

Validation of the Italian version of the Oxford Ankle Foot Questionnaire for children

Nicolò Martinelli¹ · Giovanni Romeo¹ · Carlo Bonifacini¹ · Marco Viganò¹ · Alberto Bianchi¹ · Francesco Malerba¹

Accepted: 3 July 2015 / Published online: 14 July 2015
© Springer International Publishing Switzerland 2015

Abstract

Purpose The purpose of this study was to translate the Oxford Ankle Foot Questionnaire (OAFQ) into Italian, to perform a cross-cultural adaptation and to evaluate its psychometric properties.

Methods The Italian OAFQ was developed according to the recommended forward/backward translation protocol and evaluated in pediatric patients treated for symptomatic flatfoot deformity. Feasibility, reliability, internal consistency, construct validity [comparing OAFQ domains with Child Health Questionnaire (CHQ) domains] and responsiveness to surgical treatment were assessed.

Results A total of 61 children and their parents were enrolled in the study. Results showed satisfactory levels of internal consistency for both children and parent forms. The test–retest reliability was confirmed by high ICC values for both child and parents subscales. Good construct validity was showed by patterns of relationships consistent with theoretically related domains of the CHQ. After surgery, the mean OAFQ scores improved in all the domains after treatment with the subtalar arthroereisis, for both children and parent scales ($p < 0.01$). Effect size ranged from small to moderate for almost all domains.

Conclusions The Italian version of the OAFQ might be a reliable and valid instrument in order to evaluate inter-

ventions used to treat children's foot or ankle problem, but needs further study on different clinical settings.

Keywords Italian OAFQ · Pediatric foot surgery · Flatfoot

Introduction

Proliferation of patient-focused questionnaires for measuring of health status and health-related quality of life is the result of the increasing international interest in surgical outcomes with measurements obtained from the perspective of patients or their non-clinician caregivers (e.g., parents of patients too young to self-report) for use in clinical research evaluating surgical *effectiveness*.

The field of orthopedics is concerned with a broad anatomical range and therefore with a wide range of functional problems. Although generic instruments are potentially responsive to clinically important changes in health, disease-specific measures can give a more detailed pattern of specific symptoms and impairment related to a specific disease. Challenges arise when clinical outcomes are assessed in children. Instruments that are originally developed for adults may not capture the pediatric patient's perspective and may not accurately reflect how children relate to their usual environments. While adult patient-reported outcome measures (PROMs) already exist for foot and ankle disease, only one questionnaire has been validated for pediatric foot and ankle diseases: the Oxford Ankle Foot Questionnaire (OAFQ) [1, 2].

The purpose of this study was to translate the OAFQ into Italian, to perform a cross-cultural *adaptation* and to evaluate the psychometric properties of the Italian version of OAFQ.

✉ Nicolò Martinelli
n.martinelli@unicampus.it

¹ Department of Ankle and Foot Surgery, IRCCS Istituto Ortopedico Galeazzi, Via Riccardo Galeazzi 4, 20161 Milan, Italy

Materials and methods

This study was performed according to the guidelines for Good Clinical Practice and the Declaration of Helsinki. All participants and their parents gave informed consent for participation in the study. A license for the use of the OAFQ was obtained from the copyright holder (Isis Outcomes, Oxford, UK). The translation procedure was performed following the guidelines proposed by Guillemin et al. [3] for the cross-cultural adaptation of HRQL measures with forward–backward and counter-translation. Forward translation into Italian of the OAFQ was independently performed by two informed translators, orthopedic surgeons, mother tongue Italian and fluent in English. The first version was obtained after a consensus meeting of the two translators. This provisional Italian version was translated back into English by three translators naive to the outcome measure: two mother tongue Italian subjects fluent in English, with medical background, and one mother tongue English speaker, fluent in Italian. A second consensus meeting of all translators was arranged in order to check for discrepancies or any problems. The English versions (the original and the three versions produced by translating it back into English) were found to be semantically similar and no further adjustments were required. The final Italian version was obtained after testing it on fifteen pediatric patients with foot and ankle disease and their corresponding parents to discuss perceived difficulties of comprehension. These patients were asked to rate all items concerning their importance (very important, important, unimportant and indifferent). None of the patients reported problems to complete questionnaires because of language problem or redundancy.

Patients

Children aged between 8 and 13 were considered eligible for this study when they were diagnosed with symptomatic flexible flatfoot by an orthopedic surgeon and referred for surgical correction. Clinical diagnosis of flatfoot was based on a valgus position of the heel ($>10^\circ$) and collapse of the medial longitudinal arch [4, 5]. Weight-bearing antero-posterior (AP) and lateral radiographs were used to confirm the clinical diagnosis: the talus-second metatarsal angle, the talonavicular coverage angle, the calcaneal pitch angle, the talo-first metatarsal angle and the medial cuneiform-fifth metatarsal height were assessed.

Children and parents were asked to complete separate “child/adolescent” and “parent” versions of the Italian OAFQ, the Child Health Questionnaire—Child Form 87 (CHQ-CF87) and the Child Health Questionnaire—Parent Form 50 (CHQ-PF50) [6].

All the patients underwent surgical correction of the deformity through subtalar arthroereisis [7]. During the pre-admission visit, patients and their parents were asked to fill out the Italian version of the OAFQ, the Italian version of the CHQ-CF87 and CHQ-PF50. All the questionnaires were scored as recommended in the original versions. The OAFQ includes 15 items, of which 14 are grouped into *three* subscales: physical (6 items), emotional (4 items) and school and play (4 items). Each item has a five-point Likert-type response format ranging from 0, *poor function*, to 4, *good function*. The last item provides information about patient’s satisfaction on the footwear they can wear. Missing data in the OAFQ were imputed if half or more items in a domain scale were present; in these cases, the missing items were given the mean of the subject’s own item scores in that scale [1]. The scores of all items were summed separately for the three subscales and were transformed to a percentage scale (0–100), where a higher score represents a better functioning. The CHQ is a generic health instrument developed to assess the physical and psychosocial well-being of children independently from the underlying disease. The parent version (CHQ-PF50) and the children version (CHQ-CF87) consist of 50 and 87 items, respectively. These questionnaires cover different concepts of physical, psychosocial health and disabilities associated with these. These concepts include: global health (GGH), general health perceptions (GH), physical functioning (PF), role functioning—emotional (RE), role functioning—behavioral (RB), role functioning—physical (RP), bodily pain (BP), behavior (BE), global behavior (GBE), mental health (MH), self-esteem (SE), family activities (FA), family cohesion (FC), parental impact—emotional (PE), parental impact—time (PT) and change in health (CH). The two concepts, RE and RB, are combined into one, namely role functioning—emotional and behavioral (REB) in the CHQ-PF50. Two concepts, PE and PT, are not applicable for CHQ-CF87 as the questionnaire is answered by the children themselves. These health concepts comprise multiple items or a single item. The response options of each item vary from 4 to 6 levels. For concepts with multiple items, the responses to the items are summed up and transformed to a scale that ranges from 0 (lowest possible score indicating the worst health) to 100 (highest possible score indicating the best health). The CHQ-PF50 and CHQ-CF87 include, respectively, 15 and 14 items. The CHQ-CF87 and the CHQ-PF50 contain similar but not identical items, and the use of both questionnaires is advised in order to detect children’s perspectives about their general health status.

The OAFQ was re-administered to the patients/parents the day before surgery in the hospital, approximately between 1 and 2 weeks after the pre-admission visit. One week was considered an acceptable time to ensure that the

clinical situation would not change during this period and that patients would not remember their answers to the first questionnaire.

Approximately 6 months after surgery, the OAFQ was re-administered to the patients/parents during a follow-up visit. CHQ-PF50 and CHQ-CF87 were administered once in the pre-admission visit. The OAFQ was administered three times (pre-admission visit, day before surgery and last follow-up visit) to all patients/parents.

Statistical analyses

Data were entered into a Microsoft Excel spreadsheet (Microsoft Corporation, Redmond, WA) and analyzed using SPSS 20.0 (SPSS Inc., Chicago, IL). Normal distribution of the scores was tested using the Kolmogorov–Smirnov test. All tests were two-tailed and conducted at a 5 % level of significance.

Reliability

Test–retest reliability was evaluated by comparing domain scores from questionnaires completed at baseline and again approximately within 2 weeks using the ICC (model: two-way random; type: absolute agreement). An ICC of more than 0.80 is usually considered an indicator of good reproducibility [8, 9]. The child–parent consistency for each domain score were assessed using the intraclass correlation coefficient (ICC). The smallest detectable difference (SDD) was calculated as $1.96 \times \sqrt{2} \times \text{standard error of measurement (SEM)}$, defined as $S\sqrt{1 - r}$, where S is the standard deviation of test results and r is test–retest reliability [10]. Internal consistency was assessed calculating the Cronbach's α , and the consistency of each question with total score subscale was evaluated using item-total correlation [11]. Values between 0.7 and 0.9 are considered acceptable for psychometric scales [12]. For the scales to be considered sufficiently reliable for use in groups of patients, item-total correlation should be above 0.4 [13].

Convergent validity

Convergent validity was evaluated for each domain with Spearman's correlation coefficient using a priori hypothesized correlations with CHQ domains. To demonstrate convergent validity, we assumed moderate to high correlations between the OAFQ subscales and the CHQ physical components (PF, RP, BP). To demonstrate divergent validity, we assumed lower correlations between the OAFQ subscales and the mental health components of the CHQ (GH, GGH, REB, GBE, MH, SE, FA, FC). According to Dawson et al., a coefficient correlation value of

0.60–0.79 indicated a strong correlation and 0.40–0.59 a moderate correlation [14]. Such psychometric properties were evaluated on the baseline data.

Feasibility was assessed by calculating missing responses and the floor and ceiling effects and were considered present if more than 15 % of the participants achieved the highest or lowest possible score [15].

Responsiveness

Responsiveness to surgical treatment was evaluated using Student's t test comparing the pre- and post-surgery scores. Interpretation of score changes was performed with standardized effect size (ES) and standardized response mean (SRM). The effect sizes were calculated as described by Husted et al. [16]; values >0.20 , >0.50 and >0.80 were considered small, moderate and large, respectively.

Results

The forward and back-translations of the OAFQ presented no major difficulties or problems with the language. A total of 61 consecutive patients (28 girls, 33 boys; mean age 10.9 ± 1.6 years) were enrolled in the study. Each participant and a parent completed questionnaires before surgery and at the follow-up visit (Table 1). The average time between completion of baseline and retest questionnaires was 10.9 days (SD 4.0; range 6–20). All 61 patients and parents participated in the reliability and responsiveness assessment. No missing data were observed during the three assessments.

Table 1 Baseline and follow-up scores from the Oxford Ankle Foot Questionnaire

Dimension	Baseline	Follow-up
Physical		
Child	65.6 (26.4)	76.8 (21.5)
Parent	64.2 (25.1)	74.3 (23.1)
School and play		
Child	85.9 (13.5)	90.2 (11.2)
Parent	82.6 (13.9)	88.3 (17.3)
Emotional		
Child	83.0 (14.9)	90.5 (10.6)
Parent	80.3 (17.4)	91.4 (12.3)
Footwear (single item)		
Child	70.0 (34.7)	82.7 (27.2)
Parent	65.1 (32.6)	81.9 (27.4)

Mean and standard deviations are reported for the four dimensions

Reliability

Table 2 shows reliability parameters of the questionnaire. The test–retest reliability was confirmed by high ICC values for both child and parents subscales. The SDD in child-reported scores was 17.0, 7.7 and 7.0 for physical, school and play and emotional subscale, respectively; in the parent-reported scores, it was 3.3, 12.9 and 7.8 for physical, school and play and emotional subscale, respectively.

Cronbach's α value for the study of the questionnaire was greater than 0.7 in all the OAFQ domains for both children and parent forms. All the questions in three subscales of the questionnaire had correlation greater than 0.4, except for the last question of the "school and play" subscale in the child form (Table 3).

At baseline, no floor effect was detected for all domains for the two version, but ceiling effect was observed in the school and play domain for 14 children (23 %) and 10 parents (16.4 %) and in emotional for 17 children (27.9 %).

Convergent validity

Tables 4 and 5 show the correlations between OAFQ and CHQ for children and parent form, respectively. A significant agreement between the OAFQ domains and the scales of the CHQ with related content, particularly in the areas of physical function and pain, was observed for both children and parent forms.

Responsiveness

The average follow-up from surgery was 7.2 months (SD 1.6; range 4–12). The mean OAFQ scores improved in all the domains after treatment with the subtalar arthroereisis, for both children and parent scales ($p < 0.01$). Effect size and SRM values ranged from small to moderate for almost all domains (Table 6).

Discussion

In this study, the cross-cultural adaptation of the OAFQ to the Italian language has been presented. It showed satisfactory psychometric properties in pediatric patients with

Table 3 Item-total correlations of the four domains

Subscale	Baseline			
	Mean score (SD)		Item-total correlation	
	Child	Parent	Child	Parent
Physical				
1	2.6 (1.2)	2.6 (1.2)	0.79	0.78
2	2.5 (1.3)	2.3 (1.2)	0.82	0.72
3	2.6 (1.2)	2.5 (1.2)	0.65	0.60
4	2.5 (1.2)	2.5 (1.2)	0.77	0.73
5	2.6 (1.3)	2.6 (1.2)	0.70	0.68
6	2.6 (1.2)	2.5 (1.2)	0.79	0.74
School and play				
7	3.5 (0.7)	3.3 (0.7)	0.73	0.70
8	3.3 (0.7)	3.2 (0.7)	0.71	0.64
9	3.2 (0.8)	3.2 (0.7)	0.52	0.57
10	3.5 (0.5)	3.3 (0.6)	0.15	0.42
Emotional				
11	3.2 (0.8)	3.0 (0.9)	0.61	0.57
12	3.1 (0.7)	3.0 (0.9)	0.75	0.72
13	3.4 (0.5)	3.2 (0.8)	0.79	0.62
14	3.4 (0.6)	3.4 (0.7)	0.63	0.54

flatfoot deformity, treated with subtalar arthroereisis. After adaptation, the Italian version of OAFQ seems to be a feasible instrument as illustrated by the absence of missing data that reflect the good acceptance of the questionnaire. The absence of "floor" effect in the preoperative evaluation resulted in accordance with the original version and confirms the good feasibility of this version. However, a ceiling effect was detected for the school and play domain school for children and parents and in emotional domain for children (27.9 %). The internal consistency (measured by Cronbach's α) was satisfactory for all analyzed subscale and similar to the original version. The high value of ICC, in particular for what concern child–parent and test–retest scores, indicates good concordance and high reliability, respectively. For what concern the construct validity, moderate/strong correlations were found between the subscales of OAFQ Italian version with the respective

Table 2 Reliability coefficients for internal consistency, child- and parent-reported scores and stability (test–retest)

	Internal consistency (alpha)		Child–parent (ICC)	Test–retest (ICC)	
	Child	Parent		Child	Parent
Physical	0.91	0.89	0.99	0.88	0.97
School and play	0.72	0.77	0.93	0.89	0.82
Emotional	0.84	0.80	0.87	0.90	0.88
Footwear	–	–	0.96	0.97	0.96

SD standard deviation, ICC intraclass correlation coefficient

Table 4 Correlation (Spearman's ρ) between the OAFQ (child) subscales and CHQ-CF87 subscales

CHQ-CF87 subscales	Correlation/coefficient		
	Physical	School and play	Emotional
Global health	0.20	−0.03	0.07
Physical functioning	0.63*	0.27*	0.35*
Role functioning—emotional	−0.003	0.02	−0.21
Role functioning—behavioral	0.24	0.24	0.01
Role functioning—physical	0.57*	0.28*	0.27*
Bodily pain	0.75*	0.30*	0.40*
Behavior	0.30*	0.16	−0.01
Global behavior item	0.12	−0.11	0.06
Mental health	0.20	0.11	−0.07
Self-esteem	0.20	−0.10	0.11
General health perceptions	0.15	0.09	0.16
Family activities	0.10	0.12	0.05
Family cohesion	0.25*	−0.10	0.15
Change in health	0.19	0.03	0.17

* Correlation is significant at the 0.05 level (two-tailed)

Table 5 Correlation (Spearman's ρ) between the OAFQ (parent) subscales and CHQ-PF50 subscales

CHQ-PF50 subscales	Correlation/coefficient		
	Physical	School and play	Emotional
Global health	0.01	0.07	−0.14
Physical functioning	0.60*	0.23	0.08
Role functioning—emotional/behavioral	0.31*	0.07	−0.03
Role functioning—physical	0.51*	0.24	0.07
Bodily pain	0.60*	0.22	0.09
Behavior	0.26*	0.36*	−0.22
Global behavior item	0.02	0.10	−0.13
Mental health	0.38*	0.14	0.10
Self-esteem	−0.03	−0.02	−0.10
General health perceptions	0.13	0.11	−0.25
Parental impact—emotional	0.16	0.24	0.06
Parental impact—time	0.20	0.19	−0.02
Family activities	0.32*	0.34*	0.17
Family cohesion	−0.01	0.14	−0.20
Change in health	0.14	0.21	0.22

* Correlation is significant at the 0.05 level (two-tailed)

parameters of the Child Health Questionnaire, in particular for the main physical and pain parameters, in both children and parents. *In the parent version*, lower correlations were identified between emotional subscale and all CHQ domains: This represents an important bias to take in account when evaluating the results. It is possible that parents do not feel correlations between foot/ankle condition and emotions of their children. This is also more evident when observing the correlation between pain and emotion: It is possible that these two domains correlate only in most severe cases. Interestingly, the emotional

domain for the children version was correlated higher with the CHQ domains in comparison with the school and play domain. This supports the importance of asking children about their symptom and health-related quality of life, because of the well-known discordance between child and parent reports of quality of life [17].

Further studies are necessary to better define the role of the emotional subscale evaluation in the pediatric patients affected by flexible flatfoot. Despite these observations, both convergent and divergent a priori validity hypotheses were satisfied. We were then able to confirm the hypothesis

Table 6 Mean OAFQ scores and responsiveness of Italian OAFQ subscales

OAFQ domains	Mean OAFQ score (SD)		<i>p</i> value	ES	SRM
	Baseline	After surgery			
Physical					
Child	65.6 (26.4)	76.8 (21.5)	<i>p</i> < 0.01	0.4	0.3
Parent	64.2 (25.1)	74.3 (23.1)	<i>p</i> < 0.01	0.4	0.3
School and play					
Child	85.9 (13.5)	90.2 (11.2)	<i>p</i> = 0.04	0.3	0.2
Parent	82.6 (13.9)	88.3 (17.3)	<i>p</i> = 0.02	0.3	0.2
Emotional					
Child	83.0 (14.9)	90.5 (10.6)	<i>p</i> < 0.01	0.5	0.5
Parent	80.3 (17.4)	91.4 (12.3)	<i>p</i> < 0.01	0.7	0.8
Footwear					
Child	70.0 (34.7)	82.7 (27.2)	<i>p</i> < 0.01	0.4	0.4
Parent	65.1 (32.6)	81.9 (27.4)	<i>p</i> < 0.01	0.5	0.5

SD standard deviation, *ES* effect size, *SRM* standardized response mean

of construct validity. The results obtained by the responsiveness assessment showed the ability of this tool to identify significant changes associated with the surgical treatment (subtalar arthroereisis). In the original study by Morris, patients treated for trauma reported a mean improvement of about 40 points in the physical subscale, 50 points for the school and play subscale and 20 points in emotional subscale [2]. Our results were significantly lower with respect to this patients subgroup, but they were similar to the elective patients subgroup in the original study. The changes in the scores were higher in comparison with respective SDD just for emotional subscale in children and physical subscale in parents. For this reason, we cannot exclude that the other subscales are subject to error in the measure obtained with the questionnaire. Further studies are necessary to assess responsiveness of the questionnaire with different treatments. Moreover, the low ES for physical and emotional subscales could indicate a relatively low effectiveness of surgical treatment or that the surgical indication is given in patients with relatively good preoperative clinical scores. We are well aware of the study limitations. In fact, the population considered in the present study is not representative of all pediatric patients affected by foot and ankle disorders, and the results we obtained should be confirmed in patients affected by different pathologies, in order to better define the reliability of the questionnaire. Moreover, it was not possible to sort patients for different severity of the disease, since no clinical nor radiological validated classification is available for pediatric flexible flatfoot.

Up to date, in Italy, a suitable score for the evaluation of foot and ankle in pediatric patients does not exist. The present work is the first attempt to provide an effective and reliable tool for the evaluation of these patients. Even if it

is still under investigation, this tool could be used in future researches in order to adopt a common language on the international scenario.

Acknowledgments The authors want to thank Dr. P. Martinkevich for her advises about the OAFQ and for sharing her research about the Danish version of the OAFQ.

References

- Morris, C., Doll, H. A., Wainwright, A., Theologis, T., & Fitzpatrick, R. (2008). The Oxford Ankle Foot Questionnaire for children: Scaling, reliability and validity. *Journal of Bone and Joint Surgery. British Volume*, 90(11), 1451–1456.
- Morris, C., Doll, H., Davies, N., Wainwright, A., Theologis, T., Willett, K., & Fitzpatrick, R. (2009). The Oxford Ankle Foot Questionnaire for children: Responsiveness and longitudinal validity. *Quality of Life Research*, 18(10), 1367–1376.
- Guillemin, F., Bombardier, C., & Beaton, D. (1993). Cross-cultural adaptation of health-related quality of life measures: Literature review and proposed guidelines. *Journal of Clinical Epidemiology*, 46, 1417–1432.
- Myerson, M. S. (2000). *Foot and ankle disorders* (pp. 711–727). Philadelphia: W.B. Saunders Company.
- Urry, S. R., & Wering, S. C. (2001). A comparison of footprint indexes calculated from ink and electronic footprints. *Journal of the American Podiatric Medical Association*, 91, 203–209.
- Ruperto, N., Ravelli, A., Pistorio, A., Malattia, C., Viola, S., Cavuto, S., et al. (2001). The Italian version of the Childhood Health Assessment Questionnaire (CHAQ) and the Child Health Questionnaire (CHQ). *Clinical and Experimental Rheumatology*, 19(4 Suppl 23), S91–S95.
- Metcalfe, S. A., Bowling, F. L., & Reeves, N. D. (2011). Subtalar joint arthroereisis in the management of pediatric flexible flatfoot: A critical review of the literature. *Foot and Ankle International*, 32(12), 1127–1139.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159–174.

9. Fleiss, J. L., & Shrout, P. E. (1977). The effects of measurement errors on some multivariate procedures. *American Journal of Public Health*, *67*(12), 1188–1191.
10. Ravaud, P., Giraudeau, B., Auleley, G. R., Edouard-Noël, R., Dougados, M., & Chastang, C. (1999). Assessing smallest detectable change over time in continuous structural outcome measures: Application to radiological change in knee osteoarthritis. *Journal of Clinical Epidemiology*, *52*(12), 1225–1230.
11. Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297–334.
12. Fitzpatrick, R., Davey, C., Buxton, M. J., & Jones, D. R. (1998). Evaluating patient-based outcome measures for use in clinical trials. *Health Technology Assessment*, *2*(14), 1–74.
13. Nunnally, J. C., & Bernstein, I. R. (1994). *Psychometric theory*. New York, NY: McGraw-Hill.
14. Dawson, J., Doll, H., Coffey, J., Jenkinson, C., & Oxford and Birmingham Foot and Ankle Clinical Research Group. (2007). Responsiveness and minimally important change for the Manchester-Oxford Foot Questionnaire (MOXFQ) compared with AOFAS and SF-36 assessments following surgery for hallux valgus. *Osteoarthritis and Cartilage*, *15*, 918–931.
15. McHorney, C. A., & Tarlov, A. R. (1995). Individual-patient monitoring in clinical practice: Are available health status surveys adequate? *Quality of Life Research*, *4*(4), 293–307.
16. Husted, J. A., Cook, R. J., Farewell, V. T., & Gladman, D. D. (2000). Methods for assessing responsiveness: A critical review and recommendations. *Journal of Clinical Epidemiology*, *53*(5), 459–468.
17. Eiser, C., & Morse, R. (2001). Can parents rate their child's health-related quality of life?: Results of a systematic review. *Quality of Life Research*, *10*(4), 347–357.