | The difficulties experienced in the maintenance of orthopaedic shoes (e.g. polishing, cleaning, repairing) |
| | | Cosmetic appearance | Do patients find their shoes ugly or beautiful? |