Paper:

Alderman, N., Pink, A., Williams, C., Ramos, S., Oddy, M., Knight, C., Jenkins, K., Barnes, M. & Hayward, C. (2019). Optimizing measurement for neurobehavioural rehabilitation services: A multisite comparison study and response to UKROC. *Neuropsychological Rehabilitation*, 1-30.

http://dx.doi.org/10.1080/09602011.2019.1582432

Optimizing measurement for neurobehavioural rehabilitation services: A Multisite comparison study and response to UKROC

Nick Alderman[a,b*], Aimee E Pink[b,c], Claire Williams[b], Sara da Silva Ramos[d], Michael Oddy[d], Caroline Knight[e,f,g], Keith G Jenkins[h], Michael P Barnes[i], & Chloë Hayward[c]

[a]Elysium Neurological Services, Elysium Healthcare, Badby Park, Daventry, UK, [b]Department of Psychology, Swansea University, Swansea, UK, [c]Independent Neurorehabilitation Providers Alliance, UK, [d]Brain Injury Rehabilitation Trust, The Disabilities Trust, Burgess Hill, UK, [e] The Oakleaf Group, Northampton, UK, [f]School of Psychology, University of Leicester, Leicester, UK [g]Elysium Neurological Services, Elysium Healthcare, St Neots Hospital, Cambridge, UK, [h]National Brain Injury Centre, St Andrew's Healthcare, Northampton, UK, [i]Rake House Farm, Rake Lane, North Shields, UK

nick.alderman@elysiumhealthcare.co.uk, a.e.pink@swansea.ac.uk (https://orcid.org/0000-0003-1516-7922), Claire.williams@swansea.ac.uk (https://orcid.org/0000-0002-0791-744X), Sara.DaSilvaRamos@thedtgroup.org, Michael.Oddy@thedtgroup.org, caroline.knight@oakleafcare.com, KJenkins@standrew.co.uk, m.p.barnes@btinternet.com, info@in-pa.org.uk

Corresponding Author: Professor Nick Alderman

Email Address: nick.alderman@elysiumhealthcare.co.uk

Abstract Word Count: 200 words

Manuscript Word Count (excluding title page, abstract, references, tables, figures): 9,971

Abstract

To evaluate the efficacy of neurobehavioural rehabilitation (NbR) programmes, services should employ valid, reliable assessment tools; the ability to detect change on repeated assessment is a particular requirement. The United Kingdom Rehabilitation Outcomes Collaborative (UKROC) requires neurorehabilitation services to collect data using a standardised basket of measures, but the responsiveness and usefulness of using these in the context of NbR remains unknown. Anonymous data collected at two assessments for 123 people was examined using multiple methods to determine responsiveness of four outcome measures routinely used in NbR (HoNOS-ABI, FIM+FAM UK, MPAI-4, SASNOS). Predictive validity of two measures of rehabilitation complexity (RCS-E, SRS) regarding the extent of difference scores on these outcome measures at reassessment was also determined. All four outcome measures demonstrated responsiveness, with higher levels for SASNOS and MPAI-4 when only participants categorised as "*most likely to change*" at first assessment were analysed. Predictive validity of the RCS-E and SRS in estimating the extent of change was variable. SRS was only predictive of improvement on the MPAI-4 whilst RCS-E was not predictive at all. Recommendations are made regarding ideal characteristics of NbR outcome measures, along with the need to develop measures of rehabilitation complexity specifically conceptualised for these programmes.

Key Words: Neurobehavioural Rehabilitation; Acquired Brain Injury; Outcome Measurement; Assessment Tools; Rehabilitation Complexity; Responsiveness

Neurobehavioural disability (NBD) is the product of interactions between damaged neural systems, neurocognitive impairment and premorbid personality traits, exacerbated by post-injury learning and environmental influences (Alderman, Wood, & Williams, 2011). Behaviour disorders associated with NBD are enduring (Kelly, Brown, Todd, & Kremer, 2008), impose serious constraints on psychosocial recovery (Alderman & Wood, 2013; Kreutzer, Marwitz, Seel, & Serio, 1996), and create severe difficulties for families (Tam, McKay, Sloan, & Ponsford, 2015; Winkler, Unsworth, & Sloan, 2006). Fortunately, whilst traditional psychiatric or diagnostic approaches do not readily inform interventions after acquired brain injury (ABI), there is now a good evidence base demonstrating the efficacy of neurobehavioural rehabilitation (NbR; Alderman & Wood, 2013; Oddy & Ramos, 2013; Wood, McCrea, Wood, & Merriman, 1999; Worthington, Matthews, Melia, & Oddy, 2006; Ylvisaker et al., 2007).

NbR incorporates constructs, theories and procedures from cognitive, behavioural and social psychology to promote the acquisition and use of functional and social skills to reduce social handicap after ABI (Wood, 1990a, 1990b). Personal autonomy is maximised, and learning is enabled by promoting spontaneous application of new skills at every suitable opportunity. In addition, NbR also has a range of specific skills that distinguish it from other forms of medical rehabilitation (see Worthington & Alderman, 2017); two of which are especially relevant here. First, as the primary goals of NbR are psychosocial and rely on learning methods to counter less adaptive behaviour driven by neurocognitive impairment, NbR services are typically led by clinical neuropsychologists. Second, whilst rehabilitation is traditionally delivered by interdisciplinary or multidisciplinary teams, a transdisciplinary team (TDT) approach is favoured in NbR. To optimise delivery to address the complex, heterogeneous needs of people with ABI, input is provided by multiple disciplines working together to establish a 24/7 approach to rehabilitation. Roles are shared across disciplinary boundaries, facilitating communication, interaction and cooperation, with members committed to teach, learn and work together to implement coordinated services. Consequently, TDT working allows the development of a mutual vision or "shared meaning", resulting in a process of: (1) shared assessment and goal selection; (2) pooling of skills, knowledge and expertise, and (3) role release, where the entire TDT team implements interventions under the supervision of members whose disciplines are normally accountable for those practices (King et al., 2009). Thus, all team members are responsible for the attainment of rehabilitation goals via delivery of a consistent treatment program, which is not "session bound". The ultimate goal is to help socially functional behaviours to become established as habit, increasing the likelihood that such behaviours generalise to other environments and improve potential for social independence.

Given these unique characteristics, the net results of NbR include the creation and maintenance of a positive therapeutic milieu, where an enriched environment promotes constructive engagement with programme participants, mediates realistic expectations about what can be achieved, undermines neurocognitive determinants of challenging behaviour, achieves good outcomes, and represents a paradigm shift from a medical to a neuropsychological basis to rehabilitation. Indeed, following the introduction of the first specialist unit in 1979, NbR services in the UK have now evolved to provide several hundred beds for people with ABI and NBD through multiple care pathways.

Determining outcomes from NbR is a requirement that fulfils a range of needs, including demonstrating individual change, service level effectiveness, benchmarking, and research (see Alderman & Knight, 2017; Alderman, 2003). To achieve this, services routinely complete standardised “global” measures for all rehabilitation participants (Turner-Stokes, Williams, et al., 2012), and in the case of ABI, services also typically utilise a basket of outcome and other measures that reflect their speciality (Skinner & Turner-Stokes, 2006). In each instance, measures should be psychometrically sound, conceptualised specifically for ABI, relevant to the stage of recovery the service caters for, and integrated seamlessly into the clinical fabric of the service (see Alderman & Knight, 2017). Inevitably though, measures invariably and necessarily, differ across services.

However, services can also be compelled by an external agency (e.g. regulator, commissioning body) to measure efficacy via outcome measures that lack relevance, leading to dissonance between the service provider and agency regarding what is important, relevant, and how to measure it (Alderman & Knight, 2017). For example, the UK Rehabilitation Outcomes Collaborative (UKROC) database (Turner-Stokes, Williams, Bill, Bassett, & Sephton, 2016) currently requires all specialist neurorehabilitation services in England to administer a fixed basket of prescribed measures to elucidate the “black box of rehabilitation”, by providing information about case complexity, inputs required to meet this, and outcomes achieved. Outputs from UKROC are then used to categorise services by the complexity of cases they admit, creating a complexity-weighted tariff for rehabilitation beds using a multi-level payment model (Turner-Stokes et al., 2016). Whilst there undoubtedly is merit in such an approach, several concerns have been raised concerning the adequacy of the UKROC approach for services delivering NbR.

First, even though valid, reliable, and responsive global measures of NBD and challenging behaviour exist (see Alderman, Williams, Knight, & Wood, 2017; Alderman et al., 2011; Wood, Alderman, & Williams, 2008), such measures are not captured in the UKROC dataset.

Consequently, the risk is that NbR services may not be correctly classified as specialist services, with outcomes achieved not properly captured in the data collected. Second, owing to the TDT approach and psychological model of rehabilitation inherent to NbR, therapy is delivered in this context whenever there is an opportunity to do so, and clinical inputs are made by the whole team, not just by the primary discipline involved in overseeing an aspect of care. As instruments included in the UKROC database (e.g. Rehabilitation Complexity Scale-Extended version [RCS-E]; Turner-Stokes, Scott, Williams, & Siegert, 2012; Northwick Park Nursing Dependency Scale [NPNDS]; Turner-Stokes et al., 1998; Northwick Park Care Needs Assessment [NPCNA]; Turner-Stokes, Nyein, & Halliwell, 1999) are aligned to rehabilitation services underpinned by traditional medical models, their use is not appropriate when clinical inputs are delivered in other ways. In other words, attempting to capture clinical inputs in NbR using UKROC instruments arguably parallels "*fitting a square peg into a round hole*". As an approach, it is neither valid nor reliable.

The net consequence of such limitations is apprehension amongst providers that they will be disadvantaged (financial or otherwise), undermining their ability to sustain and deliver effective NbR programmes to address the needs of those presenting with NBD. Such unease is shared by many, including member services of the Independent Neurorehabilitation Providers Alliance (INPA; https://www.in-pa.org.uk/) which provide the majority (approximately 70%) of specialist neurorehabilitation services for people with chronic conditions in the UK, with the National Health Service (NHS) being the chief consumer. In light of this, the overarching goal of the study was to explore what measures could most usefully enable transparency for NbR services, and to determine the extent that measures used in neurorehabilitation services organised and delivered using a medical model can be usefully employed in services underpinned by a neuropsychological approach to rehabilitation. Specifically, we sought to: (a) examine whether four outcome measures routinely used in NbR have the psychometric properties (e.g. convergent and divergent validity) to effectively measure expected change in symptoms of NBD over time (e.g. responsiveness), and (b) determine the extent that change in ratings on two measures of rehabilitation complexity (including supervision needs) on reassessment predict change in status on outcome measures. A number of measures were explored to enable comparison of UKROC endorsed measures versus others that have special relevance to NbR.

Method

Participants

An anonymised database containing outcomes for 299 participants with ABI and in receipt of NbR was compiled by 14 member services of INPA. However, 176 participants were subsequently removed for the following reasons: missing age of admission or time since injury data (n = 112); < 18 years of age (n = 6); extreme values for time since injury (> 25 years) on admission (n = 5); not receiving NbR at time of audit (n = 23), and did not have data for at least one measure at both Time 1 (initial assessment - T_1) and Time 2 (reassessment - T_2; n = 30).

The final database comprised anonymised data for 123 participants of whom 78.86% (n = 97) were male. The majority of participants (n = 81; 65.85%) were admitted for NbR within 12 months of injury, and mean age on admission was 43.93 years (range = 18-79 years, SD = 15.32, Median = 48.0 years). Mean time since injury was 37.23 months (range = 0-252 months, SD = 69.19, Median = 5.0 years) and length of stay from admission to T_1 was available for 75 participants (M = 24.69 weeks, SD = 15.16, Median = 23.0 weeks, range = 1-61 weeks). Causes of injury are detailed in Table 1.

<TABLE 1 ABOUT HERE>

Measures

Mayo-Portland Adaptability Inventory – 4 (MPAI-4; Malec & Lezak, 2008): Recently added to the UKROC dataset (Turner-Stokes, 2016), the MPAI-4 was designed to assist in the clinical evaluation of individuals during the post-acute period following ABI. It consists of 29 core items rated on a 5-point scale, where zero represents no limitations and four represents a severe problem interfering with activity more than 75% of the time. Core items reflect common physical, cognitive, emotional, behavioural and social impairments and/or disability associated with ABI, reflecting the World Health Organisation (WHO) distinctions between Impairment, Activity, and Participation. Self- and proxy-completed versions are available, and items are groups into three subscales: Ability (e.g. mobility, motor speech), Adjustment (e.g. anxiety, inappropriate social interaction), and Participation (social contact, self-care). Although not included in total scale scores, six additional items capture pre-existing and associated conditions. Raw scores are converted to T-Scores with a mean of 50 and standard deviation of 10, with higher scores indicative of more severe problems. Reference data can be drawn from either a "*National*" or "*Mayo*" sample, as appropriate (Malec & Lezak, 2008), and prior studies have demonstrated satisfactory levels of internal consistency, as well as good levels of construct, concurrent, and predictive validity (Malec et al., 2003; Tate, 2010).

Functional Independence Measure and Functional Assessment Measure UK – 2.2 (FIM+FAM; Turner-Stokes, Nyein, Turner-Stokes, & Gatehouse, 1999): Designed to provide a global index of disability, the UK FIM+FAM consists of 30 items evaluated on a 7-point ordinal scale, ranging from totally dependent (*1*) to completely independent (*7*), with higher total scores indicative of less disability. Items are organised across six subscales (nine self-care; seven transfers and mobility; five communication; four psychosocial, five cognition), although items tend to load onto two main factors: Motor (16 items, score range 16-112), and Cognitive (14 items, score range 14–98; Nayar, Vanderstay, Siegert, & Turner-Stokes, 2016; Turner-Stokes & Siegert, 2013). For this reason, only Total, Motor and Cognitive scores are utilised in the current study. In addition, a separate 6-item “Extended Activities of Daily Living” scale can be used (EADL; e.g. meal preparation, housework). The FIM+FAM has been part of the UKROC dataset since its inception and has robust psychometric properties (see Tate, 2010, pp. 440–441; Turner-Stokes & Siegert, 2013; Nayar et al., 2016). Moderate levels of responsiveness have also been reported in samples of patients undergoing specialist rehabilitation following stroke (Nayar et al., 2016) and in general inpatient neurorehabilitation populations (Turner-Stokes & Siegert, 2013).

Health of the Nation Outcome Scale – ABI (HoNOS-ABI; Fleminger et al., 2005): Representing an adaptation of the original Health of the Nation Outcomes Scale, the HoNOS-ABI was developed to assess the neuropsychiatric sequalae of ABI. Twelve items from four domains (behavioural, impairment, symptoms and social) are rated on a zero (“*no problem*”) to four (“*severe/very severe problem*”) point scale. Total scores range from 0-48, with higher scores indicating more severe problems. Psychometric data is somewhat limited, but adequate levels of criterion validity (Coetzer & du Toit, 2001) and inter-rater reliability (Fleminger et al., 2005) have been demonstrated.

St Andrews-Swansea Neurobehavioural Outcome Scale (SASNOS; Alderman et al., 2011): The SASNOS (self- and proxy-completed versions) was created specifically for use in ABI using a conceptual framework underpinned by the WHO International Classification of Functioning, Disability and Health. It consists of 49 items that capture five major domains of

NBD (Interpersonal Relationships, Cognition, Aggression, Inhibition and Communication), each with two- to three subdomains. Each item consists of a statement regarding a feature of NBD whose perceived prevalence is rated using a 7-point scale ("*never*" to "*always*"). Using normative data from a moderately sized sample of neurologically healthy controls as a reference group, ratings are transformed to standard scores with a mean of 50 and standard deviation of 10; higher scores reflect greater perception of ability and fewer symptoms of NBD. The SASNOS has robust psychometric properties (inter-rater and test-retest reliability; content construct, convergent, divergent, and discriminant validity, responsiveness; Alderman et al., 2017, 2011), and a supplementary scoring system overcoming difficulties in conveying the impact of context-dependent support (Alderman, Williams, & Wood, 2018).

Rehabilitation Complexity Scale – Extended (RCS-E; Turner-Stokes, Scott, et al., 2012): A straightforward measure of the complexity of rehabilitation needs and interventions, and part of the UKROC database. Items load onto five subscales (Care/Risk; Nursing; Therapy; Medical; Equipment/Facilities), with items rated on three- to five-point scales (total score range 0-20). As the RCS-E is completed by clinicians every two weeks, only assessment scores resulting from T_1 and T_2 were considered in the current analysis. Moderate levels of internal consistency and discriminant validity have been found, with clinicians also reporting favourably on utility, content and face value (Turner-Stokes, Williams, & Siegert, 2012). Scores on previous iterations of the measure (i.e. RCS version 2) have also been shown to provide a moderately responsive profile of rehabilitation interventions delivered in a tertiary post-acute rehabilitation service for younger adults with severe complex neurological disabilities (Turner-Stokes, Williams, & Siegert, 2010).

Supervision Rating Scale (SRS; Boake, 1996): The SRS is an observer rated tool to measure the level of supervision an individual receives from caregivers. Level of supervision is rated on a 13-point ordinal scale that can be optionally grouped into five ranked categories: independent; overnight supervision; part-time supervision; full-time indirect supervision, and full-time direct supervision. Ratings should reflect supervision received at the time of the assessment rather than predicted needs, and thus the cumulative impact of an individual's impairment and in turn, financial costs to the person being rated, caregivers, and funders. SRS is reported to show good-to-excellent inter-rater reliability concurrent and discriminant validity (see Tate, 2010, pp. 588-590), and can detect change beyond measurement error in the context of post-acute rehabilitation (Reed et al., 1999).

Procedure

This study retrospectively examined anonymised data from a basket of outcome measures and two measures of rehabilitation complexity routinely administered by member services of INPA to fulfil contractual, clinical, and other requirements. Participating member services were asked to consider contributing cases which had been evaluated twice, with both test (T_1) and re-test (T_2) assessments completed on any of the measures pertaining to this study. In addition, for any cases submitted, services were asked to provide standard demographic and injury related characteristics (e.g. gender, age on admission, cause of injury). Services were not required to implement measures they were not already using, only to consider submitting previously collected data that might inform study objectives. For this reason, the INPA Research and Outcomes Group (who oversaw the project) determined that the work constituted service evaluation (see: http://www.hqip.org.uk/guide-for-clinical-audit-research-and-service-review/). All participating services signed a service level agreement outlining the aims of the project, method, anticipated dissemination of outputs, and governance parameters for the work. The latter included assurances that: (a) neither service users nor services would be identifiable in the final dataset; (b) data would be stored on encrypted, password protected computers and only made available to specified members of the group with responsibility for data analysis, and (c) data would only be used for the purposes for which it was obtained.

Statistical Analysis

All Analyses were undertaken using SPSS 22.0 and Medcalc. With the exception of the SASNOS and MPAI-4, variables were not normally distributed. However, as attempts to normalise distributions can cloud subsequent interpretation (Feng et al., 2014), transformations to normalise data were not conducted. Instead, both parametric and non-parametric statistical tests were employed as appropriate. Use of parametric procedures was additionally justified through acknowledgement in the literature that the Likert type scales comprising the instruments investigated here comprises a level of measurement that falls between ordinal and interval, and that value, range and variability of data increases when indexes are derived from summing ratings across multiple items in the form of total and subscale scores (see Allen & Seaman, 2007; Warner, 2018).

Validity was explored by determining the strength of linear associations between the various measures using Spearman's rho correlations, interpreted using criteria proposed by Cohen (1988) where correlation coefficients between .10 and .29 represent a small/weak association, .30 to .49 medium, and .50-1.0 large/strong. These cut-offs have been deemed to be appropriate, if not conservative, for use in psychological research (Hemphill, 2003). Responsiveness was explored via two methods: (1) paired t-tests or Wilcoxon Signed Ranks to compare measures of central tendency at T_1 vs. T_2, and (2) measures of effect size (ES). Originally proposed by Cohen (1988), ES is routinely used to investigate responsiveness in clinical and repeated measures contexts. 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. ES is independent of sample size and overcomes many of the difficulties associated with interpreting statistical significance between measures of central tendency. For example, high significance levels can be found despite relatively small differences between means. Consequently, it is too great an assumption to infer statistical change equates to meaningful change to relevant stakeholders and results cannot be applied to the level of the individual rehabilitation participant (Alderman et al., 2017). Additionally, statistical significance implies there is a difference in scores on the same measure, which in this context is assumed to be attributable to NbR. However, statistical significance does not say anything about the size or magnitude of the difference, or in turn, the relative ability of a tool to measure change in symptoms of NBD over time; a key objective of the study.

In contrast, the magnitude of the effect can be determined with ES by applying cut-off thresholds: < .20 "*trivial*"; ≥ .20 to < .50 "*small*"; ≥ .50 to < .80 "*medium*"; ≥ .80 "*large*" ES (Cohen, 1988). However, Middel and van Sonderen (2002) noted that the strength of the correlation between T_1 and T_2 varies between samples in repeated measures designs, potentially leading to an over or underestimation of classification of the size of effects. Instead, they suggested applying an additional calculation to calibrate thresholds according to the size of the T_1–T_2 correlation, and Norman and collegaues (2003) recommended that a "*medium*" ES corresponds to clinically meaningful change when using SRM as a proxy measure of minimally important change (see Alderman et al., 2017, p.3). That is, the difference between test scores needed to indicate meaningful and practical change.

Regarded as an appropriate measure of an instrument's predictive validity (Snowden & Gray, 2010), ROC analysis was used to determine the predictive validity of the SRS and RCS-E. ROC produces a plot of the proportion of correct predictions against the proportion of "*false alarms*", utilising a binary outcome measure. In this case, a T_1-T_2 difference on any of the four outcome measures resulting in the classification of a participant as "*same/declined*" versus "*improved*" (determined when the SRM for an individual participant met or exceeded the "*medium*" ES threshold). ROC is quantified by calculating the area under the curve (AUC) and is interpreted as the probability that a randomly selected participant from the "*improved*" pool will have a lower score on the SRS and RCS-E than an individual selected at random from the "*same/declined*". An AUC of 0.50 indicates accuracy is equal to chance, with a minimum AUC of 0.70 conventionally considered the minimum value necessary for an instrument to have good predictive validity.

However, a limitation of ROC is that it predicts yes/no relationships from dichotomous data, yet both predictors in the current study (RCS-E and SRS) are incremental. Therefore, stepwise regression analysis was also undertaken to determine the best prediction models using SRS and RCS-E. That is, regression analyses were used to enable the magnitude of difference scores on the outcome measures to be evaluated using more sensitive measures of complexity of rehabilitation needs through an incremental representation of data from these measures. However, this analysis was restricted to measures that were normally distributed (i.e. MPAI-4 and SASNOS) and to Total Scores.

Results

Responsiveness of Measures

MPAI-4: scores decreased significantly from T_1 to T_2, suggesting a reduction in impairment following NbR (Table 2). SRM scores for Total, Adjustment and Participation attained the ES medium threshold associated with meaningful improvement.

<TABLE 2 ABOUT HERE>

A number of significant correlations were found between T_1 and T_2 MPAI-4 scores and principal summary scores from the other outcome measures, with the strength of correlations typically stronger at T_2. However, only small significant negative correlations were found between Adjustment, FIM+FAM and EADL scores at T_1, with no significant correlations at T_2. Correlations between MPAI-4, RCS-E and SRS scores were variable, with the majority not significant at T_1. However, these increased in size at T_2, were in the predicted direction, and achieved statistical significance (Table 3).

<TABLE 3 ABOUT HERE>

St. Andrews-Swansea Neurobehavioural Outcome Scale: Mean SASNOS Total and domain scores all increased from T_1 to T_2. However, statistically significant differences were only found for Total, Interpersonal Relationships, Inhibition and Cognition scores, and magnitude of effect size varied across SASNOS domains. Specifically, a large and medium ES was found for Cognition and Interpersonal Relationships respectively, but none of the remaining domains achieved SRMs associated with meaningful improvement (Table 2).

SASNOS Total, Interpersonal Relationships and Cognition domain scores correlated significantly with all other outcome measures (Table 4). However, correlations with the remaining three SASNOS domains were more variable. For example, significant correlations (small–large) were found for Aggression, whilst Communication domain scores did not significantly correlate with either FIM+FAM, EADL, or MPAI-4 scores at T_1. A similar pattern of results was observed at T_2, although correlations with EADL scores were generally weaker. In comparison to the MPAI-4 (Table 3), there were stronger correlations on the whole between SASNOS and measures of case complexity at both T_1 and T_2. However, no significant correlation was observed between Inhibition and Communication domain scores and RCS-E at T_1 (Table 4).

<TABLE 4 ABOUT HERE>

FIM+FAM UK (including EADL): Measures of central tendency significantly increased from T_1 to T_2, suggesting significantly greater autonomy at T_2. However, only the Cognitive and EADL subscales achieved the medium ES threshold (Table 5).

<TABLE 5 ABOUT HERE>

Significant correlations were also found at T_1 and T_2 between FIM+FAM and principal scores on other measures, and whilst typically stronger at T_2, these were all in the expected direction. The majority of correlations between FIM+FAM, SRS and RCS-E scores at T_1 and T_2 were also significant. However, at T_1, no significant correlation was found between the motor subscale of the FIM+FAM and SRS, and the significant correlation between Total FIM+FAM and SRS was small (Table 6).

<TABLE 6 ABOUT HERE>

HoNOS – ABI: Measures of central tendency decreased significantly from T_1 to T_2, suggesting some remediation in neuropsychiatric difficulties. ES also reached the medium threshold to indicate meaningful improvement (Table 5).

At both T_1 and T_2, large significant correlations were found between HoNOS-ABI and principal scores on other measures, and all were in the expected direction. Small significant positive correlations were also found between the HoNOS-ABI and both measures of rehabilitation complexity at T_1, with the strength of correlations improving at T_2 (Table 7).

<TABLE 7 ABOUT HERE>

Rehabilitation Complexity Scale – Extended: Despite only a small reduction in mean scores (range = 0.47 to 1.10) and identical median RCS-E scores at T_1 and T_2, Total, Care/Risk and Therapy mean scores were statistically lower at T_2. Wilcoxon Signed Ranks tests revealed that a greater number of patients were given lower Total ratings at T_2 for Total compared to an increase or no change in scores (Total: 52% vs. 14.7% vs 33.3%). A greater number of participants scores also remained unchanged at T_2 on the Care/Risk subscale (56.8% vs. 36.3% reduction vs. 6.9% increase); and the significance of the 'Therapy' subscale score appears to be due to the equally high proportion of participants receiving a lower rating and those whose rating did not change (45.1% vs. 45.1% vs. 9.8% increase). No other statically significant T_1 versus T_2 differences were found, and only the Therapy subscale attained the medium threshold corresponding to meaningful and practical change (Table 5).

With the exception of Equipment, no significant correlations were found between the RSC-E and MPAI-4 total or subscale scores. Additionally, apart from trivial correlations between SASNOS Total, RSC-E Equipment and Therapy items, correlations with FIM+FAM (Total and EADL) and SASNOS ranged from small to large at T_1. Four (Total, Care/Risk Nursing, Medical) of the six RCS-E items significantly correlated (large) with the SRS, and three small (Medical) to medium (Total, Nursing) significant correlations were also found with the HoNOS-ABI. A similar pattern of correlations was evident at T_2, except that stronger correlations were found between the RCS-E, FIM+FAM, HoNOS-ABI and SRS (Table 8).

<TABLE 8 ABOUT HERE>

Supervision Rating Scale: Measures of central tendency decreased significantly from T_1 to T_2, suggesting a reduction in the level of supervision required. Most participants were assigned a lower ranking at T_2 (69%) rather than the same (26.8%) or a higher ranking (4.2%). In addition, a medium ES was found, suggesting that the statistical significance attained also had practical significance (Table 5).

At T_1, SRS significantly correlated with all other outcome measures except the MPAI-4. In contrast, all correlations were significant at T_2 (Table 9).

<TABLE 9 ABOUT HERE>

Responsiveness of Outcome Measures with a Reference Group

Discriminating participants into different groups (e.g. least/most likely to change/respond to rehabilitation) is widely reported in the literature and potentially enhances responsiveness by targeting participants in which change is expected (Alderman et al. 2017; Rai, Yazdany, Fortin, & Aviña-Zubieta, 2015; Walters & Brazier, 2003). Discerning between participants in this way is usually achieved by means of a reference group, typically comprising of healthy individuals drawn from the general population or clinical participants who are stable or otherwise not expected to change. Therefore, given the availability of reference group data for both the SASNOS and MPAI-4, the responsiveness of these two measures was examined further.

SASNOS: The SASNOS was specifically designed to measure symptoms of NBD, and in turn, response to NbR. However, in our sample of NbR participants, Total SASNOS scores only demonstrated a small ES. In addition, even though cognitive impairment, difficulties with interpersonal function, deficits of communication, inhibition and aggression are all frequent legacies of ABI, responsiveness indices varied substantially across the corresponding SASNOS domains (Table 2).

Similar findings were reported by Alderman et al. (2017), who highlighted that assessments made at T_1 are not necessarily undertaken at the point of admission and that different characteristics of NBD are likely to respond to rehabilitation at different stages of participation. Indeed, evidence suggests that aggression and very challenging behaviours are most amenable to change in the early phases of rehabilitation. Consequently, by the time that T_1 and T_2 assessments were completed in the current study, some symptoms may have already responded to NbR to the point that further change would be unlikely. Additionally, given the non-homogeneous nature of ABI, some symptoms of NBD are more endemic than others (i.e. not everybody in an NbR programme will have been admitted because of aggression or lack of inhibitory control). Consequently, this might help explain why change in scores on the SASNOS Inhibition and Aggression domains in the current study only corresponded to a small or trivial ES, respectively.

To overcome such difficulties, Alderman et al. (2017) discriminated between participants "*most likely to change*" (T_1 T-score < 40; one SD or more below the mean for neurologically healthy controls) versus those "*least likely to change*" (T_1 T-score ≥ 40), finding higher effect sizes for four of the six SASNOS scores in the group *"expected to change"* compared to values obtained from the whole sample. This method was also applied here, with responsiveness recalculated for four of the five SASNOS domains (N.B. Communication was not reinvestigated as only one case scored < 40 at T_1).

Results from this analysis (see Table 10) revealed statistically significant T_1-T_2 differences across all SASNOS domains in the group "*most likely to change*", with higher scores at T_2 suggesting a reduction in NBD symptoms. Indeed, mean scores for Inhibition and Aggression fell within the expected range for neurological healthy controls at T_2 (Alderman et al., 2011), suggesting real improvement following NbR. ES remained unchanged for Cognition (large) but increased from small to large for Total (.51 vs. 1.10), medium to large for Interpersonal Relatonships (.64 vs. .73), and trivial to large for both Aggression (.19 vs. 1.01) and Inhibition (.28 vs. .67). Therefore, participants in the "*most likely to change*" group who demonstrated significant NbR symptoms at T_1 showed real improvement at T_2.

In contrast, mean scores changed very little from T_1 to T_2 in participants categorised as "*least likely to change*", and even though the Interpersonal Relationships domain met the medium threshold (SRM = .61) to indicate meaningful change, this was still lower than observed in the "*most likely to change to change*" group.

<TABLE 10 ABOUT HERE>

MPAI-4: A similar approach was employed with the MPAI-4. However, as scores on this measure are standardised against two samples ("*Mayo*'" and "*National*"; Malec & Lezak, 2008) comprised of people with ABI in rehabilitation programmes rather than a neurological healthy reference group, no clear criteria for discriminating those "*most/least likely to change*" from T_1 to T_2 was available. Instead, cut-offs were aligned with the descriptive categories for MPAI-4 T-scores. Scores <30 are associated with "*relatively good outcomes*" and arguably, less potential for further rehabilitative gains. However, only one participant in the current sample had a T_1 score <30. Consequently, the following cut-offs were used instead: (1) "*least likely to change*" (T_1 T-score < 50; mild-moderate limitations), and (2) "*most likely to change*" (T_1 T-score > 50; moderate-severe limitations).

Mean MPAI-4 scores were significantly lower at T_2, suggesting a reduction in disability across both groups (see Table 11). Within group adjusted SRMs were also generally higher than those observed for the whole sample combined, although the magnitude of ES remained unchanged in the "*least likely to change*" group. In contrast, in the group considered "*most likely to change*", ES improved from medium to large for Total (.56 vs. .85), Adjustment (.51 vs. .76), and Participation (.67 vs. .88), and from small to medium for Ability (.38 vs. .73).

<TABLE 11 ABOUT HERE>

Predictive validity of measures of case complexity, needs and supervision

All four outcome measures demonstrated some degree of responsiveness, with improved scores at T_2 suggesting positive and meaningful response to NbR. However, a commensurate reduction in case complexity was inconsistently observed across the SRS and RCS-E. Namely, even though participants were rated as requiring less support on the SRS at T_2 (medium ES), findings on the RCS-E were generally inconsistent with the various improvements evidenced across the four outcome measures. Consequently, the extent to which the RCS-E and SRS are able to predict response to NbR, as captured by change in status on reassessment across the four outcome measures was explored further.

First, we examined whether there was a commensurate reduction in scores on the SRS and RCS-E for participants classed as having made improvements from T_1 to T_2 on each outcome measure. To enable this, progress made on each outcome measure by individual participants was categorised using the criteria used by Eisen, Ranganathan, Seal, and Spiro (2007) and Alderman et al. (2017). Individual SRM's were calculated for each participant and compared on each measure to the minimum threshold (Eisen et al. 2007) corresponding to a medium ES (medium ES cut-off determined using the additional calculation that takes into account the size of the T_1-T_2 correlation advocated by Middel and van Sonderen, 2002). Individuals attaining or exceeding the minimum threshold were classed as "*improved*", with remaining participants classified as "*same/declined*". To reduce likelihood of type 1 error, only Total scores for each outcome measure were used. This method was also applied to those participants previously categorised as most likely to improve on the SASNOS and MPAI-4, where availability of reference group data made it possible to further refine responsiveness on these measures (Table 12).

<TABLE 12 ABOUT HERE>

The proportion of participants categorised as "*improved*" at T_2 varied from 29.4% to 48.3%, with this proportion increasing further for the SASNOS (74.1%) and MPAI-4 (65.7%) when only participants "*expected to change*" were considered. On the basis that participants categorised as "*improved*" should evidence a concurrent decrease in rehabilitation complexity and supervision needs, we expected this group to demonstrate a greater reduction in SRS and RCS-E scores than those categorised as "*same/declined*". Consistent with this, mean T_2-T_1 difference scores for the SRS differed significantly between the "*improved*" and "*same/declined*" groups as determined by the MPAI-4 (*mean difference* = 3.13, t (61) = 4.581, $p < .001$) and SASNOS (*mean difference* = 1.43, t (54) = 1.790, $p = .040$). In addition, significant within group differences were evident on the SRS for participants categorised as "*most expected to change*" from assessment at T_1 on the MPAI-4 (*mean difference* = 2.23, t (31) = 2.848, $p = .004$ and SASNOS (*mean difference* = 1.74, t (24) = 1.883, $p = .036$). In both instances, those categorised as either "*improved*" or "*expected to change*" evidenced a greater concurrent decrease in supervision needs. In contrast, T_2-T_1 change scores on the RCS-E did not differ across either of these grouping variables, suggesting no concurrent decrease in rehabilitation complexity in the group evidencing meaningful change from T_1-T_2 on the MPAI-4 and SASNOS.

Next, we employed ROC analysis to determine whether change in ratings on the SRS and RCS-E could successfully predict the likelihood of being improved on each of the four outcome measures at T_2 (Table 13). The non-parametric ROC method was used to calculate AUC, and in each analysis, the outcome variable was the dichotomous variable categorising participants as either "*improved*" versus "*same/declined*", and the predictor variable was the difference score on the SRS and the RCS-E from T_2-T_1. Regarding the SRS, the magnitude of the difference score was found to be "fairly' predictive of improvement on the MPAI-4 when all cases were considered and for the subgroup classed as being *"most expected to change"* at T_1. However, the SRS difference score did not attain the minimum AUC for any of the other outcome measure, including the subgroup "*most expected to change*" at T_1 on the SASNOS. Likewise, the RCS-E difference score was not found to be predictive of improvement on any of the four outcome measures.

<TABLE 13 ABOUT HERE>

Finally, stepwise regression analysis was undertaken to determine whether SRS and RCS-E scores could predict MPAI-4 and SASNOS difference scores. There was a large correlation between MPAI-4 difference scores and the SRS (.67), but not RCS-E (.11) in the whole sample combined. In line with this, stepwise regression analysis retained the SRS but not RCS-E. The model was statistically significant (F (1,42) = 33.78, p < .001), accounting for approximately 45% of the variance of the change in scores observed between T_1 and T_2 on the MPAI-4 (R^2 = .446, Adjusted R^2 = .433). In the subgroup *"most expected to change"*, there was a large correlation between the SRS and MPAI-4 difference scores (.58), but only a trivial one between RCS-E and MPAI-4 (-.07). Stepwise regression retained the rejected RCS-E (F (1,22) = 11.316, p = .003), with the model accounting for approximately 34% of the variance (R^2 = .340, Adjusted R^2 = .310). With regard to the SASNOS, difference score correlations with SRS (.04) and RCS-E (.11) were trivial in the whole sample combined, and both predictor variables were removed in the first step. A similar finding was made when the subgroup "*most expected to change*" were considered separately, with very low correlations evident between SASNOS with SRS (.16) and RCS-E (-.07), with neither of them passing step one of the regression analysis.

Discussion

Determining outcomes from NbR fulfils a number of key objectives for a range of stakeholders; with results holding potential consequences for people needing these services, including programme sustainability through funding and financing. Therefore, in any evaluation, it is imperative that services utilise a basket of outcome and other measures that are reliable and valid in the context of NbR, are appropriate for the population, its needs, and the purported aims of the service. However, tension can arise when services are compelled by external agencies, such as UKROC, to measure efficacy via outcome measures that potentially lack relevance. Consequently, the overarching goal of this study was to explore what measures could most usefully help enable transparency for NbR services, and to determine the extent that measures used in neurorehabilitation services organised and delivered using a medical model can be usefully employed in services underpinned by a neuropsychological approach to rehabilitation.

Validity and responsiveness of outcome measures routinely used within NbR

The first major aim of the study was to examine whether four outcome measures routinely used in NbR have the statistical properties to effectively measure expected change in symptoms of NBD over time. The four outcome measures examined here contained items representing the range of physical, cognitive, emotional, behavioural and social problems associated with ABI (FIM+FAM, MPAI-4), neuropsychiatric outcome (HoNOS-ABI), as well as NBD symptoms and associated social handicap (SASNOS). Satisfactory levels of convergent validity were found, with medium to large significant correlations observed between measures at T_1 (.47 to .65) and T_2 (.49 to .77). Evidence of divergent validity was also apparent when various subscales measuring specific areas of outcome were compared across measures. For example, there were trivial correlations between SASNOS Inhibition and FIM+FAM Motor scores at T_1 (.14), and between SASNOS Communication and FIM+FAM Motor scores at T_2 (-.09).

Additionally, all four outcome measures evidenced some degree of responsiveness. First, large statistically significant differences were found for 16 of the 17 comparisons made using the four outcome measures, with only the Aggression domain of the SASNOS failing to achieve statistical significance. However, methods examining statistically significant change from T_1 to T_2 are arguably of limited value, as relatively small differences between means can result in very high significance levels. In addition, statistically significant changes from T_1 to T_2 reveals very little about the magnitude of change on an individual level, and thus, do not help services to demonstrate effectiveness.

Second, we also employed a measure of ES, applicable here as it articulates the size of any differences between scores within a measure evident on reassessment, enabling comparison of its relative ability to detect change as a consequence of NbR. Specifically, attainment of a medium or higher ES reflects a clinically significant change that is meaningful and of practical benefit. Applying this method to the four measures was instructive, as despite highly significant T_1 to T_2 differences; a medium or higher ES was only found for eight of the 17 comparisons. Further examination revealed that all four measures proved to be responsive in tracking response to NbR, although this varied substantially in the three outcome measures containing various subscales. The HoNOS-ABI has only one score which attained a medium ES. In contrast, only three of the four scores from the MPAI-4, and only the Cognitive and EADL subscales from the FIM+FAM, attained the minimum medium threshold to indicate meaningful change. Although, the degree of change observed on individual subscales may naturally vary depending on the focus of NbR programmes. However, perhaps the most surprising finding was that four of the six comparisons on the SASNOS did not attain a medium ES, even though this scale was conceptualised and designed specifically to measure NBD with the intention of tracking response to NbR.

Benefits of Using Measures Which Have a Reference Group

A potential limitation associated with pooling data collected on an outcome measure for comparative purposes is the inclusion of ratings for participants who are not experiencing difficulties. Multiple outcomes are associated with ABI, with survivors constituting a complex, non-homogenous population. Thus, it should not be assumed that all rehabilitation participants present with all potential consequences, including symptoms of NBD. Consistent with this, Alderman et al. (2018) previously described ten unique combinations of SASNOS domain profile scores, confirming that not all participants in NbR have uniform difficulties. For instance, 66% had a profile characterised by difficulties with 'Interpersonal Relationships' and 'Cognition', whilst only 2% were rated as having problems across all five SASNOS domains.

In the case of SASNOS, the ability to identify symptoms that are problematic is possible by the availability of data from a neurologically healthy reference group. As is the case with cognitive function, the frequency and severity of NBD symptoms falls on a continuum and using a reference group to establish cut-offs is particularly valuable. Alderman et al. (2017) previously noted that the responsiveness of the SASNOS was poor, but argued that only assessments where T_1 scores fell below the normal range for the reference group should be included when determining responsiveness. This is because participants with scores in the normal range may have already responded positively to NbR at T_1, or may simply not have been experiencing difficulties in a particular domain as a consequence of ABI in the first place. However, it should be noted that the current study is not an investigation of the effectiveness of NbR per se, but an exploration of measures that may be appropriate to inform such an ambition. The former may be investigated by comparison of T_1 scores taken shortly after admission to a programme, with T_2 collected at discharge. In contrast, scores collected in this study were sampled at various times during admission. Nevertheless, when specifically considering participants "*most likely to change*" at T_1, responsiveness indices for the SASNOS improved, with ES estimates increasing to medium or large. In contrast, meaningful change was only attained for the SASNOS Interpersonal Relationships domain amongst the group "*least expected to change*". This likely reflects gains from the strong positive therapeutic climate in NbR, where continued exposure to rich and affirmative environments may further benefit social functioning and relationships, even if these are compatible with the expected normal range when first assessed.

Reference group data is also available for the MPAI-4, which is drawn from two samples of people with ABI participating in post-acute rehabilitation. However, the description of these samples lack detail, rendering difficulties with benchmarking compatibility of programmes and participants with those in this study. Consequently, determining a threshold for the MPAI-4 scores in order to establish when assessments of Ability, Adjustment and Participation reflect scores that might be expected in the neurologically healthy population was not possible. Instead, participants assessed as having "moderate–moderate-to-severe" limitations (T_1 T-score > 50) were categorised as those "*most likely to change*", with those with "mild–mild-to-moderate limitations" (T_1 T-score < 50) categorised as those "*least likely to change*". The extent of the T_1-T_2 difference between scores confirmed this was the case, as whilst three of four MPAI-4 scores met the minimum medium ES threshold when assessments from the whole sample were considered; all four met or exceeded the minimum threshold when examined in the group deemed "*most likely to change*". The MPAI-4 also demonstrated responsiveness in the "*least likely to change*" group, but ES levels were comparable to those obtained for the whole sample and smaller than those obtained in the group "*most likely to change*".

To conclude, participants categorised as the "*most likely to change*" at T_1 evidenced larger gains, with ES exceeding the threshold for meaningful change on both the SASNOS and MPAI-4. Appraisal of the responsiveness indices for those *"least likely to change"* is also interesting, as difference scores for only one of five SASNOS comparisons was indicative of meaningful change, whereas three of the four MPAI-4 comparisons reflected this. The greater specificity regarding expectations about responsiveness from the SASNOS reflects the clarity obtained from having a reference group comprised of neurologically healthy controls as opposed to other people with ABI for the MPAI-4.

Recommendations Regarding Outcome Measures for NbR Services

All four outcome measures examined in this study demonstrated responsiveness and are capable of capturing individual change through a reduction in impairment and increase in autonomy. Thus, all could be usefully employed in the context of NbR. However, even though all four measures overlap in what they measure, there are important differences between them that need to be considered by potential users. In addition, personal preferences will further influence choice.

For example, some measures include multiple subscales enabling a more detailed overview of outcomes arising from ABI, or a more comprehensive investigation of a specific area of interest. In contrast, HoNOS-ABI has a single output which may be particularly useful in a screening context where primary interest is in neuropsychiatric outcome, such as in a secure or forensic ABI rehabilitation service, or when time-constraints necessitate rapid assessment. That said, reviewing scores from individual items may also provide important clinical information, although the psychometric properties of these are less known and are likely to be weaker than those for the whole scale. In comparison, the other three outcome measures all have subscale/domain scores, enabling a profile of strengths and challenges to be created to inform assessment, goal setting and response to NbR. For example, the FIM+FAM considers multiple outcomes from ABI and is useful for demonstrating gains in function, skills and abilities, after any constraints to rehabilitation incurred from NBD have been removed. The MPAI-4 has a similar usage, but is articulated in a different way. Rather than measuring abilities and limitations with reference to particular areas of outcome, MPAI-4 reflects the overall impact of ABI impairments on adjustment and function. In contrast, SASNOS enables assessment of a single outcome from ABI, that of NBD. However, and unlike HoNOS-ABI, its greater number and range of items facilitates exploration of five domains and 13 subdomains of NBD, enabling very detailed investigation of a heterogeneous, multifaceted syndrome. Therefore, when NBD is the principal obstacle to engagement, a comprehensive examination of one area of interest potentially has greater value for assessment and planning in rehabilitation.

In relation to statistical investigations of responsiveness, availability of reference groups resulted in SASNOS and MPAI-4 demonstrating the greatest evidence of responsiveness and also the highest proportion of participants assessed as achieving meaningful change (74.1%

and 65.7% respectively). Although, availability of a neurologically healthy reference group for the SASNOS was particularly advantageous, especially for highlighting potential rehabilitation goals and transparency regarding interpretation of change scores. Of course, no measures are mutually exclusive, and two or more could populate a service basket.

Additional desirable characteristics which may further help guide choice and also assist in the development of new instruments include: (1) freely available in the public domain; (2) routinely used in NbR and standardised across services; (3) conceived for measuring change in symptoms and social handicap in people with ABI, and anchored to an appropriate underlying theoretical framework; (4) evidence of robust psychometric properties; (5) capable of detecting change in symptoms of NBD/reduction of social handicap when used in NbR; (6) known range of responsiveness indices to facilitate interpretation of group and individual level differences in scores on repeated assessment; (7) meaningful scores which are easy to understand; (8) availability of data from a reference group to enable discrimination between normal and abnormal assessments; (9) outputs capable of assisting with goal planning, and (10) availability of proxy- and self-completion versions.

Predictive validity of measures of rehabilitation complexity

The second substantive aim of the study was to determine the extent that change in ratings on two measures of rehabilitation complexity on reassessment predict progress made in NbR as determined by the extent of T_1-T_2 change in scores on the four outcome measures explored here. This was primarily undertaken as the UKROC initiative uses measures of rehabilitation complexity to categorise individual services according to the complexity of cases they admit, with subsequent impacts on the access to and sustainability of these services. However, and similar to outcome measures included in the UKROC database, the instruments used to determine complexity status were not created specifically for NbR despite the potential importance of results obtained.

We expected that improvements in autonomy, as captured by T_1-T_2 differences obtained on the four outcome measures, would be paralleled by concurrent reductions in rehabilitation complexity as captured by the SRS and RCS-E. However, evidence for this was relatively weak. Indeed, even though responsiveness indices for the SRS suggested that participants required less supervision at T_2, only a reduction in Therapy scores was found on the RCS-E; a measure of complexity employed by UKROC. Between groups comparisons for participants categorised as either "*improved*", based on their individual SRMs, or as "*most expected to change*", based on scores at T_1, suggested that they needed less supervision (SRS) compared to participants categorised as "*same/declined*" or "*least likely to change*". In contrast, RCS-E scores did not differ across these groups.

The extent to which change in ratings on the measures of rehabilitation complexity predicted individual change in scores on the outcome measures also proved variable. T_1-T_2 SRS difference scores were predictive of participants who "*improved*" on the MPAI-4, but RCS-E T_1-T_2 difference scores did not discriminate between "*improved*" and "*same/declined*" individuals on any of the four outcome measures. Prediction models using both the SRS and RCS-E further confirmed these findings. A reduction in SRS scores was only predictive of improvement as captured by decreased MPAI-4 scores, and the RCS-E did not contribute significantly to any of the prediction models.

Whilst the current findings are informative, they also prompt further questions. Reduction in supervision was predictive of a decline in levels of impairment and disability as measured by the MPAI-4. However, some items contributing to these subscales are not exclusively concerned with symptoms of NBD (e.g. Use of Hands, Transportation, and Mobility). Therefore, and given that the SRS did not predict specific reduction in symptoms on the SASNOS, then its predictive success is potentially attributable to improvement in other aspects of outcome that are captured on the MPAI-4. Even so, this is undermined by the finding that change in SRS status was not predictive of differences between scores on reassessment on the FIM+FAM and HoNOS-ABI either, which are also not explicitly concerned with NBD. A further explanation is that SRS ratings are made on the basis of the level of supervision received at the time of assessment, rather than what is actually required (Boake, 1996, p. 766). In NbR services, supervision is embedded in the environment. For instance, an SRS rating of 11 reflects that *"…the* (participant) *lives in a setting in which the exits are physically controlled by others (for example, a locked ward)"*. This form of environmental supervision is characteristic of many NbR units regardless of whether the individual rehabilitation participant requires this or not, especially in services that lack a graded care pathway. Therefore, it may have been that some participants in the current sample had made sufficient improvement at T_2 to enable a move from a locked environment, but at the time of the assessment, this was not yet available. An additional point is that the SRS does not include an item to provide an indication of the level of security required, and that some participants in NbR environments may be accommodated within low- or medium-secure services. However, as ratings are made on the basis of what supervision was given at the time of assessment rather than predicted levels of supervision required (Boake, 1996, p. 766), scores may overestimate the level of supervision actually required as a consequence of improvement due to NbR; these gains in autonomy were correctly reflected in the outcome measures but less well on the SRS. Clearly, rating what supervision is needed rather than what was available in the environment will weaken the predictive ability of the measure and may account for the variable results found here.

In contrast, total RCS-E scores had poor predictive power for all four outcome measures. One possible explanation for these findings is that the RCS-E was designed to be utilised in medically organised services, whereas NbR services differ from medical rehabilitation models in several important ways (Worthington, Wood, & McMillan, 2017). For example, the "skilled nursing needs" item is rated on a 5-point scale to reflect the level of intervention required. Whilst this is appropriate for other neurorehabilitation settings, in NbR qualified nurses are embedded in the TDT and are omnipresent. This characteristic may be reflected in ratings of four ("*requires… very frequent monitoring / intervention by a qualified nurse*") and three ("*requires… behavioural management*"). Further, capacity to manage challenging behaviour is a key characteristic of NbR services and qualified nurses, along with all other members of the TDT, are responsible for delivering prescribed psychological interventions to achieve this. Consequently, this may further help to explain the lack of T_1-T_2 differences on the RCS-E which was apparent on the four outcome measures. Even so, our findings ultimately question the validity of using instruments to measure case complexity in NbR that were not conceived specifically for that purpose or fail to reflect the special characteristics of such services. Taken together, these findings highlight significant need for the development of specific measures for use in this context. For example, a measure will need to take into account and capture how members of the various professions share roles and cross discipline boundaries to pool and integrate expertise and skills to enable maximum efficiency in assessment and intervention; and reflect "role release", the extent where intervention strategies are carried out by any team member, under the supervision and support of team members whose disciplines are accountable for those practices. Simply recording hours of face-to-face contact would be inadequate, and a measure will additionally need to capture the extent of consultation and direction given to the wider team by individual professionals.

In the meantime, two potential measures that capture aspects of rehabilitation complexity specific to NbR could be considered. The first is the Neurobehavioural Expectations Scale (NES; Swan & Alderman, 2004), which attempts to quantify the therapeutic load placed on individual rehabilitation participants through a single score derived from consideration of a number of items inherent to NbR (e.g. number of behavioural interventions received, items reflecting the extent of how much of the day conforms to a structure to facilitate learning and practice of social routines and functional skills). The second is the supplementary scoring system recently developed for the SASNOS (Alderman et al., 2018), where all 49 items are assigned a 3-point dependency rating to recalibrate standardised SASNOS scores to reflect the amount of support received, and by implication, how complex they are. These weighted scores are especially helpful in the case of rehabilitation participants whose standardised scores suggest NBD symptoms are in the normal range for the neurologically healthy population, and as a result, may be considered ready for discharge. However, if the lack of symptoms is a function of high levels of support received, then clearly this needs to be considered; in these cases the recalibrated scores are much lower and give some indication of the likely levels of NBD symptomatology should support be removed.

Study limitations

The study is not without limitations. First, retrospective data was drawn from a sample of convenience, resulting in inconsistency concerning when T_1 assessments were completed and varying length of time between T_1 and T_2 assessments. Consequently, exposure to NbR would have varied across participants, with consequent impacts on ratings across the various outcome measures unknown. For example, we previously highlighted how the SASNOS Aggression domain may have lacked responsiveness when examined in the whole sample because NbR may have already effectively targeted this area of outcome. A further example is the Communication domain of the SASNOS, where only one case at T_1 fell outside the normal range for neurologically healthy controls. Ideally, assessments would have been completed across all participating services on admission and at standard intervals thereafter to control for confounding effects associated with data being collected at different time points. Second, as this study considered data that had been collected through routine clinical work, not all services included all of the outcome measures of interest in their basket of outcome tools. Going forward, a prospective study would overcome any threat to validity arising from such limitations.

Conclusions

In conclusion, outcome measures not specifically designed for use in NbR demonstrated responsiveness to change in scores at two assessment points, suggesting that they provide a meaningful benchmark to assess response to rehabilitation, at both individual and service level. However, the availability of reference group data for some measures to enable discrimination between normal and abnormal assessments proved advantageous. However, even though responsiveness is an essential criterion for the within group, repeated measures context in rehabilitation, the choice over which measures to utilise is ultimately dependent on a range of factors (e.g. focus of assessment, needs, personal preference). This study also demonstrated how applying measures of rehabilitation complexity not designed specifically for use in NbR (i.e. RCS-E) is less than ideal, and highlighted how existing instruments intended for this purpose should be utilised, and further new measures developed. Importantly, the inclusion of these measures in the UKROC basket would increase the validity of the results submitted by NbR services, and in turn, reduce risks associated with continuing with the current strategy. Finally, in addition to the issues pertaining to UKROC in the UK, the findings and implications of this study have wider relevance regarding outcome measurement in NbR and specialist services broadly, regardless of the country this is undertaken and the prevailing commissioning requirements.

Acknowledgements

We would like to thank all member services of the Independent Neurorehabilitation Providers Alliance for their continued support and contribution to initial design and acquisition of study data.

Disclosure of interest

The authors report no conflict of interest.

References

Alderman, N. (2003). Contemporary approaches to the management of irritability and aggression following traumatic brain injury. *Neuropsychological Rehabilitation*, *13*(1–2), 211–240. doi:10.1080/09602010244000327

Alderman, N., & Knight, C. (2017). Keeping the ‘scientist-practitioner’ model alive and kicking through service-based evaluation and research: examples from neurobehavioural rehabilitation. *The Neuropsychologist*, *3*, 25–32.

Alderman, N., Williams, C., Knight, C., & Wood, R. L. (2017). Measuring Change in Symptoms of Neurobehavioural Disability: Responsiveness of the St Andrew’s-Swansea Neurobehavioural Outcome Scale. *Archives of Clinical Neuropsychology*, *32*(8), 951–962. doi:10.1093/arclin/acx026

Alderman, N., Williams, C., & Wood, R. L. (2018). When normal scores don’t equate to independence: Recalibrating ratings of neurobehavioural disability from the ‘St Andrew’s – Swansea Neurobehavioural Outcome Scale’ to reflect context-dependent support. *Brain Injury*, *32*(2), 218–229. doi:10.1080/02699052.2017.1406989

Alderman, N., & Wood, R. L. (2013). Neurobehavioural approaches to the rehabilitation of challenging behaviour. *NeuroRehabilitation*, *32*(4), 761–770. doi:10.3233/NRE-130900

Alderman, N., Wood, R. L., & 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*(1), 83–100. doi:10.3109/02699052.2010.532849

Allen, I. E. and Seaman, C .A. (2007). Likert scales and data analyses. *Quality Progress.* Retrieved from: http://asq.org/quality-progress/2007/07/statistics/likert-scales-and-data-analyses.html [Accessed 22 January 2019].

Boake, C. (1996). Supervision rating scale: A measure of functional outcome from brain injury. *Archives of Physical Medicine and Rehabilitation*, *77*(8), 765–772. doi:10.1016/S0003-9993(96)90254-3

Coetzer, R., & du Toit, P. L. (2001). HoNOS–ABI; a clinically useful outcome measure? *Psychiatric Bulletin*, *25*(11), 421–422. doi:10.1192/pb.25.11.421

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd Editio). New York, USA: Psychology Press.

Eisen, S. V., Ranganathan, G., Seal, P., & Spiro, A. (2007). Measuring Clinically Meaningful Change Following Mental Health Treatment. *The Journal of Behavioral Health Services & Research*, *34*(3), 272–289. doi:10.1007/s11414-007-9066-2

Feng, C., Wang, H., Lu, N., Chen, T., He, H., Lu, Y., & Tu, X. M. (2014). Log-transformation and its implications for data analysis. *Shanghai Archives of Psychiatry*, *26*(2), 105–9. doi:10.3969/j.issn.1002-0829.2014.02.009

Fleminger, S., Leigh, E., Eames, P., Langrell, L., Nagraj, R., & Logsdail, S. (2005). HoNOS–ABI: A reliable outcome measure of neuropsychiatric sequelae to brain injury? *Psychiatric Bulletin*, *29*(02), 53–55. doi:10.1192/pb.29.2.53

Hemphill, J. F. (2003). Interpreting the magnitudes of correlation coefficients. *American Psychologist*, *58*(1), 78–79. doi:10.1037/0003-066X.58.1.78

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*(6), 457–470. doi:10.1080/02699050802060647

King, G., Strachan, D., Tucker, M., Duwyn, B., Desserud, S., & Shillington, M. (2009). The Application of a Transdisciplinary Model for Early Intervention Services. *Infants & Young Children*, *22*(3), 211–223. doi:10.1097/IYC.0b013e3181abe1c3

Kreutzer, J. S., Marwitz, J. H., Seel, R., & Devany Serio, C. (1996). Validation of a neurobehavioral functioning inventory for adults with traumatic brain injury. *Archives of Physical Medicine and Rehabilitation*, *77*(2), 116–124. doi:10.1016/S0003-9993(96)90155-0

Malec, J. F., Kragness, M., Evans, R. W., Finlay, K. L., Kent Ann, & Lezak, M. D. (2003). Further psychometric evaluation and revision of the Mayo-Portland Adaptability Inventory in a national sample. *Journal of Head Trauma Rehabilitation*, *18*(6), 479–492.

Malec, J. F., & Lezak, M. D. (2008). Manual for The Mayo-Portland Adaptability Inventory (MPAI-4) for Adults, Children and Adolescents. *Test*. doi:10.1148/radiol.2493080468

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*, *2*(4). doi:10.5334/ijic.65

Nayar, M., Vanderstay, R., Siegert, R. J., & Turner-Stokes, L. (2016). The UK Functional Assessment Measure (UK FIM+FAM): Psychometric Evaluation in Patients Undergoing Specialist Rehabilitation following a Stroke from the National UK Clinical Dataset. *PLOS ONE*, *11*(1), e0147288. doi:10.1371/journal.pone.0147288

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 on JSTOR. *Medical Care*, *41*(5), 582–592.

Oddy, M., & Ramos, S. D. S. (2013). The clinical and cost-benefits of investing in neurobehavioural rehabilitation: A multi-centre study. *Brain Injury*, *27*(13–14), 1500–1507. doi:10.3109/02699052.2013.830332

Rai, S. K., Yazdany, J., Fortin, P. R., & Aviña-Zubieta, J. A. (2015). Approaches for estimating minimal clinically important differences in systemic lupus erythematosus. *Arthritis Research & Therapy*, *17*(1), 143. doi:10.1186/s13075-015-0658-6

Reed, K., Boake, C., Caroselli, J. S., Neese, L. E., Becker, C. L., & Scheibel, R. S. (1999). The supervision rating scale (SRS): A program evaluation tool for postacute brain injury rehabilitation. *Archives of Clinical Neuropsychology*, *14*(8), 795–796. doi:10.1093/arclin/14.8.795

Skinner, A., & Turner-Stokes, L. (2006). The use of standardized outcome measures in rehabilitation centres in the UK. *Clinical Rehabilitation*, *20*(7), 609–615. doi:10.1191/0269215506cr981oa

Snowden, R. J., & Gray, N. S. (2010). Helpful and unhelpful risk assessment practices reply. *Psychiatric Services*, *61*, 530–531.

Swan, L., & Alderman, N. (2004). Measuring the relationship between overt aggression and expectations: a methodology for determining clinical outcomes. *Brain Injury*, *18*(2), 143–160. doi:10.1080/02699050310001596923

Tam, S., McKay, A., Sloan, S., & Ponsford, J. (2015). The experience of challenging behaviours following severe TBI: A family perspective. *Brain Injury*, *29*(7–8), 813–821. doi:10.3109/02699052.2015.1005134

Tate, R. L. (2010). *A Compendium of Tests, Scales, and Questionnaires: The Practitioner's Guide to Measuring Outcomes after Acquired Brain Impairment*. New York, US: Psychology Press.

Turner-Stokes, L. (2016). *The UK Rehabilitation Outcome Collaborative (UKROC) Database UKROC database for specialist rehabilitation*. Retrieved from https://www.kcl.ac.uk/nursing/departments/cicelysaunders/attachments/UKROC-database-only-brief-outline-Sept-2016-rev2.pdf

Turner-Stokes, L., Nyein, K., & Halliwell, D. (1999). The Northwick Park Care Needs Assessment (NPCNA): A directly costable outcome measure in rehabilitation. *Clinical Rehabilitation*, *13*(3), 253–267. doi:10.1191/026921599677787870

Turner-Stokes, L., Nyein, K., Turner-Stokes, T., & Gatehouse, C. (1999). The UK FIM+FAM: development and evaluation. *Clinical Rehabilitation*, *13*(4), 277–287. doi:10.1191/026921599676896799

Turner-Stokes, L., Scott, H., Williams, H., & Siegert, R. (2012). The Rehabilitation Complexity Scale – extended version: detection of patients with highly complex needs. *Disability and Rehabilitation*, *34*(9), 715–720. doi:10.3109/09638288.2011.615880

Turner-Stokes, L., & Siegert, R. J. (2013). A comprehensive psychometric evaluation of the UK FIM + FAM. *Disability and Rehabilitation*, *35*(22), 1885–1895. doi:10.3109/09638288.2013.766271

Turner-Stokes, L., Tonge, P., Nyein, K., Hunter, M., Nielsona, S., & Robinson, I. (1998). The Northwick Park Dependency Score (NPDS): a measure of nursing dependency in rehabilitation. *Clinical Rehabilitation*, *12*(4), 304–318. doi:10.1191/026921598669173600

Turner-Stokes, L., Williams, H., Bill, A., Bassett, P., & Sephton, K. (2016). Cost-efficiency of specialist inpatient rehabilitation for working-aged adults with complex neurological disabilities: a multicentre cohort analysis of a national clinical data set. *BMJ Open*, *6*(2), e010238. doi:10.1136/bmjopen-2015-010238

Turner-Stokes, L., Williams, H., Sephton, K., Rose, H., Harris, S., & Thu, A. (2012). Engaging the hearts and minds of clinicians in outcome measurement – the UK rehabilitation outcomes collaborative approach. *Disability and Rehabilitation*, *34*(22), 1871–1879. doi:10.3109/09638288.2012.670033

Turner-Stokes, L., Williams, H., & Siegert, R. J. (2010). The rehabilitation complexity scale version 2: A clinimetric evaluation in patients with severe complex neurodisability. *Journal of Neurology, Neurosurgery and Psychiatry*, *81*(2), 146–153. doi:10.1136/jnnp.2009.173716

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*, *1*(1), 4. doi:10.1186/1477-7525-1-4

Warner, R. (2018). Is a Likert scale considered interval, ratio, or nominal? *Quora*. Retrieved from https://www.quora.com/Is-a-Likert-scale-considered-interval-ratio-or-nominal# [Accessed 22 January 2019].

Winkler, D., Unsworth, C., & Sloan, S. (2006). Factors that lead to successful community integration following severe traumatic brain injury. *Journal of Head Trauma Rehabilitation*, *21*(1), 8–21.

Wood, R. L. (1990a). A neurobehavioural paradigm for brain injury rehabilitation. In R. L. Wood (Ed.), *Neurobehavioural sequelae of traumatic brain injury* (pp. 3–17). London: Taylor & Francis.

Wood, R. L. (1990b). Towards a Model of Cognitive Rehabilitation. In R. L. Wood & I. Fussey (Eds.), *Cognitive Rehabilitation in Perspective* (pp. 3–26). London: Taylor Francis. doi:10.4324/9780429490088-2

Wood, R. L., Alderman, N., & Williams, C. (2008). Assessment of neurobehavioural disability: A review of existing measures and recommendations for a comprehensive assessment tool. *Brain Injury*, *22*(12), 905–918. doi:10.1080/02699050802491271

Wood, R. L., McCrea, J. D., Wood, L. M., & Merriman, R. N. (1999). Clinical and cost effectiveness of post-acute neurobehavioural rehabilitation. *Brain Injury*, *13*(2), 69–88. doi:10.1080/026990599121746

Worthington, A. D., & Alderman, N. (2017). *Neurobehavioural Rehabilitation: A Developing Paradigm*. (T. McMillan & R. L. Wood, Eds.), *Neurobehavioural Disability and Social Handicap Following Traumatic Brain Injury, Second Edition* (Second Edi). Psychology Press. doi:10.4324/9781315684710

Worthington, A. D., Matthews, S., Melia, Y., & Oddy, M. (2006). Cost-benefits associated with social outcome from neurobehavioural rehabilitation. *Brain Injury*, *20*(9), 947–957. doi:10.1080/02699050600888314

Worthington, A., Wood, R. L., & McMillan, T. M. (2017). Neurobehavioural disability over the past four decades. In T. McMillan & R. L. Wood (Eds.), *Neurobehavioural Disability and Social Handicap Following Traumatic Brain Injury* (pp. 19–30), 2nd Ed. Psychology Press.

Ylvisaker, M., Turkstra, L., Coehlo, C., Yorkston, K., Kennedy, M., Sohlberg, M. M., & Avery, J. (2007). Behavioural interventions for children and adults with behaviour disorders after TBI: A systematic review of the evidence. *Brain Injury*, *21*(8), 769–805. doi:10.1080/02699050701482470

Table 1. Acquired brain injury diagnosis frequencies.

Diagnosis		Frequency (N)	%
TBI		55	44.7
CVA		25	20.3
	Infarct	4	3.3
	Haemorrhagic Stroke	7	5.7
	Sub-arachnoid haemorrhage	3	2.4
	Other Stroke	11	8.9
Anoxia		10	8.1
Inflammation		11	8.9
Intoxication		4	3.3
Other		13	10.6
Missing		5	4.1

Table 2. Statistically significant differences between MPAI-4 and SASNOS scores and magnitude of effect size achieved at first and second assessment.

	T₁	***T₂***					
	Mean (SD)	**Mean (SD)**	***t***	***p***	r_{T1-T2}	**SRM[a]**	
MPAI-4							
Total	50.02 (9.29)	44.98 (10.74)	4.51	<.001	.60	.56	*medium*
Ability	49.03 (10.12)	46.01 (10.78)	3.19	.002	.67	.38	*small*
Adjustment	50.34 (8.47)	45.74 (9.05)	4.30	<.001	.48	.51	*medium*
Participation	55.68 (11.10)	49.32 (11.48)	5.65	<.001	.64	.67	*medium*
SASNOS							
Total	40.02 (13.82)	45.72 (11.39)	3.89	<.001	.62	.51	*small*
Interpersonal Relationships	30.28 (13.37)	38.39 (13.52)	4.79	<.001	.56	.64	*medium*
Cognition	21.40 (13.24)	32.09 (12.98)	8.52	<.001	.73	1.10	*large*
Inhibition	54.53 (10.93)	57.24 (8.79)	2.16	.035	.51	.28	*small*
Aggression	56.28 (13.58)	58.70 (12.28)	1.45	.154	.51	.19	*trivial*
Communication	63.24 (10.64)	66.41 (10.16)	2.71	.009	.62	.35	*small*

[a]Effect magnitude thresholds adjusted to take into account r_{T1-T2} strength using the Middel and van Sonderen (2002) solution.

Table 3. Correlations between the MPAI-4 total and subscales scores and the other measures at T_1 (N = 64-94) and T_2 (N = 32-42).

MPAI-4	Total	Abilities	Adjustment	Participation
T_1				
FIM+FAM	-.65**	-.71**	-.37**	-.75**
FIM+FAM EADL	-.51**	-.60**	-.22*	-.70**
SASNOS	-.47**	-.32**	-.50**	-.48**
HoNOS ABI	.54**	.65**	.46**	.48**
SRS	.13	.04	.17	.31**
RCS-E	.15	.15	.04	.34**
T_2				
FIM+FAM	-.77**	-.78**	-.30	-.89**
FIM+FAM EADL	-.68**	-.67**	-.20	-.76**
SASNOS	-.51**	-.40**	-.47**	-.49**
HoNOS ABI	.62**	.48**	.58**	.58**
SRS	.58**	.42**	.50**	.65**
RCS-E	.44**	.27*	.36**	.49**

*p<.05; **p<.01

Table 4. Correlations between SASNOS and the other measures at T_1 (N = 81-109) and T_2 (N =35-66).

SASNOS		Total	Interpersonal Relationships	Cognition	Inhibition	Aggression	Communication
T1							
	FIM+FAM	.50**	.57**	.52**	.24*	.34**	-.15
	FIM+FAM EADL	.48**	.59**	.60**	.23*	.27*	-.07
	MPAI-4	-.47**	-.48**	-.42**	-.32**	-.28**	.02
	HoNOS ABI	-.62**	-.55**	-.37**	-.45**	-.54**	-.35**
	SRS	-.40**	-.44**	-.39**	-.27**	-.35**	-.34**
	RCS-E	-.43**	-.48**	-.34**	-.16	-.32**	-.05
T2							
	FIM+FAM	.56**	.47**	.62**	.39*	.30	.23
	FIM+FAM EADL	.35*	.28	.56**	.13	-.01	-.04
	MPAI-4	-.51**	-.40**	-.55**	-.44**	-.31*	-.21
	HoNOS ABI	-.49**	-.46**	-.38**	-.32*	-.38**	-.23
	SRS	-.63**	-.54**	-.57**	-.39**	-.34**	-.37**
	RCS-E	-.58**	-.48**	.46**	-.39**	-.47**	-.33*

*p<.05; **p<.01

Table 5. Statistically significant differences between FIM+FAM, HoNOS-ABI, RCS-E and SRS scores and r
first and second assessment.

	T₁	***T₂***	***T₁***	***T₂***		
	Mean (SD)	**Mean (SD)**	**Median (IQR)**	**Median (IQR)**	***Z***	***p***
FIM + FAM						
Total	133.06 (41.42)	151.79 (44.00)	145.00 (61.25)	170.00 (59.00)	4.07	<.00
Motor subscale	80.43 (18.43)	87.18 (30.48)	94.00 (52.50)	104.50 (34.50)	3.12	.002
Cognitive subscale	48.40 (18.43)	61.65 (19.58)	47.00 (30.00)	63.50 (31.75)	4.74	<.00
EADL	11.92 (8.72)	17.25 (10.48)	8.00 (7.75)	18.50 (18.50)	3.69	<.00
HoNOS-ABI Total	16.04 (6.47)	12.70 (6.31)	16.00 (9.5)	11.00 (8.00)	3.32	.001
RCS-E						
Total	11.30 (4.20)	10.20 (4.24)	10.00 (8.00)	10.00 (7.50)	4.16	<.00
Care/Risk	2.34 (1.21)	1.87 (1.18)	2.00 (2.00)	2.00 (2.00)	4.20	<.00
Nursing	1.53 (1.38)	1.61 (1.28)	1.50 (3.00)	1.50 (3.00)	0.84	.399
Therapy	5.63 (1.12)	4.96 (1.46)	5.00 (2.00)	5.00 (2.00)	5.07	<.00
Medical	1.19 (1.24)	1.21 (1.25)	1.00 (3.00)	1.00 (3.00)	0.05	.960
Equipment	0.61 (0.66)	0.63 (0.64)	1.00 (1.00)	1.00 (1.00)	0.40	.686
SRS Total	8.90 (2.31)	6.46 (3.62)	8.00 (3.00)	7.00 (7.00)	5.83	<.00

[a]Effect magnitude thresholds adjusted to take into account $r_{T1\text{-}T2}$ strength using the Middel and van Sonderen

Table 6. Correlations between the FIM+FAM and the other measures at T_1 (N = 78-114) and T_2 (N = 37-72).

FIM+FAM	Total	Motor subscale	Cognitive subscale	EADL
T₁				
SASNOS	.50**	.38**	.58**	.48**
MPAI-4	-.65**	-.60**	-.49**	-.51**
HoNOS ABI	-.65**	-.56**	-.59**	-.50**
SRS	-.24*	-.12	-.41**	-.36**
RCS-E	-.45**	-.40**	-.51**	-.50**
T₂				
SASNOS	.56**	.34*	.64**	.35*
MPAI-4	-.77**	-.66**	-.73**	-.68**
HoNOS ABI	-.68**	-.51**	-.71**	-.58**
SRS	-.63**	-.39*	-.76**	-.45**
RCS-E	-.63**	-.53**	-.67**	-.48**

*p<.05; **p<.01

Table 7. Correlations between HoNOS ABI and the other measures at T_1 (N = 64-93) and T_2 (N = 32-61).

HoNOS-ABI	***T1***	***T2***
FIM+FAM	-.65**	-.68**
FIM+FAM EADL	-.50*	-.58**
SASNOS	-.62**	-.49**
MPAI-4	.54**	.62**
SRS	.27*	.53**
RCS-E	.31**	.53**

*p<.05; **p<.01

Table 8. Correlations between RCS-E and the other measures at T_1 (N = 77-99) and T_2 (N =32-95).

RCS-E		Total	Care/Risk	Nursing	Therapy	Medical	Equipment
T_1							
	FIM+FAM	-.49**	-.33**	-.31**	-.27*	-.28*	-.47**
	FIM+FAM EADL	-.50**	-.37**	-.35**	-.38**	-.41**	-.32**
	SASNOS	-.43**	-.40**	-.48**	-.16	-.33**	-.06
	MAPI-4	.15	.20	.07	-.07	-.04	-.44**
	HoNOS ABI	.31**	.17	.39**	.16	.25*	.17
	SRS	.63**	.63**	.56**	.44	.57**	-.14
T_2							
	FIM+FAM	-.63**	-.52**	-.53**	-.49**	-.48**	-.44**
	FIM+FAM EADL	-.48**	-.54**	.29	-.27	-.38*	-.45**
	SASNOS	-.41**	-.44**	-.41**	-.25*	-.32**	.07
	MPAI-4	.03	.09	.01	-.06	-.12	.28**
	HoNOS ABI	.53**	.51**	.48**	.27	.48**	.36*
	SRS	.70**	.67**	.71**	.40**	.66**	-.06

*p<.05; **p<.01

Table 9. Correlations between SRS and the other measures at T_1 (N = 47-62) and T_2 (N = 35-69).

SRS	T_1	T_2
FIM+FAM	-.31*	-.63**
FIM+FAM EADL	-.46**	-.45**
SASNOS	-.37**	-.62**
MPAI-4	.20	.57**
HoNOS ABI	.34**	.52**
RCS-E	.65**	.68**

*p<.05; **p<.01

Table 10. Recalculated responsiveness indices for SASNOS Total and domain scores for participants most expected to change (T_1 score <40) versus those whose initial assessment scores suggested NbR symptoms were in the expected range for neurologically healthy controls (T_1 score ≥40).

	***T*₁**	***T*₂**					
SASNOS *T*₁ <40	**Mean (SD)**	**Mean (SD)**	***t***	***p***	**r_{T1-T2}**	**SRM[a]**	
Total	27.92 (8.58)	38.67 (9.76)	5.72	<.001	.44	1.10	*large*
Interpersonal Relationships	22.88 (12.16)	33.00 (13.02)	4.89	<.001	.40	.73	*large*
Cognition	18.62 (10.80)	30.22 (12.15)	8.77	<.001	.65	1.20	*large*
Inhibition	31.00 (8.13)	44.56 (15.42)	1.65	.161	.41	.67	*medium*
Aggression	32.27 (5.28)	48.53 (16.17)	3.21	.011	.19	1.01	*large*
SASNOS *T*₁ ≥40	**Mean (SD)**	**Mean (SD)**	***t***	***p***	**r_{T1-T2}**	**SRM[a]**	
Total	50.51 (7.24)	51.85 (8.96)	0.69	.497	.17	.16	*trivial*
Interpersonal relationships	46.46 (5.50)	51.11 (6.88)	2.28	.040	.38	.61	*medium*
Cognition	46.01 (3.21)	48.63 (7.48)	1.25	.268	.83	.51	*small*
Inhibition	57.08 (7.51)	58.68 (6.52)	1.54	.130	.43	.21	*small*
Aggression	61.18 (8.54)	60.78 (10.35)	0.28	.784	.43	.04	*trivial*

[a]Effect magnitude thresholds adjusted to take into account r_{T1-T2} strength using the Middel and van Sonderen (2002) solution.

Table 11. Recalculated responsiveness indices for MPAI-4 Total and subscale scores comparing participants the most severe impairments on initial assessment (T_1 score ≥50) versus those rated with mild-medium impairments (T_1 score <50) relative to other people with ABI.

	T_1	T_2					
Score at $T_1 < 50$	**Mean (SD)**	**Mean (SD)**	***t***	***p***	$r_{T1\text{-}T2}$	**SRM**[a]	
Total	52.21 (7.42)	46.71 (9.67)	4.69	<.001	.48	.62	*medium*
Ability	51.68 (8.40)	47.72 (9.83)	3.60	<.001	.57	.47	*small*
Adjustment	52.48 (6.75)	46.80 (8.47)	5.42	<.001	.44	.69	*medium*
Participation	57.15 (10.02)	50.34 (11.24)	5.73	<.001	.59	.70	*medium*
Score at $T_1 \geq 50$	**Mean (SD)**	**Mean (SD)**	***t***	***p***	$r_{T1\text{-}T2}$	**SRM**[a]	
Total	56.89 (5.57)	50.11 (8.15)	5.01	<.001	.34	.85	*large*
Ability	57.68 (5.86)	51.53 (8.91)	4.27	<.001	.41	.73	*medium*
Adjustment	55.79 (5.31)	49.38 (7.87)	4.93	<.001	.23	.76	*large*
Participation	61.96 (8.35)	53.26 (11.38)	5.99	<.001	.54	.88	*large*

[a]Effect magnitude thresholds adjusted to take into account $r_{T1\text{-}T2}$ strength using the Middel and van Sonderen (2002) solution.

Table 12: Responsiveness indices, cut-off scores for "medium" improvement, and percent of individuals categorised as attaining a T_1-T_2 difference on the various outcome measures that is likely to be of practical significance[a].

Outcome measure	**SRM**		**ES "*medium*" cut-off[b]**	**% "improved"**
MPAI-4	.56	*medium*	.56	39.4
SASNOS	.51	*small*	.57	48.3
FIM+FAM	.66	*small*	.79	29.4
HoNOS-ABI	.53	*medium*	.51	44.0
MPAI-4$_{T1\geq50}$	.85	*large*	.45	65.7
SASNOS$_{T1<40}$	1.10	*large*	.47	74.1

[a]Individual SRM ≥ ES "*medium*" cut-off score.
[b]Effect magnitude thresholds adjusted to take into account $r_{T1\text{-}T2}$ strength using the Middel and van Sonderen (2002) solution.

Table 13: Predictive accuracy (Spearman's ρ and AUC) of SRS and RCS-E with respect to the four outcome measures.

		Likelihood of being "*improved*"		
	ρ	**AUC**	**SE**	**95% CI**
$SRS_{T1\text{-}T2\ DIFF}$				
MPAI-4	.51	.791*	.06	.67 - .91
FIM+FAM	.05	.531	.11	.31 - .75
HoNOS-ABI	.21	.622	.08	.47 - .78
SASNOS	.21	.617	.08	.47 - .77
$MPAI\text{-}4_{T1\geq50}$	.35	.715*	.09	.54 - .89
$SASNOS_{T1<40}$	.17	.605	.11	.38 - .83
$RCS_{T1\text{-}T2\ DIFF}$				
MPAI-4	.05	.528	.09	.35 - .71
FIM+FAM	.07	.544	.13	.29 - .80
HoNOS-ABI	-.17	.405	.10	.21 - .60
SASNOS	-.05	.474	.09	.29 - .66
$MPAI\text{-}4_{T1\geq50}$	.16	.588	.11	.36 - .81
$SASNOS_{T1<40}$	.05	.533	.13	.28 - .78

*AUC .70 to .80 equates with "fair" predictive ability