Measures of physical functioning after hip fracture: construct validity and responsiveness of performance-based and self-reported measures

I. FARAG1, C. SHERRINGTON1, S. J. KAMPER1, M. FERREIRA1, A. M. MOSELEY1, S. R. LORD2, I. D. CAMERON3

1George Institute for Global Health, University of Sydney, Sydney, Australia
2Neuroscience Research Australia, University of New South Wales, Sydney, Australia
3Rehabilitation Studies Unit, Sydney Medical School, University of Sydney, Sydney, Australia

Address correspondence to: I. Farag. Tel: +61 2 8238 2440; Fax: +61 2 9657 0301. Email: ifarag@georgeinstitute.org.au

Abstract

Objectives: this study aimed to investigate the construct validity and responsiveness of performance-based and self-reported measures of strength, mobility and balance after hip fracture.

Design: secondary analysis of clinical trial data.

Subjects: a total of 148 older people undergoing hip fracture rehabilitation.

Methods: correlation coefficients assessed construct validity. Internal responsiveness was assessed by calculating effect sizes (ES) I and II. Area under the receiver operating characteristic curve (AUC) assessed external responsiveness with change in EuroQol as the reference.

Results: correlations between performance-based and self-reported measures were small to medium (strength r=0.17, mobility r=0.45 and balance r=0.37). The most responsive performance-based measures included walking speed (ESI 1.7, ESII 1.2), Physical Performance and Mobility Examination (ESI 1.3, ESII 1.0) and chair-rise test (ESI 1.1, ESII 0.8). Self-reported mobility (ESI 0.8, ESII 0.6) and strength (ESI 0.8, ESII 0.6) were more responsive than self-reported balance (ESI 0.3, ESII 0.2). External responsiveness (AUC) was greatest for walking speed (0.72) and lowest for the measures of body sway (0.53).

Conclusion: self-reported and performance-based indices appear to assess different constructs and may provide complementary information about physical functioning in people after hip fracture. Measures of strength and mobility showed greater ability to detect change than measures of balance.

Keywords: psychometrics, ageing, hip fracture, self-reported questionnaires, older people

Introduction

There is an array of tools available to measure physical functioning following hip fracture. The choice of a tool relies on knowledge of the properties of the tool and the nature and type of information it provides. The standards against which the utility of the tool is measured include its validity and responsiveness. Validity is largely an indicator of the degree to which a tool measures what it is designed to measure [1]. Responsiveness is the measurement tool’s ability to detect change. There are two aspects to responsiveness; the first aspect has been identified as ‘internal responsiveness’ and is defined as the ability of a tool to detect change over a pre-specified time frame, whereas the second aspect refers to the extent to which change in a measure relates to a corresponding change in a reference measure and has been identified as the ‘external responsiveness’ [2]. The use of a measurement tool with poor responsiveness, in a clinical trial, may lead to underestimation of the effects of intervention [3]. Measurement methods can be classified into two broad categories; self-reported tools that rely on the patients’ estimation of their health status and physical performance tools requiring assessment of physical parameters such as muscle strength, balance and mobility. The increased utilisation of
self-reported questionnaires encourages the active participation of the patient in setting the goals of treatment and in defining what constitutes ‘worthwhile’ intervention [3, 4].

The literature is mixed with regard to the level of concordance between self-reported and performance-based measures. Some studies have found significant associations between them [5] and found both to be important indicators of health in old age as well as powerful predictors of mortality. However, others have found discordance between them [6], which in part could be due to the measurement of different constructs.

The aim of this study was to investigate the construct validity and ability to detect change (responsiveness) of performance-based and self-reported measures of strength, mobility and balance after hip fracture. Both ‘internal’ and ‘external’ responsiveness were investigated.

Method

Participants

Data used for this study were obtained from an assessor-blinded, randomised controlled trial [7] investigating the difference between the intervention group’s higher-dose (60 min/day) exercise programme conducted while standing and the control group’s lower-dose (30 min/day) programme conducted in sitting or supine. The programme was conducted over a 16-week period. One-hundred and sixty participants with surgical fixation for hip fracture were recruited from hospital rehabilitation units in Sydney, Australia.

Measures

Among a battery of measurement tools, the following objective assessments were conducted at baseline and 16 weeks after randomisation: knee-extensor strength, walking speed, Physical Performance and Mobility Examination (PPME), chair-rise test, maximum balance range, step test, body sway and lateral stability. Self-rated mobility, strength and balance were also assessed and a utility-based quality-of-life survey (the EuroQol 5D) was utilised. The results of the trial [7] did not indicate a difference between the groups for the two primary outcome measures of walking speed and knee-extensor strength in the fractured leg at 16 weeks. There was a significant between group difference favouring the intervention group on the chair-rise test (mean difference: 0.04; 95% confidence interval (CI): 0.01–0.08).

Test descriptors

Knee-extensor strength was measured with the participant sitting with the hips and knees at 90° of flexion and using a spring gauge fixed to a crossbar behind the subject [8]. Walking speed was measured over a 6 m distance using a stopwatch [9]. The PPME consists of six functional activities (including bed mobility, transfer skills, timed multiple stands from a chair, full tandem standing balance, step test and a timed ambulation test), with each scored as 0 (fail), 1 (low pass) or 2 (high pass) and added to give a total score out of 12 [10]. The step test consists of the number of times a subject can step onto a 5 cm block in 15 s [11], and the chair-rise test consists of sit-to-stand-to-sit five times from a 45 cm seat height. Balance tests included measures of postural sway and the maximum balance range; measured using a swaymeter attached to the participant’s waist and measuring body displacement in an anterior/posterior direction and laterally [8, 12].

The self-rated measures were scored on 5-point Likert scales (very poor, poor, fair, good and very good). Participants were asked to rate themselves on their leg strength, mobility and balance [7]. The EuroQol 5D consists of five dimensions (mobility, personal care, usual activities, pain/discomfort and anxiety/depression). The responses are rated at three levels of severity; no problem, some/moderate problem and extreme problems [13].

Statistical methods

According to the COSMIN taxonomy (Consensus-based Standards for the selection of health Measurement INstruments), one of the aims of assessing construct validity is to determine the internal relationship between tools [14]. Pearson correlation coefficients were therefore calculated to assess the construct validity between performance-based and self-reported measures at programme completion (16 weeks from programme onset). Comparisons were made between self-rated strength and knee-extensor strength and the chair-rise test; self-rated mobility and the PPME and walking speed; and self-rated balance and the maximum balance range, body sway, step test and lateral stability. Correlation values between 0.1 and 2.9 were considered to be small, 0.3 and 0.49 were considered to be medium and 0.5 and 1.0 were considered to be large, or showing a strong relationship [15].

To assess internal responsiveness, two methods of measuring effect size were calculated. Effect size I (ESI) expresses the mean change score relative to the variability in the baseline data (ESI: change/baseline SD) and effect size II (ESII) expresses the mean change score relative to the variability of the change scores (ESII: change/SD of change). Values of 0.2, 0.5 and 0.8 or greater were used to represent the lower boundaries of small, moderate and large responsiveness, respectively [2]. ESI and ESII were calculated for the total cohort and separately for the intervention and control groups to investigate the impact of intervention on the responsiveness statistic. With the calculation of effect size only the baseline standard deviation of those subjects who completed the trial was used as change scores were utilised in the calculations.

External responsiveness was measured using the area under the receiver operating characteristic curve (AUC) for performance-based and self-reported measures, with change in utility-based quality of life (EuroQol) as the reference standard. Improvement in EuroQol was dichotomised at 0.5.
SD baseline scores, a suggested approach for estimating the smallest detectable change [16]. The AUC ranges from 0.5 (no accuracy in discriminating improved from unimproved participants) to 1.0 (perfect accuracy) [17]. The statistical significance of the difference in the AUC for the various measures of strength, mobility and balance was calculated.

Results

Participant characteristics are shown in Table 1. Four-fifths of the 160 participants were female and 34% were assessed as having cognitive impairment, defined as three or more adjusted errors on the Short Portable Mental Status Questionnaire [18]. One-hundred-and-forty-eight people completed the 16-week assessment, 72 in the high-dose exercise intensity group and 76 in the low-dose exercise intensity group.

Construct validity

Correlations between performance-based and self-reported measures were small to medium (see Table 2). There was a small correlation between measurement of knee-extensor strength and self-rated strength. Medium correlations were obtained for the measurement of walking speed and PPME and self-estimated assessment of mobility. There were also medium correlations between self-rated balance and three of the performance-based tests for balance, with the step test yielding correlation values of 0.37 and lateral stability yielding a small inverse correlation (−0.14).

Table 1. Participant characteristics (n = 160)

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Mean (SD) or n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>130 (81%)</td>
</tr>
<tr>
<td>Age at injury (years)</td>
<td>84 (8)</td>
</tr>
<tr>
<td>Height (cm)</td>
<td>159.60 (10.20)</td>
</tr>
<tr>
<td>Using walking aid</td>
<td>90 (56%)</td>
</tr>
<tr>
<td>Presenting with cognitive impairment (%)*</td>
<td>54 (34%)</td>
</tr>
</tbody>
</table>

*≥3 adjusted errors on the Short Portable Mental Status Questionnaire.

Table 2. Construct validity of performance-based and self-reported measures of strength, mobility and balance at programme completion

<table>
<thead>
<tr>
<th>Self-reported measure</th>
<th>Performance-based measure</th>
<th>Correlation (n = 148)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-reported strength</td>
<td>Knee-extensor strength</td>
<td>0.17</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>Chair-rise test</td>
<td>0.22</td>
<td>0.01</td>
</tr>
<tr>
<td>Self-reported mobility</td>
<td>Walking speed</td>
<td>0.40</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>Physical Performance and Mobility Examination</td>
<td>0.45</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Self-reported balance</td>
<td>Maximum balance range</td>
<td>0.33</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>Step test</td>
<td>0.37</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>Body sway</td>
<td>0.31</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>Lateral stability</td>
<td>−0.14</td>
<td>0.13</td>
</tr>
</tbody>
</table>

Internal responsiveness

For the total cohort, the most responsive performance-based measures, with at least one large effect size (0.8 or greater), were walking speed (ESI 1.60, ESII 1.10); PPME (ESI 1.24, ESII 0.86); chair-rise test (ESI 1.12, ESII 0.74) and step test (ESI 1.0, ESII 0.7). The most responsive self-reported measures were mobility (ESI 0.81, ESII 0.63) and strength (ESI 0.81, ESII 0.55). Lower values were generally obtained for ESII comparable with ESI. As expected, only the measure that showed a significant between-group difference in the trial [7] (chair-rise test) reported a difference in ESII for the two intervention groups, there was no marked between-group discrepancy in ESII for the other measures. Baseline measures, 16-week results, change scores, ESI and ESII for the intervention and control groups and the cohort in total are shown in Table 3.

External responsiveness

The AUCs for performance-based measures were greatest for walking speed 0.72 (95% CI 0.65–0.81), chair-rise test 0.66 (95% CI 0.57–0.75) and maximum balance range 0.65 (95% CI 0.55–0.75). For the self-reported measures the AUC for strength was 0.67 (95% CI 0.59–0.76); mobility 0.59 (95% CI 0.50–0.68) and balance 0.55 (95% CI 0.46–0.64). There was a statistically significant difference in the external responsiveness of performance-based and self-reported measures for mobility (P = 0.03), but not statistically significant for strength (P = 0.18) and balance (P = 0.16).

A sensitivity analysis was conducted evaluating the correlation of the baseline data. The results were similar or lower at baseline. Similarly, an analysis conducted for the non-cognitively impaired population provided similar correlation values as for the total cohort.

Discussion

This study found the correlations between performance-based and self-reported measures to be small to medium in people after hip fracture. We therefore suggest that performance-based and self-reported tools appear to measure different constructs about physical functioning in this population.

The weakness of the association between the types of measures is understood when considering the factors that impact on the assessment of health. When a patient is asked to estimate health status or evaluate the effect of intervention, it is unlikely that they do so from a purely physical perspective. Their perception may be influenced by psychological, emotional and environmental factors. Furthermore, retrospective estimation of health status is likely to be influenced by memory, attitude and mood [19]. For example, it is acknowledged that with the global measures of self-reported
Table 3. Descriptive data, internal and external responsiveness for the low- and high-intensity exercise groups and the total cohort

| Test                      | Baseline mean (SD) | 16-week mean (SD) | Difference between baseline and 16-week score mean (SD) | AUC
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Knee-extensor strength (kg)</td>
<td>7.04 (3.65)</td>
<td>9.94 (4.59)</td>
<td>2.71 (4.68)</td>
<td>0.82</td>
</tr>
<tr>
<td>Chair-rise test (number/s)</td>
<td>0.15 (0.08)</td>
<td>0.25 (0.12)</td>
<td>0.09 (0.12)</td>
<td>0.83</td>
</tr>
<tr>
<td>Walking speed (m/s)</td>
<td>0.29 (0.19)</td>
<td>0.63 (0.31)</td>
<td>0.33 (0.28)</td>
<td>0.84</td>
</tr>
<tr>
<td>PPME</td>
<td>7.04 (1.75)</td>
<td>9.34 (2.16)</td>
<td>2.26 (2.29)</td>
<td>0.55</td>
</tr>
<tr>
<td>Maximum balance range test (mm)</td>
<td>86.31 (36.73)</td>
<td>106.23 (44.23)</td>
<td>45.19 (208.00)</td>
<td>0.26</td>
</tr>
<tr>
<td>Step test</td>
<td>0.68 (0.87)</td>
<td>5.89 (4.86)</td>
<td>5.21 (7.10)</td>
<td>0.34</td>
</tr>
<tr>
<td>Body sway (cm)</td>
<td>2.91 (0.94)</td>
<td>3.98 (0.94)</td>
<td>0.97 (1.13)</td>
<td>0.59</td>
</tr>
</tbody>
</table>

ESI, effect size I, calculated as the difference between 16-week and baseline scores/baseline SD.
ESII, effect size II, calculated as the difference between 16-week and baseline scores/SD of the difference between 16-week and baseline scores.

health the perception of health status remains stable after a certain age, whereas the observer-measured functional status may deteriorate [20]. The reverse may also hold, however, in that following a major life event such as a fall, leading to hip fracture, self-perception of deterioration may endure, despite recovery.

Or it may be that performance-based measures are assessing a defined aspect of function, such as ‘knee extensor strength’, which may be regarded as a simplification of the abilities required for ‘performance in daily activities’. Self-reported measures, on the other hand, provide a more global indication of the capacity to perform functional tasks. Performance-based measures may also reflect performance at a single point in time only [21]; and that point may not provide the most accurate indicator of ability; Guyatt et al. [22], for example, found that ‘encouragement’ impacted significantly on the test results.

Performance-based measures are supposedly more objective as trained professionals assess the degree of physical function via observation. However, while health care providers evaluate function based on established criteria that are shown to correspond to important changes [23], they tend to underestimate functional disabilities [6]. Further, such an assessment method is not designed to measure the total human experience of the disability.

There was variable internal responsiveness of the individual tools, but, on the whole, moderate-to-high values were found for strength and mobility and lower values for balance; both in the performance-based measurement and in self-report. For this older age group in particular, fear of falling may influence the assessment of function, possibly impacting on the results of the performance-based tools but also on the patients’ perception of improvement. This may account for lower responsiveness of the self-rated balance assessment tool.

The results for the different statistics used to measure internal responsiveness (ESI and ESII) correspond well. In general, lower values were obtained for ESII which may be regarded as a simplification of the abilities required for a defined aspect of function, such as ‘knee extensor strength’.

The responsiveness of the tool should not be regarded as an independent characteristic of the tool and responsiveness of the tool should be interpreted in the context of the treatment provided [2]. Our results indicate larger responsiveness values for the high-intensity exercise group compared with the low-intensity exercise group for the chair-rise test, where there was a significant difference in group difference [7], thus confirming that it is difficult to separate the concept of the responsiveness of the tool from the treatment context in which the tool is utilised. Although we have shown that several of the tools used are in fact responsive for this population group and in the context of the treatment provided, there is no suggestion that this can be generalised to other interventions, populations or even other body parts.

Studies on construct validity and responsiveness of the tools used for this population group are hampered by
Measures of physical functioning after hip fracture

Acknowledgements

We thank all those involved in the clinical trial: the Enhancing Mobility After Hip Fracture Research Group staff, participants, relatives and staff at the participating hospitals. This research was undertaken at The George Institute for Global Health, affiliated with the University of Sydney.

Conflicts of interest

S.R.L. is a company director of Balance Systems Inc., which makes equipment items used in the assessment (knee-extension strength, maximal balance range and body sway) which are commercially available through the Prince of Wales Medical Research Institute. All other authors have nothing to declare.

Funding

The study was supported by a project grant from the National Health and Medical Research Council (NHMRC), Australia. Professors Lord and Cameron, A/Prof Sherrington and Dr Kamper receive salary support from the NHMRC. The financial sponsors played no role in the design, execution, analysis and interpretation of data or writing of the study.

References


Conclusion

Self-reported and performance-based indices appear to assess different constructs and may provide complementary information about physical functioning in people after hip fracture. Moderate-to-high values of internal responsiveness were found for self-rated strength and mobility and lower values were found for balance. External responsiveness of performance-based measures was higher than for self-reported measures for mobility but not for balance or strength.

Key points

- Self-reported and performance-based indices assess different constructs.
- Performance-based measures tended to be more responsive than self-reported measures.
- Measures of strength and mobility are more responsive than measures of balance.

several issues, however. In the evaluation of responsiveness while ESI and ESII were investigated in this study, the most appropriate measure of responsiveness remains a matter of debate. In fact, the use of the same criteria for the different methods of evaluation suggests a lack of clarity in the definition [2]. Furthermore, Gatchel et al. [24] asserts that there is no consensus in the literature as to what constitutes an adequate external anchor when assessing external responsiveness. Ideally, the reference measure would be a gold-standard measure; but no such measure for evaluation of physical function following hip fracture exists. We chose EuroQol as it comprises several domains of importance to patients, has been extensively investigated and has well-established reliability and validity [19, 25–27]. The fact that EuroQol is a general quality-of-life measure and by design encapsulates more than just physical function is likely to explain the moderate association between our reference standard and the physical-function measures. With this limitation in mind, the results suggest that mobility, in this case the measure of walking speed, is more likely to identify those who have improved as measured by the EuroQol, when compared with the measures of balance and strength.

The results of this study support the combined use of performance-based and self-reported tools in the evaluation of function following hip fracture, acknowledging that different and complementary information may be provided by each. The combined use of the tools leads to better discernment of important improvement and identifies the nature of that improvement [28]. In addition, determination of the responsiveness of the tool allows for informed decisions about the choice of assessment tools for use in research and in the clinical setting for this population group.

Received 29 November 2011; accepted in revised form 17 May 2012