Rasch Analysis of Combining Two Indices to Assess Comprehensive ADL Function in Stroke Patients

I-Ping Hsueh, MA; Wen-Chung Wang, PhD; Ching-Fan Sheu, PhD; Ching-Lin Hsieh, PhD

Background and Purpose—To justify the summation of scores representing comprehensive activities of daily living (ADL) function, a Rasch analysis was performed to examine whether items of the Barthel Index (BI), assessing ADL, and items of the Frenchay Activities Index (FAI), assessing instrumental ADL, contribute jointly to a single, unidimensional construct in stroke patients living in the community. The number of scoring points of both indices was examined for their usefulness in discerning the various ability levels of ADL in these patients.

Methods—A total of 245 patients at 1 year after stroke participated in this study. The BI and FAI were administered to the patient and/or the patient’s main caregiver by interview.

Results—The initial Rasch analysis indicated that the middle scoring points for many items of the BI and FAI could be collapsed to allow only dichotomous response categories. All but 2 items of the FAI, social occasions and walking outside, fitted the model’s expectations rather well. These 2 items were excluded from further analysis. A factor analysis performed on the residuals of the Rasch-transformed scores recovered no dominant component. These results indicate that the combined 23 dichotomous items of the BI and FAI assess a single unidimensional ADL function.

Conclusions—A clinically useful assessment of the comprehensive ADL function of patients at or later than 1 year after stroke can be obtained by combining the items of the BI and FAI (excluding 2 FAI items) and simplifying the responses into dichotomous categories. It is also demonstrated that the items of the new scale measure comprehensive ADL function as a single unidimensional construct when assessed at 1 year after stroke. (Stroke. 2004;35:721-726.)

Key Words: activities of daily living cerebrovascular accident disability evaluation

Stroke is the most common cause of dependence in activities of daily living (ADL) among the elderly. ADL generally refers to basic or personal ADL, which has been widely used as a main outcome measure after stroke. However, basic ADL does not capture significant losses in higher levels of physical function or activities that are necessary for independence in the home and community (ie, Instrumental ADL [IADL]). Both the basic ADL and IADL are recommended as the primary outcome measures after stroke. Several authors have recommended combining the basic ADL and the IADL to comprehensively measure ADL function. Such a combined scale is expected to be more responsive and have a wider range than either of the individual measurements.

Previous studies using factor analysis found that the Barthel Index (BI) (measuring basic ADL) and the Frenchay Activities Index (FAI) (measuring IADL) assessed different constructs in patients with stroke. Because the BI and FAI supplemented each other with a minimum of overlap in content, it was proposed that they be combined to assess comprehensive ADL function. However, the main question in combining items of ADL and IADL indices is whether these items together contribute to a unidimensional construct, which is necessary to justify the summation of scores. Furthermore, factor analysis operates on interval scale scores, whereas the BI and FAI scores are ordinal by nature. Thus, the results of studies using factor analysis to examine whether items of both indices could be combined to measure a single construct were inconclusive.

Rasch analysis transforms ordinal scores to the logit scale and thus to an interval-level measurement. Furthermore, the Rasch model can examine whether items from a scale measure a unidimensional construct. If items do not fit the model’s expectation, unidimensionality is not preserved. Therefore, Rasch analysis is a useful method to validate the unidimensional construct of all the items of the BI and FAI and to obtain an objective, interval scale. Appropriateness of scoring points refers to whether or not participants can be differentiated by their responses as clearly as the points allow. The items of the BI are either dichotomous or on a 3- or 4-point scale; those of the FAI are on a 4-point scale. Recent studies have shown that a higher number of scoring points may not lead to finer differentiation of the participants. Given that the BI has been found to have a
Subjects and Methods

Subjects
The study sample was recruited from the registry of the Quality of Life After Stroke Study in Taiwan between December 1, 1999, and December 31, 2001. Patients were included in the study if they met the following criteria: (1) diagnosis of cerebral hemorrhage or cerebral infarction; (2) first onset of cerebrovascular accident, without other major diseases; (3) stroke onset within 14 days before hospital admission; and (4) informed consent given personally or by proxy. Further selection and exclusion criteria can be found elsewhere.15

Procedure
Data were collected at 1 year after stroke. An occupational therapist administered the BI and FAI to the patient and/or the patient’s main caregiver by face-to-face or telephone interview. The therapist tried to obtain the actual performance of the patients in daily life from all the informants available.

Measures
Two dichotomous, six 3-point, and two 4-point items constitute the BI.16 The FAI, which was developed to assess social activities or lifestyle after stroke,17 consists of fifteen 4-point items. Stroke severity at admission to the hospital was determined by the Canadian Neurological Scale.18 All 3 scales have been shown to be valid and reliable.18–22

Data Analysis
The analysis consisted of 2 parts. First, the appropriateness of the scoring points in each item of the BI and FAI was investigated with the use of the Rasch partial credit model.10 The scoring points of the items of both indices were reorganized if they were found to be inappropriate. Second, the unidimensionality of the combined BI and FAI scale with appropriate categories was examined by Rasch analysis. The WINSTEPS computer program23 was used to perform the Rasch analysis.

Table 1. Characteristics of the Patients With Stroke (n=245)

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Male/female</th>
<th>130/115</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basic characteristics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sex</td>
<td>Male/female</td>
<td>130/115</td>
</tr>
<tr>
<td>Age</td>
<td>Mean year (SD)</td>
<td>65.4 (11.2)</td>
</tr>
<tr>
<td>Diagnosis</td>
<td>Cerebral hemorrhage</td>
<td>68 (28%)</td>
</tr>
<tr>
<td></td>
<td>Cerebral infarction</td>
<td>177 (72%)</td>
</tr>
<tr>
<td>Side of hemiplegia</td>
<td>Right/left</td>
<td>101/144</td>
</tr>
<tr>
<td>Canadian Neurological score at admission</td>
<td>Median (interquartile range)</td>
<td>7 (4–9)</td>
</tr>
<tr>
<td>At 1 year after stroke</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BI</td>
<td>Median (interquartile range)</td>
<td>90 (65–100)</td>
</tr>
<tr>
<td>FAI</td>
<td>Median (interquartile range)</td>
<td>6 (0–13)</td>
</tr>
<tr>
<td>Combined ADL raw score</td>
<td>Median (interquartile range)</td>
<td>9 (5–13)</td>
</tr>
<tr>
<td>Combined ADL Rasch-transformed score</td>
<td>Mean (SD)</td>
<td>−1.17 (3.94)</td>
</tr>
</tbody>
</table>

The partial credit model can be used for polytomous items such as those of the BI and FAI. We examined the order of the step difficulties of each item of both indices to see whether their scoring points could be simplified.

To examine unidimensionality of the combined scale, infit and outfit statistics were used to examine whether the data fit the model’s expectation. The infit mean square (MNSQ) is sensitive to unexpected behavior affecting responses to items near the person’s proficiency measure; the outfit MNSQ is sensitive to unexpected behavior by persons on items far from the person’s proficiency level.10,24 MNSQ can be transformed to a Z statistic, termed the standardized Z value (ZSTD), which follows approximately the t or standard normal distribution when the items fit the model’s expectation. In this study items with both infit and outfit ZSTD beyond 2 were considered poor fitting.

When items fit the model’s expectation, the residuals (observed scores minus expected scores) should be randomly distributed. A factor analysis was conducted to verify whether any dominant component existed among the residuals. The unidimensionality assumption held if no dominant component was found.25

The sufficiency of the BI and FAI combined was verified by examining whether there were substantial gaps between the item difficulties along the item hierarchy.

Results
A total of 311 patients were registered in the Quality of Life After Stroke Study during the study period. Eight of the 311 patients declined to participate in the study, and 58 patients either did not survive another stroke or could not be contacted for follow-up. A total of 245 patients were included in this study. The Canadian Neurological Scale scores showed that the patients had a broad range of severity at admission. The characteristics of the study sample are shown in Table 1.

Appropriateness of Level of Scaling
All but 2 items (social occasions and walking outside) fitted the model’s expectation fairly well, indicating that these items measured a single construct for patients at or after 1 year after stroke. Table 2 shows that all but 2 of the 23 polytomous items of both indices exhibited disordering of the step difficulty (ie, that the difficulty of a higher step was lower than that of its adjacent lower step), indicating that the middle categories were never the most likely responses of any patient and were thus redundant. Accordingly, the response
were not very far away from the critical point.

...some degree of unexpected behavior affecting responses to

...remaining 23 items. The infit ZSTD had slightly more variations

...items were excluded from the rest of the analyses. Table 3 shows

...which were statistically significant at the 0.05 level. These 2

...of 3.4 (MNSQ = 9.90 and 4.76, respectively) and infit ZSTD statistics

...hobbies (MNSQ = 3.4, 2.5, and 2.1, respectively), those items with slightly extreme infit ZSTD were

...reliability was 0.94 (which can be similarly interpreted as

...0.55), indicating that these items yielded very

...arranged in descending order of difficulty.

...the 2 Poor-Fitting Items of the FAI Were Excluded

...variances (SD = 1.7) than the outfit ZSTD (SD = 0.5), indicating that there was some degree of unexpected behavior affecting responses to

...were not very far away from the critical point ±2 or ±3 (more

...categories were collapsed into a dichotomy. The highest score
category of the items in the BI was rescored as 1, indicating
full independence in the activity. The rest of the categories
were rescored as 0, indicating dependence in the activity. For
the FAI, all but the lowest score category (0) of the items
were collapsed into a category and rescored as 1, indicating
participation in the activity. The lowest score category of the
items in the FAI remained 0, indicating no participation in the activity.

Unidimensionality of Simplified Scale

When the Rasch dichotomous model was applied to examine the
25 dichotomous items together, 2 misfit items (social occasions and
walking outside) had outfit ZSTD statistics of 2.9 (MNSQ = 9.90 and 4.76, respectively) and infit ZSTD statistics
of 3.4 (MNSQ = 1.37) and 5.3 (MNSQ = 1.86), respectively, which were statistically significant at the 0.05 level. These 2
items were excluded from the rest of the analyses. Table 3 shows
that none of the outfit ZSTD was statistically significant for the
remaining 23 items. The infit ZSTD had slightly more variations
(SD = 1.7) than the outfit ZSTD (SD = 0.5), indicating that there was some degree of unexpected behavior affecting responses to
items near the person’s ability measure. However, because they
were not very far away from the critical point ±2 or ±3 (more

...liberal) and their corresponding outfit ZSTD statistics were not
significant, those items with slightly extreme infit ZSTD were not treated as poor-fitting items. A factor analysis on the residuals of Rasch-transformed scores showed no dominant factors. The first and second factors accounted for only 12.3% and 8.3% of the residual variance, respectively. These results indicated that there was a good model-data fit and that the assumption of unidimensionality held for these 23 dichotomous items when assessed at 1 year after stroke.

We further examined the psychometric properties of the
23-item scale. The person measures (ranging from −9.46 to 6.04) had a mean of −1.17 and a SD of 3.94. The person
reliability was 0.94 (which can be similarly interpreted as
Cronbach’s α), indicating that these items yielded very
precise estimates for the patients. The person reliability was practically identical to that under the partial credit modeling (0.93), ie, collapsing categories did not reduce the person reliability. The person measures under the Rasch dichotomous model were also highly correlated with those under the partial credit model (r = 0.97), indicating that collapsing the categories did not substantially alter ADL score in the patients.

The Figure shows the map of persons and items along the
continuum. The item difficulties were spread out across the
patients, indicating that these items could well differentiate these patients. The Figure also shows that there were obvious gaps between items BI3 (grooming) and BI8 (transfer), between BI9 (mobility) and BI4 (dressing), between BI2 (bathing) and BI10 (stairs), and between FAI6 (local shopping) and FAI10 (driving a car/bus travel).

A conversion table showing combined BI and FAI raw scores (23 items) and their Rasch-transformed scores (ie, person measure in logit) is shown in Table 4. It allows prospective users to easily derive the Rasch-transformed scores from the raw scores.

**Discussion**

Combining ADL and IADL indices can provide a measurement tool with an enhanced range and sensitivity for assessing comprehensive ADL function, which is especially useful when applied to patients with stroke who are living in the community. Using Rasch analysis, this study found that such a measurement tool can be created by combining all but 2 items of the BI and FAI and simplifying the polytomous responses into dichotomous categories. Because the items of the new scale have been shown to measure comprehensive ADL function as a single, unidimensional construct for patients at 1 year after stroke, it is valid to sum the item scores of the combined scale. The total score represents a stroke patient's ADL function on a single continuum encompassing the entire range of ADL. A higher score indicates higher independence in living in the community.

There are some advantages to using the combined scale. The items of the BI are located in the lower part of item difficulty, indicating easier activities, and those of the FAI are located in the upper part, indicating more difficult activities. The total score of the combined scale is useful in showing whether the subject has basic ADL problems or IADL problems. A conversion table was compiled showing combined BI and FAI raw scores (0 to 23) and their Rasch-transformed scores. The Rasch-transformed scores can be viewed as interval-level measurements, whereas the combined BI and FAI raw scores are ordinal data. Since most statistical techniques assume that the data are at least on an interval scale, the Rasch-transformed scores of the combined scale are recommended for future applications. In addition, both the original BI and FAI and their scoring guidelines as used in this study can be employed as usual. To use the new scale proposed in this article, the users can first obtain the combined ADL score for a patient by adding the dichotomized scores of all but the 2 poor-fitting items and then use the conversion table to look up the corresponding Rasch-transformed interval score.

There have been few empirical investigations of the appropriateness of scoring points of the BI and FAI. The items in both indices contain 2 to 4 scoring points. We found that the middle categories of both indices provide little information about the patients and are redundant. Thus, the response categories of all the items were recoded as dichotomous in this study. Given that the psychometric
properties of the dichotomous items were equivalent to those of the polytomous items and that a scale with only dichotomous items is much more convenient and efficient to administer, the dichotomous categories are recommended for use in both indices for patients at or after 1 year after stroke.

Tennant et al.26 also used Rasch analysis to examine the dimensionality of the BI and found that the item “bladder control” did not fit the model’s expectation. The discrepancy between their findings and our findings may be explained by the different scoring guidelines of the BI used in the respective studies. We adopted the scoring guidelines suggested by Wade and Collin,22 whereas they used those from Shah et al.27 Therefore, the results from this study should be interpreted with caution when other scoring guidelines are used.

Cultural or environmental factors may explain why the items “social occasions” and “walking outside” did not fit the model well.28,29 In addition, all the patients were assessed at 1 year after stroke and had been discharged from the hospital >6 months previously. Our results may not be broadly applicable to patients with different characteristics (eg, patients at a subacute stage or those who have just been discharged from the hospital).

The combined scale can be further improved by generating items to fill the gaps among pairs of current items (eg, grooming and transfer). These new items will be able to differentiate the ability levels of patients falling in the gaps of the current scale. Future studies to examine utility and other psychometric properties (eg, interrater reliability and responsiveness) of the combined scale in stroke patients are needed. Furthermore, it may be promising to extend the BI and FAI into an item bank and to use computerized adaptive testing to further elaborate ADL measurement.

In summary, this study provides good evidence for a unidimensional construct of the dichotomized and combined BI and FAI, with the exclusion of 2 FAI items. The results imply that these 2 indices can be combined to assess comprehensive ADL function in patients after 1 year after stroke.

Acknowledgments
This study was supported by research grants from the National Science Council (NSC 89–2320-B–002–006, NSC 90–2320-B–002–033, and NSC 91–2320-B–002–008) and the National Health Research Institute (HNRI-EX92-9204PP).

References
Rasch Analysis of Combining Two Indices to Assess Comprehensive ADL Function in Stroke Patients
I-Ping Hsueh, Wen-Chung Wang, Ching-Fan Sheu and Ching-Lin Hsieh

Stroke. 2004;35:721-726; originally published online February 12, 2004;
doi: 10.1161/01.STR.0000117569.34232.76

Stroke is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2004 American Heart Association, Inc. All rights reserved.
Print ISSN: 0039-2499. Online ISSN: 1524-4628

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://stroke.ahajournals.org/content/35/3/721

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in Stroke can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to Stroke is online at:
http://stroke.ahajournals.org//subscriptions/