J Rehab Med 2001; 33: 137–144

MEASURES OF ORDER CONSISTENCY BETWEEN PAIRED ORDINAL DATA: APPLICATION TO THE FUNCTIONAL INDEPENDENCE MEASURE AND SUNNAAS INDEX OF ADL

Lisbeth Claesson[1,2] and Elisabeth Svensson[3]

From the [1]Department of Neurology, Institute of Clinical Neuroscience, Göteborg University, [2]Department of Occupational Therapy and Physiotherapy, Division of Health and Caring Sciences, Göteborg University and [3]Mathematical Statistics, Göteborg University and Statistics Department, Örebro University, Sweden

The aim was to evaluate the inter-item consistency between corresponding items in the Functional Independent Measure (FIM) and Sunnaas index of ADL (SI) by means of a statistical approach that takes account of the non-metric properties of ordinal data. The ADL dependence in 204 stroke patients was assessed by interview 3 months after the onset of stroke. The result showed that the inter-item consistency between the FIM and SI were high for many items, but operational differences between some FIM and SI items were also identified. The statistical evaluation demonstrated that some of the ordered categorical levels of the seven-point item scales in the FIM could be aggregated into four levels without loss of information.

Key words: ADL, stroke, non-parametric statistics, ordinal data, inter-item consistency.

J Rehab Med 2001; 33: 137–144

Correspondence address: Lisbeth Claesson, Institute of Clinical Neuroscience, Department of Neurology, Sahlgrenska University Hospital, SE-413 45 Göteborg, Sweden. E-mail: licl@neuro.gu.se

(Accepted November 6, 2000)

INTRODUCTION

Various types of measurement instrument for assessment of dependence in daily life activities (ADL) are commonly used in rehabilitation medicine. Typical sub-variables of functional ADL assessments are eating, grooming, dressing, bathing, toileting, transfers (from bed to chair etc.) and mobility, in personal activities and housekeeping, shopping and transportation, in instrumental activities (1–4). The Functional Independence Measure (FIM) (5) and Sunnaas Index of ADL (SI) (6) are instruments used to measure ADL.

The FIM is a discipline-free generic measurement instrument used as a uniform measure of disability in terms of independence/dependence. The FIM is recommended for use in medical wards for setting goals for patients during rehabilitation (5). The FIM was originally developed to assess the burden of care by observation (5), but has also been used in telephone interviews (7, 8). The SI is a Norwegian measurement instrument for the assessment of dependence in ADL. The SI was originally constructed as an instrument for measuring personal and instrumental ADL function in stroke patients (6).

The level of dependence on help in performing each activity is recorded on a seven-point scale in the FIM (5) and on a four-point scale in the SI (6). The data from these rating scales have an ordered structure but no other mathematical properties, which means that statistical methods designed for metric, quantitative data cannot be applied to the ADL assessments (9–13).

The aim of the present study was to apply a statistical method (9, 10) that takes account of the ordered categorical properties of data for a comprehensive evaluation of the consistency between the assessments of dependence in daily life activities by the FIM and the SI items. The seven-point scales for the FIM items were also calibrated in relation to the corresponding four-point item scales in the SI in two ways: by comparing the operational criteria of the categorical levels and by minimizing the systematic disagreement between the two assessments. The cut-off points between the FIM levels that correspond to the levels in the SI were identified, and the level of order consistency between the SI and the two approaches to condensing the FIM scales were compared.

METHODS

Subjects

This study is a part of the Göteborg 70+ Stroke Study (14). In this study, 249 consecutive acute stroke patients, at least 70 years of age were randomized to care in a stroke unit or care in general medical wards. The groups were comparable at entry with regard to gender, living conditions, mean age and medical history, with the exception of angina pectoris, which was more common in the stroke unit group. The groups were also comparable with regard to neurological score, side of predominant neurological deficit, degree of pareses and speech disorder at entry. Of the 249 patients, 229 (92%) had brain infarction, 11 had intra-cerebral haemorrhage, five had transient ischaemic attack and four had other diagnoses. The predominant neurological deficit was on the right side in 107 (43%) patients and on the left side in 122 (49%); 105 (42%) patients had a slight paresis, 44 (18%) moderate and 61 (25%) severe paresis. Speech disorders were found in 120 (48%) patients (14).

The present study includes 204 patients evaluated 3 months after the onset of stroke. Of these patients, 126 (62%) were women (mean age, 80.5 years; range, 68.9–96.5 years) and 78 men (mean age, 78.3 years; range, 70.3–92.2 years).

Instruments and procedures

The original FIM (5) consists of 13 physical (or motor) items assessing *feeding*; *grooming*; *bathing*; *dressing-upper body*; *dressing–lower body*; *toileting*; *bladder management*; *bowel management*; *transfer–bed, chair, wheelchair*; *transfer–toilet*; *transfer–tub, shower*; *walk/wheelchair*; and *stairs*; and five social-cognitive items assessing *comprehension*; *expression*; *social interaction*; *problem solving*; and *memory*. The ordered

 ISSN 1650–1977

Table I. *Comparative items in the Functional Independence Measure (FIM) and the Sunnaas index of ADL (SI)*

Functional Independence Measure	Sunnaas index of ADL
Feeding	Eating
Grooming	Grooming
Bathing	Bath/shower
Dressing-upper body	Dressing–undressing
Dressing-lower body	Dressing–undressing
Toileting	Toilet management
Bladder management	Continence
Bowel management	Continence
Transfer–bed, chair, wheelchair	Transfer
Transfer–toilet	Transfer
Transfer–tub, shower	Transfer
Walk/wheelchair	Indoor mobility
Comprehension	Communication
Expression	Communication
FIM items not including	SI items not including
Stairs	Cooking
Social interaction	Housework
Problem solving	Outdoor mobility
Memory	

categories of each item scale are 'total dependence', 1; 'maximal contact assistance or the subject expends 25–49% of the effort', 2; 'moderate contact assistance or the subject expends 50–74% of the effort', 3; 'minimal contact assistance or the subject expends ≥75% of the effort', 4; 'supervision or setup', 5; 'modified independence', 6; and 'complete independence', 7. The validity (15–18) and reliability (19–25) of the FIM are well documented. A Swedish translation of the manual of September 1991 (26) was used.

The original SI (6) consists of 12 items: *eating*, *continence*, *indoor mobility*, *toilet-management*, *transfer*, *dressing–undressing*, *grooming*, *cooking*, *bath/shower*, *housework*, *outdoor mobility* and *communication*. The ordered categories of each item scale are; 'total dependence', 0; 'needs some help from other persons or can manage alone, but does not', 1; 'can manage alone and does under special conditions', 2; and 'complete independence', 3. The SI has both the wording 'can' and 'does' included in the definition of the categories, but in this study the raters evaluated how the person was performing (does). Although the SI has been used for many years in Scandinavia, there have been few studies concerning the measurement. A Finnish study reported evidence for the concurrent validity of the SI in measuring functional recovery after stroke (27). A Swedish translation of the manual of December 1992 was used.

The formulation of the operational criteria in the two instruments was the basis for the selection of the items for evaluation of the consistency order (Table I). The single items *dressing-undressing, continence* and *communication* in the SI occur as two items in the FIM. The item *transfer* is a single item in the SI but occurs as three items in the FIM (Table I, Appendix).

Assessments of ADL ratings were performed independently by two registered occupational therapists. The patients were randomly assigned to the occupational therapists, who evaluated them using a semi-structured interview based on the different items. Most of the ADL assessments [155] were carried out in the patient's home, while 29 were conducted in a geriatric clinic, 19 in a nursing home and one in a hospital. In the interview, the patients described the performance of each activity, and the interviewer scored the answer concerning dependence. When there was any doubt about a patient's ability, they were asked to perform the activity and the raters tried to clarify whether the patient performed the activity every day. The interviewer asked relatives or medical staff, if accessible.

The inter-observer agreement in ADL assessments has been evaluated in 20 patients. The median percentage agreement (PA) in the FIM items was 70% (range, 45–95%), and in the SI items 85% (range, 65–90%).

Statistical method

The statistical approach (9) for the evaluation of the order consistency between the assessments of dependence according to the FIM and the SI is described step by step as applied to the item *eating/feeding*. The joint distribution of the paired recordings from the four-point SI and the seven-point FIM is presented in a 4 × 7 contingency table (Fig. 1a).

The marginal distribution of the 204 individuals on the four levels of the SI was [9, 18, 22, 155], which means that the majority of individuals, 76% [155 of 204], were classified as 'completely independent' 3 in eating (Fig. 1a) According to the FIM assessments, 69% [140 of 204], were classified as 'completely independent' 7 in eating, as the marginal distribution of FIM was [8, 4, 1, 1, 39, 11, 140] (Fig. 1a).

The observed frequency distribution of paired assessments on the FIM and the SI is compared with the pattern of total agreement in ordering of all individuals. This pattern of complete order consistency is called the rank-transformable pattern of agreement (RTPA), as the rank ordering of all individuals is independent of the two scales (10, 28). The RTPA is defined by the two sets of marginal distributions and is constructed by pairing off the two sets of marginal distributions against each other. Figure 1b shows the RTPA defined by the marginals of Fig. 1a. The non-zero cell frequencies are easily calculated. The eight observations in the FIM 1 are paired with eight of the nine observations in the SI 0. The remaining observation in the SI 0 should be paired with one of the four in the FIM 2. The remaining three in the FIM 2 are combined with three of the 18 in the SI 1, etc.

The observed pattern of inter-scale assessments in Fig. 1a is slightly dispersed from the RTPA, which means that the ordering of at least some of the individuals differs concerning the level of dependence in eating between the FIM and the SI assessments. Dispersed observations from the RTPA appear in cells opposite the main diagonal; that is, in the upper-left and lower-right regions relative to a non-zero cell. The two pairs of observations in the cell SI 0, FIM 2 in Fig. 1a disagree in ordering with the single pair in the cell SI 1 and FIM 1, as the ADL level in the SI is higher when the FIM is lower.

The number of pairs with a disagreement in ordering out of all possible different pairs of individuals defines the measure of disorder, D (9, 29). The number of ADL levels in the SI is $m_1 = 4$, and the ordered categorical SI levels are indexed i, where i stands for 0, 1, 2, 3. Corresponding notations for FIM are $m_2 = 7$, where j represents the FIM levels 1,…,7. The ij:th cell frequency is denoted x_{ij}. For example, two individuals scored SI 0 and FIM 2, which is noted i = 0, j = 2, then $x_{02} = 2$. x_{ij}^{ul} denotes the upper-left region frequency relative to the ij:th cell, $i = 1,\ldots, m_1$ and $j = 1,\ldots, m_2$. The upper-left region relative to the nine observations scored SI 3 and FIM 5 contain the non-zero cell frequencies $x_{17} = 1$, $x_{27} = 1$ and $x_{26} = 3$. These five observations in the upper-left region relative to $x_{35} = 9$ are expressed as $x_{35}^{ul} = 5$.

The measure of disorder is defined by $D = \dfrac{2\sum_{i=1}^{m_1}\sum_{j=1}^{m_2} x_{ij}x_{ij}^{ul}}{n(n-1) - t}$, where $t = \sum_{i=1}^{m_1}\sum_{j=1}^{m_2} x_{ij}(x_{ij} - 1)$, which is the correction factor for the observations tied to the same cell. For example, in Fig. 1a there are 138 identical pairs of assessments of SI 3 and FIM 7. As there was more than one observation in most cells, which means that there are tied pairs, the number of possible paired comparisons of observations is reduced by the correction factor, $t = \sum_{i=1}^{m_1}\sum_{j=1}^{m_2} x_{ij}(x_{ij} - 1) = 7 \times 6 + 2 \times 1 + 2 \times 1 + 12 \times 11 + 18 \times 17 + 3 \times 2 + 9 \times 8 + 8 \times 7 + 138 \times 137 = 19524$.

For the frequency distribution of Fig. 1a, the nominator of D is $2\sum_{i=0}^{3}\sum_{j=1}^{7} x_{ij}x_{ij}^{ul} = 2(7 \times 0 + 2 \times 0 + 1 \times 2 + 2 \times 0 + 1 \times 0 + 1 \times 0 + 12 \times 0 + 1 \times 0 + 18 \times 1 + 3 \times 1 + 1 \times 0 + 9 \times 5 + 8 \times 2 + 138 \times 0) = 2(2 + 18 + 3 + 45 + 16) = 168$. Note that the 16 observations scored SI 1 and FIM 3, 4, 5 have no observations in the upper-left regions. The measure of disorder is $D = \dfrac{168}{204 \times 203 - 19524} = \dfrac{168}{21888} = 0{\cdot}00768$

The interpretation of this measure of disarray is that 168 of the

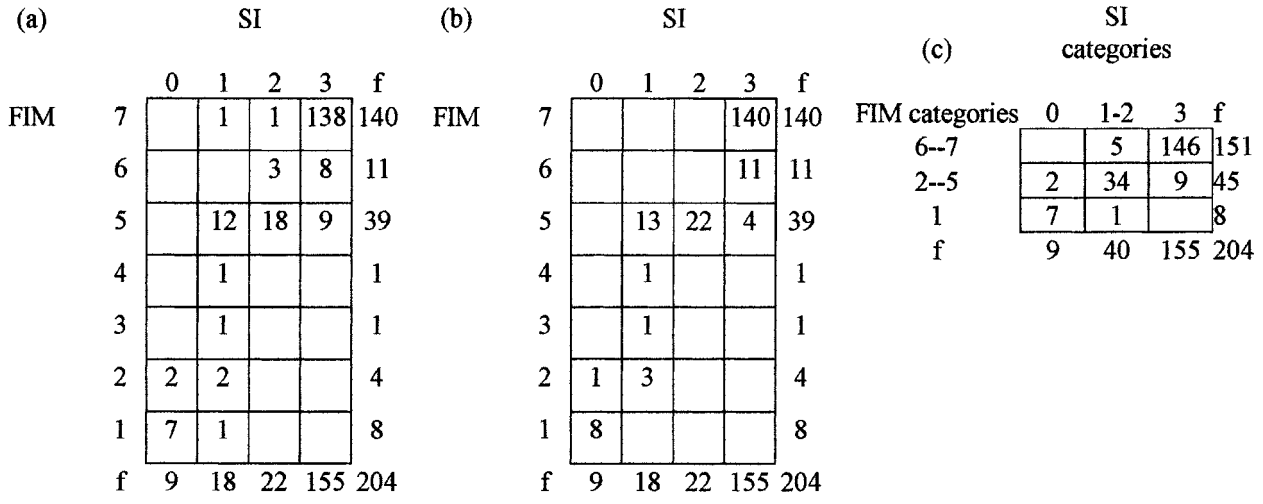

Fig. 1. (a) Joint distribution of paired assessments of the level of dependence in eating measured by FIM and SI. (b) Rank-transformable pattern of agreement defined by the marginal distributions of Fig. 1a. (c) Joint distribution of observations of the calibrated scales for the assessment of the level of dependence in eating. f = frequencies.

possible 21888 different pairs of observations are disordered, which is a small proportion (0.8%).

Svensson (9) has defined a *coefficient of monotonic agreement, MA*, as the difference between the proportions of ordered and disordered pairs of observations. MA is calculated simply by $MA = 1 - 2D$. The maximum value, MA = 1, is obtained when there is a total ordering of all observations, i.e. when D = 0, which is the case for the RTPA. The MA of the pattern in Fig. 1a is $1 - 2 \times 0.00768 = 0.985$, which is a high level of order consistency between the SI and the FIM in the assessment of the level of dependence in *eating/feeding*.

Inter-scale calibration

By means of the two sets of marginal distributions and the RTPA it is possible to reduce the seven-point FIM scale in a way that minimizes the systematic disagreement between the reduced FIM scale and the corresponding SI item scales. Identical marginal distributions mean lack of systematic disagreement or bias.

In the two assessments of dependence in *eating/feeding* (Fig. 1a), marginal homogeneity could almost be achieved by a reduction to a 3×3 scale comparison, by grouping the FIM levels into 1, 2–5 and 6–7 and the SI levels into 0, 1–2 and 3 (Fig. 1c). The level of agreement in ordering, MA, is 0.995 and the percentage agreement (PA) in categorical levels between the calibrated scales is 92%.

RESULTS

The results of the 204 paired assessments using the FIM and SI are presented at the item level.

Table II shows that the measures of MA range between 0.841 and 0.986, where *continence/bowel management* and *communication/comprehension* (MA, 0.867) represent the items with an order consistency less than 0.90. The items with a high level of order consistency between the FIM and SI assessments were *dressing-undressing/dressing-lower body* (MA, 0.986, Fig. 2), *eating/feeding* (MA, 0.985, Fig. 1a), *grooming/grooming* and *toilet-management/toileting* (MA, 0.982).

Table II also shows the MA and PA for the FIM and SI assessments, when the FIM levels are reduced to a four-point scale according to the operational definitions of the item scales. In two items, *toilet-management/toileting* and *communication/comprehension*, the level of order consistency decreased, which is a sign of inconsistency between the operational definitions of the scales. The median PA was 77% (range, 48–87%). A high level of order consistency but a low PA, as in *transfer/transfer tub, shower* (MA, 0.973; PA, 48%) indicates a systematic disagreement between the two four-point sets of data. The pairs of observations are close to the RTPA (high MA), but the RTPA deviates from the main diagonal (low PA), which means inter-scale bias. The presence of bias, irrespective of a high MA, means that the items are not interchangeable.

The results of the unbiased calibration of FIM and SI are shown in Table III a,b,c. The cut-offs in FIM are defined by the steps in RTPA and by the fact that lack of bias requires equal marginal distribution. The table shows that an unbiased calibration required a reduction of the item levels to three-point scales for the FIM and SI, except for *continence/bladder management* and *indoor mobility/walk or wheelchair*. For six of the items, the SI levels 1 and 2 were grouped, and in six other items, the SI levels 2 and 3 were grouped (Table III a,b). The increased levels of order consistency for almost all items indicate that the ADL assessments were not performed entirely according to the operational definitions. The percentage agreement increased for all items, and the median PA was 92% (range, 67–98%).

The measures of MA and PA show that both calibration approaches failed to find a correspondence between the items *continence* and *bowel management*, while there was a high level of order consistency between the items *continence* and *bladder management*. The unbiased grouping of the ordinal categorical levels of the items concerning *dressing–undressing* and *dressing-upper body* and *dressing-lower body* indicate that the two FIM items have different operational definitions in relation to the SI item. For the two FIM items of *communication*, the highest level of order consistency with the SI item concerned

Table II. *Cut-off points between the categories of corresponding operational definitions in the seven-point FIM and the four-point SI assessed 3 months after stroke (n = 204). The coefficent of monotonic agreement (MA) are given, both according to observed and calibrated scales. Percentage agreement (PA) is given for calibrated scales*

		Scale levels in SI					
Compared items in SI/FIM	Observed scales MA	Total dependence (0)	Needs some help from another person (1)	Does under special conditions (2)	Independence (3)	Calibrated scales MA	PA (%)
Eating/Feeding	0.985	1, 2	3, 4	5, 6	7	0.995	83
Grooming/Grooming	0.982	1, 2	3, 4, 5	6	7	0.982	87
Bath-shower/Bathing	0.956	1, 2	3, 4, 5	6	7	0.960	67
Dressing–undressing/Dressing–upper body	0.965	1, 2	3, 4, 5	6	7	0.967	83
Dressing–undressing/Dressing–lower body	0.986	1, 2	3, 4, 5	6	7	0.987	84
Toilet-management/Toileting	0.982	1, 2	3, 4, 5	6	7	0.969	78
Continence/Bladder management	0.964	1, 2	3, 4, 5	6	7	0.973	85
Continence/Bowel management	0.841	1, 2	3, 4, 5	6	7	0.843	65
Transfer/Transfer–bed, chair, wheelchair	0.967	1, 2	3, 4, 5	6	7	0.967	78
Transfer/Transfer–toilet	0.965	1, 2	3, 4, 5	6	7	0.971	76
Transfer/Transfer–tub, shower	0.929	1, 2	3, 4, 5	6	7	0.973	48
Indoor mobility/Walk or wheelchair	0.913	1, 2	3, 4, 5	6	7	0.921	56
Communication/Comprehensen	0.867	1, 2	3, 4, 5	6	7	0.829	64
Communication/Expression	0.951	1, 2	3, 4, 5	6	7	0.956	69

expression (MA, 0.985; PA, 93%), provided that the FIM levels 2–4 and 5–7 were grouped (Table II, III b). Table II shows that there was an inconsistency in the operational definitions between the SI and FIM item of *communication/comprehension* (MA, 0.829; PA, 64%).

The MA values of unity and very high PA (97% and 98%) indicate that the FIM items *transfer-bed, chair, wheelchair* and *transfer-toilet* corresponded well with the *transfer* item of the SI, provided grouping of the FIM levels 1–2, 3–5 and 6–7 and the SI levels 0, 1 and 2–3 (Table III b). The operational comparison between the *transfer/transfer-tub, shower* (MA = 0.973, PA = 48%) showed a high level of order consistency and a high level of systematic disagreement between the item assessments. Grouping of the FIM levels 1, 2–4 and 5–7 and the SI levels 0, 1–2, 3 increased the PA (80%), but decreased the MA (0.935), which is a sign of lack of consistency between the FIM and SI for these variables.

DISCUSSION

The statistical approach used in the present study takes account of the non-metric, ordered structure of data from scale assessments. Hence, the result of the analyses are interpretable and valid without any distributional restrictions of the data. The ability of the statistical approach to allow a comprehensive evaluation of two different scales of the same variable with the regard to operational similarities and dissimilarities has been demonstrated with regard to the interchangeability between the FIM and SI items. A high level of order of consistency would imply that the individuals being assessed have comparable levels of ADL dependence as measured by the two scales.

This study showed that the SI/FIM items *eating/feeding, grooming/grooming* and *dressing-undressing/dressing–lower body* were operational comparable. The relatively low level of order consistency between the SI and FIM items *indoor mobility* and *walk/wheelchair* and between *communication* and *expression, comprehension* could be explained by differences in the operational definitions. The SI item *indoor mobility* concern mobility at home and at work, while the FIM item *walk/wheelchair* is defined as walking indoors for a distance of 50 metres. The item *communication* in the SI includes use of a telephone, which is not included in the FIM items *expression* and *comprehensiveness*. Evaluation of the assessments indicates that the occupational therapists have taken account of the

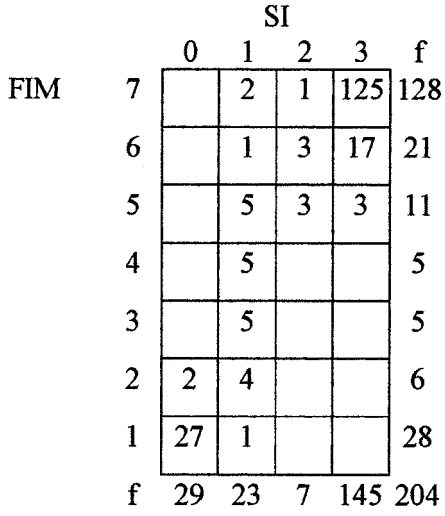

Fig. 2. Joint distribution of paired assessments of the level of dependence in dressing-undressing/dressing-lower body measured by FIM and SI. f = frequencies.

Table III. *a, b, c Unbiased calibration made by cut-off points by marginal homogeneity between the categories defined by FIM and SI assessed 3 months after stroke (n = 204). The coefficent of monotonic agreement (MA) are given, both according to observed and calibrated scales. Percentage agreement (PA) is given for calibrated scales. Tables a, b and c present different grouping of the ordinal levels in the SI*

(a) The ordinal levels in SI are the following: 0 = total dependence, 1 = needs some help from other person and 2 = does under special conditions, and the 3 = independence

		Scale levels in the SI				
Compared items in SI/FIM	Observed scales MA	Total dependence (0)	Needs some help from another person/Does under special conditions (1)/(2)	Independence (3)	Calibrated scales MA	PA (%)
Eating/Feeding	0.985	1	2, 3, 4, 5	6, 7	0.995	92
Grooming/Grooming	0.982	1	2, 3, 4, 5	6, 7	0.993	93
Dressing–undressing/Dressing–upper body	0.965	1, 2	3, 4, 5	6, 7	0.996	94
Dressing–undressing/Dressing–lower body	0.986	1	2, 3, 4, 5	6, 7	0.995	94
Continence/Bowel management	0.841	1, 2, 3, 4	5, 6	7	0.861	67
Transfer/Transfer–tub, shower	0.929	1	2, 3, 4	5, 6, 7	0.935	80

(b) The ordinal levels in the SI are the following: 0 = total dependence, 1 = needs some help from other person, and the 2 = does under special conditions and 3 = independence

		Scale levels in SI				
Compared items in SI/FIM	Observed scales MA	Total dependence (0)	Needs some help from another person (1)	Does under special conditions/Independence (2)/(3)	Calibrated scales MA	PA (%)
Bath–shower/Bathing	0.956	1, 2	3, 4, 5	6, 7	0.994	91
Toilet-management/Toileting	0.982	1	2, 3, 4, 5	6, 7	0.999	98
Transfer/Transfer–bed, chair, wheelchair	0.967	1, 2	3, 4, 5	6, 7	1.0	98
Transfer/Transfer–toilet	0.965	1, 2	3, 4, 5	6, 7	0.998	97
Communication/Comprehension	0.867	1	2, 3, 4	5, 6, 7	0.969	89
Communication/Expression	0.951	1	2, 3, 4	5, 6, 7	0.985	93

(c) The ordinal levels in the SI are the original four-point

		Scale levels in SI					
Compared items in SI/FIM	Observed scales MA	Total dependence	Needs some help from another person	Does under special conditions	Independence	Calibrated scales MA	PA (%)
Continence/Bladder management	0.964	1	2, 3, 4	5, 6	7	0.975	85
Indoor mobility/Walk or wheelchair	0.913	1, 2	3	4, 5	6, 7	0.920	73

different operational definitions of the scales in their assessments.

Comparison of the inter-item consistency showed a difference between the calibration of the FIM according to the operational criteria of categorical levels and the unbiased calibration to the SI assessments. This results showed slightly lower MA and PA values for the operational calibration. This might be explained by the dissimilarities in the operational criteria but also by the observers' experience and interpretation of the manuals. This study indicated that it is important to have a critical attitude towards ADL instruments as the instruments serves as communication tools between different health care professions.

Comparison of the operational criteria in the FIM and SI revealed some differences in definitions. The levels of independence of 6 and 7 in the FIM and of 2 and 3 in the SI correspondedin

most cases, as did the levels of dependence of 1 to 5 in the FIM and of 0 to 1 in the SI. Level 5, 'supervision', in the FIM had no directly corresponding level in the SI, but 'supervision' is often found in SI level 1. The definition in FIM level 6 includes modified independence with 'use of assistive devices to perform an activity', doing it out of concern for safety, and no more than a reasonable amount of time needed to perform the activity. The SI includes these three components in its definition of level 2 but, in contrast to the FIM, does not consistently include the factors of time consumption and concern for safety in its definitions of all items. In most cases, one of these components is represented in the criteria in the operational definition. This may be one of the reasons why FIM 6 and SI 2 do not always correspond. The FIM has a uniform construction in the seven-point rating scale, with a similar definition of the steps in the ordered categorical scale and where the level of dependence is assessed as percentage of independence (5). While this can make it easier for the rater to distinguish between the levels, using the FIM instrument demands training in the rating system to achieve good reliability (24).

The FIM is discipline-free, which means that it is a measure usable by any trained person regardless of discipline. When using the instrument, the team members use a uniform language to observe the actual need for care, and can set goals in the rehabilitation programe. To be able to assess the activities as precisely as possible, one has to observe the patient over a longer period of time that was possible in this study where assessments were made by interview. The occupational therapists were experienced in assessing ADL. However they were not specially trained in assessing FIM, although they had had instruction from an FIM trainer. The interviews were conducted in the patients' normal environment which was an advantage in assessing ADL as the questions could be individualized according to each patient's environment.

The comparisons between the instruments made by means of unbiased calibration identified three different levels of categories in most of the items. The pattern of agreement was different for the items. The results showed dissimilarity in the evaluation of modified dependence and modified independence and illustrated the difficulties in delimitation between the categories in the middle of the scales.

ACKNOWLEDGEMENTS

This study is a part of the Göteborg 70+ Stroke Study led by Professor Christian Blomstrand and Professor Björn Fagerberg. The study is made possible through grants from the John and Brit Wennerströms Foundation for Neurological Research, the Foundation for Neurological Research and the Local Association of Neurological Disabled (NHR Göteborg). We would also like to express our gratitude to occupational therapist Gunilla Gosman-Hedström for sharing the work of collecting the data and to Agneta Hallén for her valuable help with the data administration.

REFERENCES

1. Bruett TL, Overs RP. A critical review of 12 ADL scales. Phys Ther 1969; 49: 857–862.
2. Donaldsson SW, Wagner CC, Gresham GE. A unified ADL evaluation form. Arch Phys Med Rehabil 1973; 54: 175–185.
3. Eakin P. Assessments of activities of daily living: a critical review. Br J Occup Ther 1989; 52: 11–15.
4. Unswort CA. The concept of function. Br J Occup Ther 1993; 56: 287–292.
5. Hamilton BB, Granger CV, Sherwin FS, Zielezny M, Tashman JS. A uniform national data system for medical rehabilitation. In: Fuhrer M, ed. Rehabilitation outcomes: analysis and measurement. Baltimore: Paul H Brooks, 1987: 137–147.
6. Vardeberg K, Kolsrud M, Laberg T. The Sunnaas index of ADL. World Federation of Occupational Therapy Bulletin, 1991: 30–35.
7. Chang W, Slaughter S, Cartwright D, Chan C. Evaluating the FONE FIM: part 1. Construct validity. J Outcome Meas 1997; 1: 192–218.
8. Chang W, Chan C, Slaughter S, Cartwright D. Evaluating the FONE FIM: part II. Concurrent validity and influencing factors. J Outcome Meas 1997; 1: 259–285.
9. Svensson E. Comparison of assessment in using continuous and discrete ordinal rating scales. Biometrical J 2000; 42: 417–434.
10. Svensson E. Analysis of systematic and random differences between paired ordinal categorical data. Stockholm: Almqvist & Wiksell International, 1993.
11. Merbitz C, Morris J, Grip JC. Ordinal scales and foundations of misinference. Arch Phys Med Rehabil 1989; 70: 308–312.
12. Altman DG. Practical statistics for medical research. London: Chapman and Hall, 1991.
13. Tudor G, Koch GG. Review of nonparametric methods for the analysis of crossover studies. Stat Methods Med Res 1994; 3: 345–381.
14. Fagerberg B, Claesson L, Gosman-Hedström G, Blomstrand C. Effect of acute stroke unit care integrated in a care continuum vs conventional treatment—a randomized one-year study of elderly patients. The Göteborg 70+ Stroke Study. Stroke 2000; 31: 2578–2584.
15. Granger CV, Cotter AC, Hamilton BB, Fielder RC, Hens MM. Functional assessment scales: a study of persons with multiple sclerosis. Arch Phys Med Rehabil 1990; 71: 870–875.
16. Hall KM, Hamilton BB, Gordon WA, Zasler ND. Characteristics and comparisons of functional assessments indices: Disability Rating Scale, Functional Independence Measure, and Functional Assessment Measure. J Head Trauma Rehabil 1993; 8: 60–74.
17. Kidd D, Stewart G, Baldry J, Johnson J, Rossiter D, Petruckevitch A, Thompson AJ. The Functional Independence Measure: a comperative validity and reliability study. Disabil Rehabil 1995; 17: 10–14.
18. Pollak N, Rheault W, Stoecker JL. Reliability and validity of the FIM for persons aged 80 years and above from a multilevel continuing care retirement community. Arch Phys Med Rehabil 1996; 77: 1056–1061.
19. Brousseau L, Wolfson C. The inter-rater reliability and construct validity of the Functional Independence Measure for multiple sclerosis subjects. Clin Rehabil 1994; 8: 107–115.
20. Chau N, Daler S, Andre JM, Patris A. Inter-rater agreement of two Functional Independence Measure (FIM) and a subjective uniform continuous scale. Disabil Rehabil 1994; 16: 63–71.
21. Fricke J, Unsworth C, Diane W. Reliability of the Functional Independence Measure with occupational therapists. Aust Occup Ther J 1992; 40: 7–15.
22. Fricke J, Unsworth C. Inter-rater reliability of the original and modified Barthel index, and a comparison with the Functional Independence Measure. Aust Occup Ther J 1996; 43: 22–29.
23. Hamilton BB, Laughlin JA, Fiedler RC, Granger CV. Interrater reliability of the 7-level Functional Independence Measure (FIM). Scand J Rehab Med 1994; 26: 115–119.
24. Ottenbacher KJ, Mann WC, Granger CV, Tomita M, Hurren D, Charvat B. Inter-rater agreement and stability of functional assessment in the community-based elderly. Arch Phys Med Rehabil 1994; 75: 1297–1301.
25. Ottenbacher KJ, Hsu Y, Granger CV, Fiedler RC. The reliability of Functional Independence Measure: a quantitative review. Arch Phys Med Rehabil 1996; 77: 1226–1232.
26. Grimby G, Gudjonsson G, Thylander G. Functional Independence Measure, Swedish version. 1991.

27. Korpelainen JT, Niilekselä E, Myllylä VV. The Sunnaas Index of activities of daily living: responsiveness and concurrent validity in stroke. Scand J Occup Ther 1997; 4: 31–36.
28. Svensson E. Application of rank-invariant method to evaluate reliability of ordered categorical assessments. J Epidem Biostat 1998; 3: 403–409.
29. Gosman-Hedström G, Svensson E. Parallel reliability of the Functional Independence Measure and the Barthel ADL index. Disabil Rehabil 2000; 22: 702–715.

Appendix. Operational criteria in selected items in the Functional Independence Measure (FIM) and Sunnaas Index of ADL (SI). A comparison is made between these items of operational criteria in FIM and SI. The same individual items in SI are found in the comparison in two or three times.

Operational criteria in Functional Independence Measure		Operational criteria in Sunnaas Index of ADL	
Feeding	Use of suitable utensils to bring food to mouth, chewing and swallowing, once meal has been appropriately prepared	**Eating**	Cutting food, eating, chewing and swallowing
Grooming	Oral care, hair grooming, washing hands and face, shaving or applying makeup	**Grooming**	Wash oneself (including administrating face cloths, towels and taps), brush teeth, comb hair, shave and check buttocks for pressure-sores. Adequate hygienic standard
Bathing	Bathing the entire body from the neck down (tub, shower, or bed bath)	**Bath/shower**	Take a bath/shower and using the taps, drying afterwards, undressing a dressing
Dressing–upper body	Dressing above the waist as well as putting on and removing prosthesis or orthosis when applicable	**Dressing/undressing**	Dressing and undressing everyday clothes, (including bra, socks, shoes and outer garment) and an acceptable standard
Dressing–lower body	Dressing from the waist down as well as putting on and removing prosthesis or orthosis when applicable	**Dressing/undressing**	Dressing and undressing everyday clothes, (including bra, socks, shoes and outer garment) and an acceptable standard
Toileting	Maintaining perineal hygiene and adjusting clothing before and after toileting	**Toilet-management**	Toilet-transfer, cleaning oneself and dressing when visiting the toilet. Hygienic standard satisfactory
Bladder management	Complete and intentional control of the bladder and use of equipment or agents necessary for bladder control. Two variables, frequency of incontinence and level of assistance required for bladder management. Score both part I and part II. Then record the lower score	**Continence**	Control over urine and faeces
Bowel management	Complete and intentional control of bowel movements and use of equipment or agents necessary for bowel control. Two variables, frequency of incontinence and level of assistance required for bowel management. Score both part I and part II. Then record the lower score	**Continence**	Control over urine and faeces
Transfer–bed, chair, wheelchair	Management of all aspects of transferring to bed from bed chair, or wheelchair, or coming to a standing position, if walking is the typical mode of locomotion	**Transfer**	Transfer from bed to chair, chair to chair, wheelchair to toilet. In and out of bed. To the toilet or shower–chair/commode
Transfer–toilet	Getting on and off toilet	**Transfer**	Transfer from bed to chair, chair to chair, wheelchair to toilet. In and out of bed. To the toilet or shower–chair/commode
Transfer–tub, shower	Getting into and out of tub or shower	**Transfer**	Transfer from bed to chair, chair to chair, wheelchair to toilet. In and out of bed. To the toilet or shower–chair/commode
Walk/wheelchair	Walking once in a standing position, or using a wheelchair once in a seated position, indoors	**Indoor mobility**	Being mobile at home and at work
Comprehension	Clear comprehension either auditory or visual communication	**Communication**	Ability to convey and receive information, use telephone and gate-phone. Ability to call for help
Expression	Clear expression of either verbal or non verbal language	**Communication**	Ability to convey and receive information, use telephone and gate-phone. Ability to call for help