Development and validation of the patient-reported “Facial Function Scale” for facioscapulohumeral muscular dystrophy

Karlien Mul, Feri Wijayanto, Tom G. J. Loonen, Perry Groot, Sanne C. C. Vincenten, Simone Knuijt, Jan T. Groothuis, Thomas J. J. Maal, Tom Heskes, Nicol C. Voermans, and Baziel G. M. van Engelen

Department of Neurology, Donders Institute for Brain, Cognition, and Behaviour, Radboud University Medical Center, Nijmegen, The Netherlands; Institute for Computing and Information Sciences, Radboud University, Nijmegen, The Netherlands; Department of Informatics, Universitas Islam Indonesia, Yogyakarta, Indonesia; Department of Oral and Maxillofacial Surgery, Radboudumc 3D-Lab, Radboud University Nijmegen Medical Center, Nijmegen, The Netherlands; Department of Rehabilitation, Donders Institute for Brain, Cognition, and Behaviour, Radboud University Medical Center, Nijmegen, The Netherlands

ABSTRACT

Purpose: Facial weakness and its functional consequences are an often underappreciated clinical feature of facioscapulohumeral muscular dystrophy (FSHD) by healthcare professionals and researchers. This is at least in part due to the fact that there are few adequate clinical outcome measures available.

Methods: We developed the Facial Function Scale, a Rasch-built questionnaire on the functional disabilities relating to facial weakness in FSHD. A preliminary 33-item questionnaire was created based on semi-structured interviews with 16 FSHD patients and completed by 119 patients. For reliability studies, 73 patients completed it again after a two-week interval. Data were subjected to semi-automated Rasch analysis to select the most appropriate item set to fit model expectations.

Results: This resulted in a 25-item unidimensional, linear-weighted questionnaire with high internal consistency (person separation index = 0.92) and test–retest reliability (patients’ locations ICC = 0.98 and items' locations ICC = 0.99). Good external construct validity scores were obtained through correlation with the Communicative Participation Item Bank questionnaire, examiner-reported Facial Weakness Score and facial weakness subscale of the FSHD evaluation score (respectively $r = 0.733$, $r = -0.566$, and $r = 0.441$, all $p < 0.001$).

Conclusions: This study provides a linear-weighted, clinimetrically sound, patient-reported outcome measure on the functional disabilities relating to facial weakness in FSHD, to enable further research on this relevant topic.

IMPLICATIONS FOR REHABILITATION

- Facial weakness and its functional consequences are an often underappreciated clinical feature of facioscapulohumeral muscular dystrophy (FSHD), both in symptomatic treatment and in research.
- To enable the development and testing of therapeutic symptomatic interventions for facial weakness, clinical outcome measures are required.
- This study provides a linear-weighted, clinimetrically sound, patient-reported outcome measure on the functional disabilities relating to facial weakness in FSHD patients.

Introduction

Facioscapulohumeral muscular dystrophy (FSHD) is a hereditary muscle disorder. The disease affects many muscles in the human body including upper and lower extremity as well as trunk muscles, but one of its most characteristic symptoms is weakness of the facial muscles [1].

The facial weakness leads to functional impairments in, among others, speaking and swallowing [2], and to a reduction in facial expression hindering non-verbal communication and negatively affecting participation. It is therefore experienced as a debilitating symptom by FSHD patients [3–5]. In spite of its clinical relevance, facial weakness is an often underappreciated symptom both in symptomatic treatment by healthcare professionals and in research on FSHD.

To enable the development and testing of therapeutic symptomatic interventions for facial weakness, clinical outcome measures are required. These could include impairment measures that grade the degree of facial weakness like the Iowa Oral Pressure Instrument or a physician-reported score [2,6]. However, both in FSHD and other medical conditions affecting the face, it was...
shown that the degree of weakness does not necessarily correlate with the experienced symptom burden [5,7]. Therefore, alongside more objective outcome measures, patient-reported measures of facial functioning would be an outcome of interest for the assessment of clinical relevance of (the degree of) facial weakness and for future intervention studies.

Ideally, such an outcome measure should be an interval scale that provides a numerical value, in contrast to ordinal scales that only provide a rank order [8,9]. Rasch analysis, a mathematical model based on the assumption that a patient with a high overall ability will have a higher probability of fulfilling any single task compared to a patient with a lower overall ability, can be used to create interval scales from ordinal data [10]. A crucial advantage of interval scales is that the difference or distance between scores on the scale is equal and therefore allows for parametric statistical testing and accurate comparisons of changes across the scale [9,11]. Additionally, Rasch analysis can provide plots to evaluate how well the test items are capturing the variable of interest. For example, Rasch analysis of the Motor Function Measure in FSHD patients and of the Muscular Impairment Rating Scale in myotonic dystrophy revealed that these outcome measures on motor function were not able to capture patient abilities sufficiently across the entire disease spectrum [12,13]. For chronic inflammatory demyelinating polyneuropathy (CIDP), it was shown that a Rasch-built interval scale was more responsive compared to a widely used ordinal scale [14].

The aim of this study was to develop and validate a Rasch-built patient-reported outcome measure on the functional disabilities relating to facial weakness in FSHD patients.

Materials and methods

Preliminary questionnaire development

We obtained qualitative data through semi-structured in-depth interviews with 16 FSHD patients on the functional and psychosocial consequences of facial weakness in FSHD [5]. Next, a team of caregivers experienced in the management of FSHD patients including neurologists, speech- and language pathologists, occupational therapists, and rehabilitation specialists, drafted a set of items comprising the entire spectrum of possible functional consequences of facial weakness based on a combination of data from the qualitative study and their clinical experience with FSHD.

Through an iterative process, items were selected and rephrased by the team to create a preliminary version of the questionnaire consisting of 33 items (Supplemental Table 1). For 27 of the items, the team deemed four response categories appropriate, and for six items three response categories were chosen, with higher scores indicating a higher symptom burden. Two FSHD patients checked the preliminary questionnaire for clarity.

The questionnaire was written in Dutch. To create the English version included in this manuscript, the Dutch version was translated to English by two researchers independently and then back translated to Dutch by two other researchers [15]. All translations were reviewed and discrepancies were resolved to obtain an English version semantic and conceptual equivalent to the Dutch version.

Data collection and ethical approval

The preliminary questionnaire was sent out in 2019 and 2020 to 155 participants of an observational cohort study at the Radboud University Medical Center, The Netherlands, that included Dutch-speaking genetically confirmed FSHD patients aged 18 years and older (FSHD-FOCUS study) [16]. All participants visited the Radboudumc for extensive clinical testing, including assessment of facial weakness. Questionnaires belonging to the study were sent out to the participants digitally. The response rate was 77%. To assess test–retest reliability patients were invited to complete the questionnaire twice, 2 weeks apart. Standardized written instructions were given to patients to complete the questionnaire.

This study was conducted according to the principles of the Declaration of Helsinki (version October 2013) and in accordance with the Medical Research Involving Human Subjects Act (WMO). The study protocol was approved by the regional medical ethics committee. All patients signed informed consent.

Rasch analysis

A thorough description of Rasch analysis for clinicians can be found elsewhere [17]. There were no missing data. As the likelihood-ratio test was significant, the use of the rating scale model that assumes equal thresholds across all items was excluded and the partial credit model was set as default [18]. An initial manual analysis was performed using Rasch unidimensional measurement methods software (RUMM2030) to assess the fit to the Rasch model of the preliminary questionnaire including checks for individual item and person fit, disordered thresholds, differential item functioning (DIF) for sex and age (three groups with roughly equal numbers of patients: 18–50 years; 51–65 years; ≥66 years), unidimensionality (independent t-test approach [19]), and response dependency [20]. Individual item misfit was defined as a fit residual value exceeding 2.5, significant chi-square probability value after Bonferroni’s adjustments or a combination of both.

Next, disordered thresholds were restored and the presence of DIF was checked manually. The resulting dataset was subjected to a semi-automated Rasch analysis by optimizing a criterion called the in-plus-out-of-questionnaire log likelihood (IPQQ-LL) using the autoRasch R package [21]. In brief, this analysis searches for the most informative set of items to measure differences in the trait of interest, and by doing so naturally caters for typical Rasch criteria such as item goodness of fit, unidimensionality, and (to some extent) response dependency. Finally, the results of the semi-automated analysis were manually checked and fit to the Rasch model optimized.

Convergent validity and test–retest reliability

For convergent validity purposes, other measures on facial weakness and/or function were included. The total scores of our newly developed Facial Function Scale (FFS) were correlated to the following outcome measures using the Pearson correlation coefficient.

The Communicative Participation Item Bank (CPIB) is a questionnaire that measures communicative participation [22]. The total score from 10 items ranges from 0 to 30, in which higher scores indicate better communicative interaction. The examiner-reported Facial Weakness Score (FWS) assesses the degree of facial weakness in a set of facial movements bilaterally [6]. The total score spans from 0 to 126, with higher scores indicating more severe weakness. A facial subscale of the FSHD evaluation score was used. It assesses facial weakness on a three point scale (0 = no weakness; 1 = mild weakness; 2 = severe weakness) [23].
Test–retest reliability was determined using a one-way random effects analysis of variance model for group comparison to calculate intraclass correlation coefficients (ICCs).

*p* Value cut-offs of <0.05 were considered statistically significant for correlations analyses. For Rasch analyses, Bonferroni correction for multiple testing was applied [24].

### Results

#### Study population

The preliminary questionnaire was completed by 119 FSHD patients (60 males) aged 23–82 years (mean 56.0 ± 14.4 years). The cohort encompassed the entire severity spectrum (FSHD evaluation score range 0–15, mean 7.8 ± 4.4 points).

#### Rasch analysis

The initial analysis showed that the preliminary questionnaire did not meet the Rasch model expectations (Table 1). Differential item functioning was checked for age and sex, and was not present. Disordered thresholds, most often occurring due to patients having difficulty discriminating between response options, were found for 17/33 items. For nearly all of these items, there was a relative infrequency of the most difficult response option (i.e., the highest score). Due to the high number of items with disordered thresholds, for uniformity purposes, thresholds for all items were collapsed from four to three, of from three to two, depending on the original range of response options. The exact collapsing of the response categories was based on the distribution of response options per category.

Next, the automated analysis was performed and the highest IPOQ-LL was obtained on an itemset containing 27 items (Supplemental Figure 1). In this itemset, one item was included that still had disordered thresholds (item 12 on the preliminary questionnaire: duration of meals), despite the collapsing from four to three response categories. This item did, however, show a good outfit, infit and predictability (α) value and because of its importance in maintaining a more optimal items’ and thresholds’ distribution it was kept in the final model.

In the 27-item dataset obtained by the automated analysis, there were multiple item pairs that showed high residual correlations of >0.3, indicating local dependency. The two item pairs with the highest residual correlations were items 18 (item numbers from preliminary questionnaire: “speaking unclear due to facial weakness” and 19 “speaking unclear when fatigued” (residual correlation 0.508) and items 7 “choking on liquids” and 8 “choking on foods” (residual correlation 0.457). For both pairs, the item showing the least clinical relevance, and the most over- or underdetermination on its item characteristic curve, was removed. Removing these items improved the unidimensionality of the scale (Table 1).

The final questionnaire consisted of 25 items and showed acceptable fit to the Rasch model (Table 1). One item pair with a high residual correlation (item 29 “ability to whistle” and item 30 “ability to blow up a balloon”, residual correlation 0.455) was kept in the model, in order to maintain a more optimal person-item threshold distribution. The item difficulty of the final scale ranged from −2.37 to 1.13 logits. Patient’s location ranged from −4.64 to 2.15 logits (Table 2).

#### Questionnaire results

The mean score on the final FFS was 30.0 ± 15.4, range 0–72 (centile metric score, see Table 3). Eight patients (7%) had the lowest possible score (floor effect), while none had the maximum score (ceiling effect). The item “blowing up a balloon” was the most challenging item and item “fatiguing of the face during eating” the least challenging item to fulfill (Figure 1). Among the items at the lower end of the scale (i.e., the items that patients found less challenging to fulfill) were many items on the topic of eating.

In the literature, problems with the duration of meals and fear of choking have been reported in FSHD patients. On these two items (duration of eating and fear of choking) respectively 24.3% and 20.2% of patients reported difficulty.

#### Convergent validity

The newly developed 25-item Rasch-built FFS demonstrated a strong correlation to the total CPIB score (*r* = 0.733, *p* < 0.001), and moderate correlations to the examiner-reported FWS and FSHD evaluation score facial weakness subscale (respectively *r* = −0.566 and *r* = 0.441, *p* < 0.001), indicating adequate convergent validity.

#### Reliability

Seventy-three patients completed the questionnaire a second time, two weeks after initial completion. The scale showed excellent test–retest reliability scores for both patients’ locations (ICC 0.978) and items’ locations (ICC 0.995).

### Discussion

This study provides a patient-reported outcome on facial function, designed specifically for FSHD. The development of this scale paves the way for more extensive research on facial weakness and function in this disease.

We applied a semi-automated approach to do a Rasch analysis. The substantial benefit of this approach is the capability to lessen the burden of the time-consuming and laborious tasks of the typical Rasch analysis. Accompanied with pre- and post-analysis and employing criterion called the IPOQ-LL, this approach automates most of the iterative parts in the typical Rasch analysis and aims for a set of items that best represent the patients’ ability. Having

---

Table 1. Results of Rasch analyses.

<table>
<thead>
<tr>
<th>Analysis</th>
<th>Item location (mean, SD)</th>
<th>Item fit residual (mean, SD)</th>
<th>Person location (mean, SD)</th>
<th>Person fit residual (mean, SD)</th>
<th>Person trait interaction df; p value</th>
<th>Item trait interaction df; p value</th>
<th>PSI</th>
<th>Unidimensionality t-test % (95%CI)</th>
<th>Floor effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial (33 items)</td>
<td>0.00 ± 0.99</td>
<td>−0.32 ± 1.23</td>
<td>−2.27 ± 1.29</td>
<td>−0.27 ± 0.91</td>
<td>66; 0.019</td>
<td>0.94</td>
<td>13.3% (93–17.3)</td>
<td>5%</td>
<td></td>
</tr>
<tr>
<td>After reordering thresholds (33 items)</td>
<td>0.00 ± 0.95</td>
<td>−0.34 ± 1.25</td>
<td>−1.86 ± 1.45</td>
<td>−0.29 ± 0.87</td>
<td>66; 0.037</td>
<td>0.94</td>
<td>10.6% (6.6–14.6)</td>
<td>5%</td>
<td></td>
</tr>
<tr>
<td>Semi-automated analysis (27 items)</td>
<td>0.00 ± 0.97</td>
<td>−0.31 ± 1.16</td>
<td>−1.82 ± 1.45</td>
<td>−0.25 ± 0.88</td>
<td>54; 0.166</td>
<td>0.92</td>
<td>8.0% (4.0–12.1)</td>
<td>6%</td>
<td></td>
</tr>
<tr>
<td>Final (25 items)</td>
<td>0.00 ± 0.99</td>
<td>−0.33 ± 1.16</td>
<td>−1.84 ± 1.45</td>
<td>−0.27 ± 0.87</td>
<td>50; 0.678</td>
<td>0.92</td>
<td>6.3% (2.3–10.4)</td>
<td>7%</td>
<td></td>
</tr>
</tbody>
</table>

PSI: person separation index; SD: standard deviation; CI: confidence interval.

*Proportion of participants achieving the lowest score possible.
Table 2. Final Rasch-built “Facial Function Scale”.

<table>
<thead>
<tr>
<th>Number</th>
<th>Question</th>
<th>Often (2)</th>
<th>Sometimes (1)</th>
<th>Never (0)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Do you suffer from dry eyes or the feeling of grains of sand in your eyes?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Do your eyelids get fatigued during the day?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Do you experience difficulty swallowing drinks?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Do you choke while eating?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Do you experience difficulty with the control of your saliva (does saliva run out of your mouth or do you choke on saliva)?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Do you spill food from your mouth?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Does your face get fatigued during eating?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Does eating take longer due to weakness of the muscles around your mouth?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Do you experience difficulty biting off foods?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Do you experience difficulty chewing foods?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Are you afraid of choking when eating and/or drinking?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Does your face get fatigued during speaking?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Is your speech more slurred due to facial weakness?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Does your face get fatigued during laughing?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Do others think you come across as angry or indifferent?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>Do others ask you if you are having a good time?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>Do you feel like you are misunderstood due to changed facial expressions?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>Are social situations less satisfactory to you due to facial weakness or changed facial expressions?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>Does facial weakness reduce your self-confidence?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>Are you able to wink?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>Are you able to whistle?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>Are you able to blow up a balloon?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>Are you able to drink through a straw?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>Are you able to smile?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>Are you able to swallow with your lips put together?</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The presented raw scores in this table should be summed to obtain a total score, that is then transformed into an interval score (Table 3, nomogram) for proper application. Permission to use this scale should be obtained through the authors.

been proved to cater to the typical criteria of Rasch analysis theoretically and empirically, this procedure has succeeded to obtain a clinimetrically sound set of items.

The final FFS is an interval scale from which the raw ordinal based sum-scores can be transformed into interval scores, rendering its results suited for parametrical statistical testing. The scale demonstrated a high internal consistency, indicating that the scale has a good discriminatory capacity between different levels of facial (dis)ability. The small floor effect and absence of a ceiling effect in this cohort, showed that the scale is capable of capturing nearly the entire spectrum of abilities. The proportion of 7% of patients reporting no difficulties at all (score of zero) is in line with the finding of very mild facial involvement, or even a facial sparing phenotype, in approximately 10% of all FSHD patients [6,25,26]. An earlier study found that patients with FSHD reported more difficulty with eating duration and fear of choking compared to healthy individuals, which is consistent with our finding that 24% and 20% of patients had an abnormal score on these two tasks, respectively [2].

The questionnaire was developed based on input from FSHD patients with the specific aim to capture their symptom burden due to facial weakness. This emphasis on clinical relevance renders the scale highly suitable to test symptomatic interventional strategies to improve facial function, such as, but not limited to, speech therapy, facial physical therapy, or cognitive behavioral therapy.

To determine the value of this questionnaire as an outcome measure in clinical trials on targeted therapies intended to halt progression of muscle weakness and wasting, additional research is required to further explore the relation between the degree of facial muscle weakness and symptom burden. The finding of moderate correlations between the patient-reported and physician-reported scales suggests that the relation between the severity of weakness of the facial muscles and the symptom burden is not straight-forward. Further studies on this topic should
preferably be longitudinal studies, as it is still both unknown how facial muscle weakness progresses over time, as well as how facial functioning changes with age and different disease stages.

This study has some limitations. The final scale included an item pair with high residual correlation. This impacts the unidimensionality of the scale, as was shown by improvement of its unidimensionality upon removal of other items with high residual correlations. However, keeping the concerning item pair with high residual correlation in, resulted in a better thresholds’ distribution, and acceptable unidimensionality. Additionally, although for a rare disease as FSHD this study included a sizeable

Figure 1. Distribution of facial function evaluation (ability location) as assessed with the Facial Function Scale and threshold map of the final 25-item scale.
cohort, ideally approximately 10 observations per category are obtained to provide a stable model, thus requiring a larger sample size [18].

The questionnaire was originally written in Dutch. Although the questionnaire was translated and back-translated according to the international standards [15], cross-cultural and -language validation was not performed, as the questionnaire was only completed by Dutch patients. Next steps would be to translate the questionnaire into additional languages, have native speakers complete the questionnaire, and then run DIF analyses to account for cultural and/or linguistic differences.

This study provides a linear-weighted, clinimetrically sound, patient-reported outcome measure on the functional disabilities relating to facial weakness in FSHD patients. The responsiveness of this scale is currently being assessed in a longitudinal study. Establishing its sensitivity to change and determining the ‘minimally clinically important difference’ will aid in optimizing the scales’ use in future studies, especially in clinical trials.

Acknowledgements

Several authors of this publication are members of the Radboudumc Center of Expertise for neuromuscular disorders (Radboud-NMD), Netherlands Neuromuscular Center (NL-NMD) and the European Reference Network for rare neuromuscular diseases (EURO-NMD).

Disclosure statement

The authors report no conflicts of interest relevant to the current research.

Funding

This project was funded by the Princes Beatrix Spierfonds, Grant Number W.OR18-07, and the FSHD Stichting, Grant Number WP29.

Data availability statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References