The Oxford ankle foot questionnaire for children

SCALING, RELIABILITY AND VALIDITY

We developed the Oxford ankle foot questionnaire to assess the disability associated with foot and ankle problems in children aged from five to 16 years. A survey of 158 children and their parents was carried out to determine the content, scaling, reliability and validity of the instrument. Scores from the questionnaire can be calculated to measure the effect of foot or ankle problems on three domains of children’s lives: physical, school and play, and emotional. Scores for each domain were shown to be internally consistent, stable, and to vary little whether reported by child or parent. Satisfactory face, content and construct validity were demonstrated. The questionnaire is appropriate for children with a range of conditions and can provide clinically useful information to supplement other assessment methods. We are currently carrying out further work to assess the responsiveness of questionnaire scores to change over time and with treatment.

Typical clinical assessment methods such as range of movement, clinical rating scales, radiographs and gait analysis do not capture the patient’s perspective and may not accurately reflect how children function in their normal environment. A reliable validated method is therefore needed to measure the severity of children’s problems from the perspective of both the child and the parent.

Patient-reported outcome measures in other areas of orthopaedics have been shown to be useful for research, benchmarking and clinical audit. Although these have been developed for adults with foot and ankle problems, they have not been shown to be valid and reliable for use with children. To date, only two such instruments are available for children, the Juvenile Arthritis Foot Disability Index, but this is condition-specific for juvenile arthritis and the Clubfoot Disease Specific Instrument which was devised for parent completion. Existing generic instruments for children and their families which measure physical function, such as the Activities Scale for Kids, include a range of items that are not specific to the lower limb and so are less appropriate for children with foot and ankle problems. Currently, there is no family-assessed instrument which can measure how children’s lives are specifically affected by foot and ankle problems.

The first stage in developing a patient-reported instrument is to gain an understanding of the patients’ experience of their condition through qualitative research. Focus groups of children and adolescents between the ages of five and 15 years with a variety of foot and ankle problems were held, as were separate focus groups with their parents. This work identified a list of issues that the children themselves felt to be important, and the results obtained formed the basis of the Oxford ankle foot questionnaire for children.

The aim of the present study was to examine the scaling, reliability, face, content, and construct validity of the questionnaire. Scaling aims to retain the items which are most useful in the instrument in homogenous subscales. Reliability assesses to what extent scores are reproducible whether reported by the child or their parent, and are reproducible on different occasions. Validity indicates that the scales measure what they intend to and that the content is appropriate and meaningful to respondents.

Methods

A total of 25 questions were constructed on the basis of our previous research. Three additional items were suggested by clinicians: (i) concerns about the appearance of the foot; (ii) concerns about the way in which the child walked; and (iii) swelling, giving 28 items in total (Table I). The scoring system assessed how frequently each issue was a problem (0, ‘always’; 1, ‘very often’; 2, ‘quite often’; 3, ‘rarely’; 4, ‘never’). The second half of the
questionnaire was the Kidscreen (27-item version), a generic measure of a child’s quality of life which we used to assess the construct validity of the new questionnaire. Supplementary questions were asked about the time taken to complete the questionnaire, and whether there were any relevant issues that respondents felt had been omitted. Separate ‘Child and Adolescent’ (for the purpose of this paper henceforth ‘Child’ and ‘Parent’) versions were produced using the same items but reworded accordingly.

Survey. Over a six-month period, a group of orthopaedic surgeons, rheumatologists, orthotists and physiotherapists identified 500 patients under the age of 16 who attended a regional orthopaedic hospital with a foot or ankle problem. This included 140 children under the age of five who could not be expected to complete the questionnaire, and who were consequently excluded. The study population was therefore a convenience sample comprising 360 children aged between five and 16 at the time of survey. For scale development it is generally accepted that there should be at least five times the number of respondents as there are questions, and at least 100 in total. As there were 28 items in our test questionnaire we needed a minimum of 140 children to meet this criterion. The most frequent diagnoses were congenital deformity (commonly clubfoot), planovalgus flatfoot due to benign joint hypermobility, acute pain or trauma (such as a sprain or fracture) and cerebral palsy. The children with cerebral palsy were all ambulant and two-thirds had hemiplegia. Less frequent diagnoses included idiopathic toe-walking, syndromes (including Marfan’s and Down’s), hereditary motor sensory neuropathy, juvenile idiopathic arthritis, osteochondrosis and apophysitis.

An introductory letter and the two questionnaires, were sent by post to each family, requesting that both the parent and the child questionnaires be completed. To maximise the

<table>
<thead>
<tr>
<th>Table I. All tested questionnaire items showing domain structure from factor analysis and subsequently excluded items (child version is shown; the parent version is reworded using the prefix ‘Has your child......’)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Physical</strong></td>
</tr>
<tr>
<td>Have you found walking difficult because of your foot or ankle?</td>
</tr>
<tr>
<td>Have you found it difficult to run because of your foot or ankle?</td>
</tr>
<tr>
<td>Has it been difficult to stand up for long periods?</td>
</tr>
<tr>
<td>Have you had pain in your foot or ankle?</td>
</tr>
<tr>
<td>Have your legs been sore or ached after walking or running?</td>
</tr>
<tr>
<td>Have you felt tired because of your foot or ankle?</td>
</tr>
<tr>
<td>* Has tripping over your feet been a problem when getting about?</td>
</tr>
<tr>
<td>* Has your foot or ankle felt stiff in the morning or after resting?</td>
</tr>
<tr>
<td><strong>School and play</strong></td>
</tr>
<tr>
<td>Has your foot or ankle stopped you joining in with others in the playground?</td>
</tr>
<tr>
<td>Has your foot or ankle stopped you playing in the park or outside?</td>
</tr>
<tr>
<td>Has your foot or ankle stopped you taking part in PE lessons?</td>
</tr>
<tr>
<td>Has your foot or ankle stopped you taking part in any other lessons at school?</td>
</tr>
<tr>
<td>* Have you had to take any time off school because of your foot or ankle?</td>
</tr>
<tr>
<td>* Has your foot or ankle stopped you from taking part in any trips or outings?</td>
</tr>
<tr>
<td>* Has your foot or ankle stopped you from taking part in any clubs or activities after school?</td>
</tr>
<tr>
<td>* Has your foot or ankle stopped you from doing any jobs or chores?</td>
</tr>
<tr>
<td>* Have you had difficulty getting around your school?</td>
</tr>
<tr>
<td><strong>Emotional</strong></td>
</tr>
<tr>
<td>Has the way you walk bothered you?</td>
</tr>
<tr>
<td>Have you been bothered by how your foot or ankle looks?</td>
</tr>
<tr>
<td>Has anyone been unkind to you because of your foot or ankle?</td>
</tr>
<tr>
<td>Have you been embarrassed because of your foot or ankle?</td>
</tr>
<tr>
<td><strong>Footwear and clothing</strong></td>
</tr>
<tr>
<td>Has your foot or ankle stopped you wearing any shoes that you wanted to wear?</td>
</tr>
<tr>
<td>* Have your shoes worn out quickly because of a foot or ankle problem?</td>
</tr>
<tr>
<td>* Has anything to do with your foot or ankle affected what clothes you have chosen to wear?</td>
</tr>
<tr>
<td>* Has the skin on your foot or ankle been sore from rubbing on your shoes or splints?</td>
</tr>
<tr>
<td><strong>Other items</strong></td>
</tr>
<tr>
<td>† Have you found climbing difficult because of your foot or ankle?</td>
</tr>
<tr>
<td>† Has your foot or ankle been swollen?</td>
</tr>
<tr>
<td>† Have you felt sad because of trouble with your foot or ankle?</td>
</tr>
</tbody>
</table>

* excluded after Rasch analysis
† excluded after factor analysis
response rate it was stated that if the child was unable or
did not wish to complete one, only the parent questionnaire
required completion. If no response was received after the
initial mailing, a reminder letter and a second copy of the
questionnaires was sent after six weeks. Families who
returned completed questionnaires promptly within the
first month were sent a second set of questionnaires after
two weeks to assess the stability of the responses. Data
from the completed questionnaires were entered twice into
a custom database by administrators and the consistency of
the two entries was audited; discrepancies between the two
entries were corrected. The survey procedures were
approved by the local Ethics Committee.

**Statistical analysis.** Data were analysed primarily using
SPSS version 15.0 (SPSS Inc., Chicago, Illinois) and pre-
senting using mean with SD, median with range, or number
(%) as appropriate. Although the rating scales of individual
questionnaire items were ordinal, total scores were more
continuously distributed, making parametric statistical anal-
yses appropriate. Factor analysis with a principal compo-
ents extraction (i.e. retaining all the variability in each item)
and varimax rotation was used to explore whether distinct
(i.e. uncorrelated), and similar (across the parent and child
datasets) dimensions of groups of items (factors) could be
extracted from the two datasets. Factors (or domains) were
extracted if their eigenvalue (the proportion of total variance
explained by the factor relative to that explained by a single
item) was > 1.9,10 The structure of the identified dimensions
and the functioning and fit of individual items was examined
using item response theory, with the fitting of a one-parame-
ter Rasch model.11 Rasch analysis is recommended as the
best contemporary method for scale development.12 Item
response theory assumes that as the trait being measured
increases in this instance the child’s disability, the probability
of a maximum score on the item also increases. For the item
response theory analyses, responses to all items were
recorded into three rather than five, categories (0, ‘always’,
‘very often’ or ‘sometimes’; 1, ‘rarely’; and 2, ‘never’) in
order to create groups of more equal size. Domain scores
were transformed to a percentage score, where a higher score
indicates better health for ease of interpretation. All other
analyses used un-recoded item and domain scores. Statistical
significance was taken at the 5% level (p < 0.05).

**Item reduction.** An item was considered for exclusion from
the candidate questionnaire if it met any of three predeter-
mined criteria: (i) if, in terms of its response distribution,
there were particularly high ceiling or floor effects (where
at least 90% of the responses to the item fell into either of
the most extreme response categories); (ii) if, during factor
analysis, the item cross-loaded (loading > 0.40 on more
than one factor), indicating that it was not unique to one
factor; or (iii) if, on fitting the Rasch model to the domain
scales, the item showed a particularly poor fit to the model,
thereby adversely affecting the scale.

**Reliability.** The frequency of response to individual ques-
tionnaire items was reviewed for each of the child and
parent samples. Chance-corrected agreement between
child and parent responses to the same items were assessed
using the κ statistic,13 where values of 0.21 to 0.40 repre-
sent fair agreement, 0.41 to 0.60 moderate agreement,
and 0.61 to 0.80 good agreement over that expected by
chance.14 The internal consistency of the domain scales
was assessed using Cronbach’s α statistic and item-total
correlations. The effect of deleting items on internal reli-
ability was systematically assessed. Scales with an α statist-
ic between 0.7 and 0.9 are considered to be composed of
items that adequately measure a uniform construct.15 The
reliability of child- and parent-reported domain scores
was assessed using an intra-class correlation coefficient.
Test-retest reliability was assessed by calculating an intra-
class correlation coefficient for the subset of children and
parents for whom a second rating was received one month
after the initial questionnaire. An intraclass coefficient of
> 0.7 is considered adequate for use in populations and a
reliability greater than 0.9 good enough for use with
individuals.16 The statistical significance of differences in
child- and parent-reported mean scale scores were exam-
ined using paired t-tests. The effect of the age of the child
on both parent-child and test-retest reliability was exam-
ined in three ways: first, by grouping the children into
quartiles by age, and second, by dividing into three pre-
deﬁned groups (less than eight years, eight to 11 years,
and 12 years and older); ﬁnally, we investigated the effect
of excluding children under eight years old from the
analysis. The statistical significance of any differences
between the groups was assessed using analysis of vari-
ance (ANOVA).

**Construct validity.** Construct validity was assessed by test-
ing two hypotheses. First, given that all children had been
assessed and some had received treatment, we hypothesised
that questionnaire scores would be lower (more severe dis-
ability) for children with neurological conditions and
chronic syndromes than for those with more benign diag-
noses and acute or fluctuating conditions. Children with
diagnosis of cerebral palsy, hereditary motor sensory
neuropathy and named clinical syndromes were grouped
together; all other diagnosis were classified as benign, acute
or fluctuating. The statistical significance of any difference
in scores was assessed using ANOVA.

Second, as the severity of a foot and ankle problem
increased, both Kidscreen and questionnaire scores
would decrease (i.e. get poorer) as indicated by positive
correlations between the scale scores. The Oxford ankle
foot questionnaire scores were correlated (using Pear-
son’s correlation coefficient, r) with scores from the Kid-
screen domains Physical Wellbeing, Psychological
Wellbeing, Autonomy and Parents, Peers and Social Sup-
port, and School Environment.10 Scales measuring similar
constructs were expected to show a higher correlation
than those measuring dissimilar constructs (where r < 0.3
indicates a low, 0.3 to 0.49 a moderate and > 0.50 a high
correlation).17
Results

Survey response and sample characteristics. The survey was mailed to the families of 360 children; 15 packages were returned undelivered. Valid responses were received from families of 158 children, giving a response of 46% including 142 child-completed and 154 parent-completed questionnaires. Children who self-reported were marginally older (mean age 11.1 years, SD 2.9) than those for whom a parent-completed questionnaire was returned (mean age 10.8 years, SD 3.1). Participants commented that the questionnaire was both quick and easy to complete. Although there were slightly more boys than girls, the proportion of participants within each diagnostic category was representative of the distribution among the survey population.

Item functioning. There were very few items for which a response was missing; there were only four such instances in all the children’s responses. For the majority of items the most frequent response was that the issue was ‘never’ a problem (range of ‘never’ responses: child 22% to 79%, parent 26% to 77%). The median response was ‘never’ for 16 of 28 child items and 14 of 28 parent items. For most items the response frequency reduced with increasing severity of response; very few responses were ‘always’ (child: 2% to 17%; parent: 0% to 24%). Over 70% of participants gave the response ‘never’ for six child (and four parent) items. Over 60% gave the response ‘never’ for 11 child (and nine parent) items. The mean item scores ranged from 2.3 to 3.7 for the child items and 2.2 to 3.7 for the parent items. The items perceived to be the least problem (> 70% indicating ‘never’) were difficulty walking, foot or ankle pain, and soreness or aching after walking or running.

Creating the scales. Initial factor analyses on child and parent datasets produced six and four factor solutions, respectively, but with one factor explaining most of the variance. Three items (‘swelling’, ‘climbing’, and ‘sad’) were removed at this stage because they loaded with dissimilar items or cross-loaded on more than one factor. A coherent four-factor solution was then found in which 25 items were grouped together in dimensions that we titled ‘Physical’ (eight items), ‘School and Play’ (nine items), ‘Emotional’ (four items), and ‘Footwear and Clothing’ (four items) (Table I). Rasch analyses on each dimensional scale resulted in two items being dropped from the Physical domain and five items from the School and Play domain; no items were dropped from the Emotional scale. The footwear and clothing scale was dropped altogether because of its inadequate measurement properties, as demonstrated by item response theory analysis.

Child and parent reliability. Agreement between parents and children for individual items was fair to moderate (κ from 0.3 to 0.5). However, correlations between the child and parent domain scale scores were fairly good. In paired t-tests, the differences between child and parent-reported mean scores were not statistically significant. Excluding children less than eight years old from the analyses increased reliability only slightly.

Test-retest reliability. The stability of domain scores over one month was assessed in 38 children. The child-reported scores for the School and Play and Emotional domains were very stable, whereas the Physical domain scores were less stable. Parent-reported scores were stable for the Physical and Emotional domains but less so for School and Play. The child-reported Physical domain was the only one in which scores differed significantly between tests (p = 0.03; all other domains p > 0.1). Scores were slightly higher at the second assessment, except for children under the age of eight, where the scores tended to be slightly lower. Reliability increased only slightly when responses from children under the age of eight were excluded. The domain scores of those who were invited to take part in the test-retest study, that is, the early responders to the survey, and those who took part did not differ from the scores of the later responders, who did not form any part of the retest study group (p > 0.2 for all domains, whether child- or parent-reported).

Construct validity. Domain scores for the group with neurological conditions and syndromes were statistically significantly lower than those for the other group in all domains, whether child- or parent-reported (17% to 28% lower, p < 0.001). Correlations between the questionnaire and Kidscreen scores were mostly statistically significant at p < 0.01. The correlation between the Physical and Kidscreen Physical Wellbeing domains was moderate (child-reported) or high (parent-reported); The questionnaire Physical scores correlated poorly with the Kidscreen Peers and Social Support domain. The questionnaire School and Play domain scores correlated well with Kidscreen’s Physical Wellbeing scale and less well with the Kidscreen Peers and Social Support. The correlation between the Oxford ankle foot questionnaire School and Play and the Kidscreen School Environment domains was higher using child- than parent-reported scores. The Oxford ankle foot questionnaire Emotional domain scores correlated moderately with the Kidscreen Psychological Wellbeing and Physical Wellbeing scores. Both the child- and the parent-reported questionnaire Emotional domain scores correlated moderately with the Kidscreen Autonomy and Parents scores. However, whereas the parent-reported School and Play and Physical domain scores correlated poorly with the Kidscreen Autonomy and Parents scores, the child-reported scores for these same domains correlated highly.

Discussion

The Oxford ankle foot questionnaire offers a means to measure the effects of foot and ankle conditions in children using issues that are considered important by children. Each domain scale was developed using a series of classic and contemporary test criteria (including item response theory) which ensure that each scale is internally consistent, i.e. measures a single trait and that each item has different levels of difficulty or severity. The final instrument com-
prises 15 questions of which 14 can be divided into three subscales which measure problems in domains titled Physical (six items), School and Play (four items), and Emotional (four items). Raw domain scores can be transformed into percentage scores to make them easier to interpret; lower scores indicate more severe disability. Although the item asking about ‘Being able to wear the shoes you want’ did not fit with the scales, it was salient to many children and to most girls. The item has strong face validity and is included as a categorical descriptive variable but not allied to any domain scale.

A balance must be struck between providing a quick and expedient method to measure the construct of interest while also covering all the relevant issues. The few additional activities that were suggested by participants (dancing, ice-skating and cycling) were relevant to the Physical scale, but our analyses already suggested some item redundancy in this domain. Clinicians recommended three additional items to those identified from the focus groups. These were included in the initial questionnaire, and two are retained in the final instrument because children rated them as important. We are therefore confident that the questionnaire has satisfactory face and content validity.

Discordance between child and parent reports of quality of life is well recognised. Although the scores derived from children and parents using the questionnaire did not correlate perfectly, they rate extremely favourably in comparison with child-parent reliability in other instruments. This is probably because many of the issues that the questionnaire explores are readily observable, or issues about which parents are likely to be well informed. Both child and parent perspectives should be assessed whenever possible, although children are best placed to report on their emotional and social functioning and wellbeing.

The consistency of domain scores when measured a few weeks apart supports the stability of scores. The exception was the child-reported Physical scores, which might be expected to vary with time. The questionnaire asks children to consider ‘the last week’ so a shorter interval of perhaps a few days between questionnaire administration would be a better test of score stability. A limitation to the test-retest survey was that it was only conducted with the early responders, and not all of those respond. However, we can be confident that there was no response bias in our results because the mean domain scores of the early and late responders were similar. Nevertheless, further work is warranted to assure the stability of domain scores. Others have suggested that children under eight years old are less reliable than children eight years and older. Our results supported this view, as the youngest children showed a different pattern of response in the test-retest study, and greater variation between tests. Therefore, although children under the age of eight years might complete the questionnaire, their scores should be interpreted with caution.

On testing our hypothesis of discriminant validity, children with pervasive neurological diagnoses and syndromes scored statistically significantly lower than children with benign, acute or fluctuating conditions and on testing the convergent validity of the questionnaire we observed higher correlations between the similar physical and psychological domains scores from the questionnaire and Kidscreen. Lower correlations were observed between the physical and emotional domains and how children felt about their relationships with friends and peers (divergent validity), measured using the Kidscreen peers and social support. Correlations between school-related domains of each instrument were significant but moderate. There was a notable difference in correlation between the child- and the parent-reported domains of the questionnaire and the Kidscreen scale measuring ‘autonomy and relationship with parents’. The children’s scores implied a statistically significant pattern of feeling disempowered with increasing disability, whereas the parents’ scores did not reflect this trend so strongly. This emphasises the importance of asking children about their health and experiences whenever possible. Taken together, the evidence of appropriately discriminating between groups of diagnoses, and correlations with Kidscreen, support the construct validity of the questionnaire.

Limitations to this study include the fact that the children being assessed were already receiving treatment: it is therefore unsurprising that mean and median scores were low. However, despite this, we still observed scores over the full range of each scale. Also, there were insufficient numbers of children within each response category for Rasch analysis, which we were obliged to address by recoding the responses into three categories. We also acknowledge the moderate response to the postal survey, although those who participated in our survey were representative of the target population in terms of diagnosis and treatment received. We do not know why some families did not respond, but the response rate was typical for this type of postal survey. We would expect a better response if the questionnaire was issued in a clinical setting, or as part of a trial in which the child or family had already consented to take part.

The Oxford ankle foot questionnaire is now available for use in research and clinical practice with children aged between five and 16 years. It has been validated to discriminate between children reporting different degrees of disability from foot or ankle problems. We included children with a range of diagnoses in the survey. The qualitative component of our research suggested that the ways in which children are affected by their foot or ankle problems are similar whatever the diagnosis. The severity of these effects are reflected by different scores for different diagnoses. The instrument is not suitable for those who are unable to walk, or who have a significant proximal component to their disability. Although we conducted the survey by post, we believe the instrument can be useful in the clinical setting and may be administered by telephone. Respondents should be encouraged to complete all the items, as missing responses undermine the validity of scale scores. We found few instances of missing data, which suggests that the questionnaire is easy to complete.
when at least half the items in any one domain are completed, it is common for the average of the other item scores in that scale to be imputed as a best estimate. We were unable to assess the effect of this technique because of our lack of missing data, but recommend this approach should missing responses occur.8

Future research in to the development and validation of the questionnaire will involve assessing whether the instrument is responsive to change. We will administer the questionnaire to children attending clinics for the first time and who have not received recent treatment for their foot or ankle problem. We will monitor changes in the OAFQ domain scores, as well as the change in other related scores, after any intervention. We will reassess test-retest reliability, and continue to monitor the dimensionality and scaling of the instrument as more data become available.

We are grateful to the children and families who took part, and to our colleagues at the Nuffield Orthopaedic Centre who identified the survey population, notably orthotists, orthopaedic surgeons, physiotherapists, and staff at the Oxford Paediatric and Adolescent Rheumatology Clinic. We are also grateful to A. King for assistance with programming and data management, and to H. Hay for helping to administer the survey.

This study was funded by the UK Medical Research Council through a Research Training Fellowship in Health Services Research held by Dr Morris.

No benefits in any form have been received or will be received from a commercial party related directly or indirectly to the subject of this article.

Supplementary material

A further opinion by Mr A. Younger is available with the electronic version of this article along with the final instrument, further details of all item response theory analyses and tables showing characteristics of survey population and participants, domain scale scores, reliability coefficients and Pearson’s correlation between scores on our website at www.jbjs.org.uk

References