|  |
| --- |
| Effects of Balance Training on Functional Independence Measure (FIM) scores |
| Role of psychometric properties in interpretation of FIM scores |
| Cathryn Ghena SPT Outcome Measures University of North Carolina Department of Physical Therapy |

11/14/2012

Article 1: Eser F, Yavuzer G, Karakus D, Karaoglan B. The effect of balance training on motor recovery and ambulation after stroke: A randomized controlled trial. *Eur J Phys Rehabil Med*. 2008;44(1):19-25.

Article 2 (Psychometric assessment): Stineman MG, Shea JA, Jette A, et al. The functional independence measure: Tests of scaling assumptions, structure, and reliability across 20 diverse impairment categories. *Arch Phys Med Rehabil*. 1996;77(11):1101-1108.

Article 3 (Reliability assessment): Ottenbacher KJ, Hsu Y, Granger CV, Fiedler RC. The reliability of the functional independence measure: A quantitative review. *Arch Phys Med Rehabil*. 1996;77(12):1226-1232. doi: 10.1016/S0003-9993(96)90184-7.

Article 4 (Responsiveness assessment): Beninato M, Gill-Body KM, Salles S, Stark PC, Black-Schaffer RM, Stein J. Determination of the minimal clinically important difference in the FIM instrument in patients with stroke. *Arch Phys Med Rehabil*. 2006;87(1):32-39. doi: 10.1016/j.apmr.2005.08.130.

**Article 1 Summary**

After suffering a stroke, many aspects of health and wellness have been potentially impacted. While the effects of the stroke will vary person to person based on the type, severity, and lesion location, studies have shown that there are a few factors that can be reliable predictors of rehab potential i.e. digit extension and shoulder shrug. Previous studies cited in this literature review have suggested that balance abilities post stroke may also be a predictor of rehab potential. Specifically, increased balance post stroke seems to have a significant correlation with: 1) length of stay in an inpatient facility, 2) locomotor function, and 3) functional abilities. If the ability to achieve good balance does in fact have such correlations, then it makes sense to be very intentional in the inclusion of balance training during post-stroke inpatient rehab. However, there are many ways in which balance can be incorporated into rehab. The purpose of this study was to look at the effect of a very specific balance protocol—force plate biofeedback balance training—on recovery of hemiparetic patients post stroke.

In order to assess the effect of biofeedback balance training on stroke recovery, a randomized controlled trial (RTC) with 41 patients was conducted. Demographics are as follows: men=25 women=16, average age=60.9, average time since stroke=6mo, right hemiparesis=15, left hemiparesis=26, ischemic stroke=31, and hemorrhagic stroke=10. All patients presented with a resulting hemiparesis. Inclusion criteria required that all patients 1) had the ability to understand and follow simple verbal instructions, 2) were ambulatory prior to the stroke, 3) could currently stand and take at least one step with or without assistance, 4) had no medical contraindications to walking, and finally 5) this was their first episode of unilateral stroke with resulting hemiparesis. The exclusion criteria stated that patients were not eligible to be participants in this study if they had 1) a prior history of any type of neurological pathology, 2) any type of prior condition affecting balance, neglect, impairing vision, or conscious levels, or 3) another medical condition unrelated to the stroke that effected the lower extremity. Such criteria pulled in a very specific post stroke population while still ensuring they would be able to participate in the intervention. The exclusion criteria did a good job of making sure that the balance impairments would not be a result of anything besides the stroke. These participants were randomly assigned to two groups with the control group n=19 and the experimental group n=22. Both groups participated in a ‘conventional’ post stroke rehab program for 8weeks which consisted of individualized rehab based on patient needs i.e. physical therapy (ROM, strengthening, NDT, PNF, positioning, PRE, etc.), occupational therapy, and/or speech therapy. The experimental group then had additional balance training on a Nor-Am Target Balance Training System 15minutes a day, 5days/wk for 3weeks. The subjects had to stand on the force plates, but were able to receive assistance either from a device or person if necessary in order to remain safe. These supplement sessions required the patient to respond to different tasks requiring weight shifting in all directions. They were able to watch a line on the screen that represented their center of mass move towards different targets.

In order to measure change, assessments were done at the start and finish of the 8 weeks. There were several different outcome measures used in this study to assess the progress made by both groups. Brunstromm staging was used to assess changes in motor recovery, the Rivermead Mobility Index (RMI) was used to assess mobility, and the Functional Independence Measure (FIM) was used to assess changes in functional activity level. Both groups significantly improved in all measures at the end of treatment, suggesting the effectiveness of the conventional rehab. However, the experimental group with their biofeedback balance training did not improve significantly more than the control group. This suggests that while it does not seem to have detrimental effects, supplemental balance in the form of biofeedback may not have as great as an effect on rehab prognosis as the literature suggests. However, that doesn’t mean balance should be completely neglected either. In fact, because there are so many ways to address balance, it would make sense to do further research. Considering the population, it may make sense that instead of doing biofeedback, the balance training could have consisted of more task specific training. Practicing tasks (perhaps self-care, mobility, etc. for example) in a way that challenged balance would not only be more functional than biofeedback, but may also lead to more significant improvement. It is also important to take into consideration the amount of practice that is necessary for this population. 15minutes for 15 sessions is not a lot of practice; perhaps a significant change would have been seen with a longer duration of biofeedback training. Regardless, the role of balance training in a post-stroke inpatient population is something that deserves further research.

**The Psychometric Properties of the FIM**

*Material for Table 1 was collected from articles 2, 3, and 4. These are separate studies that looked into the psychometric properties of the FIM. Most of the material is gathered from article 2 unless stated otherwise.*

**Table 1**

|  |  |
| --- | --- |
| Type of patients considered | 93,289 pts all of whom had been discharge from 252 rehabilitation hospitals across the country; all of these patients were then classified into one of twenty categories in agreement with the Functional Related Groups including  Neurological pts: stroke, non-traumatic brain injury, TBI, non-traumatic SCI, traumatic SCI, Guillain-Barre, general neuro  Musculoskeletal pts: LE fracture, joint replacement, other orthopedic, lower limb amputation, other amputation. OA, RA  Miscellaneous pts: cardiac, pulmonary, pain, major multiple trauma (MMT), MMT with brain/spin, other.  Stroke, LE fracture, and joint replacement were most prevalent in this study.  Only hospitals that had FIM credential criteria were allowed to participate in this study. |
| Type of patients excluded | Pts were excluded if:   * there was any missing data in their chart * data irregularities (i.e. miscoded values) * atypical stays (i.e. unusually long or short, admitted eval only, or death) * >16y.o. |
| Measurement level | Ordinal data  The FIM is made up of 18 items, all of which are score on an ordinal scale from 1 to 7 with 1 indicating dependence and 7 indicating independence. |
| Content validity | While most health care professionals can technically perform the FIM, there is a push to be properly trained and certified before administering. The fact that the FIM is recognized by CMS has been largely responsible in such accreditation. This specific study required that all rehabilitations hospitals had to require FIM credentialing criteria. This means all administered FIM measures were done according to a specific set of standards, helping to increase content validity. |
| Test-retest reliability | Ottenbacher KJ, Hsu Y, Granger CV, Fiedler RC. The reliability of the functional independence measure: A quantitative review. *Arch Phys Med Rehabil*. 1996;77(12):1226-1232. doi: 10.1016/S0003-9993(96)90184-7.  The FIM is commonly used to compare changes in functional abilities from time of admittance until the time of discharge. The test-retest reliability, however, was not addressed in this study. However, a supplemental article by Ottenbacher, a coauthor on the article used for the rest of this chart, reviewed 11 studies all of which reported their gathered reliability of the FIM. These studies used the interclass correlation coefficient (ICC) to compute reliability. The mean interrater reliability across all 11 studies was 0.95 for the total FIM. Reliability of the subscales as well as individual FIM items were also computed . The subscales reliability ranged from0 0.78 (social cognition) to 0.95 (self-care) while the individual tasks ranged from 0.61 (comprehension) to 0.90 (toilet transfer). This suggests that the motor components of the FIM have a bit more reliability than the cognitive items.  The studies used in this article consisted of a varying patient population, settings, and of course raters. The total FIM’s test-retest reliability of 0.95 shows that this measure has acceptable reliability despite being used on such a wide population in a wide variety of environments. This is one of the reasons that the FIM is so widely used today. |
| Internal consistency | Cronbach’s alpha was used to determine the homogeneity for each impairment category for the total FIM, the motor-FIM, and the cognitive-FIM. The lowest internal consistency was a Cronbach’s alpha of 0.88, suggesting that there is good internal consistency for not only all impairments but also all versions of the FIM.  Example:   |  |  |  |  | | --- | --- | --- | --- | | Impairment | 18-Items | Motor-FIM | Cognitive-FIM | | Stroke | 0.94 | 0.93 | 0.93 | |
| Construct validity – hypothesis | Based on previous studies, it was hypothesized that all 18 items of the FIM played an important role in providing necessary information about a patient’s functional independence and thus their burden of care. Construct validity was determined by assessing range of item means and standard deviations. This was done for all FIM items and then compared across impairment category. It was also hypothesized that the range of item means would be similar across each impairment category.  “Item response distributions and standard deviations should be roughly symmetrical in items measuring the same construct.”  “Each item in a hypothesized group (motor-FIM vs. cognitive-FIM) should be substantially linearly related (*r*>.40) to the total score compute from other items in that group with that item removed.” |
| Construct validity – collection | All analyses were run twice to help determine item level analysis. Essentially they were looking to see if the 7-level ordinal scale could be collapsed into more concise levels. When such analyses were done, there was no gain suggesting that the 7 levels is necessary for maintaining good clinical meaning.  Item response variability: 360 item-impairment combinations. Standard deviations ranged from 0.9 to 2.6. This suggests that the questions included in the FIM truly help to determine functional levels.  Item total correlations: 90% of the motor-FIM items *r* ranges from 0.40 to 0.91 which meets the hypothesis that the items of the subtests were substantially linearly related to the total score. All test items help to provide a patient’s level of function.  Discriminant validity: all items in the motor and cognitive subscales were more highly correlated with their own scale than with the other. This was true for all impairments as well as all test items. This means that the tasks are appropriately placed in the correct subcategory in order to identify functional level and that there were no distracting questions.  Overall, this study as well as many others have determined that the FIM has good validity making it a valuable and readily used outcome measure across the globe. |
| Responsiveness – hypothesis | Article 4 (Responsiveness assessment): Beninato M, Gill-Body KM, Salles S, Stark PC, Black-Schaffer RM, Stein J. Determination of the minimal clinically important difference in the FIM instrument in patients with stroke. *Arch Phys Med Rehabil*. 2006;87(1):32-39. doi: 10.1016/j.apmr.2005.08.130.  In order to assess responsiveness, it must be determined who really changed and who really did not change. This was not reviewed in the original psychometric article by Stineman, but has been heavily researched by other authors. In a study by Beninato, the authors were looking to determine what level of increase in FIM scores was necessary in order for the change to be classified as significant. Once again, because the FIM is so readily used across the globe, it makes sense that this MCID number has been heavily researched. In this study, the authors set out to determine MCID based on an external criterion (physician perceived change) compared to change in FIM scores. The authors hypothesized that they would find similar MCIDs as found in other studies. They were also looking to see the effect of FIM scores ad admission on MCID (i.e. how does an initial low FIM score at admission effect MCID?) |
| Responsiveness – collection | In order to determine the MDIC, 4 attending physicians, all of whom were board certified in physical medicine and rehabilitation with specializations in stroke rehab as well as between 3-14years of experience, were used as the external criterion of pt change. A 15 pt likhert scale with -7 (a very great deal worse) to 7 (a very great deal better); they were blinded from the FIM scores of the pts. The change of the pts based on this scale was then analyzed along with admission and d/c FIM scores in SPSS. Significant correlations were found between the physicians’ assessments and changes in the total FIM (r=.44, P<.001) as well as between physician’s assessments and changes in motor-FIM (r=.51, P<.001).  By determining area under the ROC curve, sensitivity, specificity, positive likelihood ratio, and negative likelihood ratio (all with a 95% confidence interval), this study was able to determine an MCID value for the total FIM, motor-FIM, and cognitive-FIM.  Total-FIM=22  Motor-FIM=17  Cognitive-FIM=3      Effect size:  Total FIM= 0.82; motor-FIM=0.91; cognitive-FIM=0.61 |
| Floor and ceiling effects? | No items had high outliers so no evidence of ceiling effects.  The stair item had low outliers as well as small standard deviations suggesting that this item had a floor effect across multiple patient types. This makes sense because if you can’t walk then you can’t do stairs (i.e. LE amputation, SCI, stroke, etc.) |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Experimental Group** | | | | **Control Group** | | | |
|  | *Pre-Tx* | *Post-Tx* | *Z* | *P Value* | *Pre-Tx* | *Post-Tx* | *Z* | *P value* |
| **FIM Score** | 81.4 +/- 16.0 | 92.1+/- 12.2 | -3.9 | 0.001 | 85.4+/-20.3 | 96.9+/-16.0 | -3.6 | 0.001 |

***Statistical Significance of pre and post Total-FIM scores* Table 2**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Pretest | Post-test | Difference | Exceed MCID? |
| Control | 85.4+/-20.3 | 96.9+/-16.0 | -11.5 | No |
| Intervention | 81.4 +/- 16.0 | 92.1+/- 12.2 | -10.7 | No |
| Difference | - 4 | -4.8 | 0.8 | No |

***Clinical Significance of pre and post Total-FIM scores* Table 3**

***Breakdown of FIM MDIC*  Table 4**

|  |  |
| --- | --- |
| *Test* | *MDIC* |
| Total-FIM | 22 |
| Motor-FIM | 17 |
| Cognitive-FIM | 3 |

*Interpretation of the Eser et al. article “The effect of balance training on motor recovery and ambulation after stroke” based on the psychometric properties of the FIM.*

This original study looked at the effect of biofeedback balance training on a post-stroke population. Three different measures were used to assess the level of change in the subjects—Brunnstrom stage, the Rivermead Mobility Index (RDI), and the Functional Independence Measure (FIM). These measures were trying to track changes from admission to discharge in both the control and experimental groups. No significant changes occurred in either group according to the Brunnstrom staging. However, significant changes from admission to d/c, not between groups (significant defined as P <0.001), occurred in the control and experimental groups according to both the RDI and FIM (see table 2). The authors interpreted this information saying that their conventional stroke rehab was able to make significant changes in functional ability from day of admission to day of discharge, but there was no added benefit from incorporating biofeedback balance training. When I first read this article, I agreed with the results and though that it would be a good idea to conduct further research trying different types of balance training. While I still think that this is a good idea, further investigation of the FIM has caused me to ask several more questions about this study. According to Beninato et al., the MDIC of the total-FIM states that there must be an increase in FIM score of 22-points in order for the change to be considered clinically significant (see Table 4). The study by Eser et al. certainly reported a *statistically* significantchange in FIM scores, however they did not report a *clinically* significant change. In this study, all of the point changes were lower than what the MCID required for a progression to be considered clinically effective (see table 3).

From a research point of view, it is usually good for your study to find these statistically significant changes. The fact that there was no statistically significant difference between the control and experimental groups in this study means that biofeedback balance produces no change and they should keep researching. If there had been a significant difference (so biofeedback balance actually improved level of function more than conventional rehab), then we would want to see if the changes were *clinically* significant in accordance with the MDIC. If the patients had improved more than 22-points in the total-FIM as a result of biofeedback balance, then we would want to think about using this as an intervention with our patients. The role of the MDIC is really to make what we find both in research and from our outcome measures more practical. An intriguing definition of MDIC found in Beninato et al. stated “[MDIC] is the smallest difference in score in the domain of interest which patients perceive as beneficial and which would mandate, in the absence of troublesome side effects and excessive cost, a change in the patient’s management.” As researchers, the p value might tell you a lot, but as therapists we should really be interested in the MDIC because this value gives us insight into how we can adjust our intervention and management in an effort to give our patients the best quality of care possible. The fact that even the conventional protocol for the control group didn’t have a big enough change to satisfy the FIM’s MDIC causes me to question what activities are involved. It would be hoped that that specific inpatient hospital’s rehab protocol for stroke patients would be able to guide the patients towards making clinically significant changes. Other questions to ask include how much time spent in therapy is expected before you meet the MDIC? What types of activities help guide you on this path? Of course this is all variable patient to patient.

While the changes in FIM score were not clinically significant in this study, it was appropriate to use the FIM in this population. Extensive research has been done validating the FIM for post-stroke patients in an inpatient setting. In fact, it is appropriate to use the FIM in a myriad of populations because of the extensive research and validation that has been done on it. According to Stineman et al, in 1996 the FIM was used in more than 60% of rehabilitation hospitals nationwide. Its use has since increased as seen in its role as tool for documenting change for reimbursement through Medicare. Because of its extensive use and research, the only problems I ran into while gathering data for this project was the timeliness of the material. Most of the studies and reviews validating the FIM were conducted in the mid-90’s. Perhaps though, the longstanding role of the FIM in rehabilitation only further reinforces why the FIM is still used today.