The Functional Independence Measures (FIM) reliability, validity and responsiveness in spinal cord injury: literature review

Reprodutibilidade, validade e responsividade da escala de Medida de Independência Funcional (MIF) na lesão medular: revisão da literatura

Daniela de Campos Barbetta¹, Marcos Renato de Assis²

ABSTRACT
The Functional Independence Measures scale is one of the most frequently used tools for the assessment of functional capacity of individuals with spinal cord injuries. Measurements of the functional outcome in these individuals are essential in the rehabilitation process. This study aimed at reviewing the scientific literature regarding the reliability, validity and responsiveness of FIM in individuals with spinal cord injuries. The results showed good internal consistency, inter-rater reliability, validity and responsiveness, except for the assessment of the cognitive dimension and the locomotor domain.

KEYWORDS
spinal cord injuries /rehabilitation, disability evaluation, review literature as topic

RESUMO
A escala de Medida de Independência Funcional é um dos instrumentos mais utilizados na avaliação da capacidade funcional dos indivíduos com lesão medular. O registro da evolução funcional desses indivíduos é fundamental no processo de reabilitação. Este estudo teve como objetivo revisar na literatura científica a reproducibilidade, validade e responsividade dessa escala em indivíduos com lesão medular. Os resultados encontrados mostraram alta concordância interna e reproducibilidade inter-avaliador; validade e responsividade, exceto para avaliação da dimensão cognitiva e do domínio locomoção.

PALAVRAS CHAVE
traumatismos da medula espinhal/reabilitação, avaliação da deficiência, literatura de revisão como assunto
INTRODUCTION

The spinal cord injury is a disabling condition caused by trauma or disease that affects the motor, sensitive and autonomic functions. It has an incidence of 15 to 40 cases in a million, resulting in 11 thousand new cases a year in the United States alone, with a prevalence of 250 thousand individuals.

The functional impairment caused by the spinal cord injury depends on the extension and severity of the lesion and interferes directly on the individual’s capacity to perform the activities of daily living (ADL). The assessment of the degree of functional impairment is defined as any systematic attempt to objectively measure the level of functionality in several aspects that include physical health, self management capacity, activity performance quality, intellectual status, social activity and emotional status.

In 1983, Granger and cols, supported by the Congress and by the American Academy of Medicine and Rehabilitation, collected a set of data to measure disability and to evaluate the outcome of rehabilitation programs. In this project, 36 tools of functional assessment, published and non-published, were revised with the objective of identifying the most common and useful items for the construction of a scale that could assess function under several conditions of disability, such as the cerebrovascular accident, multiple sclerosis, spinal cord injury, in addition to other diseases. Thus, the Functional Independence Measure (FIM) appeared, a multidisciplinary tool that consisted of a set of items, with fast and uniform use, as well as consistent and reliable measures.

In 1984, pilot studies of the FIM scale were started and at the end of the experimental phase, the current tool was attained, with 18 items. These items evaluate self-care activities (eating, getting ready, washing oneself, dressing the upper part of the trunk, dressing the lower part of the trunk and grooming); transferences or mobility (bedchair/wheelchair, toilet, bathtub/shower); locomotion (wheelchair/gait and stairs); sphincter control (bladder and intestine); communication (comprehension and expression) and social cognition (social interaction, problem solving and memory). Each one of these activities receives a score that goes from 1 (total dependence) to 7 (complete independence), with a total score that ranges from 18 to 126.

The first studies on the psychometric properties of the FIM were performed when it was still under construction and were carried out in populations with different degrees of disability, including individuals with spinal cord injury. However, the results have shown to be controversial in the latter group. On the other hand, to evaluate the functional potential and record the evolution of individuals with spinal cord injury throughout time, using a reliable tool, is essential for the adequate measurement of functional impairment caused by the spinal cord injury. Therefore, the aim of the present study is to revise in the scientific literature the reproducibility, validity and responsiveness of the FIM scale in the assessment of individuals with spinal cord injury.

METHOD

The literature search was carried out using the following databases: MedLine PubMed with the key words Reproducibility of Results or Validation Studies, Spinal Cord Injuries and Disability Evaluation available at MeSH and at the Lilacs database, with the key words Validação (Validation) and Reabilitação (Rehabilitation), obtained from the Key Words in Health Sciences—DeCS of Bireme database. The studies published up to June 2007 and their references were selected, with no lower date limit, including only those that evaluated the properties of the FIM scale in individuals with spinal cord injury and excluding those that did treat separately the data concerning this population.

RESULTS

A total of 16 studies were found. The data concerning the design, statistical analysis and results (reproducibility, validity and responsiveness) were organized by date of publication of the articles (Table 1) and the review results were grouped by psychometric properties. Most of the studies evaluated the properties of the FIM scale items, separating motor and cognitive dimensions.
The Functional Independence Measures (FIM) reliability, validity and responsiveness in spinal cord injury: literature review

### Description of the studies on the reproducibility, validity and responsiveness of the Functional Independence Measure (FIM) scale.

<table>
<thead>
<tr>
<th>Author/Year</th>
<th>Sample</th>
<th>Statistical Analysis</th>
<th>Reproducibility</th>
<th>Validity</th>
<th>Responsiveness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Davidoff et al.</td>
<td>Case Group (SO): N=41; Controls (S): N=22; age: 30.4±1.7 mean; 85% of schooling; 12.9±0.4 years; level C=41%, T=39%</td>
<td>Scatter Plots</td>
<td>-</td>
<td>Cognitive score: ceiling effect</td>
<td>-</td>
</tr>
<tr>
<td>Dodds et al.</td>
<td>Evaluation of hospital admission and release; N=786 patients registered at UDSMR</td>
<td>Cronbach’s ω</td>
<td>Low internal consistency in locomotion item</td>
<td>Good construct validity</td>
<td>Responsive.</td>
</tr>
<tr>
<td>Segal et al.</td>
<td>N=57 patients</td>
<td>Pearson’s Coefficient of correlation</td>
<td>High concordance (0.74-0.87) for complete lesions and incomplete paraplegia; low concordance for incomplete tetraplegia (0.49).</td>
<td>-</td>
<td>Responsive, except for the cognitive domain</td>
</tr>
<tr>
<td>Marino et al.</td>
<td>N=22 men with lesions from C4-C7 up to 72h of SCI; age: 18-63</td>
<td>Spearman’s Coefficient of correlation</td>
<td>-</td>
<td>FIM and EMMS: good correlation (0.84); FIM and QIF: very good correlation (0.93).</td>
<td>-</td>
</tr>
<tr>
<td>Grey et al.</td>
<td>N=40 men; 85% paraplegic; 67.5% age of moment SCI; 29.6±9.57; mean time between data collection of both groups: 7.25 weeks (SD=1.93)</td>
<td>Unspecified Coefficient of correlation</td>
<td>Total score (0.83); per domains (&gt;0.71); Cognitive dimension (&lt;0.45).</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Shineman et al.</td>
<td>N=4440 individuals with traumatic (1831) and non-traumatic (2609) SCI registered at UDSMR</td>
<td>Cronbach’s ω</td>
<td>Good internal consistency; total, motor, cognitive FIM</td>
<td>Stairs floor effect</td>
<td>-</td>
</tr>
<tr>
<td>Ota et al.</td>
<td>N=100 paraplegic and tetraplegic patients; Frankel A and B; men: 84; mean time of lesion: 29 months</td>
<td>Comparison of scores without statistical analysis</td>
<td>-</td>
<td>Validity among tetraplegic and paraplegic individuals. No validity among paraplegic individuals &gt; T5 and &lt; T16. SCI &gt; T5: 6 months. SCI &lt; T6: 3 months. Ceiling effect: tetraplegics: 10 months.</td>
<td>-</td>
</tr>
<tr>
<td>Heinemann et al.</td>
<td>N=53 patients registered at UDSMR</td>
<td>Paired T Test</td>
<td>-</td>
<td>Educational activities and administration of drugs: negative correlation between FIM and time spent. Indirect activities: negative correlation only at hospital release.</td>
<td>Responsive for total FIM.</td>
</tr>
<tr>
<td>Karamahrangalis et al.</td>
<td>N=38 patients; men: 76%; age: 33.94±1.59; paraplegics: 78%</td>
<td>Spearman’s Coefficient of correlation</td>
<td>Questionnaire versus perfunctorial; good responsiveness. Higher correlations: comprehension and memory. Lower correlation: LLLL dressing.</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Shineman et al.</td>
<td>N=84387 patients registered at UDSMR</td>
<td>Cronbach’s ω</td>
<td>Good correlation; Total, motor and cognitive FIM.</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

N: Number of individuals; SCI: spinal cord injury; C: cervical; T: thoracic; L: lumbar; UDSMR: Uniform Data System for Medical Rehabilitation; FIM: Functional Independence Measure; ULMS: upper limb motor score. QIF: quadriplegia index of function; UULL: upper limbs and LLLL lower limbs; SD: standard deviation. * Transversal study; † longitudinal study; ‡ case-control study; § Evaluation at hospital admission and hospital release; ⎯ Does not apply
the same characteristics, capacities and qualities.\textsuperscript{20} The items that are considered redundant at the internal consistency evaluation must be grouped. This measure is carried out through Cronbach’s $\omega$, which, when $>0.7$, reflects a good internal consistency.\textsuperscript{20} The articles found in the literature search showed good internal consistency in the motor and cognitive dimensions and in the total FIM score.\textsuperscript{12,15}

Other models for the grouping of items from the FIM score (conceptual models) were proposed and organized with the objective of exploring certain domains, of which validity of content is evaluated by a group of specialists with clinical experience and who are involved in research.\textsuperscript{19} At the evaluation of the consistency of these models, a low correlation was observed between the items: LLLL dressing,\textsuperscript{24} stairs, comprehension, expression and memory\textsuperscript{12} and locomotion,\textsuperscript{27} which varied among the studies and depended on the level and severity of the lesion\textsuperscript{23} and the type of lesion (traumatic or non-traumatic).\textsuperscript{15}

The clinical usefulness of these divisions is the possibility of detecting functional changes that can occur in a set of items and not in others. For instance, individuals who are independent in self-care, but dependent in mobility receive the same motor score than those who are independent in both.

### Chart 1

<table>
<thead>
<tr>
<th>AUTHOR/YEAR</th>
<th>DESIGN</th>
<th>STATISTICAL ANALYSIS</th>
<th>REPRODUCIBILITY</th>
<th>RESULTS</th>
<th>RESPONSIVENESS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Middleton JW and cols.\textsuperscript{1/2} 1998*</td>
<td>N=167</td>
<td>ANOVA</td>
<td>-</td>
<td>Cognitive score no difference. Motor score higher at more caudal levels. Significant between C5 and C7-L. Self-care and sphincter management and mobility differentiates tetraplegics from paraplegics. Mobility differentiates lesions below and above T7. Locomotion: ceiling effect in gait/CR; floor effect in stairs.</td>
<td>-</td>
</tr>
<tr>
<td>Hall KM and cols.\textsuperscript{v} 1999*</td>
<td>N=3971 (admission); 4033 (hospital release); 903 (1 yr); 714 (2 yrs); 90 (3 yrs); 570 (5 yrs)</td>
<td>Spearman’s Coefficient of correlation; Mann-Whitney</td>
<td>Unspecified Coefficient of correlation; dressing UULL and LLLL</td>
<td>Good Correlation ($&gt;0.72$); transfers; eating and getting ready; Cognitive FIM: no difference. Motor FIM: higher score in lower lesions. Feeding, bathing, locomotion: tetraplegic different from paraplegic. Toilette: cervical different from lumbar. Sphincter management/shower transfer: lumbar differs from cervical and thoracic. Mobility: difference among groups.</td>
<td>Escera cognitiva: efecto techo. Escera motor MIF: Responsa aho o segundo año.</td>
</tr>
<tr>
<td>Ribeiro M and cols.\textsuperscript{1/2} 2004*</td>
<td>N=93 patients from HC-FMUSP and57 from Umarizal Rehabilitation Center; men: 72%; age: 33.8 ± 13.9; Traumatic SCI: 78.7%; mean time of lesion: 9.7 ± 13.6 months; lesion level: C: 30.5%; T: 52.7%; L: 17.0%.</td>
<td>Paired non-parametric tests</td>
<td>ANOVA</td>
<td>Good intra-observer reproducibility. Locomotion domain presents good correlation ($&gt;0.7$) with the scales: WISCI; RMI; BI; SCIM.</td>
<td>Responsive for total FIM and in items separately.</td>
</tr>
<tr>
<td>Morganti B and cols.\textsuperscript{1/2} 2005*</td>
<td>N=184 men and 100 women; Age: 50.4 ± 19.3 yrs; time of lesion: 56.9 ± 43.9 days; Traumatic: 50.3%; Non-traumatic: 44.7%; Level of lesion C: 81/284 Level C: 81; T1-T6: 50.4 ± 19.3 yrs; Time of lesion: 56.9 ± 43.9 days T: 148; L/S: 55. ASIA: A: 84; B: 19; C: 129; D: 52.</td>
<td>Spearman’s Coefficient of correlation</td>
<td>Good intra-observer reproducibility.</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Lugo LH and cols.\textsuperscript{1/2} 2007*</td>
<td>N=42; men: 88%; Age: 44 yrs; lesion level: C: 27.8%; T1-T6 = 40.9%, T6-L6 = 31.2%.</td>
<td>Kalmogorov-Smirnov</td>
<td>Mann-Whitney</td>
<td>T Test</td>
<td>FIM score increases according to lesion level and ASIA.</td>
</tr>
</tbody>
</table>

\textsuperscript{N} Number of individuals; SCI: spinal cord injury; C: cervical; T: thoracic; L: lumbar; FIM: Functional Independence Measure; UULL: upper limbs and LLLL lower limbs; WISCI: Walking Index for Spinal Cord Injury; RMI: Rivermead Mobility Index; BI: Barthel Index; HC-FMUSP: Hospital dos Clínicos of the School of Medicine of the University of Sao Paulo; h: hours; ASIA: American Spinal Cord Injury Association.  
\textsuperscript{*} Transversal study; \textsuperscript{†} longitudinal study; \textsuperscript{‡} case-control study; \textsuperscript{§} Evaluation at hospital admission and hospital release. - Does not apply
in the opposite situation, i.e., dependent in self-care, but independent in mobility. The regrouping of the items can increase the tool responsiveness.

II Validity

The validity is the property that evaluates the capacity of the tool to measure what it proposes to do. For instance, the assessment of functionality must include items such as personal care, mobility and locomotion. It must be related a particular situation, a proposal or a population in which the tool will be applied to.20

II.I Construct Validity

The construct validity evaluates the degree of compliance of a tool with the theory through the associations between relevant parameters.19,20

The studies that evaluated the construct validity of the FIM scale in spinal cord injury usually associated the obtained scores with the level of the lesion. It is expected that individuals with higher level lesions, with more severe involvement of the muscular strength of the upper limbs be more dependent, i.e., present lower scores than those that present a lower neurological involvement.

The comparison among studies is difficult to carry out, due to the lack of sample characterization. Additionally, the number of individuals is small and the stratification varies between: two groups (paraplegic and tetraplegic individuals);28 three (cervical, thoracic and lumbar);16 and four groups, based on the preserved musculature,6,17,29,30 or severity of lesion.72 The division of a small number of patients in groups can lead to a type II error in the analysis of results (not finding a difference when there is one).

The studies demonstrated a good construct validity in the assessment of the total score16,18,27,25,30 and the motor dimension,16,17 showing a difference among the patients, as individuals with higher lesions presented lower scores.

At the assessment of the cognitive dimension, all studies showed low construct validity, as no significant difference was found in the score between the paraplegic and tetraplegic individuals, regardless of the lesion severity.10,16,17 In this case, issues such as comprehension, expression, memory, problem-solving and social interaction are not relevant in the functional assessment, as these individuals maintain the superior brain functions intact, except when there is associated head trauma. Thus, they receive the maximum score, which characterizes the ceiling effect between the groups.

At the assessment of each item of the FIM scale, separately, most studies presented low sensitivity in the detection of differences between the groups in the locomotion domain (gait/wheelchair and stairs).15,17,28 The “climb stairs” item showed a floor effect in most studies, with a score of 1,12,17,28 and the gait/wheelchair item, a ceiling effect with a score of 6,17,28 In this last case, paraplegic individuals in manual wheelchairs as well as tetraplegic individuals in motor wheelchairs can receive the same score, in spite of their functional differences. Regarding the stairs, most individuals with spinal cord injury use the wheelchair exclusively for locomotion, which makes them dependent for this task. Only one study carried out in Brazil found a difference regarding the locomotion item between paraplegics and tetraplegics; however, no difference was observed in the score between those that presented thoracic lesion and the ones that presented lumbar lesion.16

The difference found might be due to the difficulty that the tetraplegic living in Brazil have of acquiring a motor wheelchair, thus receiving a lower score than the paraplegic individuals, as they present more difficulty in locomotion in a manual wheelchair.

At the assessment of the mobility, self-care and sphincter management, we observed a difference among the studies that can be attributed to the criterion adopted for the sample stratification.16,17,28

II.II Concurrent Validity

The concurrent validity can be called instrumental validity or criterion-related validity. It is used to demonstrate the accuracy of a tool through the comparison with the gold-standard.15,31 Two studies compared the FIM scale with other tools, presenting good correlation among them.1,31 Another study evaluated the concurrent validity of the FIM scale, comparing its score with hours of nursing care required by the patients per day presenting good correlation. That is, the individuals that need more hours of care are those that present low scores.32

III Responsiveness

The responsiveness is the capacity of a tool to measure changes in a pre-established period of time31. In the studies that evaluated the responsiveness of the FIM scale in spinal cord injury, the time interval between the hospital admission, hospital release and post-hospital release reassessment varied broadly and was not always described. Several studies observed a good responsiveness at the admission and the release.6,16,17,28 The responsiveness at three different times: admission, release and reassessment was also good for the total score after three months6 and for the motor dimension after one27 and two years16, regardless of the level and severity of the lesion. Ota et al.29 in a study carried out in Japan, identified that the responsiveness varies according to the level of the injury. However, this study evaluated a small number of patients in each group and did not consider an adequate statistical test to evaluate the possibility of the occurrence of type I error. Lugo et al.30 demonstrated that individuals ASIA A and B presented a ceiling effect one year and a half after the injury, whereas individuals ASIA C, D and E reached the maximum score within one year.

A low responsiveness was observed for the items transference in tetraplegic individuals28 and for the item locomotion and cognitive dimension, regardless of the level of the injury.28

One of the studies collected data from different samples, at three moments, which is not an adequate design for the assessment of responsiveness. The losses were considered, but the reason was not described.5

CONCLUSION

The studies found in the search presented different results regarding the validity and the responsiveness in the assessment of the total and motor score and of the 13 motor items evaluated separately. It is interesting to observe that most of the studies were developed in English-speaking countries and only one was carried out in Brazil by Riberto et al.16

Lawton et al.33 stated that the differences found in the results might be due to factors such as the degree of difficulty presented by the patients when performing the tasks, which varies among populations from different countries; differences in translation, in the training given to those applying the scale and the versions used in the assessment.

Questions related to the translation and training influence the
psychometric quality of the tool, decreasing its reproducibility and making comparisons difficult. However, a quantitative review on this subject, comparing several studies with different populations, showed good equivalent reproducibility, as well as intra and interobserver reproducibility.

The floor effect was observed in both the validity and the responsiveness assessment and the ceiling effect in some items. The observation of these effects is the result of the lack of precision and capacity of a tool to detect significant changes in the functional levels of individuals located at the extremities. If an individual has a maximum or a minimum score in one measurement, the occurrence of improvement or worsening, respectively, cannot be identified. It is important to mention that there are methodological restrictions in the data analysis, as an adequate statistical test was not always used. Many studies used parametric tests to analyze the ordinal variables, with a small sample and distribution that differs from the normal one. In these cases, the use of non-parametric tests is indicated.

In general, the literature review allowed us to verify that the FIM scale presents high internal concordance and good interobserver reproducibility. The scale presented good results regarding the validity and responsiveness.

The FIM is a generic tool, created to evaluate functional independence through its total score. Although the fragmented analysis of this tool is not formally recommended by the UDSMR, the division of the scale into motor and cognitive dimensions has been often employed in the literature. Many studies evaluated the properties of different combinations of FIM items and most of them observed low validity and responsiveness, in the locomotion domain as well as in the cognitive dimension. Individuals with spinal cord injury can present cognitive impairment, depending on the etiology of the injury or comorbidities and, but for exceptional cases, they present locomotion impairment, which is broadly variable according to the lesion characteristics. Therefore, if one wants to focus on one of these aspects, other tools must be used, as the FIM scale, although largely employed, presents low sensitivity in this context. One alternative would be to maintain the use of this scale, which is broadly used, and aggregate items that can increase its sensitivity in certain situations.

REFERENCES