Clinimetric evaluation of a new overall disability scale in immune mediated polynuropathies

I S J Merkies, P I M Schmitz, F G A van der Meché, J P A Samijn, P A van Doorn, for the Inflammatory Neuropathy Cause and Treatment (INCAT) group

Objective: To determine the validity, reliability, and responsiveness of a new overall disability sum score in immune mediated polynuropathies.

Methods: Three impairment measures (MRC sum score, sensory sum score, grip strength (Vigorimeter)) and three disability scales (an overall disability sum score (ODSS), Hughes' functional scale (f score), Rankin scale) were assessed in a cross sectional group of 113 clinically stable patients (83 with Guillain-Barré syndrome, 22 with chronic inflammatory demyelinating polyneuropathy (CIDP), eight with gamopathy related polyneuropathy). The ODSS was also used serially in 20 patients with recently diagnosed Guillain-Barré syndrome (n = 7) or CIDP (n = 13) and changing clinical conditions. Multiple regression studies were performed to compare the impact of impairment disturbances (independent variables) on the various disability scales (dependent variables).

Results: Moderate to good construct validity (stable group: Spearman's rank test (absolute values), r = 0.41–0.79; longitudinal group: multiple correlation coefficient, R = 0.69–0.89; p < 0.006 for all associations) and reliability (intraclass correlation coefficient, R = 0.90–0.95; p < 0.0001) were demonstrated for the ODSS. Its SRM values were high (> 0.8), indicating good responsiveness. Impairment measures accounted for a higher variance proportion of the ODSS compared with the f score and Rankin (R = 0.64 v. 0.56 and 0.45, respectively).

Conclusions: All clinimetric requirements were met by the overall (arm and leg) disability sum score in immune mediated polynuropathies. Its use is therefore suggested in evaluating immune mediated polynuropathies.

Clinical assessment in patients with Guillain-Barré syndrome, chronic inflammatory demyelinating polynuropathy (CIDP), and polynuropathy associated with a monoclonal gamopathy of undetermined significance (MGUSP) has traditionally been focused on impairment and less often on disability. The most commonly used impairment scales are the motor scales based on the Medical Research Council (MRC) grading system, and different sensory outcome measures including various modes of sensation. Disability, on the other hand, has primarily been evaluated using the (modified) Hughes' functional grading scale (f score) and the (modified) Rankin scale. The clinimetric properties of the f score have been demonstrated in patients with Guillain-Barré syndrome, but no formal clinimetric evaluation of the Rankin scale has been undertaken in patients with polynuropathies. Despite their simplicity and obvious face value, clinical use of the Hughes' and Rankin scales is somewhat limited, as their emphasis is strongly directed towards mobility and they provide little information about arm function. Arm dysfunction in immune mediated polynuropathies may contribute to disability; so disability measures ought to address the arms as well as the legs in these conditions. Nevertheless, only a few studies have evaluated arm disabilities in these disorders.

In a recent study, a disability scale was applied in these conditions that described functional disturbances of both the legs and the arms. This overall disability sum score (ODSS)—forming part of the Guy's neurological disability scale—showed a significant correlation with the "INCAT" sensory sum score, thereby proving its validity (table 1). However, further clinimetric evaluation of this comprehensive disability scale is required before it is recommended for general use in patients with immune mediated polynuropathies.

Prompted by these observations, we investigated the construct validity, reliability, and responsiveness of the ODSS. Validity is defined as the relation between the concept to be measured and the scale used to assess that concept. It usually relies on expert judgments (content and face validity), by demonstrating a high correlation between the scale and a gold standard (criterion validity) or, in the absence of a gold standard, by examining the degree of association between a scale and other widely used measures (construct validity).

Reliability addresses the internal consistency in multi-item scales and the ability of a scale to demonstrate reproducibility of the scores by the same (intraobserver) or a different examiner (interobserver), or by the same patient (test–retest reliability) in the case of self rating scales.

Responsiveness is defined as the ability of a scale to detect meaningful clinical changes over time when evaluating the benefits of a medical intervention. A statistic and heuristic approach in examining responsiveness has been described. Statistical responsiveness captures the ability of an instrument to measure any change, irrespective of its relevance; this form of responsiveness can be assessed within a group of patients receiving the same treatment or between groups of patients being treated in different ways. Heuristic techniques are based on comparing changes as assessed by an outcome measure with an external indicator—for example, patients judging their clinical condition as being worse, stable, or improved.

Abbreviations: CIDP, chronic inflammatory demyelinating polyneuropathy; GBS, Guillain-Barré syndrome; INCAT, inflammatory neuropathy cause and treatment group; MGUSP, monoclonal gamopathy of undetermined significance; MRC, Medical Research Council; ODSS, overall disability sum score; SRM, standardised response mean.
Finally, the ODSS was compared with the f score and Rankin scale to determine which of these disability measures captures most adequately the impairments leading to disability.

**METHODS**

**Patients**

A cross sectional group of 113 patients (83 with Guillain–Barré syndrome, 22 with CIDP, and eight with MGUSP) in stable clinical condition were recruited from the Rotterdam immune mediated polyneuropathy databank and the Dutch GBS study group (stable group). Patients with Guillain–Barré syndrome, CIDP or MGUSP were recruited, as it was argued that these disorders represent parts of a continuum with respect to their pattern of neuromuscular dysfunction. These selected patients had residual symptoms or signs of their illness, representing a broad range of disability. Nine CIDP selected patients had residual symptoms or signs of their illness, with intravenous immunoglobulin. On this treatment their clinical condition has been stable for more than six months, with intravenous immunoglobulin. On this treatment several patients met the international criteria for their illness.

All Guillain–Barré and CIDP patients met the international criteria for their illness. The diagnosis MGUSP was established after excluding all other known causes of gammopathy and polyneuropathy.

**Assessment tools/scales**

The **MRC sum score** is a summation of the MRC grades (range, 0–5) given in full numbers of the following muscle pairs: upper arm abductors, elbow flexors, wrist extensors, hip flexors, knee extensors, and foot dorsal flexors. The MRC sum score ranges from 0 (“total paralysis”) to 60 (“normal strength”). Good validity and interobserver reliability for this scale have been demonstrated in patients with Guillain–Barré syndrome.

The **“INCAT” sensory sum score** was recently introduced and extensively evaluated clinimetrically in patients with immune mediated polyneuropathies. In brief, this sensory scale comprises pin prick and vibration sense plus a two point discrimination value in the arms and legs, and ranges from 0 (“normal sensation”) to 20 (“most severe sensory deficit”).

The **Vigorimeter** (Martin Co, Tuttlingen, Germany) is a hand held dynamometer for measuring grip strength. Good clinimetric properties have been demonstrated in patients with immune mediated polyneuropathies.

The **overall disability sum score** (ODSS) is a composed of a recently published arm and leg disability scale with a total score ranging from 0 (“no signs of disability”) to 12 (“most severe disability score”) (Table 1). The ODSS comprises a good functional description of the arms and legs in a checklist form suitable for interviewing patients. Daily arm activities such as dressing the upper part of the body, doing and undoing

---

**Table 1** The overall disability sum score (ODSS)

<table>
<thead>
<tr>
<th>Arm disability scale – function checklist</th>
<th>Not affected</th>
<th>Affected but not prevented</th>
<th>Prevented</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dressing upper part of body (excluding buttons/zips)</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>Washing and brushing hair</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>Turning a key in a lock</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>Using knife and fork (/spoon—applicable if the patient never uses knife and fork)</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>Doing/undoing buttons and zips</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
</tbody>
</table>

**Arm grade**

0 = Normal
1 = Minor symptoms or signs in one or both arms but not affecting any of the functions listed
2 = Moderate symptoms or signs in one or both arms affecting but not preventing any of the functions listed
3 = Severe symptoms or signs in one or both arms preventing at least one but not all functions listed
4 = Severe symptoms or signs in both arms preventing all functions listed but some purposeful movements still possible
5 = Severe symptoms and signs in both arms preventing all purposeful movements

**Leg disability scale – function checklist**

<table>
<thead>
<tr>
<th>No</th>
<th>Yes</th>
<th>Not applicable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Do you have any problem with your walking?</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>Do you use a walking aid?</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>How do you usually get around for about 10 metres?</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>Without aid</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>With one stick or crutch or holding to someone’s arm</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>With two sticks or crutches or one stick or crutch and holding to someone’s arm</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>With a wheelchair</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>If you use a wheelchair, can you stand and walk a few steps with help?</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>If you are restricted to bed most of the time, are you able to make some purposeful movements?</td>
<td>O</td>
<td>O</td>
</tr>
</tbody>
</table>

**Leg grade**

0 = Walking is not affected
1 = Walking is affected but does not look abnormal
2 = Walks independently but gait looks abnormal
3 = Usually uses unilateral support to walk 10 metres (25 feet) (stick, single crutch, one arm)
4 = Usually uses bilateral support to walk 10 metres (25 feet) (sticks, crutches, two arms)
5 = Usually uses wheelchair to travel 10 metres (25 feet)
6 = Restricted to wheelchair, unable to stand and walk few steps with help but able to make some purposeful leg movements
7 = Restricted to wheelchair or bed most of the day, preventing all purposeful movements of the legs (eg, unable to reposition legs in bed)

Overall disability sum score = arm disability scale (range 0–5) + leg disability scale (range 0–7); overall range: 0 (no signs of disability) to 12 (maximum disability). For the arm disability scale: allocate one arm grade only by completing the function checklist. Indicate whether each function is “affected,” “affected but not prevented,” or “prevented.” For the leg disability scale: Allocate one leg grade only by completing the function questions.
buttons and zips, washing and brushing hair, using a knife and fork, and turning a key in a lock are scored as being “not affected,” “affected but not prevented,” or “prevented.” Subsequently, these results are translated into an arm grade (score range, 0 (normal arm abilities) to 5 (severe symptoms and signs in both arms preventing all purposeful movements)). The leg scale highlights problems of walking, taking into account the use of a device. The results are also translated into a leg grade (score range, 0 (walking is not affected) to 7 (restricted to wheelchair or bed most of the day, preventing all purposeful movements of the legs) (table 1). The selected arm and leg disability scales are subsets of a more comprehensive Guy’s neurological disability scale. Good clinimetric requirements have recently been demonstrated for all components of the Guy’s scale in patients with multiple sclerosis.

The modified Hughes’ functional grading scale (f score) assesses the functional ability of the patient, with a strong emphasis on mobility. The f score of the patients included in this study ranged from 0 (no symptoms or signs) to 5 (requiring artificial ventilation for at least part of the day).

The Rankin scale has primarily been used in patients with stroke. The grades of this scale range from 0 (no symptoms at all) to 5 (severe disability, bedridden, incontinent, and requiring constant nursing care and attention). No formal clinimetric evaluation of this scale has been undertaken in patients with immune mediated polyneuropathies.

Test procedures
General aspects
All participants gave their informed consent and were studied in a quiet and comfortably warm room at our outpatient clinic. The assessments were done in a random order. For the assessment of general strength, joint and limb position was standardised, and the point at which counterforce was applied was defined before the start of the study and used when examining each muscle group. Sensory function was examined in triplicate under prescribed standardised conditions, with the patient lying in a supine position. For the assessment of grip strength with the Vigorimeter, all patients were examined under standardised conditions.

The study took place between December 1998 and January 2000 and was performed on behalf of INCAT, a group of senior European neurologists with a special interest in neuroimmunological illnesses.

Validity and reliability
We undertook several different forms of validation of the ODSS. “Validity by assumption” (that is, face and content validity) involved a review of all aspects of the ODSS by the INCAT expert panel. The correlation between the ODSS and other outcome measures was investigated by correlation and regression studies (construct validity). To assess the reliability and validity of the ODSS in the stable group of 113 patients, two neurologists and six experienced residents in neurology formed 28 different couples. Preceding the study, all investigators received instructions in assessing the outcome measures. Twenty seven (“variable”) couples investigated a total of 68 patients (two to three patients for each couple). The remaining 45 stable patients were examined by the “experienced” couple (1M + 1S). The latter couple was formed to examine the effect of training (and thus a possible increase in responsiveness) resulting from frequent use of the ODSS.

The patients were examined on two different occasions in our outpatient clinic. During the first visit the two members of an appointed pair performed their scores independently and consecutively (usually within two hours) (intrabrowser measures). Within two to four weeks, the patient returned for a second visit and only one investigator of the earlier assigned pair examined the patient again (intraobserver values), without having access to previous results. The assessment sequence at entry and the examination at the second visit were equally distributed among the members of an assigned couple. Eventually, each member of a couple examined approximately the same number of patients. With the exception of the f score, all scales were assessed at each visit in all patients. For the validity and the regression model studies, only the recruited scales’ values at one examination were used.

Responsiveness
Twenty consecutive patients were examined longitudinally by the same clinician (1M). The ODSS was assessed at study entry and at the weeks 2, 4, 8, 12, 16, 21, 26, 32, 40, and 52 of follow up, with additional clinical investigations if necessary. At each visit, the patients were requested to judge whether their clinical condition had deteriorated (grade 1), remained stable (grade 2), or improved (grade 3) when compared with the last visit (“clinical judgment scores”). At study entry, the patients compared their clinical condition against their physical status in the two weeks before the start of the study.

Statistics
Validity and reliability
In the cross sectional stable group, the correlation between ODSS and the other outcome measures was analysed using Spearman’s rank correlation test. Random effects linear regression analyses were also performed between the ODSS and the other scales in the longitudinal group, taking into account the associations caused by the longitudinal structure. The linear regression analyses were done using the program “xreg” in STATA 6.0 (Stata Corporation, 1997. Stata Statistical Software: release 6.0. College Station, Texas, USA), which is based on a cross sectional time series regression model as described by Dwyer and Feinleib. The scores obtained are presented as multiple correlation coefficients. The inter-rater and intrarater reliability of the ODSS was quantified by estimation of the intraclass correlation coefficient using one way random effects analysis of variance model for the two investigator groups (“experienced” and “variable”).

Responsiveness
Responsiveness was examined in both statistical and heuristic forms. Statistical responsiveness was investigated by calculating the standardised response mean (SRM) scores for the ODSS at various arbitrarily chosen occasions during follow up (weeks 12, 26, 40, and 52). The SRM is equal to the mean change in score divided by the standard deviation of the change in score:

\[ SRM = \frac{\mu_1 - \mu_0}{SD(\mu_1 - \mu_0)} \]

where \(\mu_1\) = mean ODSS value of the longitudinally examined group at week = 1, and \(\mu_0\) = mean ODSS value at week = 0 (entry). A value between 0.5 and 0.8 is considered moderate responsiveness, and 0.8 or greater as high responsiveness.

For each patient, the differences between every pair of consecutive ODSS values were calculated (= ODSS value at visit i minus ODSS value at visit (i-1) = ODSS changes). These differences were associated with the corresponding clinical judgment scores using “xreg,” thus investigating the heuristic form of responsiveness.

All analyses were performed using Stata 6.0 for Windows 97. A probability (p) value of < 0.05 was considered statistically significant.

Comparative study
In the stable group, univariate and multivariate linear regression analyses were done to determine which disability measure (ODSS, f score, or Rankin scale—dependent variables) had the strongest association with a group of
impairment measures (MRC sum score, INCAT sensory sum score, grip strength by the Vigorimeter—individual variables). If necessary, a transformation of the dependent variable (for example, by logarithmic conversion) was done to obtain a normal distribution. We concentrated on univariate regression studies, aiming for the best fit between dependent and independent variables. This was achieved through systematic evaluation of the graphs constructed from the linear regression studies, which included a restriction cubic spline function on the independent variable. Subsequently, multivariate linear regression analyses were performed, with a backward stepwise eliminating strategy to construct the final models. The strength of association between the dependent variable and explanatory variables was presented as $R^2$—the fraction of variance explained by the independent variable from a regression model. Only the results that included the right hand grip strength values are presented, as these findings were similar to those incorporating the left hand values.

### RESULTS

#### General aspects

All eight examiners concluded that the ODSS was easily applicable and required less than two minutes for completion. The stable group of patients (54 women and 59 men) had a median duration of symptoms before the start of the study of 5.1 years. Seven of these patients were bed bound and 14 required assistance or a device to walk short distances. The remaining 92 patients could walk independently. The corresponding median values and ranges for all scales in these patients are presented in table 2.

#### Validity and interobserver/intraobserver reliability

The correlation studies comparing the ODSS with the other scales and the reliability values for the ODSS in the stable group of patients are presented in table 3. Significant validity and good reliability were demonstrated for the ODSS by the “experienced” and “variable” couples of investigators. In the longitudinal group, significant associations were also obtained between the ODSS and other measures: ODSS vs MRC sum score: random effects analysis of variance, $R^2 = 0.89$; vs sensory sum score: $R^2 = 0.74$; vs grip strength: $R^2 = 0.72$ (right hand) and $R = 0.69$ (left hand); vs f score: $R = 0.86$; vs Rankin: $R = 0.88$ ($p < 0.0001$ for all associations).

#### Responsiveness

Eight women and 12 men were examined longitudinally. At study entry, four were bed bound, one requiring artificial ventilation, and nine were unable to walk independently. All patients experienced general loss of strength, sensory disturbances, and deficit in their daily functional activities. We completed 201 visits during a follow up period of 40 to 58 weeks (median 52). Nineteen patients completed a one year follow up. With the exception of one patient with Guillain–Barré syndrome who only experienced mild symptoms, all patients had received initial treatment with intravenous immunoglobulin (0.4 g/kg body weight/day for five consecutive days). All but one patient with CIDP showed good functional improvement on intravenous immunoglobulin during follow up. The non-responder received a treatment...
course of oral prednisone and also showed improvement. The patients with Guillain–Barré syndrome did not show any deterioration during follow up and improved gradually over time.

After initial improvement, all 12 CIDP patients who were responsive to intravenous immunoglobulin needed continued interval treatment (0.4 g/kg/d for one to two days at intervals of 12 to 14 weeks) to maintain their earlier improvement. Ultimately, all patients showed a decrease in their degree of impairment and an improvement in functional ability during follow up. Improvement in the longitudinal group resulted in a general reduction in the ODSS values, indicating improvement, with median values of 3, 2.5, 2, and 2 at weeks 12, 26, 40, and 52, respectively, compared with the median entry value of 5 (Wilcoxon signed-rank test: $p \leq 0.0008$ for all comparisons).

Good SRM scores were obtained for the ODSS in these patients (1.2, 1.5, 1.4, and 1.4 at weeks 12, 26, 40, and 52, respectively). The patients graded their clinical condition 53 times as “deteriorating,” 38 times as “stable,” and 110 times as “improving.” These values were significantly associated with the ODSS changes obtained serially in these patients (random effects linear regression analyses: $R = 0.66; p = 0.008$).

### Comparative study

The MRC sum score was the strongest predictor of disability compared with grip strength (Vigorimeter) and the sensory sum score. Univariate linear regression analyses were as follows:

- on ODSS: MRC sum score, $R^2 = 0.45$; grip strength, $R^2 = 0.40$; sensory sum score, $R^2 = 0.21$;
- on f score: MRC sum score, $R^2 = 0.43$; grip strength, $R^2 = 0.34$; sensory sum score, $R^2 = 0.16$;
- on the Rankin scale: MRC sum score, $R^2 = 0.34$; grip strength, $R^2 = 0.24$; sensory sum score, $R^2 = 0.14$.

Overall, a higher proportion of variance in disability, explained by impairment measures, was captured by the ODSS than by the f score and the Rankin scale (fig 1).

### DISCUSSION

In this study, we show that clinimetric requirements—such as being easily applicable, valid, reliable, and responsive to clinical changes over time—were met by the ODSS in patients with immune mediated sensory-motor polyneuropathies. This scale highlights problems not only with walking but also with daily arm activities. Its concept is therefore more comprehensive than the widely used Hughes’ functional grading scale (f score) and the Rankin scale, which are mainly directed towards mobility and do not provide information on arm function.

In addition, general loss of strength and sensory deficits leading to disability were better monitored by the ODSS than by the f score or the Rankin scale. It therefore appears that the ODSS is to be preferred for evaluating disability in these disorders.

The impairment variables recorded by the ODSS explained two thirds of the disability. This finding implies that other forms of impairment should be considered in future studies as contributors to disability. For example, variables such as general fatigue and depression have been suggested as important events in patients with immune mediated polyneuropathies which may lead to functional deficit. Lennon and associates reported six reasons for persistent disability in patients with Guillain–Barré syndrome. These were: muscle weakness, sensory dysfunction, contractures, fatigue, other medical conditions, and psychological factors such as anxiety, depression, and lack of motivation.

In our study, weakness—as measured by the MRC sum score—was the most important independent explanatory factor in the patients’ level of disability. This finding is consistent with a recent paper addressing outcome in various forms of polyneuropathy.

With respect to the aims of the current study, some further methodological issues should be addressed. First, the SRM scores obtained for the ODSS only showed within-group responsiveness. It is not clear whether substantial discriminative responsiveness scores would be obtained for the ODSS when evaluating various groups of patients—for example, in a trial setting comparing a placebo versus a treated group.

Second, univariate and multivariate linear regression analyses of the f score and Rankin scale were performed, despite the fact that these outcome measures are ordinal constructs. An ordinal logit estimation model—as described by the program “ologit” in Stata 6.0—was also applied on these ordinal variables, but as the description of these analyses was rather complex and the results quite similar to the linear regression studies, we decided to present the data in the current form for clarity. Third, despite having good validity, reliability, and responsiveness, future studies should determine whether the ODSS has greater responsiveness scores than other validated and reliable disability measures.

This is important, as the power of a study is directly linked to the responsiveness of the applied measure (greater responsiveness corresponding to greater power).

### Conclusions

We have demonstrated the simplicity, validity, reliability, and responsiveness of the overall (arm plus leg) disability sum score in patients with immune mediated polyneuropathies. Impairment leading to disability was better monitored by the ODSS than by the other disability measures tested. Thus we
suggested that the ODSS should be used to monitor disability in immune mediated polyneuropathies.

**ACKNOWLEDGEMENT**

This study was part of the Biomed project (No BMH4-CT96 0324).

**Authors’ affiliations**

I S J Merkies, F G A van der Meché, J P A Samijn, P A van Doorn, Department of Neurology, University Hospital Rotterdam/Erasmus University, Dr Molewaterplein, Rotterdam, Netherlands

P I M Schmitz, Department of Statistics, Daniel den Hoed Cancer Centre, University Hospital Rotterdam/Erasmus University

**REFERENCES**


