O merskih značilnostih dveh ocenjevalnih lestvic za oceno prizadetosti pri pacientih po ishemični možganski kapi

On metric characteristics of two disability measures applied to acute ischaemic stroke patients

asist. Gaj Vidmar, dipl. psih.\(^{(1)}\), Jelka Janša, MSc, del. terap.\(^{(2)}\)

\(^{(1)}\) University of Ljubljana, Department of Psychology, Ljubljana, Aškerčeva 2, Slovenia; gaj.vidmar@ff.uni-lj.si
\(^{(2)}\) Medical Centre, Department of Neurology, Ljubljana, Zaloška 2, Slovenia

Izvleček

Sistematično ocenjevanje pacientovega stanja se v okviru rehabilitacije pacientov po možganski kapi že dolgo uporablja. Funkcionalne posledice se ocenjuje predvsem preko stopnje fizične neodvisnosti, pri čemer je kriterij neodvisnosti zmožnost izvajanja opravil, potrebnih za vsakdanje življenje. Tako se je razvil koncept dnevnih aktivnosti (DA).

Raziskava je primerjala dve tovrstni ocenjevalni lestvic: Barthelov indeks (BI) in Razširjeni Barthelov indeks (RBI). BI je standarden instrument za ocenjevanje DA v okviru raziskav s področja rehabilitacije, RBI pa se od njega razlikuje po razširjenem točkovanju fizične prizadetosti in dodanem kognitivnem delu. V raziskavi je sodelovalo 33 pacientov s prebolelim prvim cerebrovaskularnim inzultom ishemične etiologije, oskrbovanih na nevrološkem oddelku, njihovo stanje pa je bilo obravnavano 1, 3 in 6 tednov po kapi.

Prispevek obravnava štiri glavna vprašanja: občutljivost RBI v primerjavi z BI ("učinek stropa"), zmožnost obeh lestvic pokazati rehabilitacijski trend, zanesljivost obeh lestvic ter kriterijalne veljavnosti RBI (povezanost z motorni lestvicino in spominskim testom). Raziskava je na vsa vprašanja razen prvega odgovorila pozitivno, poudarek prispevka pa je na uporabi različnih metod analize podatkov: inferenčne statistike in vizualizacije podatkov, bivariatnih in multivariatnih statističnih metod ter parametričnih in neparametričnih testov.

Abstract

Systematic assessment of patient performance has a long tradition and it is being intensely developed within stroke rehabilitation. Functional outcome has been assessed primarily in terms of degree of physical independence, the criteria of independence being performance-oriented and conceptualised in the ADL (Activities of Daily Living) approach.

Two ADL-scales were compared in the study: the Barthel Index (BI) and the Extended Barthel Index (EBI). BI has become the standard ADL measure for rehabilitation research, while EBI is a modification of the BI with extended scoring of physical disability and addition of a second, cognitive part. Thirty-three patients with first ischaemic stroke, all having received early rehabilitation service within the neurology department, were assessed at week 1, week 3 and week 6 post stroke.

Four principal issues were addressed: sensitivity of EBI to clinically important changes compared to BI (the question of ceiling effect); ability of both scales to show the rehabilitation trend; reliability of both scales; and criterion validity of EBI (relation to a motor scale and a memory test). While all the questions were answered positively bar the first one, emphasis was also put on combining diverse data-analysis methodologies: inferential statistics and data visualisation, bivariate and multivariate methods and parametric as well as nonparametric tests.
1. Introduction

Stroke is the third leading cause of death and a major cause of hospital admission and long-term disability in most industrialised populations (Bonita, 1992). Occupational therapists, trying to prevent disability and helping people develop independent function through purposeful activity, should therefore be able to choose the most appropriate activities according to the needs and abilities of every individual stroke patient (Thompson & Morgan, 1990).

The outcome of patients after stroke has been assessed primarily in terms of the degree of physical independence, whereby the criteria have been performance-oriented and conceptualised in the Activities of Daily Living (ADL) approach. Since the ADL concept was introduced more than half a century ago, a number of well-validated scales have been developed, including the Barthel Index, the Katz Index, the Northwick Park ADL Index, the Functional Observation Test, and the Edm ands ADL. ADL can be divided into two categories: Basic ADL comprise the patient's capacity for basic survival in managing basic hygiene, self-care and mobility (Murdock, 1992), while the term Instrumental ADL refers to more demanding tasks and involves interaction with the physical, social and cultural environment (e.g. using telephone or going shopping). ADL measures are predominantly based either on the dependence or the functional model of disability. The former assesses the amount of help needed by patients to complete certain basic tasks of self-care (Fletcher & Bulpitt, 1985), while the later concentrates on how the patient performs on certain tasks. In this study\(^1\) two scales, based on the dependence model, were of primary interest, namely the Barthel Index (BI) and the Extended Barthel Index (EBI).

The BI consists of 10 activities and it is an ordinal scale with minimum value of 0 and maximum value of 20. The patient's ability to perform self-care is evaluated by observation of actual behaviour. The BI has been used in more than 200 scientific studies in a wide range of research setting and has been recommended as a standard measure of disability. Its construct validity has also been well established. However, it shares a common problem of ADL scales: the maximum and minimum score do not represent patient's best and worst possible functional condition, respectively. Sensitivity to change at the least disabled and at the most disabled end of the scale is small and this is usually referred to as the ceiling effect (Ebrahim, 1990) and the floor effect (Shah, Vanclay & Cooper, 1990), respectively.

An answer to this problem is the concept of the extended BI, which is the focus of the study. The EBI was originally constructed for use with multiple sclerosis patients within the European multiple sclerosis rehabilitation group (now called MARCH). It is used routinely at the Neurologische Krankenhaus in Munich, Germany, for stroke, Parkinson's disease and multiple sclerosis patients, and it is also used in some other German neurology clinics, as well as in Switzerland and Belgium. The EBI consists of 16 items, 10 on physical disability (the same as those of BI, only scored 0 to 4 instead of 0 to 3) and 6 on cognitive functioning. Interrater reliability of the EBI has been established (Prosiegel, Bottger & Schenk, 1996), as well as its concurrent validity to the Functional Independence Measure (FIM, an American BI based scale, somewhat more demanding for use). Reliability was measured in terms of Kappa

\(^1\) The paper is based on a study, supported by a TEMPUS PHARE Scholarship and a British Council grant, both awarded to the second author, and the corresponding project submitted by the second author as part of MSc course in Occupational Therapy at the University of East London. The data were collected at the Department of Neurology of the Medical Centre in Ljubljana.
statistics for individual items and the cognitive part turned out to be less reliable than the physical one.

2. Methods

Subjects were 33 acute ischaemic stroke patients, admitted to the Neurology Department during the period October 1995 - May 1996, receiving there the medical treatment and the usual comprehensive inpatient rehabilitation services.

There were four instruments used: in addition to the BI and the EBI also the Fugel-Meyer Motor Impairment Scale (FM) and the Rivermead Behavioral Memory Test (RBMT). The BI can be thought of as a "gold" standard for assessing disability of stroke patients, while the EBI was translated into Slovene for the purpose of the study. The FM (maximum possible score 226) was used for addressing the concurrent validity issue, since previous studies identified early post stroke motor impairment as a predictor of long-term functional disability (Chae et al., 1995). The RBMT, yielding either a standardised score (0-24) or a profile score (0-12), consists of eleven subtests and as a valid measure of memory performance translated into Slovene, it was an appropriate reference point for the EBI memory item.

The procedure was the following: each of the patients was assessed on the same day at week 1, 3 and 6 post stroke. Patients were always scored on the BI and the EBI (reversing the order on the next occasion to minimise order effects), as well as on the FM, while the RBMT was performed only at week 6. Patients were followed up if they were transferred or discharged.

The study was approved by the Ethical Committee of the Medical Centre in Ljubljana and each patient confirmed the willingness to participate in the study in writing.

3. Results and Discussion

The distributions of the scales applied at week 1, 3 and 6 are summarised by the box-whiskers plots in Figure 1 (EBI-P and EBI-C stand for the physical and the cognitive part of EBI, respectively). A common scale is obtained by dividing the scores by the respective maximum attainable values (BI:20; EBI:64; EBI-P:40; EBI-C:24). The rehabilitation trend of BI and EBI is stressed by the two dashed lines, connecting the relative medians at week 1, 3 and 6.

The principal criticism of the BI is that a patient can achieve the maximum score without being able to function in the environment, which is called the ceiling effect. EBI represents an attempt to tackle this problem and one of the aims of our study was to test EBI's performance in this respect. However, it can not be said that the results indicate a ceiling effect of either BI or EBI. Nevertheless, the ceiling effect of EBI might exist when assessing ischaemic stroke patients, only that too few patients achieved the maximum functional independence in the first six weeks post stroke (thus scoring a maximum or close to that) to allow for a statistical demonstration/test. On the other hand, EBI is clearly manifesting less spread at week 3 and 6 and a higher relative average at week 3, which is therapeutically important and may speak for choosing EBI as a measure that enables the rehabilitation team to make relevant decisions and prognoses.
As far as the general rehabilitation trend is concerned, i.e., the decrease in patients' dependence over time, it is evident from Figure 1 that both scales convey such information. Another demonstration of the same fact is provided by the mean ranks listed in Table 1. Due to skewness of the distributions (right at week 1, left at week 6) and the ordinal level of measurement, the Page test for ordered alternatives is an appropriate nonparametric test in this case (Siegel & Castellan, 1988) and the results given in Table 1 prove that the medians increase significantly across repeated samples.

Table 1: Analysis of rehabilitation trend — mean ranks and summary of Page test

<table>
<thead>
<tr>
<th>Scale</th>
<th>Week 1</th>
<th>Week 3</th>
<th>Week 6</th>
<th>L</th>
<th>z</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>BI</td>
<td>1,00</td>
<td>2,12</td>
<td>2,88</td>
<td>458</td>
<td>7,63</td>
<td>&lt; 0,0001</td>
</tr>
<tr>
<td>EBI</td>
<td>1,00</td>
<td>2,09</td>
<td>2,91</td>
<td>459</td>
<td>7,78</td>
<td>&lt; 0,0001</td>
</tr>
</tbody>
</table>

As the first step in addressing the issues of reliability and validity of EBI and its two subscales, dimensionality of the EBI should be analysed. Had there been a sufficiently large number of subjects sampled at each week or had there been three independent samples of subjects rated, factor analysis (FA) would have been the obvious choice for studying dimensionality, especially because EBI is far better at meeting the criteria for an interval scale than BI. Meaningful insight can actually be obtained by pooling the data from the three weeks and performing FA of EBI items — despite the fact that the fundamental FA assumption of independence (Fulgosi, 1988) is thus violated — in accordance with the principle that data analysis should only be an aid to thought rather than its substitute (Breakwell, Hammond & Fige-Schaw, 1995).

Here we present the results of cluster analysis (CA) and multidimensional scaling (MDS), addressing the issue of dimensionality (and therefore also construct validity) of EBI (and its physical and cognitive parts as separate scales). CA results are presented first, whereby
Ward's algorithm was used in all instances and the analyses based upon inter-item similarity, measured by Pearson's correlation coefficient (which, as it is sometimes overlooked, requires normality only when used for inference/hypotheses testing purposes). Horizontal icicle plots with rectangular branches in Figure 2 depict amalgamation schedules for data from week 1, 3 and 6, while the vertical icicle plot with oblique branches in Figure 3 shows clustering of pooled data (the 'overall' similarity matrix was obtained by averaging similarity matrices from individual weeks by means of Fisher's $z$ transformation).

\[ \text{Week 1} \]

\[
\begin{array}{cccccccccccccc}
\text{Item #} & 8 & 6 & 5 & 7 & 4 & 15 & 14 & 13 & 12 & 11 & 16 & 10 & 9 & 3 & 2 & 1 \\
\end{array}
\]

\[ \text{Week 3} \]

\[
\begin{array}{cccccccccccccccc}
\text{Item #} & 16 & 15 & 14 & 13 & 12 & 11 & 10 & 9 & 8 & 7 & 6 & 5 & 4 & 3 & 2 & 1 \\
\end{array}
\]

\[ \text{Week 6} \]

\[
\begin{array}{cccccccccccccccc}
\text{Item #} & 16 & 15 & 14 & 13 & 12 & 11 & 10 & 9 & 8 & 7 & 6 & 5 & 4 & 3 & 2 & 1 \\
\end{array}
\]

\text{Figure 2: Dendrograms depicting clustering of EBI items at week 1, 3 and 6 (similarity measure: Pearson's r; algorithm: Ward's method)}
As the emphasis of this paper is on data-analysis methods rather than on the neuropsychological or rehabilitation science contents, an in-depth discussion and explanation of inter-item co-variation is not given here. However, taking into account the fact that all the inter-item correlations were positive a very general judgement of two-dimensionality can be expressed, the last two items of the physical part (#9 and #10 — controlling bowels and bladder) 'belonging' to the cognitive cluster and the last item of the cognitive part (#16 — vision/neglect) being somewhat 'unique'. Without going into any detail we must mention that there is substantial theoretical background and a body of empirical studies supporting the 3-factor and 4-factor alternatives to the 2-factor model (e.g., see Hartman-Maier & Katz, 1995, regarding visual neglect, Jitapunkul, Kamolratanakul & Ebrahim, 1994, regarding incontinence and Stineman et al., 1996, for a discussion of FIM dimensionality).

For the same reason as stated above interpretation of various MDS solutions is omitted and the only MDS results presented are those in Table 2, summarising the change in adequacy with which sample space is represented (as measured by stress — the most widely used measure for this purpose) as a function of dimensionality. Unlike in CA, Euclidean distance was the distance measure of choice (the matrix for pooled data was obtained by simple averaging). Zero epsilon was used and after finding the 1-D solution, where the standard Guttman-Lingoes initial configuration was requested, the final configuration obtained was always used as the initial configuration for the following analysis (increased dimensionality). The vertical bars separating cells within rows indicate preferable 'breaks' (the solid bar indicating the first choice and the dotted one an eventual second choice) observed if a procedure analogous to Cattell's Scree test in FA (i.e., looking for major changes in trend slope an a stress vs. dimensionality plot) were used to determine the optimal solution.

**Table 2: Summary of multidimensional scaling results — stress as a function of solution dimensionality (solid and dotted vertical bars indicating major leaps in stress magnitude)**
The question of reliability, which is a necessary condition for validity (Nunnaly & Bernstein, 1994), was addressed next. Since inter-rater reliability, as measured by kappa coefficient, has already been established for EBI (refer to the introduction) and BI (Collin et al., 1988, and other studies) and test-retest reliability of BI has also been successfully tested (Gompetz, Pound & Shah, 1994), our focus was on reliability in terms of internal consistency. The results — Chronbach's alpha coefficients — are presented in Table 3.

Table 3: Chronbach's alpha for the scales applied

<table>
<thead>
<tr>
<th>Scale</th>
<th># of items</th>
<th>α Week 1</th>
<th>α Week 3</th>
<th>α Week 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>EBI</td>
<td>16</td>
<td>0.897</td>
<td>0.934</td>
<td>0.948</td>
</tr>
<tr>
<td>EBI-P</td>
<td>10</td>
<td>0.901</td>
<td>0.957</td>
<td>0.969</td>
</tr>
<tr>
<td>EBI-C</td>
<td>6</td>
<td>0.846</td>
<td>0.803</td>
<td>0.793</td>
</tr>
<tr>
<td>BI</td>
<td>10</td>
<td>0.857</td>
<td>0.922</td>
<td>0.946</td>
</tr>
</tbody>
</table>

When examining these results one should bear in mind Cortina's (1993) appeal that absolute alpha should be interpreted with caution, because even though alpha does increase with item intercorrelation and does decrease with multidimensionality, alpha values can be high in spite of low item intercorrelations and multidimensionality. The relatively high alphas obtained therefore do not at all prove that the scales are homogeneous, which is particularly the case with EBI total score, where (at least) two dimensions are undoubtedly involved in addition to a relatively large number of items. It can merely be said that there are few or no 'solitary' items, meaning that within a scale each item correlates significantly with at least one other item. On the other hand, it may be argued that regardless of interpretational difficulties the total score of a multidimensional test with high alpha is meaningful because it is free of error associated with the use of different items. Another 'comforting' observation to be derived from Cortina's findings on the claim that alpha is a lower bound of reliability is that due to differences in item variance (or, more generally, the fact that items are not essentially tau-equivalent), reliability might notably exceed the alpha values obtained.

Table 4: Correlations (Spearman's rho) of EBI, EBI-P and BI with FM at week 1, 3, and 6

<table>
<thead>
<tr>
<th>Scale</th>
<th>ρ Week 1</th>
<th>ρ Week 3</th>
<th>ρ Week 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>EBI</td>
<td>0.658</td>
<td>0.821</td>
<td>0.866</td>
</tr>
<tr>
<td>EBI-P</td>
<td>0.704</td>
<td>0.829</td>
<td>0.879</td>
</tr>
<tr>
<td>BI</td>
<td>0.725</td>
<td>0.854</td>
<td>0.886</td>
</tr>
</tbody>
</table>

Concurrent validity — a type of criterion validity — of EBI was checked by correlating it and its physical part to the Fugel-Meyer Motor Impairment Scale (a validated instrument, regularly used within physiotherapy practice). As EBI (and even more so EBI-P) correlates almost perfectly with BI and BI is considered a 'gold standard', a meaningful comparison can be made with the BI's correlation with FM. BI has been declared an ordinal scale, though, so a
correlation measure suitable for ordinal variables is required, such as Spearman's rho (requiring, of course, the scores to be ranked first). The results are presented in Table 4 ($p<0.001$ for all cells), confirming EBI's concurrent validity with FM.

Table 5: Correlations (gamma) of the EBI memory item with the two RBMT forms

<table>
<thead>
<tr>
<th>RBMT score</th>
<th>$\gamma$</th>
<th>max $p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standardised</td>
<td>0.736</td>
<td>$&lt;0.0001$</td>
</tr>
<tr>
<td>Profile</td>
<td>0.773</td>
<td>$&lt;0.0001$</td>
</tr>
</tbody>
</table>

Special attention was paid to the EBI item on memory, learning and orientation (#15), comparing the ratings with RBMT scores (results are presented in Table 5). In addition to the positive findings, the correlation measure is noteworthy: since the item can be regarded as an ordered categorical variable and the RBMT scores can be transformed into ordered categories, whereby many tied observations are encountered on both variables, the use of gamma coefficient is advisable (Siegel & Castellan, 1988). The significance level is reported as the upper limit of the $p$ value, resulting from the large-samples normal approximation.

### 4. Conclusion

The EBI did not demonstrate what is supposed to be one of its principal advantages compared to BI, i.e., the absence of ceiling effect, since the patients' scores continued to increase during the follow-up period on both scales (however, EBI did prove to be somewhat more sensitive). Both scales clearly showed the rehabilitation trend, which was also confirmed by the nonparametric test of trend monotony. Construct validity of EBI was studied through dimensionality analysis, performed by means of cluster analysis and multidimensional scaling (pointing to the existence of at least subscales). Reliability of the EBI and the BI was assessed in terms of internal consistency and the Cronbach's alpha values obtained were very high for the total EBI, the two EBI subscales and the BI. Criterion validity, more specifically concurrent validity of EBI was confirmed through very high and statistically significant correlation of the total score and of the physical part with the FM motor impairment scale. Finally, the item of the cognitive part measuring everyday memory problems was separately validated by correlating highly and significantly with the RBMT memory test.

Along with the results, some strictly methodological issues are discussed, particularly regarding the application of cluster analysis and multidimensional scaling, as well as the meaning and properties of coefficient alpha. A general conclusion can be drawn that the diversity of data analytical approaches proved useful.

The study could be extended and perfectioned by further research, incorporated into ordinary therapeutic activity, including the use of other disability measures (such as FIM). As far as statistical analysis is concerned, it might be desirable to test unidimensionality of various scales by means of scalogram analysis (using the technique for polychotomous responses). The issue of cocalibration of different functional assessment systems by means of Rasch-type probabilistic models (as advocated in Fisher et al., 1995) might also be addressed.

As only disability measures were considered in the study, it is necessary to mention the handicap issue. Extending beyond the performance of certain activities, assessed by disability
measures, handicap can be defined as a disadvantage for an individual, resulting from an impairment or disability, that limits or prevents fulfilment of the individual's normal role (Badley, 1995). Disability measures are thus not sufficient, although one can hardly contest the assumption that if disability is minimised, handicap will consequently be reduced. Individual's experience of a handicap depends on almost limitless number of variables and compared to disability handicap is more difficult to measure; however, there are psychological measurement instruments addressing the area of handicap. A translation/adaptation, accompanied by adequate quantitative methodology, could certainly be considered, and it might prove helpful in following the present tendency to create a more client-oriented approach in medical treatment and rehabilitation.

Acknowledgements

The authors are grateful to dr. Blaž Zupan for patiently encouraging the writing of this paper, to doc.dr. Valentin Bucik for giving valuable psychometric advice, and to asist. Gregor Sočan for reading the manuscript and providing insightful comments.

5. References


