Measuring change in symptoms of neurobehavioural disability: responsiveness of the St Andrew's – Swansea Neurobehavioural Outcome Scale

Nick Alderman [12], Claire Williams [2], Caroline Knight [34] Rodger Ll Wood [5]

[1]Priory Brain Injury Services, The Priory Group, Grafton Manor, Grafton Regis, Northants, UK

[2]Department of Psychology, College of Human and Health Sciences, Swansea University, Swansea, UK

[3]Priory Brain Injury Services, The Priory Group, Burton Park, Melton Mowbray, Leicestershire, UK

[4]School of Psychology, University of Leicester, Leicester, Leicestershire , UK

[5]Neuropsychology Clinic, College of Medicine, Swansea University, Swansea, SA2 7EP, UK

Corresponding Author: Professor Nick Alderman, Clinical Director, Priory Brain Injury Services, Grafton Manor, Grafton Regis, Northants NN12 7SS, UK; email NickAlderman@priorygroup.com; telephone (+44)-1908-300204.

This is a pre-copyedited, author-produced version of an article accepted for publication in *Archives of Clinical Neuropsychology* following peer review.

To Cite: Nick Alderman, Claire Williams, Caroline Knight, Rodger Ll. Wood; Measuring Change in Symptoms of Neurobehavioural Disability: Responsiveness of the St Andrew's-Swansea Neurobehavioural Outcome Scale. *Arch Clin Neuropsychol* 2017 1-12. doi: 10.1093/arclin/acx026

ABSTRACT

Objective

Neurobehavioural disability (NBD) after acquired brain injury (ABI) is often associated with poor outcome. The 'St Andrew's-Swansea Neurobehavioural Outcome Scale' (SASNOS) was developed to measure NBD in a range of applications. Two of the 'holy trinity' of psychometric properties, reliability and validity, have been comprehensively mapped, but the extent to which SASNOS meets the third, responsiveness, has not been investigated. Demonstrating responsiveness is essential in instruments employed in repeated measurement scenarios to confirm their ability to discriminate real change from error. However, there is no single agreed method for determining responsiveness. For some instruments this property remains unexplored. A difference in scores attaining statistical significance for aggregate data is frequently cited as support for this construct, but this approach remains heavily criticised. This study explores responsiveness of SASNOS.

Method

Consecutive SASNOS assessments completed over varying times for 145 individuals participating in neurobehavioural rehabilitation, drawn from multiple services, were compiled into a retrospective sample of convenience. Multiple methods were employed to confirm internal responsiveness, including those identifying statistically significant change, minimally detectable change and minimally important change.

Results

All methods confirmed responsiveness as a psychometric property of SASNOS; the extent depended on method used and NBD domain investigated. A number of indicators are presented which equip clinicians and researchers with options to interpret results from repeated assessments, including the individual level in the context of rehabilitation.

Conclusions

SASNOS reliably measures change over time in NBD symptoms, further confirming its suitability as an instrument for investigating multidimensional outcomes of ABI.

KEY WORDS

Head injury, traumatic brain injury; assessment; norms/normative studies; rehabilitation; practice effects/reliable change; statistical methods

INTRODUCTION

Neurobehavioural disability (NBD) is the product of interactions between damaged neural systems and neurocognitive functions, further modified by premorbid personality traits and post-injury learning (Wood, 2001). It comprises elements of executive and attentional dysfunction, labile mood, altered emotional expression, poor impulse control, poor insight, problems of awareness and social judgement, and a range of personality changes that impede psychosocial recovery (Kreutzer Marwitz, Seel and Serio, 1996). Behaviour disorders associated with NBD are enduring (Kelly, Brown, Todd and Kremer, 2008) and impose serious long-term social handicap (Burke, Wesolowski and Lane, 1988; Alderman and Wood, 2013). Fortunately, post-acute neurobehavioural rehabilitation (NbR) has shown that many behaviour problems can be ameliorated to reduce their psychosocial impact (Wood, Alderman and Worthington, in press). There is now a good evidence base that demonstrates efficacy of NbR programmes, at the level of the individual case, group and service level, both clinically and economically (see for example Ylvisaker, Turkstra, Coehlo, Yorkston, Kennedy, Moore et al, 2007; Alderman and Wood, 2013; Alderman, Knight and Brooks, 2013; Worthington, Matthews, Melia & Oddy, 2006; Oddy and da Silva Ramos, 2013; Wood, McCrea, Wood & Merriman, 1999). However, in order for NbR to successfully target these behaviours, reliable and valid methods of assessing them must first be available.

The 'St Andrews – Swansea Neurobehavioural Outcome Scale' (SASNOS) was developed to meet this need (Alderman, Wood and Williams, 2011). Underpinned by the World Health Organisation International Classification of Functioning, Disability and Health, the SASNOS has robust psychometric properties, inter-rater and test-retest reliability has been established, and multiple

indicators of validity have been demonstrated. Recent research has also demonstrated that SASNOS is able to capture the impact of context on ratings, aiding short-term goal setting and clinical decision making (Alderman, Williams and Wood, submitted). However, whilst two of the 'holy trinity' of psychometric properties of outcome measures, reliability and validity, have been comprehensively mapped, the extent to which SASNOS meets the third, responsiveness, has yet to be determined.

Broadly defined, responsiveness is the ability to detect change when it has occurred. It has two major aspects: 1) 'external' responsiveness, which reflects the extent to which change in a measure relates to corresponding change in a standard health status measure; and 2) 'internal' responsiveness, which characterises the ability of a measure to detect change over a pre-specified time frame. The latter aspect of responsiveness is therefore particularly important to consider when standardised outcome measures are used in repeated measures contexts, such as NbR. Whilst it may be relatively easy to determine a statistically significant change in scores over time, determining when this indicates a clinically meaningful change is more difficult. Indeed, because of its partial dependency on sample size, statistical significance does not always correspond to the clinical relevance of the observed effect (de Vet, Terwee, Ostelo, Beckerman, Knol and Bouter, 2006; Eisen, Ranganathan, Seal and Spiro, 2007). For this reason, determining the extent an instrument measures change over time not attributable to error, versus its ability to capture change that is clinically meaningful, has heavily influenced investigations of internal responsiveness. In turn, this has resulted in two main approaches to determining the smallest amount of change on a measure over time that is likely to be of importance.

The first is Minimally Detectable Change (MDC), which refers to the smallest difference between scores at Time 1 and Time 2 ($T_1 - T_2$) that falls outside the measurement error of an instrument (Terwee, Dekker, Wiersinga, Prummel and Bossuyt, 2003). Two main methods of calculation fall

under MDC. The first is Standard Error of Measurement (SEM) where differences between T_1 and T_2 smaller than one SEM are considered a consequence of error rather than real change. The second is MDC confidence intervals. This is where confidence intervals (typically 90-95%; Copay, Subach, Glassman, Polly and Schuler, 2007; Donoghue and Stokes, 2009; Walton, Macdermid, Nielson, Teasell, Chiasson and Brown, 2011) can be constructed around SEM to further inform judgements regarding the minimum difference in scores that is not a result of measurement error within a given level of assurance. Absolute reliability from a statistical perspective is determined from the SEM, whilst MDC provides a means of applying a degree of certainty in the clinical context that a change in scores is not attributable to measurement error. However, whilst MDC is useful for contexts where assurance is needed that any difference in scores in unlikely to be attributable to error within an instrument, it does not necessarily equate to meaningful and important change.

Minimally Important Change (MIC) is used for this purpose and can be defined as "(the difference in score) that corresponds to the smallest change in status that stakeholders (persons, patients, significant others, or clinicians) consider important" (Eisen, Ranganathan, Seal and Spiro, 2007; p.273). As with MDC, there are multiple methods to determine MIC. These include distribution-based methods that utilise the statistical characteristics of a sample; anchor-based methods which compare change in scores with a minimal important change defined on an objective (for example, change in medication usage, change in access to health services); subjective methods (for example, patient self-report on a global health measure), and the Delphi method, where a panel of experts reach consensus on what constitutes MIC following several rounds of consultation. However, each approach is not without limitations (Walters and Brazier, 2003), leading to the suggestion that distribution-based methods should be used alongside a meaningful exterior anchor wherever possible (Wyrich, Tierney, Babu, Kroenke and Wolinsky, 2005).

However, identifying an external anchor in the context of NbR can be difficult. First, a single health related measure will almost certainly fail to adequately capture the complexity of ABI. Second, lack of homogeneity within ABI and the presence of impairment from multiple sources, not just behaviour, complicate interpretation of the impact of change between different measures. Third, variable insight and awareness undermines validity of self-report, weakening any anchor derived from patients in rehabilitation. Fourth, discharge from rehabilitation may provide an external anchor testifying to reduction of NBD; however, multiple reasons influence discharge, not just remediation of NBD so using this may result in error. Finally, in ABI studies where data has been harvested from convenience samples, availability of additional variables which may constitute anchors may be limited. To overcome these difficulties, researchers exploring relationships between distribution-based approaches and MIC (as defined using external anchors) have proposed using the following thresholds to determine proxy indicators of meaningful change.

Standard Error of Measurement: Various threshold values of the SEM of changes in scores in stable reference groups have been shown to correspond with MIC (Rai, Yazdany, Fortin and Avina-Zubieta, 2015), with a SEM of 1 advocated as the minimum that reflects this criterion (Wyrwich and Wolinsky, 2000; Wyrwich, Tierney and Wolinsky, 1999; Wyrwich, Nieraber, Tierney and Wolinsky, 1999).

0.5 Standard deviation (0.5 SD): In a systematic review of thirty-eight studies for health-related quality of life instruments, Norman, Sloan and Wyrwich (2003) found that MIC estimates were close to one half of a SD in all but 6 studies. On this basis, they argued that 0.5 SD could be used in most circumstances to determine meaningful change.

Standardised Response Mean (SRM): Originally proposed by Cohen, Effect Size (ES) has been frequently used to determine responsiveness in clinical investigations conducted in a repeated measures context. Most relevant here is the standardised response mean (SRM), a version of ES

applicable within groups where change scores for the same individuals at T_1-T_2 are of interest. The magnitude of the size of the effect is determined by applying cut-off thresholds: <20 'trivial'; ≥20 to <50 'small'; ≥50 to <80 'medium'; ≥80 'large' effect size. However, Middel and van Sonderen (2002) note that the strength of the correlation between T_1 and T_2 varies between samples in repeated measures designs. Consequently, application of ES thresholds at the levels proposed by Cohen may lack validity, as variable correlation size results in over or underestimation of classification of the size of the effects. Instead, Midell and von Sonderen suggest applying an additional calculation to calibrate thresholds according to the size of the T_1-T_2 correlation, and Norman, Sloan and Wyrwich (2003) recommended that a 'medium' ES corresponds to clinically meaningful change when using SRM as a proxy measure of MIC.

However, despite the availability of these methods, internal or external responsiveness is still not habitually investigated in ABI. For example, in a review of ABI tests, scales and questionnaires, Tate (2010) found no corresponding information for responsiveness for 10 out of 27 instruments reportedly measuring NBD or related symptoms. In addition, in studies of measures where responsiveness was investigated, most utilised statistical significance (31.8%) or ES (27.2%) alone, or a combination of the two methods (41%). In light of this, the aim of the current study was to comprehensively map the internal responsiveness of the SASNOS because it is routinely used to assess response to rehabilitation and this had not been determined previously. Information about this psychometric property is essential in order to properly interpret any variance in scores arising in the context of repeated measurement. Specifically, the study sought to (a) determine whether SASNOS has the statistical properties to effectively measure expected change in symptoms of NBD over time; (b) identify cut-off scores for clinicians and researchers to employ across a range of contexts to reliably discriminate genuine improvement from those due to error in the instrument. A number of distribution-based methods were employed to meet the study objectives because of lack of consensus surrounding the conceptualisation and assessment of responsiveness, (see the

recommendations of Wyrwich, Tierney, Babu, Kroenke and Wolinsky 2005), and difficulty determining an external anchor in the context of NbR.

METHOD

Participants

An anonymous database containing ABI outcomes for 542 participants in residential NbR programmes across the UK was consulted. The database was originally compiled as part of a larger multisite study investigating time of onset and types of NBD arising from ABI. It contained data from a basket of ABI outcome measures, including the SASNOS, as well as some standardised demographic and clinical information. All assessments contained on the database were undertaken as part of routine clinical practice and completed by the clinical team responsible for the delivery of each participants NbR programme.

Participants were included in the current study if the outcomes of two consecutive (T_1 and T_2) SASNOS assessments were available. 145 participants met this criterion, of whom 71% were male. There were multiple causes of ABI: trauma (44%), cerebral anoxia (17.1%) and cerebrovascular accident (15.6%) comprised the majority, accounting for over 75% of all cases. Mean age at injury was 42.5 years of age (SD = 18.2, range 4.7-78.8) and on admission 46.2 years of age (SD = 14.8; range = 14-79). Mean time since injury at T_1 was 40.9 months, although there was significant variability (SD = 10.5, range = 0-516). Time spent in rehabilitation at T_1 was also variable as the sample included both acute rehabilitation and long-stay, slow-stream rehabilitation care pathways (mean = 45.9 weeks, SD = 52.3, range 1.30 – 339.9). Consequently, and as could be expected, the T_1-T_2 time interval was not standardised, although 75% of participants were rated at T_1-T_2 within twelve months. No significant gender differences were found on any of these variables.

Ethical approval for the study was granted by NRES Committee East Midlands - Leicester (reference 11/EM/0283).

Measures

The St Andrews-Swansea Neurobehavioural Outcome Scale (Alderman, Wood and Williams, 2011): This consists of 49-items that capture five major domains of NBD (interpersonal behaviour, cognition, aggression, inhibition and communication), each with 2-3 subdomains. Each item consists of a statement regarding a feature of NBD whose perceived prevalence is rated using a seven point scale ('never' to 'always'). Ratings are based on observations of the person being assessed over a two week period. A major strength of SASNOS over existing measures of NBD is availability of normative data from a moderate sized group of neurologically healthy controls. Total rating scores pertaining to the measure as a whole, five domains and 12 subdomains, are transformed to standardised scores derived from the T-distribution, with a mean of 50 and standard deviation of 10 enabling meaningful comparison between categories of NBD symptoms, and the general population. Transformations are constructive, so higher scores reflect greater independence. The SASNOS has known psychometric properties, sufficient items to capture the diverse range of NBD signs and symptoms, satisfactory test-re-test (0.82-0.96) and inter-rater reliability (0.59-0.83), and multiple indicators of validity have been demonstrated. However, as stated earlier, responsiveness has not been investigated to date.

Total sum of ratings from the SASNOS, and those for each of the five domains, were transformed to T-scores. First, for each rehabilitation participant a Z-score was calculated for the five domains and overall sum of ratings by subtracting their individual sum of ratings from the mean for neurologically healthy controls and dividing the result by the standard deviation for neurologically healthy controls

(see table XIII, p94, Alderman, Wood and Williams, 2011). Second, individual T-scores were obtained in the usual way ($T = 50 + 10z$). It is these T-scores that are reported in the subsequent evaluation described here.

A further point of relevance regarding scores for this investigation is the distinction between those who present with symptoms of NBD that are atypical of those observed in the neurologically healthy population. The number of standard deviations from the mean provides one method of highlighting potential areas of concern. Whilst scores in excess of 2 SD from the mean have statistical significance, those exceeding 1 SD also have been advocated within neuropsychology as being clinically of interest (Heaton, Grant and Matthews, 1991). In an earlier SASNOS study, Alderman, Wood and Williams (2011) demonstrated cut-off scores based on one (40) and two (30) SD discriminated between neurologically healthy and ABI participants. For example, 86% of neurologically healthy controls achieved a total score of more than 40 and 92% greater than 30. In contrast, nearly all people with ABI achieved a total T-score that fell below 40, whilst 77.9% achieved scores below 30. As these results suggest NBD is highly atypical of neurologically healthy people, and highly characteristic of ABI, the recommendation of the authors was that T- scores less than 40 should be considered exceptional and consequently an indicator for potential rehabilitation goals.

Data Analysis

To determine the responsiveness of the SASNOS, data was analysed using the following methods:

1. Statistical significance: Paired sample *t* tests were performed to compare SASNOS ratings at T_1 and T_2 (alpha level p<.05).

2. Minimally detectable change: SEM was obtained from $SD \times \sqrt{(1 - ICC)}$ ((Rehabilitation Institute of Chicago, 2010). SD was from T_1, and ICC the reliability estimate of the test, in this case the consistency intra-class correlation of SASNOS scores over time, as calculated and reported by Alderman, Wood and Williams (2011). MDC at the 90th confidence interval (MDC_{90}) was derived from

$SEM \times 1.65 \times \sqrt{2}$. In this equation, 1.65 is the z-score at the 90th confidence level, and the square root of 2 is used as a multiplier to account for error associated with repeated measurements (Hayley and Fragala-Pinkham, 2006; Stratford, 2004). Similarly, substituting 1.65 with 1.96 gives MDC at the 95th confidence level (MDC_{95}) $SEM \times 1.96 \times \sqrt{2}$.

3. Minimally important change proxies: SEM was calculated as above. The 0.5 SD proxy MIC was derived using the standard deviations for the six SASNOS scores at T_1 (Norman, Sloan and Wyrwich, 2003). SRM was derived by dividing the T_1-T_2 mean change score by the SD of that change score $(\bar{x}\ change)/(SD\ change)$.

4. Categorising the extent of change: Following the criteria adopted by Eisen, Ranganathan, Seal and Spiro (2007), participants were categorised on the four distribution-methods as being 'improved', 'same' or 'declined'. For SEM, participants were classed as 'improved' if the $T_1 - T_2$ difference exceeded 1 SEM, 'same' if the difference was less than 1 SEM, and 'declined' if it fell more than 1 SEM. The same method was used to categorise participants for MDC and 0.5 SD. In the case of SRM, participants were defined as 'improved' if recalibrated SRM values were 'moderate' or more using $SRM \geq 0.50$, 'same' for $-0.50 \geq SRM < 0.50$, and 'declined' when $SRM \leq -0.50$. (Cohen, 1988; Eisen, Ranganathan, Seal and Spiro, 2007).

5. Agreement between MDC methods: The weighted kappa statistic was used to determine levels of agreement between the distribution-based methods and the proportion of participants categorised as 'improved', 'same' and 'declined' at T_2 compared to baseline scores at T_1 as a consequence of the criteria described above. Weighted kappa was employed following the precedent using an identical methodology described by Eisen and colleagues (2007). In both studies, results of distribution-based methods were employed to populate categorical items and kappa provides an appropriate statistic for this purpose. Weighted kappa was utilised to take into account that categories are ordered and

to reflect the extent of any disagreement arising from classification using scores from the five distribution-based methods.

Statistical analysis was undertaken using SPSS v22 (IBM Corp., 2013), with the exception of weighted kappa values which were calculated using MedCalc v16.8 (MedCalc Software, 2016).

RESULTS

Aggregate change over time

Summary statistics for SASNOS total and domain scores at T_1 and T_2 are summarised in table 1, along with the results of statistical significance testing and magnitude of effect size (SRM).

< TABLE 1 ABOUT HERE >

With the exception of 'Aggression' domain scores, paired samples t-tests revealed highly significant ($p<.002$ or better) increases in SASNOS total and domain scores from T_1 to T_2. Differences in mean T-scores varied from 8.32 scale points ('Cognition') to 1.27 ('Aggression'). However, the magnitude of effect size, as calculated using SRM, varied across SASNOS domains. Consistent with the result from the *t* test, a 'large' effect size was found for 'Cognition'. In contrast, the size of the effect was 'medium' for the total SASNOS score and 'Interpersonal Behaviour' domain; and 'small' for the remaining three domains. Whilst there was no statistically significant difference between mean T-scores for 'Aggression', lack of convergence between this method and SRM is evident from the 'small' effect size for 'Inhibition' and 'Communication' despite highly statistically significant differences ($p=.002$ and $p<.001$ respectively).

Values achieved for the MDC and MIC measures for SASNOS total and domain scores are shown in table 2. Table 3 summarises the percentage of participants whose $T_1 - T_2$ scores suggest differences in ratings above what would be considered a consequence of error within the instrument or

individual variability (MDC - SEM, MDC_{90}, MDC_{95}). Table 3 also shows the percentage of participants who exceeded a responsiveness threshold found in other studies to equate to important, meaningful change (MIC proxies - SEM, 0.5 SD, SRM).

< TABLES 2 & 3 ABOUT HERE >

The percentage of participants whose $T_1 - T_2$ scores indicate real change rather ranged from 32-64% based on SEM, 14-46% based on MDC_{90}, and 12-44% based on MDC_{95}. As both MDC methods are confidence intervals are derived from the SEM, it is not surprising that SEM identified the most individuals as achieving scores that fall above the minimum required for absolute reliability. Using SASNOS total scores, just over half the sample showed a reduction in symptoms of NBD at T_2 based on SEM, and just over a third based on both MDC_{90} and MDC_{95}. 'Cognition' domain scores showed the greatest change using all three methods (44.1 – 64.8%), and the least change was observed for 'Inhibition' domain scores based on MDC (14.5 and 12.4%) and 'Aggression' scores based on SEM (32.4%).

Across all SASNOS scores, the proportion of participants showing clinically meaningful improvement ranged from 32-64% for SEM, 24-46% for 0.5 SD, and 27-54% for SRM. SEM categorised the largest percentage of participants as having achieved minimally important change at T_2 using SASNOS total scores (53.1%), followed by SRM (46.9%) then 0.5 SD (40.7%). SEM also identified the greatest proportion of participants as meaningfully improved using SASNOS domain scores (see Table 3). With the exception of 'Inhibition' domain scores, SRM categorised the lowest percentage. Across all three methods, the domain showing the most change from $T_1 - T_2$ was 'Cognition' (46.2-64.8%), with 'Aggression' showing the least (24-32%).

Discriminating rehabilitation participants expected to make the most change

Examination of table 1 indicates that the smallest difference between means and smallest effect sizes were evident on the 'Inhibition', 'Aggression' and 'Communication' domains, with overall mean T-scores at T_1 falling above the normative reference mean (50). In contrast, mean SASNOS total, 'Interpersonal Relationships' and 'Cognition' scores fell more than one SD below the reference mean. Consistent with the non-homogeneous nature of ABI, this finding suggests that some symptoms of NBD are more endemic than others. In addition, people in the current sample were rated at different times in their rehabilitation journey, and therefore assessment at T_1 may reflect a reduction of NBD symptoms from admission.

Consequently, the responsiveness data described above may be better scrutinised by applying criteria to discriminate at T_1 between participants whose domain scores suggest an absence of problematic behaviour, versus those whose scores fall one SD or more below the mean for neurologically healthy controls (e.g., a domain score of less than 40 at T_1), thus suggesting the presence of NBD. The assumption tested by categorising participants in this way is that participants whose SASNOS scores suggest NBD symptoms are problematic at T_1 should show greater responsiveness than those which are more characteristic of the general population, given the reported efficacy of NbR programmes. Therefore, it was predicted that the proportion of participants exceeding the multiple thresholds reflecting both MDC and MIC for T_1-T_2 will be higher for participants with domain scores of less than 40 at T_1 than for those whose scores are equal to or greater than 40.

Discriminating participants into different groups is widely reported in the responsiveness literature. A 'control' group is sometimes utilised directly for comparison purposes with an intervention or treatment group; it is also used to generate statistics that are then utilised in the calculation of certain responsiveness indices. A control group does not necessarily employ healthy individuals

drawn from the general population. In other studies, clinical subjects are employed for these purposes when they are stable or otherwise not expected to change (Walters and Brazier, 2003; Rai, Yazdany, Fortin and Avina-Zubieta, 2015). Given lack of homogeneity within the ABI population and the variable proportions of individuals presenting with symptoms of NBD, as measured by the SASNOS, there was no expectation that all rehabilitation participants would necessarily present at T_1 with symptoms of all these, either because they have not been a feature of their post-injury presentation or have already been positively remediated by NbR received to date. In this case, discriminating between participants in this way advantages the analysis in two ways. Firstly, for comparison reasons, contrasting differences in T_1-T_2 scores; secondly, and more importantly, enhancing calculation of responsiveness indices by targeting participants in which change is expected.

< TABLE 4 ABOUT HERE >

Results of categorising participants are shown in table 4. This confirms the assumption that responsiveness is most evident amongst those whose SASNOS assessment suggests NBD symptoms are more prolific than the general population. In all the comparisons made, a greater percentage of participants with a T_1 score of less than 40 achieved a T_1-T_2 difference that exceeded thresholds for responsiveness across all the methods used on SASNOS total and domain scores. This finding is consistent with the hypothesis that those participants whose scores suggested they will benefit the most from NbR do so.

In relation to MDC methods, SEM categorised the most people who were expected to improve as improving (63-77%). For those rated in the normal range or higher at T_1, 25-44% achieved a difference in ratings at T_2 greater than the SEM threshold for MDC. As expected, both MDC_{90} and

MDC_{95} provided more conservative results. 37-61% of those expected to improve did so based on MDC_{90}, whilst 37-55% were categorised as having done so based on MDC_{95}.

Of the three proxy indicators of MIC, the SEM method also classified the highest percentage of participants (63-77%). The 0.5 SD method categorised 47-66% as making important, meaningful change in cases where initial SASNOS scores were less than 40, with 18-29% improved who were in the normal range or higher at T_1. The SRM method classified 54-66% expecting to improve as subsequently doing so, with 16-42% of those whose scores suggested symptoms of NBD were in the expected range at first assessment also making improvement.

Finally, magnitude of the size of the effect was examined by calculating SRM's for the group expected to show the most change, comparing them to the values obtained for the whole sample. Higher adjusted SRM's were evident for four of the six SASNOS scores in the reduced sample. With regard to classification of the magnitude of the size of the effect, this remained unchanged for the total score ('moderate': .57 vs. .71), 'Interpersonal Behaviour' ('moderate': .75 vs. .67) and 'Cognition' ('large': .92 vs. .91). However, whilst the magnitude of the effect size was originally 'small' for the remaining thee SASNOS domains when all cases were considered, this became 'large' when cases expected to show the most potential for change were considered separately ('Inhibition': .26 vs. .82; 'Aggression': .15 vs. 1.05; 'Communication': .40 vs. .93).

Agreement among methods

Agreement among the five distribution-based methods varied across SASNOS total and domain scores (see table 5). Regarding total score, weighted kappas ranged from .50-.95 ('moderate' to 'almost perfect' agreement). Agreement between domain scores was also variable, ranging from .40-.89 for 'Interpersonal Behaviour', ,36-.98 for 'Cognition', .26-1.00 for 'Inhibition', .58-.92 for 'Aggression' and .33-1.00 for Communication. Agreement was more consistent between the three

MIC methods, ranging from .66-.83 for total SASNOS scores, .78-.81 for 'Interpersonal Behaviour', .47-.70 for 'Cognition', .85-1.00 for 'Inhibition', .80-90 for 'Aggression' and .73-1.00 for 'Communication.

< TABLE 5 ABOUT HERE >

Possible effects of variability within the sample on the results obtained

As noted, some variability was evident within the sample of convenience utilised for this study and possible effects of this on the results were explored. Firstly, there was no evidence that the amount of NbR provided, based on the length of time spent in neurobehavioural rehabilitation between T_1-T_2, was related to the amount of change observed through differences on the SASNOS total score, and scores on each of the five domains (Pearson correlations ranged from -.04 to .17). Similarly, the extent of NBD symptoms measured at T_1 was not found to be related to how much time had elapsed since injury. For each of the six SASNOS scores, rehabilitation participants were categorised into two groups using the criteria specified before (T-score $<$ 40 vs. T-score $\geq$ 40). Using independent t-tests, there were no between-group differences for mean time since injury, age at injury and length of stay up to T_1 ($p > .05$). Age on admission did not discriminate between the two groups with the exception of 'Aggression'. Rehabilitation participants who scored below 40 on this domain were younger on admission (mean age = 38.6 years, SD = 14.6) than those rated within the expected range or higher for neurologically healthy controls (mean age = 47.4 years, SD = 14.7; t = 2.36, p = .02).

DISCUSSION

The overarching goal of this study was to investigate internal responsiveness of the SASNOS which had previously remained unexplored. In previous studies of health-change related instruments, multiple methods have characteristically taken the form of examining change in repeated measurements relative to change in an external anchor, most typically a general measure of physical

or mental health, or quality of life. In this study, an appropriate external anchor could not be identified and thus, internal responsiveness was examined. A lack of a 'gold standard' or universally accepted methodology for determining internal responsiveness resulted in several methods being employed, including methods examining statistical significance of change between repeated SASNOS scores, identifying change beyond that associated with error, and ascertaining important and meaningful change.

The first aim of the study was to determine whether the SASNOS has the statistical properties to effectively measure expected change in symptoms of NBD over time. Using a sample of participants in residential NbR, responsiveness was demonstrated using multiple methods. Although this varied according to the method used and the SASNOS score considered. Consistent with previous literature (de Vet, Terwee, Ostelo, Beckerman, Knol and Bouter, 2006; Eisen, Ranganathan, Seal and Spiro, 2007) methods examining statistical significance of change between repeated SASNOS scores were of limited value. Indeed, although the current sample size was not very large, comparatively small differences between means nevertheless resulted in very high significance levels using paired samples *t* tests. In addition, it's unclear how the results of the tests of statistical significance map onto real change, or can be applied to understand magnitude of change in scores on an individual level. Consequently, five other distribution based methods were used to further inform clinicians and researchers about internal responsiveness as a psychometric property of the SASNOS, and provide better information regarding the meaning of any change in ratings between assessments.

The second aim of the study was to determine the magnitude of change from $T_1 - T_2$ to determine what aspects of NBD are most amenable to NbR. Magnitude of the size of effects calculated suggested that a 'moderate' reduction in NBD symptoms was achieved across the sample overall, with between a third and a half of participants exceeding change thresholds. For example, a 'moderate' reduction in interpersonal problems was found, which is unsurprising given existing

evidence of the effectiveness of NbR for the management of social behaviours. However, a more surprising finding was that the greatest change in scores was found in the 'Cognition' domain (44-64% improving), despite the belief that neurocognitive functions tend to be fixed or static in samples participating in NbR many years beyond the time in which spontaneous recovery is expected. In addition, even though the literature emphasises that referrals to NbR services are often driven by challenging behaviour associated with lack of inhibition and aggression, effect sizes for these SASNOS domains were 'small' (ranging from 14-43%, less than half the sample). An explanation for this apparent anomalous finding rests in the lack of consistency in the time at which T_1 was administered. Few participants in the current sample were initially assessed on admission, when these challenging behaviours might have been most apparent; and given the known effectiveness of NbR in managing difficulties with inhibition and aggression, by T_1 rehabilitation may already have succeeded in reducing symptoms associated with these domains. This explanation is partially supported by the analysis in which responsiveness was determined for rehabilitation participants with SASNOS scores below the level associated with neurologically healthy controls. When this was done, the size of the effect for the 'Inhibition', 'Aggression' and 'Communication' domains increased from 'small' to 'large'. For example, with regards to aggression, 55-77% of rehabilitation participants expected to make the most change did so, compared to only 18-32% in the sample as a whole. In addition, across all domains and methods, more people improved who were expected to improve than those whose symptoms were in the normal range for neurologically healthy controls at T_1.

These findings suggest that the effectiveness of NbR as assessed using SASNOS extends across all symptom domains, with perhaps the greatest challenge being the remediation of NBD symptoms observed in the context of social interaction and relationships. It also suggests that examination of responsiveness needs to be conducted in a targeted way, by considering participants who are expected to benefit. The cut-off score of 40 used in the current study provides clinicians with one means of discriminating in this way. A future study which synchronised SASNOS assessments so they

captured NBD on admission and regular intervals thereafter would clarify the relative speed in which the different domains are successfully targeted. Even so, on the basis of the current findings, we advocate that NbR is effective across all major categories of NBD as captured by the SASNOS.

The final goal was to identify cut-off scores that would allow clinicians and researchers to reliably discriminate between differences in SASNOS scores that indicate genuine improvement rather than error in the instrument. A range of indicators were identified that could be used for this purpose (table 2), depending on the questions asked. For example, researchers requiring higher confidence levels to determine numbers of people responding using cohort data may be drawn to MDC, as the principal need will be to ensure discrimination of cases whose difference scores exceed thresholds for error. When the question asked necessitates specifying an overall index of the magnitude of change, SRM would be appropriate. SRM is simple to calculate using repeated SASNOS assessments and the magnitude of the effect size determined using Middel and van Sonderen's (2002) solution. In contrast, when the main goal is to determine individual change in scores, less conservative thresholds may be appropriate, especially in a clinical context. Although there was high agreement between the three MIC methods regarding categorisation of the extent of change, SEM identified the highest proportion of individuals in the sample as improved, a finding that is consistent with other studies of responsiveness of health related outcome measures (Eisen, Ranganathan, Seal and Spiro, 2007). Therefore, for evaluating change in SASNOS scores for clinical purposes, SEM is recommended for the following reasons. First, it has previously been argued SEM is a better method for determining MIC because it is independent of SD; consequently, it does not vary between samples and provides a more stable method for determining meaningful change (Wyrwich and Wolinsky, 2000; Eisen, Ranganathan, Seal and Spiro, 2007). Second, the SEM criterion has been widely reported in the assessment of MIC in individuals with both chronic medical and mental health conditions, and long-term behavioural issues (Hays, Brodsky, Johnston, Spritzer and Hui, 2005; Wyrwich, Tierney and Wolinsky, 1999; Eisen, Ranganathan, Seal and Spiro, 2007); as outcome from

ABI is also associated with long-term impairment, SEM may be similarly indicated in determining change to NBD symptoms. Finally, SEM has been cited as both a method used to discriminate change beyond that expected from error, and as a proxy that reflects minimally important change, increasing the range of applications and questions regarding change and responsiveness it can readily be employed with.

SEM can also be used to determine individual change. Probabilities of the normal curve can be applied to SEM values, making it easily applicable in determining the extent to which a difference in SASNOS ratings represents meaningful and important improvement (Domholft, 2005; Ries, Echternach, Nof and Gagnon Blodgett, 2009). SEM values in table 2 can be used for this purpose: there is a 68% probability that a difference in scores on repeated measurement will fall within ±1 SEM and a 96% probability it will be within ±2 SEM. To illustrate this, an individual rehabilitation participant achieving a difference in 'Interpersonal Behaviour' scores at repeated assessment of 6 points exceeds the 68% probability threshold that this is a meaningful change (±3.77, 1 SEM) but not at the 96% level (±7.54, 2 SEM). The same participant obtains a difference in 'Aggression' scores of 11 points. As there is a 96% probability that a repeated measure of this domain will be within ±7.16 (2 SEM), the size of the difference obtained can reasonably be interpreted as representing change beyond the expected variability. Also as 1 SEM has been cited as a proxy measure of MIC, criteria for attainment of change that is meaningful and important is also met at both the 68% and 96% probability levels. Consequently, a difference in repeated measures of more than 2 SEM is likely to reflect real variation that is both beyond that associated with error and indicative of meaningful change.

However, there are some weaknesses to the current study, principally arising from use of retrospective data drawn from a convenience sample, together with suggestions regarding how these can be addressed and potential goals for further research. One suggestion concerns the need

to identify an anchor to benchmark change in SASNOS against, thus allowing an investigation of external responsiveness. The main characteristic this should have is that at T_1 it needs to reflect access to activities restricted because of NBD, and at T_2 a reduction or absence of such restrictions. For example, a change in classification on the Supervision Rating Scale (Boake, 1996) that reflects increased autonomy (results correspond with the level of supervision required, these being independent, overnight, part-time, full-time indirect, and full-time direct supervision). A related issue that could also be addressed in this way is that the proxy measures of MIC determined from the systematic reviews of the literature were identified using health-related quality of life instruments in non-ABI populations. Aligning the size of a difference in repeated SASNOS assessments with an independent indicator of meaningful change relevant to ABI will help substantiate validity of the current MIC thresholds.

Further points to note are that SASNOS assessments were not synchronised and responsiveness data did not cover the entire duration of NbR, only a fixed period within this. Even so, the current results do nevertheless support the effectiveness of NbR in reducing NBD. However, investigating a sample in which SASNOS assessments were repeated on admission, at standard intervals thereafter, and at discharge, would potentially answer some of the questions asked earlier regarding the overall impact of NbR from admission to discharge, particularly regarding aggression, and difficulties with inhibition and communication.

Despite variability within the control group regarding timing of SASNOS assessments, analysis did not suggest these made a significant impact on the results obtained or conclusions reached. At the level of the group, time spent in NbR was not associated with change in NBD symptoms. Although detailed information regarding the sample was lacking, there were no obvious differences between rehabilitation participants rated as having more symptoms than observed in the general population form those that did not (with the exception of the 'Aggression' domain where there was a difference

in age at admission). However, given that the ABI population is so non-homogenous, including the extent of NBD and its multiple drivers, lack of linear relationships between time spent in NbR and other factors is not surprising as NbR interventions are necessarily highly configured to meet individual needs (Alderman, Knight and Brooks, 2013).

A final point concerns the lack of a reference group at this stage to enable tracking of the natural progression of symptoms of NBD amongst people with ABI who are not in receipt of specialist rehabilitation using the SASNOS. The existing literature suggests the natural evolution of NBD is associated with chronic difficulties. In contrast, reduction of symptoms was evident amongst rehabilitation participants, presumably because they were engaged in NbR. Availability of responsiveness data confirming the expected deteriorating trajectory evidenced from repeated measurement on the SASNOS amongst people with ABI who have not been in receipt of NbR would further affirm both the measure and effectiveness of NbR. It is also worthy of note that splitting cases within domains into those expected to change the most based on a score of less than 40 does provide some extra evidence regarding the approach, as the magnitude of the effect size and number of people improving was consistently greater amongst this group as opposed to those with NBD symptoms comparable at T_1 with neurologically healthy controls. NBR symptoms are also evident in the general population and it would be similarly useful to determine base rates of change in this group for comparison purposes.

In conclusion, this study presents evidence that SASNOS reliably measures change over time in NBD symptoms. This further extends the psychometric properties reported by Alderman, Wood and Williams (2011), confirming a role for use of the instrument by clinicians and researchers investigating the multidimensional outcomes arising from ABI.

REFERENCES

Alderman, N., & Wood, R.Ll. (2013). Neurobehavioural approaches to the rehabilitation of challenging behaviour. *NeuroRehabilitation*, 32, 761-770.

Alderman, N., Knight, C., & Brooks, J. (2013). Rehabilitation approaches to the management of aggressive behaviour disorders after acquired brain injury. *Brain Impairment (Special Issue: state of the art reviews on mental health in traumatic brain injury)*, 14, 5-20.

Alderman, N., Wood, R.Ll., & Williams, C. (2011). The development of the St Andrew's-Swansea Neurobehavioural Outcome Scale: validity and reliability of a new measure of neurobehavioural disability and social handicap. *Brain Injury*, 25, 83-100.

Boake, C. (1996). Supervision rating scale: a measure of functional outcome from brain injury. *Archives of Physical Medicine*, 77, 765-772.

Burke, H.H., Wesolowski, M.D., & Lane, I. (1988). A positive approach to the treatment of aggressive brain injured clients. *International Journal of Rehabilitation Research*, 11, 235-241.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences (*2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

Copay, A.G., Subach, B.R., Glassman, S.D., Polly, D.W. Jr., & Schuler, T.C. (2007) Understanding the minimum clinically important difference: a review of concepts and methods. *The Spine Journal*, 7, 541–546.

de Vet, H.C., Terwee, C.B., Ostelo, R.W., Beckerman, H., Knol, D.L., & Bouter, L.M. (2006). Minimal changes in health status questionnaires: Distinction between minimally detectable change and minimally important change. *Health and Quality of Life Outcomes*, 4, 54.

Domholft, E. (2005). *Rehabilitation research: principles and applications*. Elsevier Saunders, St. Louis, Missouri.

Donoghue, D., & Stokes, E.K. (2009). How much change is true change? The minimum detectable change of the Berg Balance Scale in elderly people. *Journal of Rehabilitation Medicine*, 41, 343-6.

Eisen, S.V., Ranganathan, G., Seal, P., & Spiro, A. (2007). Measuring Clinically Meaningful Change Following Mental Health Treatment. *Journal of Behavioral Health Services and Research*, 34, 272-289.

Hayley, S.M., & Fragala-Pinkham, M.A. (2006). Interpreting change scores of tests and measures used in physical therapy. *Physical Therapy*, 86, 735-743.

Hays, R.D., Brodsky, M., Johnston, M.F., Spritzer, K.L., & Hui, K. (2005). Evaluating the statistical significance of health-related quality-of-life change in individual patients. *Evaluation and the Health Professions*, 28, 160–171.

Heaton, R.K., Grant, I., & Matthews, C.G. (1991). *Comprehensive norms for an expanded Halstead-Reitan Battery: Demographic corrections, research findings, and clinical applications*. Odessa, Florida: Psychological Assessment Resources.

IBM Corp. (2013). *IBM SPSS Statistics for Windows*, Version 22.0. Armonk, NY: IBM Corp.

Kelly, G., Brown, S., Todd, J., & Kremer P. 2008. Challenging behaviour profiles of people with acquired brain injury living in community settings. *Brain Injury*, 22, 457-470.

Kreutzer, J. S., Marwitz, J. H., Seel, R., & Serio, C. D. (1996). Validation of a neurobehavioural functioning inventory for adults with traumatic brain injury. *Archives of Physical Medicine and Rehabilitation*, 77, 116-124.

MedCalc Statistical Software version 16.8. (2016). MedCalc Software bvba, Ostend, Belgium. https://www.medcalc.org

Middel, B., & van Sonderen, E. (2002). Statistical significant change versus relevant or important change in (quasi) experimental design: some conceptual and methodological problems in estimating magnitude of intervention-related change in health services research. *International Journal of Integrated Care*, 17, 1-18.

Norman, G.R., Sloan, J.A., & Wyrwich, K.W. (2003). Interpretation of changes in health related quality of life: The remarkable universality of half a standard deviation. *Medical Care*, 41, 582-592.

Oddy, M., & da Silva Ramos, S. (2013). The clinical and cost-benefits of investing in neurobehavioural rehabilitation: a multi-centre study. *Brain Injury*, 27, 1500-1507.

Rai, S.K., Yazdany, J., Fortin, P.R., & Avina-Zubieta, J.A. (2015). Approaches for estimating minimal clinically important differences in systemic lupus erythematosus. *Arthritis Research and Therapy*, 17, 143.

Rehabilitation Institute of Chicago (2010). *Rehabilitation measures database*. Retrieved from http://www.rehabmeasures.org/rehabweb/rhstats.aspx.

Ries, J.D., Echternach, J.L. Nof, L., & Gagnon Blodgett, M. (2009). Test-retest reliability and minimal detectable change scores for the timed "Up & Go" Test, the Six-Minute Walk Test and gait speed in people with Alzheimer Disease. *Physical Therapy*, 89, 569-579.

Stratford, P.W. (2004). Getting more from the literature: estimating the standard error of measurement from reliability studies. *Physiotherapy Canada*, 56, 27–30.

Tate, R.L. (2010). *A compendium of tests, scales, and questionnaires: The practitioner's guide to measuring outcomes after acquired brain impairment*. New York, NY: Psychology Press.

Terwee, C.B., Dekker, F.W., Wiersinga, W.M., Prummel, M.F., & Bossuyt, P.M. (2003). On assessing responsiveness of health-related quality of life instruments: guidelines for instrument evaluation. *Quality of Life Research*, 12, 349-362.

Walters, S.J., & Brazier, J.E (2003). What is the relationship between the minimally important difference and health state utility values? The case of the SF-6D. *Health and Quality of Life Outcomes*, 11, 1-4.

Walton, D.M., Macdermid, J.C., Nielson, W., Teasell, R.W., Chiasson, M., & Brown, L. (2011). Reliability, standard error, and minimum detectable change of clinical pressure pain threshold testing in people with and without acute neck pain. *The Journal of Orthopaedic and Sports Physical Therapy*, 41, 644-650.

Wood, R.Ll. (2001). Understanding neurobehavioural disability. In R.Ll.Wood & T.McMillan (Eds.), *Neurobehavioural disability and social handicap following traumatic brain injury* (pp.3-27). Hove, Psychology Press.

Wood, R.Ll., McCrea, J.D., Wood, L.M., & Merriman, R.N. (1999). Clinical and cost effectiveness of post acute brain injury rehabilitation. *Brain Injury*, 13, 69-89.

Wood, R.Ll., Alderman, N., & Worthington, A. (in press). Neurobehavioural rehabilitation. In M. Bodani, R. Faruqui and N. Agrawal (Eds.), *Oxford Textbook of Neuropsychiatry*. Oxford University Press, Oxford, UK.

Worthington, A.D., Matthews, S., Melia, Y. & Oddy, M. (2006). Cost-benefits associated with social outcome from neurobehavioural rehabilitation. *Brain Injury*, 20, 947-957.

Wyrwich, K.W., Tierney, W.M., Babu, A.N., Kroenke, K., & Wolinsky, F.D. (2005). A comparison of clinically important differences in health-related quality of life for patients with chronic lung disease, asthma or heart disease. *Health Services Research*, 40, 577–92.

Wyrwich, K.W., & Wolinsky, F.D. (2000). Identifying meaningful intra-individual change standards for health-related quality of life measures. *Journal of Evaluation in Clinical Practice*, 6, 39-49.

Wyrwich, K.W., Nienaber, N.A., Tierney, W.M., & Wolinsky, F.D. (1999). Linking clinical relevance and statistical significance in evaluating intra-individual changes in health-related quality of life. *Medical Care*, 37, 469–478.

Wyrwich, K.W., Tierney, W.M., & Wolinsky, F.D. (1999). Further evidence supporting an SEM-based criterion for identifying meaningful intra-individual changes in health-related quality of life. *Journal of Clinical Epidemiology*, 52, 861–873.

Ylvisaker, M., Turkstra, L., Coehlo, C., Yorkston, K., Kennedy, M., Moore Sohlberg, M., & Avery, J. (2007). Behavioural interventions for children and adults with behaviour disorders after TBI: A systematic review of the evidence. *Brain Injury*, 21, 769-805.

TABLES

SASNOS	T_1 $\bar{X}_1$ (SD)	T_2 $\bar{X}_2$ (SD)	t	p	r_{T1-T2}	SRM[1]
Total Score	37.93 (11.76)	43.18 (11.93)	6.053	<.001	.62	.57 *moderate*
Interpersonal Behaviour	26.37 (14.24)	34.36 (14.90)	7.454	<.001	.66	.75 *moderate*
Cognition	21.08 (13.33)	29.40 (15.26)	8.809	<.001	.69	.92 *large*
Inhibition	53.33 (10.74)	55.94 (9.24)	3.132	.002	.50	.26 *small*
Aggression	56.00 (12.65)	57.27 (12.74)	1.370	.173	.60	.15 *small*
Communication	60.86 (11.42)	63.53 (10.91)	3.724	<.001	.70	.40 *small*

Table 1: statistically significant differences in mean SASNOS T-scores and magnitude of effect size achieved at first and second assessment [[1]effect magnitude thresholds adjusted and shown in brackets to take into account r_{T1-T2} strength using Middel and van Sonderen (2002) solution].

SASNOS	MDC_{90}	MDC_{95}	SEM	0.5 SD	SRM[1]
Total Score	6.66	7.96	2.88	5.88	.57
Interpersonal Behaviour	8.72	10.42	3.77	7.12	.75
Cognition	6.17	7.38	2.67	6.67	.92
Inhibition	10.54	12.60	4.56	5.37	.26
Aggression	8.28	9.89	3.58	6.33	.15
Communication	8.76	10.47	3.79	5.71	.40

Table 2: MDC and MIC proxies calculated for the SASNOS total score and five principal domains [[1]adjusted effect magnitude thresholds shown in brackets].

SASNOS	MDC_{90}	MDC_{95}	SEM	0.5 SD	SRM
Total Score	38.6	34.5	53.1	40.7	46.9
Interpersonal Behaviour	37.9	34.5	59.3	42.8	49.7
Cognition	46.2	44.1	64.8	46.2	54.5
Inhibition	14.5	12.4	35.9	28.3	28.3
Aggression	23.4	18.6	32.4	24.8	27.6
Communication	17.9	14.5	43.4	30.3	43.4

Table 3: percent of individuals categorised at T_2 as achieving minimal detectable change and clinically meaningful improvement on the SASNOS using the various responsiveness measures.

SASNOS	MDC_{90}	MDC_{95}	SEM	0.5 SD	SRM
Total Score					
< 40	50.0	42.5	63.7	52.5	58.8
≥ 40	24.6	24.6	40.0	26.3	32.3
Interpersonal Behaviour					
< 40	43.5	40.0	64.3	47.8	54.8
≥ 40	18.5	14.8	44.4	25.9	33.3
Cognition					
< 40	49.3	47.0	68.2	49.3	58.2
≥ 40	16.7	16.7	33.3	16.7	16.7
Inhibition					
< 40	37.5	37.5	68.8	56.3	56.3
≥ 40	11.5	9.2	31.5	24.6	24.6
Aggression					
< 40	61.1	55.6	77.8	66.7	66.7
≥ 40	18.0	13.3	25.8	18.8	21.9
Communication					
< 40	50.0	33.3	66.7	50.0	66.7
≥ 40	16.2	13.4	41.5	28.9	42.3

Table 4: percent of individuals categorised at T_2 as achieving minimal detectable change and clinically meaningful improvement on the SASNOS using the various responsiveness measures. A between groups comparison has been added by categorising rehabilitation participants using their T_1 ratings (< 40 = T-score less than 1SD below the normative mean T-score of neurologically healthy controls, ≥ 40 = T-score within $\pm$1SD or greater than the normative mean T-score of 50).

		Total Score				Interpersonal Relationships				Cognition				Inhibition				Ag
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1
1	SEM																	
2	MDC_{90}	.61				.48				.49				.40				.72
3	MDC_{95}	.50	.87			.40	.89			.36	.84			.26	.78			.58
4	0.5 SD	.66	.95	.82		.60	.86	.75		.47	.98	.87		.85	.50	.35		.80
5	SRM	.82	.78	.66	.83	.81	.64	.55	.78	.75	.72	.57	.70	.85	.50	.35	1.00	.90

Table 5: weighted kappa values showing agreement amongst the five distribution based methods for the total SASN
the five domains.