ACRM
AMERICAN CONGRESS OF REHABILITATION MEDICINE

Archives of Physical Medicine and Rehabilitation

journal homepage: www.archives-pmr.org

Archives of Physical Medicine and Rehabilitation 2018;99:629-34

ORIGINAL RESEARCH

Psychometric Evaluation of the Brachial Assessment Tool Part 1: Reproducibility

Bridget Hill, PhD,[a,b] Gavin Williams, PhD,[b] John Olver, MBBS,[b] Scott Ferris, MBBS,[c] Andrea Bialocerkowski, PhD[a]

From the [a]Menzies Health Institute, Brisbane, QLD; [b]Epworth Monash Rehabilitation Medicine Unit Epworth HealthCare, Melbourne, VIC; and [c]The Alfred, Melbourne, VIC, Australia.

Abstract

Objective: To evaluate reproducibility (reliability and agreement) of the Brachial Assessment Tool (BrAT), a new patient-reported outcome measure for adults with traumatic brachial plexus injury (BPI).

Design: Prospective repeated-measure design.

Setting: Outpatient clinics.

Participants: Adults with confirmed traumatic BPI (N=43; age range, 19−82y).

Interventions: People with BPI completed the 31-item 4-response BrAT twice, 2 weeks apart. Results for the 3 subscales and summed score were compared at time 1 and time 2 to determine reliability, including systematic differences using paired *t* tests, test retest using intraclass correlation coefficient model 1,1 ($ICC_{1,1}$), and internal consistency using Cronbach α. Agreement parameters included standard error of measurement, minimal detectable change, and limits of agreement.

Main Outcome Measure: BrAT.

Results: Test-retest reliability was excellent ($ICC_{1,1}$=.90−.97). Internal consistency was high (Cronbach α=.90−.98). Measurement error was relatively low (standard error of measurement range, 3.1−8.8). A change of >4 for subscale 1, >6 for subscale 2, >4 for subscale 3, and >10 for the summed score is indicative of change over and above measurement error. Limits of agreement ranged from ±4.4 (subscale 3) to 11.61 (summed score).

Conclusions: These findings support the use of the BrAT as a reproducible patient-reported outcome measure for adults with traumatic BPI with evidence of appropriate reliability and agreement for both individual and group comparisons. Further psychometric testing is required to establish the construct validity and responsiveness of the BrAT.

Archives of Physical Medicine and Rehabilitation 2018;99:629-34

Traumatic brachial plexus injury (BPI) is a serious condition that generally affects previously healthy younger people.[1] People with BPI present with an extremely wide range of ability to use their arm based on the site and severity of the initial injury. They may undergo many months if not years of expensive and time-consuming surgery and ongoing therapy to reanimate their arm with varying degrees of success.[2-5] Historically, outcome assessment after BPI has been primarily impairment based.[6-8] Day-to-day use of the affected limb has not been routinely assessed despite this being key to the long-term outcome and overall satisfaction for the person with BPI.[9-12] Where activity has been assessed, the measures have not been psychometrically evaluated for BPI.[7] The most commonly used patient-reported outcome measure is the Disabilities of the Arm, Shoulder and Hand (DASH).[6,7] However the DASH has been shown to be multidimensional so total scores must be viewed with caution. Further, the DASH may not contain items that truly reflect how people with BPI use their affected limb[13] and are likely to address compensation or adaption rather than actual use of the affected limb.[14]

The Brachial Assessment Tool (BrAT) is a new unidimensional 31-item 4-response patient-reported outcome measure designed to address some of these issues. Based on the *International Classification of Functioning, Disability and Health* definition of activity,

Disclosures: none.

0003-9993/17/$36 - see front matter © 2017 by the American Congress of Rehabilitation Medicine
https://doi.org/10.1016/j.apmr.2017.10.015

"execution of a task or action by the individual,"[15(p.5)] items for inclusion were generated by experts in the field, including people with BPI.[13] Developed using Rasch analysis, the BrAT is a unidimensional measure assessing solely "activity after adult traumatic BPI."[16] To assess actual day-to-day use of the arm, responses are attributed directly to the affected limb. The BrAT may be used as 3 separate subscales: (1) 8 dressing and grooming items, (2) 17 whole arm and hand items, and (3) 6 no hand items; or alternatively, all 31-items may be added to produce a summed score. The BrAT item responses are scored as 0 (cannot do now), 1 (very hard to do now), 2 (a little hard to do now), and 3 (easy to do now).

Recovery from BPI occurs over a prolonged period of time and has a significant effect on a person's psychological and emotional state.[9,11,12] Further, people with a BPI often report ongoing severe pain.[17,18] These variables may influence how the person with a BPI perceives the day-to-day use of their affected limb and be a source of random error that may affect the reliability of the BrAT.[19] The BrAT was designed using Rasch analysis and has appropriate evidence supporting content validity and unidimensionality (ie, all the items appear to be measuring the same underlying construct).[16] To further support the use of the BrAT for adults with BPI in the clinical setting and to aid in the interpretation of BrAT scores, evidence of additional psychometric properties is required. All outcome measures must be reproducible (ie, people who are stable will obtain similar results from repeated assessment).[20] Reproducibility is fundamental to all aspects of measurement, and proof of reproducibility can ensure confidence in the data from which rational conclusions can be drawn.[21]

Reproducibility is comprised of 2 different but essential components: reliability and agreement.[20,22,23] Reliability addresses how stable a measure is over repeated use and how well people can be differentiated despite measurement error.[21,24,25] Measures of reliability include test-retest and intrarater reliability, defined as "the degree to which one rater can obtain the same rating on multiple occasions of measuring the same variable."[21(p.870)] Internal consistency indicates how interconnected the items are (ie, all the items appear to be related to each other and measuring something similar).[21,26] Agreement is related to absolute measurement error (ie, how close repeated-measure scores are), expressed in the actual units of the measure. In essence, reliability coefficients enable discrimination of people, whereas agreement addresses how scores differ.

The purpose of this article was to investigate the 2 parameters of reliability (test-retest reliability and internal consistency) and 3 parameters of agreement (standard error of measurement, minimal detectable change [MDC], and Bland-Altman limits of agreement [LoA]). A priori hypotheses were established based on the COnsensus-based Standards for the selection of health status Measurement INstruments (COSMIN) guidelines. We expected that (1) the BrAT will demonstrate high test-retest reliability with an intraclass correlation coefficient (ICC) of >0.8, (2) the BrAT will demonstrate high internal consistency with a Cronbach α of $\geq$0.7 and, (3) 95% of the Bland-Altman LoA scores will fall within 2 SDs above and below the mean difference score.

List of abbreviations:

BPI	**brachial plexus injury**
BrAT	**Brachial Assessment Tool**
CI	**confidence interval**
COSMIN	**COnsensus-based Standards for the selection of health status Measurement INstruments**
DASH	**Disabilities of the Arm, Shoulder and Hand**
ICC	**intraclass correlation coefficient**
LoA	**limits of agreement**
MDC	**minimal detectable change**
MDC_{90}	**minimal detectable change based on a 90% confidence interval**

Methods

This project used a multicenter, prospective repeated-measure design. Ethical approval was gained from 3 human research and ethics committees (Griffith University PES_12_13_HREC, Alfred Health 425/11, and Melbourne Health 2011.220), and all participants provided signed informed consent prior to commencement of the project.

Participants

Participants comprised a convenience sample recruited from the 106 people with BPI who participated in the Rasch analysis arm of a previously reported study. Data were collected concurrently.[16] Participants were recruited to the reproducibility arm if they had a diagnosis of traumatic BPI confirmed by magnetic resonance imaging, nerve conduction studies, intraoperative findings, or clinical assessment, and were >18 years of age at the time of recruitment. To ensure participants to this arm of the project remained stable during the assessment period, only those >12 weeks postinjury and who had not undergone surgery to reanimate the upper limb within the previous 2 years were invited to take part. Therefore, the function of their arm was likely to remain stable for the duration of this project because minimal recovery may be expected. Exclusion criteria included inability to provide informed consent, preexisting upper limb conditions that affected day-to-day activity, evidence of spinal cord injury confirmed by magnetic resonance imaging, or a diagnosis of brachial plexus birth injury.[16]

Data collection

Once participants consented to participate, they were mailed a copy of the questionnaire used for the Rasch analysis together with a reply, paid envelope. Two weeks after its return, a second identical questionnaire was mailed to them to complete. A 2-week period was selected to prevent recall bias while participants would not be expected to show any change in the day-to-day use of their arm.[26,27] To determine whether participants felt that the use of their affected limb remained stable during the study period, a 5-point global change score was used as a reference criterion.[28,29] Response options were attributed directly to the affected limb and were scored as 1 (much less than last time), 2 (a little less than last time), 3 (no change to last time), 4 (a little better than last time), and 5 (much better than last time).

Data analyses

All statistical analyses to address the a priori hypotheses were undertaken using SPSS Statistics version 22.0.[a] On the basis of recent tabled calculations, to have 90% probability or assurance of obtaining a 95% confidence interval (CI) with a precision of .15 (ie, a total width of .30), for an intraclass correlation of .80, a sample size of 41 participants is required.[30] To allow for