Rasch analysis of Dutch-translated version of the Foot Impact Scale for rheumatoid arthritis

James Woodburn¹, Thea P. M. Vliet Vlieland², Marike van der Leeden³ and Martijn P. M. Steultjens¹

Abstract

Objective. To translate the Foot Impact Scale for RA (FIS-RA) to the Dutch target language and to evaluate its internal construct validity using Rasch analysis.

Methods. Forward and backward translations of the original English version of the FIS-RA scale, combined with synthesis techniques and expert committee review, were undertaken to produce a final Dutch version with two subscales for impairment/footwear (FIS-RAIF) and activity/participation (FIS-RAAP). The pre-final version was field tested in RA patients to investigate face and content validity. FIS-RA questionnaires were completed by 207 Dutch RA patients. Rasch analysis tested the data for overall fit to the model, item and person fit, unidimensionality, differential item function (DIF) by age, gender and disease duration, targeting, reliability and local response dependency. Item deletion and re-analysis were planned, where Rasch model assumptions were violated.

Results. The FIS-RAIF ($P < 0.0001$) and FIS-RAAP ($P < 0.0001$) subscales did not fit the overall Rasch model. Misfitting items, DIF by age, gender and disease duration, and local response dependency were observed in both subscales. Item thresholds showed good coverage over both subscales although a floor effect was observed for the FIS-RAAP subscale. The person separation index was 0.81 and 0.92 for the FIS-RAIF and FIS-RAAP subscales, respectively. Both subscales were not unidimensional. Item deletion and repeat Rasch analysis produced two subscales that fitted the Rasch model and were unidimensional.

Conclusion. A Dutch language version of the FIS-RA questionnaire was successfully developed using Rasch analysis. Subscales for impairment/footwear and activity/participation showed good construct validity and were unidimensional.

Key words: Rheumatoid arthritis, Foot impairment disability, Rasch analysis.

Introduction

The prevalence of foot-related impairment and disability in RA ranges from 35 to 70% across the disease spectrum [1–3]. The Foot Impact Scale for RA (FIS-RA) was developed to measure foot-related impairment and disability at the individual and group level [4]. This multi-domain, self-administered questionnaire comprises two subscales for impairment/footwear (21 items) and activities/participation (30 items). Good psychometric properties, external validity and test-retest reliability for the original English version have been demonstrated [4, 5]. The FIS-RA has been used in a number of cross-sectional cohort, intervention and audit studies [6–8]. The FIS-RA focuses on relevant clinical issues that may not be part of standard RA care. Clinicians are using the tool to assess foot-specific levels of impairment and disability and to assess the effectiveness of treatment both in early and established RA clinics.

Across Europe there are a number of groups interested in understanding the burden of foot disease in RA, as well as the development and testing of drug, surgical and non-pharmacological interventions. A requirement therefore
exists to adapt measures beyond the source language. Successful cross-cultural validation of other questionnaires in RA, such as the work instability scale, has been reported [9]. Moreover, the use of modern psychometric techniques such as the Rasch measurement model provides robust analysis of internal construct validation [9]. Therefore, the aim of this study was to translate the FIS-RA scale into Dutch and evaluate its internal construct validity and unidimensionality using Rasch analysis.

Patients and methods

Translation of the FIS-RA

Translation of the FIS-RA into Dutch was undertaken at the Department of Rheumatology at Leiden University Medical Center (LUMC). The protocol followed the published guidelines for the process of cross-cultural adaptation of self-reported measures [10]. Briefly, this comprised: (i) initial forward translations from the English language version to Dutch by two bilingual translators, native to the Dutch language; (ii) synthesis of the translations to resolve discrepancies between translators; (iii) backward translations of the new Dutch FIS-RA scale to English by two lay translators working independently of Stage 1; (iv) expert committee review to reach consensus and produce the per-final version; and (v) field testing in the target population to test face and content validity for the new scale. Field testing was conducted on 14 RA patients with a median (range) age of 48.5 (31–86) years and disease duration of 18.5 (1–27) years with mixed educational level. The study was judged to be a non-medical research according to the Medical Research involving Human Subjects Act by the Medical Ethics Review Committee of the centres in both Leiden and Amsterdam. The original English, translated Dutch and revised translated Dutch versions of the FIS-RA are available as supplementary data online.

Rasch analysis

The Dutch FIS-RA was administered to 207 RA patients from two Rheumatology centres in Leiden and Amsterdam. Subjects had a median (range) age of 60 (23–89) years and disease duration 8.5 (1–25) years. Thirty-four patients were male and 176 were female. The Rasch analysis approach has been described in detail elsewhere [11]. Briefly, fit to the model was tested at the scale and individual item level along with differential item function (DIF) by gender, age (<50/>50 years) and disease duration (<5/>5 years). Unidimensionality was examined using the independent subject t-test procedure. Here, two item subsets are created from items loading negatively and positively on the first residual factor in the principal component analysis, and persons estimates derived from each set are compared using t-tests. The assumption of local independence is supported when <5% of tests are shown to be significant. Local response dependency indicates that the response to one item determines the response to another. Correlation between item residuals >0.3 was used to indicate dependency. Estimates of the internal consistency reliability of the subscales were based on the person separation index, where 0.7 is considered a minimum value for group use. If fit to the Rasch model was not achieved, misfitting items and those with DIF were removed and the analysis repeated. All analyses were conducted with RUMM2020 software version 4.0 (Rumm Laboratory, Perth, WA, Australia). P-values were Bonferroni adjusted.

Results

Translation of the FIS-RA

The process of forward and backward translation, synthesis and expert committee review went well. Discussions were generated on colloquial expressions in items such as, ‘for a bit’ and ‘in the background’ and item grammatical sentence structure. Field testing revealed the scale to be acceptable and easy to understand. Some patients identified overlapping items and others had difficulty relating item responses to one moment in time as instructed.

Rasch analysis

Rasch analysis of the 21-item FIS-RA subscale for impairment/footwear (FIS-RAIF) demonstrated lack of fit to the model, with a significant item–trait total chi-square interaction (P < 0.0001). The mean (s.d.) fit residual for items was −0.13 (1.42). The mean (s.d.) fit residual for persons was −0.31 (0.98), which indicated no serious misfit among subjects in the sample. Analysis of the 30-item FIS-RA subscale for activity/participation (FIS-RAAP) demonstrated lack of fit to the model with a significant item–trait total chi-square interaction (P < 0.0001). The mean (s.d.) fit residual for items was −0.42 (1.08) and for persons was −0.26 (0.74). The five best and worse fitting items are shown in Table 1.

For the FIS-RAIF subscale, two items (5: pain and tension in feet, fit residual −2.6; and 17: I need a lot of padding under my feet, fit residual 3.2) had residual fit values greater than ±2.5, but no statistically significant misfit. Neither of these items had disordered thresholds. For the FIS-RAAP subscale, Item 9, ‘walk on cobbles’ had a residual fit value greater than ±2.5 (2.6) and showed significant misfit (Table 1). Item 12, ‘longer to do things’ and Item 3, ‘can’t run’ had residual fit values greater than ±2.5, but showed significant misfit. None of these three items had disordered thresholds.

For the FIS-RAAP subscale, Item 5 showed DIF by gender (P = 0.0003) and age (P = 0.0003), with female patients and those aged >50 years more likely to affirm the item. For the FIS-RAIF subscale, there was DIF by gender and age on Item 3, ‘can’t run’ (gender: P < 0.0001; age: P < 0.0001); Item 12, ‘longer to do things’ (gender: P < 0.0001; age: P < 0.0001); and Item 19, ‘plan everything out’ (gender: P = 0.0002; age: P < 0.0001). Item 19 (P = 0.0006) and Item 20, ‘can’t get any shoes on’ (P = 0.0003) had DIF by disease duration. Item 3 was more likely to be affirmed by female patients and those aged >50 years; Item 12 by males patients and those aged <50 years; Item 19 by male patients and those aged >50 years with disease duration >50 years; and the item with DIF on age was Item 19, where male patients and those aged >50 years more likely to affirm the item.
The targeting of the item thresholds for subjects in both subscales is shown in Fig. 1. There was good coverage of thresholds over the range of foot-related impairment subscales. The mean (S.D.) person location value \([FIS-RA IF = 1.27 (1.90)]\) of subjects, respectively, had significantly different person estimates based on two item subsets. Both revised subscales were unidimensional. The FIS-RAIF subscale showed that 4.3% (95% CI 1.5, 7.1) of subjects and the FIS-RAAP subscale showed that 3.9% (95% CI: 1.3, 6.5) of subjects, respectively, had significantly different person estimates based on two-item subsets.

### Discussion

Despite its reported good psychometric properties in the initial development of the original English version of the FIS-RA scale, the translated Dutch version did not fit Rasch model expectations. The FIS-RAIF subscale performed worse showing lack of unidimensionality, lack of fit to the model, item misfit, item DIF by age and gender and local response dependency. In the development of the original English version, items related to impairment and footwear were combined into a single scale but not tested using the independent t-test method [4]. Lack of unidimensionality is a fundamental violation of the Rasch model [11]. However, through an iterative analysis process, item deletion resulted in two revised subscales, which showed good fit to the Rasch model and unidimensionality. Interestingly, items that were deleted, which showed misfit and DIF, tended to be those identified during the translation process that were either badly constructed, ambiguous or duplicated. For example, in the impairment subscale, some Dutch patients expressed some difficulty in understanding the meaning of tension within the item, ‘At the end of the day there is pain and tension in my feet’. Coupled with pain, two concepts are presented in this poorly constructed item that subsequently demonstrated misfit and DIF by age and gender. Dutch patients were also confused by Item 17, ‘I need a lot of padding under my feet’; some assuming the padding was bodily soft-tissue and others insole or cushioning material. This item also demonstrated misfit. Further, items are candidates for deletion due to local response dependency. For example, in the footwear construct, it is intuitive to see how a positive response to either Item 14, ‘I need more things so quickly’; and Item 24, ‘feel isolated because I can’t go very far’ and Item 27, ‘can’t go for walks with the people close to me’. Independent t-tests indicated that both subscales were not unidimensional. The FIS-RAIF subscale showed that 19.8% (95% CI 14.3, 25.3) of subjects and the FIS-RAAP subscale showed that 7.2% (95% CI 5.4, 10.7) of subjects, respectively, had significantly different person estimates based on two item subsets.

### Table 1

<table>
<thead>
<tr>
<th>Item</th>
<th>Location parameter</th>
<th>SE</th>
<th>Fit residual</th>
<th>Chi-square</th>
<th>Chi-square probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>FIS-RA impairment/footwear subscale</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>–0.61</td>
<td>0.16</td>
<td>0.14</td>
<td>0.10</td>
<td>0.75</td>
</tr>
<tr>
<td>15</td>
<td>–1.99</td>
<td>0.18</td>
<td>–1.32</td>
<td>0.27</td>
<td>0.60</td>
</tr>
<tr>
<td>8</td>
<td>0.95</td>
<td>0.19</td>
<td>1.07</td>
<td>0.44</td>
<td>0.51</td>
</tr>
<tr>
<td>9</td>
<td>1.05</td>
<td>0.19</td>
<td>–0.47</td>
<td>0.45</td>
<td>0.50</td>
</tr>
<tr>
<td>12</td>
<td>1.78</td>
<td>0.23</td>
<td>–0.52</td>
<td>0.59</td>
<td>0.44</td>
</tr>
<tr>
<td>16</td>
<td>–1.45</td>
<td>0.17</td>
<td>–0.11</td>
<td>4.80</td>
<td>0.03</td>
</tr>
<tr>
<td>3</td>
<td>0.34</td>
<td>0.17</td>
<td>–1.89</td>
<td>5.76</td>
<td>0.02</td>
</tr>
<tr>
<td>2</td>
<td>–1.38</td>
<td>0.17</td>
<td>–1.86</td>
<td>7.08</td>
<td>0.007</td>
</tr>
<tr>
<td>21b</td>
<td>2.13</td>
<td>0.26</td>
<td>1.57</td>
<td>7.73</td>
<td>0.005</td>
</tr>
<tr>
<td>5b</td>
<td>–1.56</td>
<td>0.17</td>
<td>–2.63</td>
<td>8.21</td>
<td>0.004</td>
</tr>
<tr>
<td>FIS-RA activity/participation subscale</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>1.06</td>
<td>0.22</td>
<td>–0.67</td>
<td>0.04</td>
<td>0.85</td>
</tr>
<tr>
<td>14</td>
<td>–1.72</td>
<td>0.18</td>
<td>0.33</td>
<td>0.07</td>
<td>0.89</td>
</tr>
<tr>
<td>4</td>
<td>1.19</td>
<td>0.22</td>
<td>–0.67</td>
<td>0.07</td>
<td>0.79</td>
</tr>
<tr>
<td>13</td>
<td>–0.30</td>
<td>0.18</td>
<td>–1.04</td>
<td>0.16</td>
<td>0.69</td>
</tr>
<tr>
<td>29</td>
<td>1.55</td>
<td>0.24</td>
<td>–0.36</td>
<td>0.22</td>
<td>0.64</td>
</tr>
<tr>
<td>20</td>
<td>–0.55</td>
<td>0.18</td>
<td>–2.03</td>
<td>7.06</td>
<td>0.008</td>
</tr>
<tr>
<td>19</td>
<td>0.88</td>
<td>0.21</td>
<td>–2.00</td>
<td>8.56</td>
<td>0.003</td>
</tr>
<tr>
<td>9c</td>
<td>–1.15</td>
<td>0.18</td>
<td>2.55</td>
<td>10.63</td>
<td>0.001</td>
</tr>
<tr>
<td>12c</td>
<td>–2.19</td>
<td>0.19</td>
<td>–1.65</td>
<td>11.24</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>3c</td>
<td>–1.89</td>
<td>0.18</td>
<td>1.83</td>
<td>26.83</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

*Bonferroni-adjusted probability value. bFit residual greater than ±2.5. cMisfitting (i.e. significantly deviating from Rasch model expectations according to \(P\)-value).
shoes with plenty of room in them’, Item 15, ‘I am limited in my choice of shoes’, or Item 16, ‘I need a wider fit of shoes’ influences the response to the others.

In the development of the original FIS-RA scale, items related to activity limitation and participation were combined to a single 31-item FIS-RAAP subscale [4]. The psychometric properties of this subscale were better in the Dutch-translated version than the FIS-RAIF subscale. Strictly speaking, the scale is not unidimensional although the lower 95% CI value (5.4%) narrowly exceeds lower 5% threshold in the binomial distribution for the proportion of patients demonstrating statistically significant t-tests. Further work is necessary to explore the influence of individual items given the findings of misfit, DIF and local item dependency.

There are now ranges of generic and disease-specific foot scales available for research use [5]. The foot function index, for example, was originally developed for use in RA although it has been subsequently revised for more generic use [12]. It has, however, undergone successful cross-cultural validation in several languages [13–15]. This encourages us to pursue further revision, adaptation and cross-cultural validity of the FIS-RA scale, as it possesses some important qualities; chiefly excellent face validity, since items were generated from qualitative patient interviews. Further work is required to understand psychometric properties such as measurement error, for example, the minimum important clinical difference, responsiveness and interpretability through the development of Rasch transformed scores. The latter permits parametric statistical analyses in studies using the FIS-RA, including power analyses for sample numbers. The FIS-RA scale also allows the measurement constructs for impairment and activity/participation to be aligned.

---

**Fig. 1** Distribution of subjects and 21-item FIS-RAIF subscale (a) and 30-item FIS-RAAP subscale (b) item thresholds on the FIS-RA logit scale for foot-related impairment and disability.
to an internationally accepted framework for functioning, disability and health [4].

In conclusion, in its original form, the Dutch-translated FIS-RA had poor construct validity and was not a unidimensional scale. However, item deletion and re-analysis provided a revised version of both subscales for impairment/footwear and activity limitation/participation restriction with excellent psychometric properties. European researchers have expressed interest in the FIS-RA as an assessment tool for use in early RA and as an outcome tool for a range of non-pharmacological (footwear, orthotics and splints, and exercise) and surgical intervention studies. Future adaptation and cross-cultural validation analyses are planned that will add UK, German and Hungarian data. Currently, the Dutch scale can be used as a research tool for group-level analyses.

**Rheumatology key message**

- A Dutch-translated version of the FIS-RA has been developed and validated.

**Acknowledgement**

Funding: This work was supported by a European league Against Rheumatism (EULAR) Health Professionals Research Grant.

Disclosure statement: The authors have declared no conflicts of interest.

**Supplementary data**

Supplementary data are available at Rheumatology Online.

**References**


