Validation of the World Health Organization Disability Assessment Schedule II (WHODAS-II) in Multiple Sclerosis

Giuseppe Magistrale1, Valerio Pisani1, Ornella Argento1, Chiara C Incerti1, Marco Bozzali2, Diego Cadavì3, Carlo Caltagirone1,4, Rossella Medori3, John DeLuca5,6, Ugo Nocentini1,4

1 Neurology and NeuroRehabilitation Unit, I.R.C.C.S. Neuromed, Pozzilli, Italy
2 Department of Systems Medicine, University of Rome "Tor Vergata", Rome, Italy
3 NeuroRehabilitation, Foundation IRCCS Santa Lucia, Rome, Italy
4 Neurology Clinical Development Group, Biogen Idec, Cambridge, MA, USA
5 Department of Neurology, University of New Jersey Medical School, New Brunswick, New Jersey, USA

Introduction:
• Multiple sclerosis is a chronic and disabling inflammatory disease that involves the CNS and one of the most widely spread neurological diseases affecting young adults (Compston et al., 2008).
• Due to its peculiar pathophysiology and the unpredictability of the disease course, MS limits patients’ autonomy and thus threatens their independence and self-respect.
• The framework of the International Classification of Functioning, Disability and Health (ICF) could represent a useful looking glass for the understanding of the impact of MS in the context of complex interactions between bodily, individual and societal perspectives (Grimby et al., 2001).
• The WHODAS-II was developed by the World Health Organisation (WHO) in order to assess patients’ self-perceived disability regardless of the diagnosis and according to the bio-psycho-social conceptual framework of the ICF.
• The self-administered version of the WHODAS-II consists of 36 items that assess functioning and disability during the 30 days prior to testing. The WHODAS-II covers six domains of functioning, i.e., mobility, self-care, getting along, life activities and social participation. Two other components have been proposed by gathering the items in the first ("activities") and last three subscales ("participation"). A shorter 32 items version (without work items) can also be administered.
• This study is aimed at the specific assessment of the psychometric properties of the WHODAS-II in MS with modern Rasch-model analyses.
• Rasch analysis is used for the assessment of clinical outcome measures as it has the property to assess several aspects of a scale, such as the functioning of the scoring method, pattern of dependency in responses as well as the dimensionality of the chosen scale (Tennant et al., 2004). Two assumptions postulated by the Rasch model are unidimensionality and local independence. The former assumes the property that a set of items measures a single latent trait and can thus generate a valid ordinal score, while the latter relates to the independence of each item in order to give an individual score on the latent variable.

Methods: one hundred and thirty-six patients were consecutively recruited at a single centre. A total of 110 patients (80.9%) had the relapsing–remitting form of MS, while 26 (19.1%) had progressive forms. Ninety-seven participants were female (71.3%) and 39 were male (28.6%). The mean age of the participants was 42.94 years (range 19–72; standard deviation (SD) ± 11.18) and mean education was 12.6 years (range 5–21; SD = 3.9). Mean Expanded Disability Status Scale (EDSS) score was 3.5 (range 1–5; SD = 1.6). All patients were administered the WHODAS-II.

Data analysis: reliability of the WHODAS-II total score and subscales was tested by internal consistency reliability, the computation of the Person Separation Index (PSI) from Rasch analysis. Item sets were assessed for threshold ordering of response categories, person and item fit, model fit, differential item functioning (DIF) and dimensionality. Item and person fit are presented as residuals and transformed to a z score: Values of approximately 0 and 1, respectively, for mean and SD indicate a good fit to the model. Individual items were indicated by fit residual values comprising between ±2.5 and chi-square probabilities above 0.05, adjusting with Bonferroni correction for the number of items. A breach to unidimensionality can be represented by the presence of correlation patterns in the items residuals after the extraction of the Rasch factor (the main scale). This could be due to redundancy in the item set or to the presence of multiple dimensions. Items showing response dependency can be combined into testlets, which are “super-items” whose score is equivalent to the summed score of single items. DIF is performed in order to assess whether gender and age groups within the same sample respond differently to individual items, despite equal levels of the latent characteristic being measured. A principal component analysis (PCA) of the residuals is performed, and the person estimates of the set of items with the highest positive and the set of the ones with the highest negative loadings on the first residual factor are contrasted through a series of independent t-tests for each individual. The percentage of tests outside the range of ±1.96 is then computed with the respective binomial confidence interval (CI). For a confirmation of strict unidimensionality, the binomial CI is expected to overlap the 5% for a nonsignificant test.

Results: When analyzing the response category structure, the threshold map revealed that the original five points scoring structure showed thresholded disorders for the vast majority of the items, overall misfit (Table 2, analysis 1) and poor fit to the model. As a result, we opted for a global rescoring of the item sets. Through this process, local dependency within each domain were combined into testlets, and all items in the life activities subscale were combined into one testlet either for the 36 and 32 items (work items removed) versions. This way we obtained a set of 34 items (including testlets) showing strict unid- mensionality (6.06% [2.6–11.5]) and no further local dependence. Both scales (with and without work items) showed good fit to the Rasch model and unidimensionality (Table 2, analysis 2 and 3) although PSI reliability decreased from 0.88 to 0.83. All items were found free from DIF and showed a nonsignificant p<0.0003 (after Bonferroni adjustment of the p value for 36 items to <0.0003), indicating individual item fit and absence of bias for gender and age. Analysis of single domains could resolve four subscales out of seven (Table 3: Getting around, life activities (household), life activities (work/school) and participation in society) and the activities and participation components. The “understanding and communicating” scale (66 valid cases), the “self-care” scale (32 valid cases) and the “getting along” scale (42 valid cases) could not be resolved because of high floor effect.

Conclusions:
• In order to achieve a good fit to the Rasch model, response categories had to be collapsed for each item of the scale.
• Another important phenomenon observed in these analyses was the presence of redundancy in items, which was resolved by combining items into testlets.
• Two working scales of the WHODAS-II (with and without work items) were obtained after these adjustments. Consequently, PSI fell from 0.88 to 0.83, indicating that the reliability index of the WHODAS-II is inflated by redundant items.
• Analysis of the subscales could resolve four working subscales out of seven. These subscales adequately fit the Rasch model with a PSI ≤ 0.70. The “activities” and “participation” components showed a satisfying fit. We could not resolve the three remaining subscales because of a high rate of extreme scores in the sample, indicating substantial floor effect for items in these specific subscales.
• Notably, no group bias (DIF) was observed in items in any of the analyses.
• Rasch analyses on specific and larger samples are required in order to confirm the reliability of unresolved subscales, and allow the establishment of a solid scoring context. The presence of substantial local dependency between several items suggests that the length of the WHODAS-II could be reduced.
• Despite these limitations, we can recommend the WHODAS-II total scale and the activities and participation components as reliable instruments for the assessment of patient-reported disability in MS.

References:

Table 1. Descriptive statistics for the WHODAS-II

<table>
<thead>
<tr>
<th>Outcome Measures</th>
<th>n</th>
<th>Means (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Understanding and communicating</td>
<td>130</td>
<td>8.2±1.13</td>
</tr>
<tr>
<td>Getting Around</td>
<td>130</td>
<td>31.8±29.9</td>
</tr>
<tr>
<td>Self-care</td>
<td>124</td>
<td>8.6±1.77</td>
</tr>
<tr>
<td>Getting along with people</td>
<td>123</td>
<td>6.0±1.06</td>
</tr>
<tr>
<td>Life activities (Household)</td>
<td>112</td>
<td>26.2±31.3</td>
</tr>
<tr>
<td>Life activities (School/Work)</td>
<td>86</td>
<td>14.3±2.5</td>
</tr>
<tr>
<td>Participation in society</td>
<td>132</td>
<td>21.6±20.5</td>
</tr>
<tr>
<td>Total score</td>
<td>97</td>
<td>18.2±16.1</td>
</tr>
</tbody>
</table>

*Higher is worse

Table 2. Fit statistics for original and adapted model of the total WHODAS-II subscale

<table>
<thead>
<tr>
<th>Analysis number</th>
<th>Action</th>
<th>Overall fit</th>
<th>Item fit residual (mean ± SD)</th>
<th>Person fit residual (mean ± SD)</th>
<th>PSI</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Overall</td>
<td>χ²=123.2, df=72, p&lt;0.001</td>
<td>0.18±1.05</td>
<td>0.21±1.10</td>
<td>0.86</td>
</tr>
<tr>
<td>2</td>
<td>Toilet adjustment</td>
<td>χ²=53.41, df=48, p=0.38</td>
<td>0.15±0.81</td>
<td>0.23±0.98</td>
<td>0.83</td>
</tr>
<tr>
<td>3</td>
<td>Work items removed</td>
<td>χ²=53.01, df=48, p=0.28</td>
<td>0.19±0.85</td>
<td>0.21±0.94</td>
<td>0.86</td>
</tr>
</tbody>
</table>

*percentage shown where the binomial CI overlaps 95%