Underestimating Physical Function Gains: Comparing FIM Motor Subscale and interRAI Post Acute Care Activities of Daily Living Scale

Christine Glenny, MSc, Paul Stolee, PhD, Mary Thompson, PhD, Janice Husted, PhD, Katherine Berg, PhD, PT


Objective: To investigate the construct validity of the activities of daily living (ADLs) sections of 2 major systems developed to measure functional ability in rehabilitation settings. Health assessment systems can inform care planning as well as policy decision-making on service effectiveness. Frailty, comorbidity, and heterogeneity make it difficult to accurately measure health outcomes for older adults. Objective investigation of the value of geriatric rehabilitation services requires assessment systems that are comprehensive, reliable, valid, and sensitive to clinically relevant changes in older patients.

Design: Trained health care workers assessed patients with both tools at admission and discharge. We used Rasch analysis to compare the instruments’ dimensionality, item difficulty, item fit, differential item function, and number of response options.

Setting: Musculoskeletal and geriatric rehabilitation units in 2 Ontario hospitals.

Participants: Older adults receiving rehabilitation (N=209; mean age ± SD, 78.5±9.3; 67% women).

Interventions: Not applicable.

Main Outcome Measures: FIM and the interRAI Post Acute Care Assessment (interRAI PAC).

Results: For both the FIM motor and the interRAI PAC ADLs items, the difficulty level of the items was much lower than the participant’s level of ability, resulting in a large ceiling effect. Also, on both scales, less actual change in functional ability was required to move between the midlevel response options.

Conclusions: Both scales have limited ability to discriminate between subjects with higher functional ability, which indicates that they may underestimate the effectiveness of inpatient rehabilitation for this group of patients when used alone.

Key Words: Activities of Daily Living; Aged; Rehabilitation.

INPATIENT REHABILITATION is designed to increase functional capacity and improve patient quality of life. For young, healthy patients, rehabilitation is perceived as a beneficial and accepted treatment; however, despite strong clinical support, the effectiveness of geriatric interventions remains controversial. While a variety of rehabilitation programs have been suggested with the potential to improve outcomes for older adults, there are limited data available for identifying successful programs. This may be attributable, in part, to inadequate outcome measures for this population.

Assessment instruments are often designed and validated with samples that are relatively uniform, healthier, and younger than the typical geriatric inpatient. Older patients represent a diverse population that may possess comorbid illnesses. The heightened clinical complexity in this population may require more comprehensive assessment strategies when evaluating programs or interventions that are capable of tracking longer-term outcomes including: rehospitalization, emergency department use, and longer-term function in the community. It is also important to examine the measurement properties of the scales used to examine the causal pathways of how better long-term outcomes were achieved; accordingly further research is needed that focuses specifically on the reliability, validity, and responsiveness of outcome measures in this population.

The FIM8 and the interRAI Post Acute Care Assessment (interRAI PAC)9 are instruments used to assess functional status in older rehabilitation patients. Efforts have been made to evaluate the equivalency of the basic activities of daily living (ADLs) items on the FIM and an earlier version of the interRAI PAC. These studies primarily focused on creating a crosswalk between FIM and interRAI PAC items;10-12 and provided evidence to support the compatibility of the scales with correlations ranging from .72 to .85. Conversely, Buchanan et al11 found that despite high item level correlations, the 2 scales had poor agreement regarding classification into prospective payment cells and concluded that the basic ADLs items on the FIM
and interRAI PAC were not equivalent. Our previous work examined the psychometric properties of the FIM motor and basic ADLs items on the interRAI PAC and suggested that their responsiveness may be relatively equivalent; however, additional research is needed to investigate their construct validity in this population.13,14

Rasch analysis15 is a statistical technique used increasingly in health research.16,17 It uses data from classically designed ordinal scales and forces it into a linear model in order to evaluate how well the empirical data corresponds to or fits the ideal model. The power of this tool applies to instrument development and the evaluation of psychometric properties of existing scales because (1) you can quantify the assumption that the items can be measured at the interval level and (2) it provides information on how to modify the instrument to become a more accurate estimate of the ideal interval scale.18 We used Rasch analysis to examine the construct validity of the FIM motor subscale and the basic ADLs items on the interRAI PAC in a sample of older adults receiving inpatient rehabilitation. The focus of this investigation was to examine the ability of each instrument to differentiate between levels of functional ability for this population. More specifically we were interested in answering the following 3 questions: (1) Does the difficulty level of the items on the FIM and the interRAI PAC correspond with the level of functional independence in older rehabilitation patients? (2) Is the amount of improvement necessary to increase 1 point on the FIM and the interRAI PAC consistent across different levels of functional independence? (3) Are the number of response options on the FIM and the interRAI PAC appropriate for defining different levels of functional independence in this population? To the best of our knowledge, this is the first study to compare the FIM motor subscale with the revised interRAI suite of ADLs items.

METHODS

Study Sample

Data were collected between October 2005 and April 2006 at London Parkwood Hospital and the Toronto Rehabilitation Institute. Patients were recruited from musculoskeletal (MSK) and geriatric rehabilitation units (GRUs) at both institutions as part of an observational cohort study which compared the various scaling properties of the FIM and the interRAI PAC (eg, responsiveness). Two trained nurses at each site assessed a total of 208 patients (n=93 GRU, mean age ± SD, 81.4±6.7, 65% women; and n=115 MSK, mean age ± SD, 76.4±10.2, 67% women) with both instruments on admission and at discharge. The focus of this investigation was to examine the ability of each instrument to differentiate between levels of functional ability for this population. More specifically we were interested in answering the following 3 questions: (1) Does the difficulty level of the items on the FIM and the interRAI PAC correspond with the level of functional independence in older rehabilitation patients? (2) Is the amount of improvement necessary to increase 1 point on the FIM and the interRAI PAC consistent across different levels of functional independence? (3) Are the number of response options on the FIM and the interRAI PAC appropriate for defining different levels of functional independence in this population? To the best of our knowledge, this is the first study to compare the FIM motor subscale with the revised interRAI suite of ADLs items.

Instruments

The FIM, developed in 1983, was designed to assess burden of care and measure physical and cognitive disability.8 The FIM contains 18 items which are scored on a 7-point ordinal scale based on the amount of assistance required to perform the activity.8 Thirteen of the items compose the motor subscale (FIM motor), and the remaining 5 items make up the cognitive subscale (FIM cognitive). In previous literature, the FIM motor and FIM cognitive have been found to define separate constructs.26-28; therefore, it was not appropriate to combine these 2 subscales in a single Rasch analysis. For this investigation, we focused on the FIM motor subscale only.

Currently, 10 interRAI instruments have been developed, which focus on designing comprehensive assessment tools for use across the health care continuum.29-31 Intended uses include care planning, outcome measurement, and quality indicators.29-31 The interRAI instruments, including the interRAI PAC, measure function in multiple domains including cognition, mood, psychosocial, nutrition, and physical functioning including ADLs and instrumental activities of daily living (IADLs). This study focuses on ADLs items, which like the FIM, are scored based on the amount of assistance required. Because the interRAI instruments have been designed for different populations and settings over time, there are varying ADLs subscales.13,32-35 The interRAI PAC items included in this study were: all ADLs self-performance items, stairs (performance), and bowel and bladder continence (interRAI PAC scanable version 07). Though additional items on the interRAI PAC may be relevant to assessing basic ADLs, they were not included in this analysis because our aim was to evaluate comparable subscales from the FIM and the interRAI PAC instruments. The full 7-point ordinal scale was used to assess all interRAI

### Table 1: Sample Characteristics for GRU and MSK Patients

<table>
<thead>
<tr>
<th>Mean Subject Characteristics</th>
<th>GRU (n=93)</th>
<th>MSK (n=115)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age*</td>
<td>81.4±6.7</td>
<td>76.4±10.2</td>
</tr>
<tr>
<td>Sex (women:men)</td>
<td>60:33</td>
<td>77:38</td>
</tr>
<tr>
<td>Functional Comorbidity Index</td>
<td>2.4±1.4</td>
<td>1.2±1.1</td>
</tr>
<tr>
<td>Length of stay (d)</td>
<td>54:0±33.4</td>
<td>23.0±15.1</td>
</tr>
<tr>
<td>Prehospital residential status*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Home</td>
<td>81</td>
<td>112</td>
</tr>
<tr>
<td>Home care/community support</td>
<td>12</td>
<td>3</td>
</tr>
<tr>
<td>Assisted living/LTC</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Other</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Prehospital living arrangements*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alone</td>
<td>42</td>
<td>52</td>
</tr>
<tr>
<td>With spouse</td>
<td>38</td>
<td>44</td>
</tr>
<tr>
<td>With other relatives</td>
<td>11</td>
<td>14</td>
</tr>
<tr>
<td>Other</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Postdischarge residential status*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Home</td>
<td>46</td>
<td>97</td>
</tr>
<tr>
<td>Home care/community support</td>
<td>34</td>
<td>13</td>
</tr>
<tr>
<td>Assisted living/LTC</td>
<td>12</td>
<td>5</td>
</tr>
<tr>
<td>Other</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>FIMmotor admission</td>
<td>44.9±12.4</td>
<td>55.7±10.8</td>
</tr>
<tr>
<td>FIMmotor discharge</td>
<td>65.7±16.9</td>
<td>78.6±7.2</td>
</tr>
<tr>
<td>interRAI PAC admission</td>
<td>40.8±12.9</td>
<td>29.8±13.5</td>
</tr>
<tr>
<td>interRAI PAC discharge</td>
<td>19.6±17.0</td>
<td>8.6±8.5</td>
</tr>
<tr>
<td>FIMcognitive on admission</td>
<td>27.3±5.6</td>
<td>34.2±1.8</td>
</tr>
<tr>
<td>interRAI PAC CPS on admission</td>
<td>1.17±1.09</td>
<td>0.35±0.75</td>
</tr>
</tbody>
</table>

**NOTE.** All data in the table refer to mean ± SD unless stated otherwise. Abbreviations: CPS, Cognitive Performance Scale; LTC, Long Term Care.

*These data refer to frequency counts.
Masters’ Partial Credit model for polytomous scales. A shared linear continuum, measured in logits (log-odds units, or single Rasch units), is developed by characterizing subjects based on their performance scores on the tool (referred to as the subject ability) and the items on the instrument (referred to as the item difficulty). Rasch models are log-linear models based on log-odds transformations. It uses ordinal data from classically designed measurement tools to construct interval measures. 

A principal component analysis of the residuals was completed for both instruments prior to the full statistical investigation, and consistent with previous literature, supported that for both instruments the analysis was done and the interRAI PAC, as well as the codes that were used in subsequent figures.

**Statistical Analysis**

Rasch analysis is based on log-odds transformations. It uses ordinal data from classically designed measurement tools to construct interval measures. A shared linear continuum, measured in logits (log-odds units, or single Rasch units), is developed by characterizing subjects based on their performance on the tool (referred to as the subject ability) and the items on the tool (referred to as the item difficulty). Rasch models are log-linear models based on the probability $P(n)$ that a subject $n$ with the ability $B_n$ will succeed on item $i$ with difficulty $D_i$, or achieve category $j$ in an item $i$ for which this achievement relative to category $j-1$ has threshold level $D_{ij}$. The analysis was done using Winsteps version 3.70.0. For this study, we applied the Masters’ Partial Credit model for polytomous scales.

Dichotomous model: $\log \left( \frac{P_n(1)}{P_n(0)} \right) = B_n - D_i$

Polytomous “Partial Credit” model:

$$\log \left( \frac{P_{ni}}{P_{n(i-1)}} \right) = B_n - D_i - F_{ij} = B_n - D_{ij}$$

A principal component analysis of the residuals was completed for both instruments prior to the full statistical investigation, and consistent with previous literature, supported that each set of items was defined by a unidimensional construct indicating that they were appropriate for further Rasch analysis. The interRAI PAC was also reverse scaled prior to the analysis (0=total dependence, 6=independence) to fit the conventional Rasch analysis.

Distribution maps were constructed separately for the MSK and GRU patients for the admission (fig 1) and discharge (fig 2) data. These are graphical illustrations of a hierarchy of items and subjects on the logit scale. The graph is centered so that the mean difficulty for the items is at 0 logits, such that items that have a positive logit value are more difficult and items with a negative logit value are less difficult than the mean. The left side represents a frequency distribution of subjects by ability, and the right represents a frequency distribution of the items by difficulty. The relative placement of the samples’ mean performance scores and the vertical spread of the subjects in relation to the items were used to illustrate the appropriateness of the tool for this sample.

Because both instruments were designed for subjects with a wider range of physical functioning than either the admission or discharge groups separately, for the remaining analysis we combined these data to better represent this population. Item characteristic curves (ICCs) were used to depict the relationship between the response option selected for each item and the participant’s total score on the latent variable. Each line on the graph represents 1 item on the instrument. Category probability curves (CPCs) were used as an additional strategy to investigate the relationship between item scores and the total score. For an item to be able to differentiate between subjects at the level of their present category structure, each item response option must be the most probable at some point along the latent variable. Disordered thresholds were used to examine how the items and response options on the FIM motor and the interRAI PAC relate to the total score for the subjects.

This study received ethics clearance from the University of Waterloo, the University of Western Ontario, the University of Toronto, and the Toronto Rehabilitation Institute.

**RESULTS**

Figures 1 and 2 show the distribution maps for both instruments on admission and at discharge, separately for the MSK and GRU groups. For the GRU group on admission (see fig 1), the spread of item difficulty (right side of graph) corresponded relatively close to the performance scores of the subjects (left side of graph) for both instruments. In the MSK group on admission (see fig 1), the mean item difficulty was approximately equal to the mean subject performance scores for the FIM motor; however, the mean item difficulty was highly skewed by the large vertical gap between STAIRS and the remaining items. At discharge (see fig 2), the mean subject performance scores were much higher than the mean item difficulty for both instruments.

Figure 3 shows the ICC graphs for the FIM motor and the interRAI PAC ADLs items in both study samples. The graphs illustrate that the majority of items in both instruments have a steep sigmoidal structure. Similarly, the CPC curves (fig 4) showed that for both instruments, the difficulty spread between...
Eleven of 13 items were clustered between 1 SD above and below the mean, and managing STAIRS item did not have enough variability to contribute to the development of the model, so it was automatically removed during the analysis.

Fig 1. Distribution (variable) maps admission data. NOTE. Table 2 lists the items included in the analysis for the FIM and the interRAI PAC as well as the codes that were used in this figure. Eleven of 13 items were clustered between 1 SD above and below the mean, and managing STAIRS was even more difficult than 2 SDs above the mean. For the interRAI PAC, the mean subject ability was more than 1 SD above the mean item difficulty and many of the items clustered in a narrow region of low subject ability. For this group, the managing STAIRS item did not have enough variability to contribute to the development of the model, so it was automatically removed during the analysis.
**Fig 2.** Distribution (variable) maps discharge data. NOTE. Table 2 lists the items included in the analysis for the FIM and the interRAI PAC as well as the codes that were used in this figure. In the GRU group, mean subject scores were approximately 1 SD higher than mean item difficulty and this difference increased to greater than 2 SDs in the MSK group. For both instruments, the majority of the items formed a redundant horizontal cluster in the low ability region where there were few subjects relative to the number of items.
categories was wider for more extreme options (0–1, 5–6–7) than the middle options (2–3–4). Also, the response options in the middle region contained unordered thresholds for nearly every item on the FIM motor and the interRAI PAC. For most items in both subject groups, the threshold probabilities were clustered in the middle region of the latent variable. Although overlapping categories were most common in the central response region, within these middle categories there was little consensus on the specific location of the overlap across items. For many of these response options, the peak threshold probability was enclosed by neighboring curves. This pattern was especially prominent in the continence, transfer, and locomotion (including locomotion, walk, and stairs) items on both instruments. Figure 4 shows an example CPC curve for the WALK item on each scale as it depicts the common pattern seen for many of the items; CPC curves for all additional items are available elsewhere.40

DISCUSSION

Overall, the results suggest that the majority of the items on the FIM motor or the interRAI PAC ADLs were able to discriminate between subjects in a very narrow and low range of physical functioning. The scales were unable to accurately differentiate between subjects once they reached the middle and higher ranges of physical functioning and therefore may underestimate change in this population. Because most patients were expected to improve in the rehabilitation setting, it is logical to assume that many will approach the higher regions of the scale at discharge. These items may also lack the ability to discriminate in the highest functioning individuals as they only aim to measure basic ADLs. However, we found that subjects were being admitted to rehabilitation with greater performance scores than many of the items on the scales and an unusually high proportion were concentrated within a few points of a perfect score, though they continued to receive rehabilitation for long periods. Previous investigations of the FIM motor have suggested that a potential ceiling effect may prevent the scale from detecting change between admission and discharge.41,42

New more difficult items may need to be added to both scales or the existing items may need to be altered to become more difficult in order to measure physical functioning accurately in this population. In all but 1 group, managing STAIRS was the most difficult item; however, lack of response vari-
ability and large gaps between the difficulty of this item, and those of the remaining scale, suggests that STAIRS (as currently measured) may not belong on either the FIM motor or the interRAI PAC ADLs scales. Prior investigations of the FIM have suggested that this item may need to be altered or removed based on low item total correlation, poor fit, disordered thresholds, and floor effects. However, Lina-cre suggested that the lack of consistency of this item with the rest of the FIM motor may be due to its greater difficulty relative to the remaining items. In the interRAI PAC, the managing STAIRS item has a coding scheme consistent with IADLs. The scoring permits capacity assessment of stair management. In addition, the interRAI PAC physical function section contains items on distance and timed walk. It is possible that increasing the difficulty of existing items or adding new more difficult items may improve the fit of the managing STAIRS item on these scales and enhance the ability of the scale to discriminate between more functionally able subjects.

The sigmoidal pattern of the ICC graphs indicates that the amount of improvement in physical functioning necessary to move between response options was greater for the extreme categories relative to the middle categories. Due to the ordinal nature of these scales, this pattern was expected at the outset of the investigation; however, the extreme slope of these graphs and additional support provided by the CPCs implies that a 1-point change on the FIM motor or the interRAI PAC ADLs does not represent the same increase in physical functioning for all categories. A much larger improvement in physical functioning was required to achieve the most independent option relative to moving between the midrange response options on both scales. This bigger step required to move between more extreme response categories may be clinically intuitive based on the category descriptions; however, it may result in misinterpreting the total score for both scales. Disordered and underused categories in these middle response options may indicate that the scales could be a more accurate estimate of a true interval scale if midcategory items were condensed. However, the lack of consistency of the overlapping response options between items could lead to difficulties in the attempt to decrease the number of response options in order to enhance scale functioning for all items. It is likely necessary to address issues around scale targeting prior to altering the structure of the response options, because modifications to item difficulty or newly added items may influence the ideal solution.

Previous literature, existing scales, and clinical opinion are important to consider when determining what items to add to the FIM motor or the interRAI PAC ADLs scales to improve their ability to detect clinically relevant change in geriatric patients. Wells et al engaged rehabilitation teams from 2 urban academic GRUs to explore current and potential uses of FIM data. The clinicians felt that items were needed which better reflected important increases in functioning for older patients, such as gait quality. Other items to consider for addition in the future may also include gait speed, distance walked, hours of physical activity, and going outdoors. The full interRAI PAC and other interRAI instruments should also be considered as potential sources of new items. For example, Armstrong et al found that the premorbid ADLs and IADLs
statuses, as well as a variable associated with delirium (easily distracted), were significant predictors of rehabilitation outcomes in older rehabilitation patients. It will be important to monitor the dimensionality of the scales as new items are added because it is possible that items may not be contained in the same construct as existing items. In this case, additional sub-scales or a battery may need to be considered. To successfully measure physical functioning, it may be necessary for rehabilitation settings to view outcomes more broadly and for rehabilitation professionals to consider choosing measures beyond basic ADLs to assess patient progress and response to treatment.

Study Limitations

A number of limitations should be considered. Though there were adequate numbers of participants to investigate clinically relevant change, there may have been too few subjects to infer conclusive results from the Rasch analysis. The frequency distributions by response options showed that some categories for some items contained fewer than 10 participants. However, this lack of variability in the data may not have been prevented by increasing the sample size, because there were enough participants in each group for the potential of 10 subjects per category. Also, by combining the admission and discharge data to investigate the response options, we artificially developed a population of subjects that may have had less variability than a true population composed of all independent samples. Because the results for this portion of the analysis were neither obvious nor definitive, different conclusions may have been reached with a more variable group of subjects.

CONCLUSIONS

Both the FIM motor and the interRAI PAC ADLs scales suffer from a similar limitation related to their inability to discriminate between levels of functional ability for older rehabilitation patients. When this limitation is considered regarding the effectiveness of rehabilitation programs for geriatric patients, it is likely that both instruments will provide an underestimate of program benefits. Therefore, even though older adults may have the potential to improve their functional status through inpatient rehabilitation, the use of these scales alone limits the evidence for the change.

References


