A Subarachnoid Haemorrhage-Specific Outcome Tool

Adrian Pace[1,*], Sophie Mitchell[2,*], Lizzie Casselden[2*], Ardalan Zolnourian[3], James Glazier[2], Lesley Foulkes[3], Diederik Bulters[3,#], Ian Galea[2,3,#]

From:

1. Greater Manchester Neurosciences Centre, Salford Royal NHS Foundation Trust, UK
2. Clinical Neurosciences, Clinical & Experimental Sciences, Faculty of Medicine, University of Southampton
3. Wessex Neurosciences Centre, University Hospital Southampton NHS Foundation Trust

* These authors contributed equally to the work

[#] Joint senior authors

Corresponding author:

Dr Ian Galea
Associate Professor in Experimental Neurology
Clinical Neurosciences, Clinical & Experimental Sciences
Faculty of Medicine, University of Southampton
Mailpoint 806, Level D, Southampton General Hospital
Southampton SO16 6YD
UK
Tel no - 0044 2381205340
Fax no - 0044 2381206085
E-mail - I.Galea@soton.ac.uk

Running title: SAH Outcome Tool

Tables & Figures: 6 tables & 2 figures

Keywords: Subarachnoid haemorrhage, Outcome studies, Stroke: other, Cerebral aneurysm Stroke: rehabilitation, Clinical practice, Neurosurgery

Word count: 5797

Abbreviations: SAH = subarachnoid haemorrhage; mRS = modified Rankin Scale; GOS = Glasgow Outcome Score; GOSE = Glasgow Outcome Score Extended; SAHOT = subarachnoid haemorrhage outcome tool; DIF = differential item functioning; MCID = minimal clinical important difference

ABSTRACT

Functional outcome after subarachnoid haemorrhage (SAH) has traditionally been assessed using scales developed for other neurological conditions. The modified Rankin score (mRS) and Glasgow Outcome Scale (GOS) are most commonly used. Employment of these scales in SAH is hampered by well-recognised limitations. We set out to develop and validate a new condition-specific SAH Outcome Tool (SAHOT). Items addressing diverse aspects of the impact of SAH were collected during focus groups involving patients, next-of-kin and multidisciplinary professionals involved in SAH management. After a series of iterative revisions, the resultant questionnaire was applied to patients and their next-of-kin at one, three and six months post-SAH. Rasch methodology was utilized to finalize the structure of the questionnaire and explore the extent to which SAHOT scores met Rasch-based criteria of successful measurement. The SAHOT was further assessed using traditional scale evaluation techniques, and validated in a second separate SAH patient cohort. The final SAHOT included 56 items dealing with cognitive, physical, and behavioural/psychological consequences of SAH. Rasch analysis indicated the scale successfully measured functional outcome post-SAH. Three item scoring categories produced the best scale performance. There was no evidence of differential item functioning between patients and next-of-kin. The SAHOT was found to be acceptable, have good convergent and divergent validity, good discrimination and excellent responsiveness. It was successfully validated in a second SAH patient cohort. The SAHOT offers the first SAH-specific scientifically robust outcome measure with potential utility in neurovascular clinical services and research studies.

Introduction

Non-traumatic subarachnoid haemorrhage (SAH) is a devastating neurological condition with an immediate mortality rate of about 12% (Huang and van Gelder, 2002). Survivors experience significant neurological sequelae, such that recovery to their previous level of functioning is uncommon. Most also experience persistent cognitive problems which interfere with return to employment (Al-Khindi *et al.*, 2010).

Clinical trials in SAH have most commonly utilized the modified Rankin score (mRS) (Farrell *et al.*, 1991) and the Glasgow Outcome Scale (GOS) (Jennett and Bond, 1975) or its extended version (GOSE) (Jennett *et al.*, 1981; Wilson *et al.*, 1998). These scales were developed to categorise outcomes following ischaemic stroke and head injury respectively, and have well-described limitations when applied to SAH (Schweizer and Macdonald, 2010; Macdonald *et al.*, 2013b). Although the exact sources of these limitations are not known, there are some clues. It has limited responsiveness in stroke (Dromerick *et al.*, 2003), and yet it is applied to SAH, where the multiplicity of pathological processes triggered by ictus complicate recovery compared to stroke. Half the patients with a post-SAH mRS of zero exhibit significant cognitive impairment when evaluated neuropsychologically (Kreiter *et al.*, 2002), one in three are unable to return to work and one in six have a mood disorder (Quinn *et al.*, 2014). Hence a mRS of zero after SAH cannot be used to define a good outcome, i.e. it lacks sensitivity, since patients with a mRS of zero have a markedly heterogenous outcome. The GOS has shown similar problems (Hutter and Gilsbach, 1993).

Although SAH is a form of cerebrovascular insult, it cannot be equated with ischaemic stroke since it has unique pathophysiology and anatomical localization. SAH occurs more frequently in people of working age with young families, so potential loss of quality-adjusted life years

and economic impact are disproportionately high (Taylor *et al.*, 1996; Rivero-Arias *et al.*, 2010). There is also a much higher incidence of post-traumatic stress symptoms after SAH (Powell *et al.*, 2002; Visser-Meily *et al.*, 2013) compared to stroke (Sembi *et al.*, 1998) and trauma (Shalev *et al.*, 1998). Clearly, the SAH patient experience does not necessarily mirror that of patients with ischaemic stroke and head injury survivors.

A study of six outcome measures developed for conditions other than SAH indicated that none adequately captured the morbidity occurring after SAH (Kim *et al.*, 2005). Besides the mRS and GOS, these included the Barthel Index, the National Institutes of Health Stroke Score, the Short Form-36 and the Mini-Mental Status Examination. The BICRO39 (Powell *et al.*, 2002), CLCE-24 (Passier *et al.*, 2010), EQ5D (Meyer *et al.*, 2010), Stroke-Specific Quality of Life (Boosman *et al.*, 2010) and Quality of Life after Brain Injury (Wong *et al.*, 2014) have also been applied to SAH after their development for other conditions, due to the absence of a SAH-specific outcome measure. In an attempt to encompass as many features of SAH impact as possible, a recent study used a booklet of questionnaires including the Self-Report Dysexecutive Questionnaire, the Everyday Memory Questionnaire, Stroke Symptom Checklist, Wimbledon Self-Report Scale, and the needs-based Stroke-Specific Quality of Life scale, to derive a summed patient-reported score (Quinn *et al.*, 2014). Despite these efforts, a SAH-specific outcome scale is still lacking.

The critical need for a better and more specific outcome measure has been articulated by the Subarachnoid Hemorrhage International Trialists (SAHIT), an international network of SAH investigators (Macdonald *et al.*, 2013a). To address this unmet need, we set out to develop a condition-specific scale for use in patients who experience SAH, hereafter referred to as the SAH Outcome Tool (SAHOT). The SAHIT recognized that information on cognitive

outcome needs to be well represented but the multiplicity of cognitive tools available render it difficult to decide which to use. We therefore decided to interview a cross-section of patients and find out which patient-reported outcomes, including those in the cognitive domain, were important. At the outset, we set the following requirements for the tool and its development: (1) specific to SAH; (2) primarily an interval (scalar) measure; (3) able to assess the full spectrum of disability; (4) scale development process involving classical and modern psychometric approaches; (5) iterative scale modifications and improvement; (6) minimal training, cost and instrumentation involved, (7) patient-reported, (8) high acceptability.

Materials and Methods

Setting

In a tertiary neurovascular centre in southern England, a multidisciplinary working group was established to develop the tool. The team consisted of experienced professionals and researchers from neurosurgery, neurology and neurovascular nursing backgrounds. The project was covered by National Research Ethics Committee approval (NRES 12/SC/0666) and local institutional approval. All data was collected in a link-anonymized fashion.

Scales used

Since mRS (Farrell *et al.*, 1991) and GOSE (Jennett *et al.*, 1981) are the most widely used outcome measures after SAH, they were considered as gold standards for scale validation in the absence of a SAH-specific measure. The modified version (Farrell *et al.*, 1991) of the Rankin Scale (Rankin, 1957) was used, ranging from 0 (no symptoms at all) to 5 (severe disability); a grade of mRS 6 (death) was added to include mortality (Quinn *et al.*, 2009). The

extended version (GOSE) (Jennett *et al.*, 1981; Wilson *et al.*, 1998) of the GOS (Jennett and Bond, 1975) was used, ranging from 1 (dead) to 8 (upper good recovery). The BICRO-39 (Powell *et al.*, 2002), CLCE-24 (Passier *et al.*, 2010), Fisher (Fisher *et al.*, 1980) and WFNS (Teasdale *et al.*, 1988) were used to describe the populations and further examine SAHOT validity.

Item generation and scale development

SAH-relevant symptoms were first identified using information collected in neurovascular clinics. This was supplemented by an English literature search over the last twenty years, using Google Scholar and Web of Science. Search terms included “subarachnoid haemorrhage”, “symptom”, “need”, “score”, “scale”, “tool” and “outcome”. A focus group of 30 SAH patients and their next-of-kin met with the team to discuss the items identified, add any other items deemed important and finalize a first draft of the questionnaire. This draft was piloted by two researchers and further refined in response to feedback with respect to instrument design and usability. It consisted of 60 items covering daily life, cognitive, physical, and behavioural/psychological consequences of SAH. For each item, respondents were asked to grade perceived change compared to their status immediately prior to the SAH using a 5-point Likert scale (0=no change; 1=small change; 2=moderate change; 3=large change; 4=complete change). A “not applicable” option was available. The direction of change (better or worse) was also recorded.

Clinical implementation

The questionnaire was deployed in the clinical neurovascular service, across inpatients and follow-up clinics. The work was divided into two phases: a development phase (n=113; 69% clinic attenders and 31% unselected consecutive admissions) and a validation phase (n=75,

100% unselected consecutive admissions). The questionnaire was administered during a face-to-face consultation, by telephone interview or postally, to both patients and their next-of-kin at one, three and six months post-SAH. In the validation phase, GOSE and mRS were co-administered with the SAHOT. BICRO39 and CLCE-24 were administered in a subset of patients at 6 months (n=19). Scales were administered at 1 month ± 1 week, 3 months ± 2 weeks and 6 months ± 1 month. The assessor did not know prior scores during subsequent time points since scores were calculated and/or tabulated at the end of the follow-up. Scales were administered by the author team consisting of three clinical trainees (SM, LC, AZ), a neurovascular research nurse (LF), a vascular neurosurgeon (DB) and a neurologist (IG). Several of the team had formal mRS and GOSE training and this was cascaded within the team. Trainees were closely supervised by senior members of the team. In most cases, the same person administered all the scales to an individual patient. The order of the tools was not specified.

Rasch analysis

The performance of the SAHOT scale was tested against the principles of Rasch measurement theory (Rasch, 1980; Tesio, 2003; Tennant *et al.*, 2004; Tennant and Conaghan, 2007). Briefly, Rasch methodology uses total scores obtained from a scale to rank examinees by their relative ability, and scale items by their relative difficulty, on a single linear continuum or 'ruler', separately and independently from each other. Linear estimates of patient ability and item difficulty are expressed in log odds ratios or logits.

For the purposes of the analysis, data across time points and examinees (patients and next-of-kin) were stacked into a single dataset. Questions left unanswered and 'not applicable' replies were treated as missing data. In a few instances the impact of SAH on certain items was

reported as positive rather than negative. These positive changes were rescored as 'no change' as they implied that no change for the worse occurred. Data was analysed using the Rasch-based software RUMM2030 (Andrich, 2012).

Volumetric assessment of bleed size

In the development phase, volumetric blood clot volume was used for the sole purpose of testing the SAHOT's convergent validity. Computed tomographic (CT) imaging of the head was available for 60 patients; there was no significant difference in the demographics and baseline characteristics of these 60 patients compared to the whole development cohort. Volumetric blood clot volume was quantified using MIPAV (Medical Image Processing, Imaging and Visualization) v7.2. CT images were only included in the analysis if acquired using the same imaging protocol within the first 3 days post-SAH, using contiguous slices. Image radiodensity threshold was set between 50 and 80 Hounsfield units, and converted to a binary mask. Regions of interest representing subarachnoid and total blood clot were drawn manually on each slice, and grouped into single three-dimensional volumes. To check reliability, the image processing operator was presented with a test set of CT scans consisting of five control cases and five SAH cases which had been previously quantified, in a blinded fashion. In the control cases, blood volume was zero, while in the SAH cases, the intra-operator coefficient of variation was 10.2%.

Statistics

Excel v15 and SPSS v22 were used. Data distribution was explored graphically and using the Kolmogorov-Smirnov test, and parametric or non-parametric statistical tests were employed accordingly. Since there was no evidence of differential item functioning (DIF) by respondent, construct validity and discrimination were studied using patient scores and, when

not available, next-of-kin scores to maximize sample size. Since there was evidence of DIF by time point, construct validity was studied at the middle time point, except for BICRO39 and CLCE-24, which were only administered at 6 months. Rasch-derived person location estimates were used during exploration of validity, discrimination, responsiveness and sample size calculations. Linear, logistic and ordinal regressions were performed with the following independent variables: age, gender, Fisher grade, WFNS, presence or absence of aneurysm, clipping *versus* coiling, and time from SAH, first in univariate then in multivariate mode. For simulation, we created three synthetic datasets, each consisting of two groups (placebo and treatment with a 20% difference in proportion of mRS=0, GOSE=8 or ordinal SAHOT=1). This was done using random sampling with replacement. As an example based on mRS=0, the treatment group was generated by the adjusting proportion of sampling within the mRS=0 stratum to achieve the desired difference in the proportion of mRS=0 compared to the placebo group. The placebo group was generated by sampling without adjustment. For each dataset, the sampling was performed 10,000 times to ensure constancy in representation of scores of the other two scales (in this example, GOSE=8 and ordinal SAHOT=1). The same was performed for a second and third simulation based on a 20% difference in proportion of GOSE=8 and ordinal SAHOT=1.

Results

Cohort characteristics and SAHOT usability

Demographics and SAH characteristics of the populations examined by the SAHOT are summarized in Table 1. 28% were non-aneurysmal after extensive investigation culminating with catheter angiography. All patients and next-of-kin who were approached completed the

scale, which took about 30 minutes to finish. The final versions of the SAHOT used in both development and validation phases are in supplementary material.

Development phase: Rasch-based SAHOT development

Rasch analysis was performed in the development phase to guide construction of the questionnaire. Initial analysis indicated that four items did not contribute to measurement of SAH impact, with poor item fit statistics and poor discrimination between groups with different degrees of SAH impact. These items were as follows: “Work (i.e. number of working hours and how much one can do at work)”, “Income (gross income)”, “Driving”, and “Headaches”. The items were discussed in a subsequent focus groups with SAH patients and professionals, where it was felt that the low prevalence of “Work”, “Income” and “Driving” amongst participants prior to SAH, and the ubiquitious nature of “Headaches” were the likely reasons underlying the items’ poor performance. The four items were therefore removed prior to further analyses, leaving a scale of 56 items on which the results presented are based. While analysis outcomes for the 60-item scale are not presented, all aspects of scale performance remained unchanged or improved following their removal.

Unidimensionality

The person separation index (Andrich, 1982) was 0.94, above the threshold of 0.9 indicative of unidimensionality. Item fit residuals and Chi square probabilities are presented in the second column of Table 2. For fit residuals we used the threshold value ≥ +/-2.5 (~ 99% significance) to determine misfit to model expectations. Chi-Square probabilities are reported with Bonferroni adjustments to 0.0025.

Targeting

There were no ceiling effects from severely disabled patients in this dataset. Seven datapoints from three patients and four next-of-kin manifested floor scores (indicating no SAH impact), representing 3% of all datapoints. This left 228 scores for analysis (scoring range 1 – 195, mean 43; SD 38). Extreme (minimum or maximum) scores correspond to indefinite measurement values outwith the measurement range of a scale, but their logit measures can be imputed once the measures for non-extreme scores have been estimated using the "only barely extreme" method (Wright, 1998). Most total scores (208/228; 91.2%) fell in the lower half of the scoring range, and the mean Rasch-derived person location estimate (-1.41 logits; SD 0.96) was distant from the mean item estimate (set to 0 for analysis purposes), indicating suboptimal targeting. The range of estimated person locations was wide, extending over 5 logits (-4.29 to +1.44). In comparison, item location estimates ranged from -1.17 to +1.66 logits. Relative person and item location distributions are graphically mapped on the same continuum of SAH impact in Supplementary Figure 1.

Differential item functioning

In order to study whether patients responded differently from their next-of-kin, and whether items performed differently at the time points examined, differential item functioning by these factors was explored. There was evidence of differential functioning by time point for five of the scale's 56 items based on month of completion: "Household chores", "Tolerance of crowded, busy or noisy places", "Basic self care (e.g. ability to wash, dress)", "Learning a new skill", and "Navigational skills (i.e. getting lost)". There was no evidence of differential functioning by respondent (patient vs next-of-kin).

Response dependency

Of 1540 correlations between item residuals, 34 (2.2%) were greater than the conventionally accepted threshold of ±0.3, indicating a small degree of non-random item covariance and possible redundancy between some items.

Performance of item scoring categories

Using the original five scoring categories, category thresholds were disordered (Linacre, 1999) in 45/56 items (80.4%). Supplementary Figure 2A illustrates the pattern of ordered thresholds expected for the five SAHOT categories. Supplementary Figure 2B illustrates the category probability curves for one of the disordered items. The probability curves for categories 1 and 2 are flat, indicating that the likelihood of an individual being assigned grade 1 or 2 is extremely low. Furthermore, at no point in the continuum are grades 1 or 2 the most likely received for this question. Similar disordering occurs in the same categories for all other items. The problem of disordered categories was addressed by collapsing the second and third categories ('small change'; 'moderate change'), and the fourth and fifth categories ('large change'; 'complete change'), producing a 3-category scoring system for analysis purposes. Collapsing categories significantly reduces disordered thresholds at the potential expense of reducing both the scale's discriminatory ability and responsiveness. However, although the scoring range decreased to 0-112 and scale-to-sample targeting deteriorated, person separation index (0.944) improved slightly. Other statistical indices of scale performance remained unchanged (Table 2, third column).

Development phase – the SAHOT ordinal scale

Due to its interval nature, a Rasch-based scale is unable to include mortality within its spectrum, since the interval between mortality and the datapoint representing the poorest outcome cannot be measured on the same scale. In order to be able to include mortality

within the same tool, the SAHOT logit metrics were subsequently graded into categories, based on their standard errors, according to the method of Wright (Wright, 2001). Eight statistically different separation units were identified; in addition to death this resulted in nine ordinal categories. A nomogram is provided for transformation of raw scores to logits and ordinal categories (Table 3).

Development phase: traditional scale analyses of SAHOT

Convergent and discriminant validity

Convergent validity of the SAHOT (Cronbach and Meehl, 1955) was studied via its correlation with the mRS and GOSE, and with prognostic variables (baseline clinical status and blood clot volume). The SAHOT correlated highly with the mRS and GOSE (Table 4A). Although not commonly used in SAH, the cognitive scale CLCE-24 has been shown to correlate with GOS after SAH (Passier *et al.*, 2010), and it also correlated with the SAHOT in this study. The SAHOT correlated better with CLCE-24 ($r=0.813$, $p<10^{-4}$) than either mRS ($r=0.325$, $p=0.174$) or GOSE ($r=-0.622$, $p<10^{-2}$). The SAHOT also correlated positively with prognostic variables, including the WFNS scale (measure of baseline clinical status), Fisher grade and volumetric measurements of baseline blood clot volume. In multivariate linear regression ($r^2=0.337$, $p<10^{-7}$), the interval SAHOT was sensitive to WFNS ($\beta=0.322$, $p10^{-3}$) and aneurysmal *versus* non-aneurysmal aetiology ($\beta=0.178$, $p=0.038$), but not to age and Fisher score. A similar pattern was seen with dichotomized mRS (0-2 *versus* 3-6) and GOSE (1-4 *versus* 5-8) in multivariate logistic regression. It was not possible to accommodate the three scales (ordinal SAHOT, mRS and GOSE) on the same multivariate ordinal regression model due to violation of the proportional odds assumption.

Discriminant validity of the SAHOT (Cronbach and Meehl, 1955) was explored via its correlations with the BICRO39 (a scale that measures psychosocial functioning) and age, and

as expected the scale failed to correlate meaningfully with either factor (Table 4B). BICRO39 is a very specific scale and does not correlate with overall outcome after SAH as measured by the GOS (Powell *et al.*, 2004).

Discrimination between individuals

There was a wide range of scores for each mRS and GOSE category, suggesting that the SAHOT may tell apart individuals with the same mRS or GOSE scores on the basis of measured SAH impact (Figure 1). This effect was more pronounced in patients with an apparently good outcome as indicated by low mRS and high GOSE categories (Figure 1).

In order to study further the discrimination between patients with an apparently good outcome, a treatment trial resulting in a 20% difference in the proportion of the best outcome (mRS=0) was simulated (Table 5). The difference in proportion of the best outcome between the two arms was 12.2% and 9.1% as measured by the ordinal SAHOT=1 and GOSE=8 respectively. In a similar simulation ie modelling a 20% difference in the proportion of GOSE=8, the corresponding difference in ordinal SAHOT=1 and mRS=0 was 7.8% and 5.9% respectively. Finally in a third simulation ie with a 20% difference in the proportion of ordinal SAHOT=1, the difference in proportion between the two groups was 13% and 12.9% for ordinal mRS=0 and GOSE=8 respectively.

In keeping with the observed incongruency between the ordinal SAHOT, mRS and GOSE, the frequency histograms of the three scales show a markedly different distribution (Figure 2).

Responsiveness

Responsiveness of the SAHOT was assessed by measuring its ability to detect improvement with time, between the three and six month time points, using Wilcoxon Signed Ranks test

and effect size (Kazis *et al.*, 1989) (Table 6). Conventionally, effect size values of 0.2–0.49 are defined as small, 0.5–0.79 as moderate and ≥0.8 as large (Cohen, 1960). Results were compared with indices obtained for the mRS and GOSE. The SAHOT and GOSE were significantly responsive, and their effect sizes were similar. The mRS was not able to detect pairwise change between the two time points.

Sensitivity analysis: aneurysmal cases

The SAHOT was developed as a tool for spontaneous (ie non-traumatic) SAH, irrespective of presence or absence of an aneurysm. Drug treatment studies have so far tended to exclude non-aneurysmal SAH, partly since the effect of aneurysm repair procedures are an important determinant of clinical outcome after SAH. Therefore, a sensitivity Rasch analysis of the SAHOT was performed using data from patients with aneurysmal SAH only. The baseline demographic and clinical characteristics of this sub-population are shown in Supplementary Table 1. The same four SAHOT items described above again exhibited poor item fit statistics and poor discrimination between groups with different degrees of SAH impact, resulting in the same 56-item questionnaire. There were no ceiling effects from disabled patients. Based on the 56-item format, six datapoints from three participants manifested floor scores (indicating no SAH impact), representing 4% of all datapoints ie similar to Rasch analysis of the whole population. This left 160 scores for analysis (scoring range 2 – 100, mean 30.5; SD 22). The person separation index was 0.945, above the threshold of 0.9 indicative of unidimensionality, and similar to Rasch analysis of the whole population. Item fit residuals and Chi square probabilities are presented in Supplementary Table 2. The mean Rasch-derived person location estimate (-1.25 logits; SD -1.25), improved after removing the non-aneurysmal cases, though it still indicated suboptimal targeting. Response dependency was similar. Relative person and item location distributions are graphically mapped on the same

continuum of SAH impact in Supplementary Figure 3. Traditional psychometric evaluation of the SAHOT in aneurysmal-only cases is presented in Supplementary Tables 3-4 and Supplementary Figure 4. The validity (Supplementary Table 3), discriminatory ability (Supplementary Figure 4) and responsiveness (Supplementary Table 4) was similar to the SAHOT as applied to the whole population. The only differences were that correlation of the SAHOT with blood clot volume (Supplementary Table 3) and GOSE responsiveness (Supplementary Table 4) lost significance, though a trend persisted in both cases. Since the number of patients is smaller in this sensitivity analysis (*versus* the whole study population), and it was performed post-hoc, caution needs to be exercised when making comparisons of scale performance metrics.

Validation

To validate the optimized version of the SAHOT in a busy clinical setting, it was administered to a separate cohort of unselected SAH patients admitted to our Centre (n=75), together with the mRS and GOSE. The stems of eight questions were rephrased after feedback received from patients during the work above (three reworded, explanation contracted in three, and explanation expanded in two stems). In keeping with the above-described Rasch-based optimization findings, a 3-point Likert scale (0=no change; 1=some change; 2=large or severe change) was used, together with a "not applicable" option. Also, the four items which did not contribute to measurement of SAH impact were excluded, resulting in a 56-item questionnaire (see www.institutional repository.co.uk, currently in supplementary material). Ordinal SAHOT scores, derived using the nomogram in Table 3, correlated highly with the mRS ($r=0.764$, $p<10^{-19}$) and GOSE ($r=0.793$, $p<10^{-21}$).

Discussion

The aim of this study was to address limitations of instruments including the mRS and GOSE in measuring SAH impact, by developing a new condition-specific instrument which measures change in the patient's level of functioning. The SAHOT was founded on a core set of changes in physical, cognitive and social functioning reported by SAH patients and their next-of-kin, refined using Rasch methodology, and validated using both Rasch-based and classical approaches.

Comparison of SAHOT to GOS and mRS

Currently, the GOS and mRS are the most commonly used outcome scales in SAH. The SAHOT represents a substantial improvement on these scales. SAHOT represents the cognitive element better, correlating more closely than the GOSE with the CLCE-24, a cognitive assessment tool (van Heugten *et al.*, 2007); the mRS did not correlate with the CLCE-24. All three scales correlated with the WFNS, with correlation coefficients of 0.345 (p=0.025), 0.381 (p= 0.013) and 0.486 (p=0.001) for the six month SAHOT, mRS and GOSE respectively. In multivariate regression, all three scales were sensitive to the WFNS and the presence or absence of an aneurysm, which are known prognostic factors; no other prognostic factors were significant for all three scales in this dataset. The SAHOT could discriminate patients within individual mRS or GOSE categories (Figure 1). The SAHOT, mRS and GOSE do not measure the same construct since there was substantial overlap in SAHOT scores across mRS and GOSE categories (Figure 1). When a 20% difference in proportion in best outcome (mRS=0, GOSE=8 or SAHOT=1) was simulated, the three scales did not agree with each other, in all three scenarios (Table 5). This in keeping with the fact that the scales have been developed for different conditions. Only the SAHOT is SAH-specific.

The SAHOT has a number of significant advantages over the mRS and GOSE, namely: (1) it is specific to SAH; (2) the starting point for its development was as close as possible to SAH patients and their carers; (3) it incorporates assessment of cognitive, emotional and social domains; (4) scale development involved both classical and modern psychometric approaches; (5) iterative scale modifications incorporating patient and next-of-kin led to improvement of the scale; (6) the scale is administered with minimal training, cost and instrumentation; (7) it is patient-reported; (8) it has high acceptability; (9) it is an interval scale enabling linear regression statistical analysis in survivors, yet can include mortality in its ordinal form; (10) it is more responsive than the mRS; (11) it can discriminate patients with notable differences in SAH impact who otherwise fall within the same mRS and GOSE categories; (12) it represents the cognitive element better than mRS and GOSE.

Rasch-based validation and limitations

Rasch analyses indicated that the scale acceptably meets criteria for successful measurement incorporated within the Rasch model. Limitations will now be discussed.

The range of SAH impact across our cohort was narrower compared to that represented by the scale. One reason could be that up to 28% had non-aneurysmal SAH (Table 1), a rate which is remarkably similar to the UK national rate at the time of writing (27.94% over the last 3 years; source UK National SAH database). This rate has probably been increasing due to improved detection of small SAH. Although non-aneurysmal SAH patients do not have as good an outcome as previously thought (Al-Khindi *et al.*, 2010), they still had better Fisher, WFNS, mRS, GOSE and SAHOT scores, and improved more during the first six months post-SAH (data not shown). Rasch analysis excluding non-aneurysmal cases was performed and showed some improvement in targeting, but it is difficult to draw conclusions since the

number of datapoints was less compared to the primary analysis. Another reason for mistargeting could be under-representation of highly disabled, institutionalized patients due to difficulty with attending follow-up clinics. Although this study may not have fully reflected the entire breadth of SAH impact, it represents the typical inclusion criteria for SAH clinical studies. Despite these limitations, SAHOT achieved good Rasch measurement qualities, and better targeting by inclusion of more severely disabled patients might have a positive impact on the Rasch metrics.

Some correlations between SAHOT item residuals exceeded the conventional threshold of ±0.3 above which pairs of residuals are not considered independent. This may potentially inflate scale reliability estimates. However the number was small (2.2%) and most values only marginally exceeded the threshold, so no items were removed as this would have resulted in loss of information on the measured construct and therefore scale validity.

Some items manifested DIF by time, which potentially weakens the possibility of meaningfully comparing scores within / across people at different time points. DIF does not equate to responsiveness to time; DIF occurs when the location estimate of a particular item changes as a result of other factors unrelated to the entity being measured, so that item measures differentially from the rest. This finding needs confirmation and its implications understood better before considering changes to scale content. If a minimal clinically important difference (MCID) for the SAHOT is determined in future studies, its magnitude could be compared against difference in test scores resulting from DIF by time. If the bias effect is much smaller than the MCID, there would be little risk of confounding the two. Conversely, if the bias approached or exceeded the MCID in magnitude, it could be mistaken for meaningful change.

Knowledge of the mRS and GOSE could not have influenced the SAHOT score for two reasons. First, the SAHOT was directly patient-reported with no operator interference in item scoring, while the mRS and GOSE was administered by the researchers. In cases where the operators were present to observe as part of research, or assist with reading or writing, they were instructed not to assist in interpretation of the questions, although they could make notes if patients asked for clarification to assist the scale development. Second, the overall SAHOT score was computed from the patient responses to the tool items at a different time, in a different location, and not by the same person who administered the mRS and GOSE.

The exact linguistic formulation of SAHOT stems was designed and iteratively refined with patients, so that they are brief and do not require additional explanation over and above the text accompanying them in the tool. Rasch analysis then tested the performance of these items. Translation of this questionnaire into other languages would therefore require revalidation using Rasch analysis.

Traditional scale evaluation

The SAHOT was found to have good convergent and divergent validity, and therefore construct validity. Correlations with Fisher grade and blood clot volume were modest, in keeping with the fact that blood clot volume is only part of a complex set of factors affecting outcome after SAH. When compared to the mRS and GOSE, the SAHOT had a desirable sensitivity, being able to discriminate patients with better outcome on the mRS and GOSE. The SAHOT had excellent responsiveness.

Usability

The use of Rasch-derived person location estimates may be considered a disadvantage in clinical practice due to reliance on Rasch expertise. However there are several ways by which the use of the SAHOT in clinical services or trials may potentially be facilitated. One can integrate Rasch analysis into an automated data processing pipeline that links primary data input to output of person location estimates. Alternatively, nomograms may be used to enable derivation of person location estimates from summed raw scores, as done previously (Mills *et al.*, 2013), assuming that the population studied is similar to the population sample used to develop the nomogram.

This scale was used in a busy clinical service, and was found to be usable. Its usability when completed by patients or next-of-kin on their own was not formally tested, but we did not encounter issues which suggest this will be problematic. The SAHOT is patient-reported while the mRS and GOSE are scored by the assessor. Even if the SAHOT needs to be administered to the patient by a clinician or carer, there is no training involved. Rasch analysis did not detect significant differential item functioning between patients and next-of-kin. This may prove useful in several situations, for example if the patient is unable to complete the assessment due to physical, emotional or speech problems. This design has led to remarkable acceptability.

An important issue to consider is the reachability of patients. We retrospectively assessed the proportion of patients who could not be reached with this tool at three months after SAH, in a nine month period within the development phase, in a systematic fashion. Only five out of 71 patients during this period, ie 7%, were judged unreachable, because they were in a district general hospital (n=3), nursing home (n=1) or rehabilitation unit (n=1). Ten patients (14%)

had died; these patients could be scored on the ordinal SAHOT, but not the Rasch-based interval SAHOT.

Research tool

While the SAHOT is less usable than the mRS and GOSE in highly disabled (eg vegetative) SAH survivors, its strength lies in its sensitivity and responsiveness to change in patients who are not highly disabled. This holds promise for future clinical studies, where the SAHOT should be able to detect changes in patient outcome amongst survivors which have so far evaded the mRS and GOS.

Although Rasch-based scales are superior to ordinal scales, they cannot deal with mortality, which is an important aspect of outcome in clinical trials. One possible solution is to report mortality separately. If mortality rate is lower in the experimental *versus* control arms, a situation may arise whereby the Rasch-based SAHOT outcome of the experimental arm may be paradoxically reduced as a result of the addition of poorly-performing survivors to that arm. In this situation, sensitivity analysis with an ordinal scale able to incorporate mortality is called for. Analysis of the SAHOT as an ordinal scale showed that it had excellent construct validity and was responsive, so that the same questionnaire can double up as an interval and ordinal scale.

Post-traumatic stress disorder

Post-traumatic stress disorder (PTSD) affects up to a third of patients after SAH (Berry, 1998; Noble *et al.*, 2008; Visser-Meily *et al.*, 2013). Studies have shown that this is unrelated to the clinical characteristics of SAH (Noble *et al.*, 2008; Visser-Meily *et al.*, 2013). Instead PTSD is related to the stress occurring post-ictally secondary to maladjustment (Baisch *et al.*,

2011), and can be explained by coping style (Noble *et al.*, 2008; Visser-Meily *et al.*, 2013) and perhaps prior psychiatric morbidity (Hedlund *et al.*, 2011). It is possible that PTSD may influence SAH outcome as measured by SAHOT, mRS and GOSE scales. This hypothesis can be tested in future studies by specifically measuring PTSD, with for example the Impact of Event Scale (Witteveen *et al.*, 2005), to determine the proportion of variance in SAHOT, mRS and GOSE which can be explained by PTSD. If the influence of PTSD is significant and substantial, it should perhaps be considered as a covariate in clinical trials of drugs targeting the biology of SAH, since PTSD would usually not be amenable to modification by such drugs.

Future directions

The SAHOT is the first SAH-specific outcome measure, developed with modern Rasch techniques, with acceptable scale properties. Future research efforts will concentrate on developing the SAHOT into an electronic format for self-reporting, external validation in other centres and across different languages and cultures, demonstrating responsiveness of the scale to effective treatments, and working with regulatory authorities and decision-makers towards its acceptance as a clinical trial outcome.

Acknowledgments

None

Funding

University of Southampton, National Institute of Health Research

Conflicts of Interest

None

Table 1. Baseline characteristics of patients. Mean and range[a], number and %[b]. There were no aneurysmal SAH patients who did not have their aneurysm secured.

	Development phase	**Validation phase**
Number	113	75
Age (years)[a]	58, 21-84	60, 25-84
Fisher grade[b]		
1	6, 5%	6, 8%
2	11, 10%	8, 10%
3	36, 32%	22, 30%
4	60, 53%	39, 52%
WFNS[b]		
1	71, 63%	47, 62%
2	19, 17%	14, 19%
3	5, 4%	2, 3%
4	14, 12%	9, 12%
5	4, 4%	3, 4%
Gender[b]		
male	28, 25%	22, 29%
female	85, 75%	53, 71%
Intervention[b]		
coiled	69, 61%	43, 23%
clipped	13, 11%	17, 57%
no aneurysm identified	31, 28%	15, 20%
Aneurysm location[b]		
anterior	66, 58%	54, 72%
posterior	16, 14%	6, 8%
no aneurysm identified	31, 28%	15, 20%

Table 2. Rasch-based scale properties of the 56-item SAHOT analysed as 3 or 5 item-scoring categories

	All patients	
	5 item-scoring categories	**3 item-scoring categories**
Scoring range	0-224	0-112
Person separation index	0.938	0.944
Mean person location (SD)	-1.41 (0.96)	-1.44 (1.30)
Item category thresholds	80.4% disordered	16.6% disordered
Item fit	Fit residuals >±2.5 in 17.9%	Fit residuals >±2.5 in 14.3%
	X^2 probability <Bonferroni adjustment in 3/56 items	X^2 probability <Bonferroni adjustment in 2/56 items
Excess correlation between item residuals (>0.30)	2.2%	1.5%

Table 3. Nomogram for transformation of raw SAHOT scores to ordinal categories

Raw score*	**SAHOT category**
0 - 7	1 – Best outcome
8 - 17	2
18 - 29	3
30 - 42	4
43 - 56	5
57 - 73	6
74 - 89	7
90 - 112	8
N/A	9 - Death

***** 56 questions, possible score 0-2, maximum total score = 112

Table 4. Convergent and discriminant validity

	Interval SAHOT		Ordinal SAHOT	
Tool	**Correlation coefficient**	**p**	**Correlation coefficient**	**p**
A. Convergent validity: outcome				
mRS	0.625	$<10^{-4}$ **	0.775	$<10^{-9}$ **
GOSE	-0.734	$<10^{-7}$ **	-0.812	$<10^{-11}$ **
CLCE-24	0.813	$<10^{-4}$ **	0.795	$<10^{-4}$ **
A. Convergent validity: prognosis				
WFNS	0.345	0.025 *	0.553	$<10^{-4}$ *
Fisher score	0.211	0.179	0.345	0.017 *
Blood clot volume: total	0.444	0.011 *	0.467	0.007 *
Blood clot volume: subarachnoid	0.381	0.032 *	0.392	0.026 *
B. Discriminant validity				
BICRO39	0.167	0.495	0.211	0.386
Age	-0.186	0.238	-0.094	0.531

Table 5. Discrimination between individuals in the best outcome category across the ordinal SAHOT, GOSE and mRS. A treatment trial resulting in a 20% difference in the proportion of the best outcome (first column: mRS=0, GOSE=8, or ordinal SAHOT=1) was simulated. The second to fourth columns show the resulting difference in the proportion of the best outcome (mRS=0, GOSE=8, or ordinal SAHOT=1).

	Difference in the proportion of the best outcome		
20% difference in the proportion of the best outcome as determined by	**mRS = 0**	**GOSE = 8**	**Ordinal SAHOT = 1**
mRS = 0	20%	9.1%	12.2%
GOSE = 8	5.9%	20%	7.8%
Ordinal SAHOT = 1	13%	12.9%	20%

Table 6. Responsiveness to time

Tool	3 month score median (range)	6 month score median (range)	p	Effect size
Interval SAHOT	-1 (-4.2 to 1.2)	-1.4 (-3.6 to 0.9)	0.02 *	0.24
Ordinal SAHOT	3 (1 to 6)	2 (1 to 6)	0.02 *	0.20
mRS	1 (0 to 5)	1 (0 to 4)	0.46	0.14
GOSE	5 (3 to 8)	7 (3 to 8)	0.04 *	0.32

Figure 1. Discrimination between individuals. A-B: Interval SAHOT *versus* mRS (A) and GOSE (B). C-D: Ordinal SAHOT *versus* mRS (C) and GOSE (D)

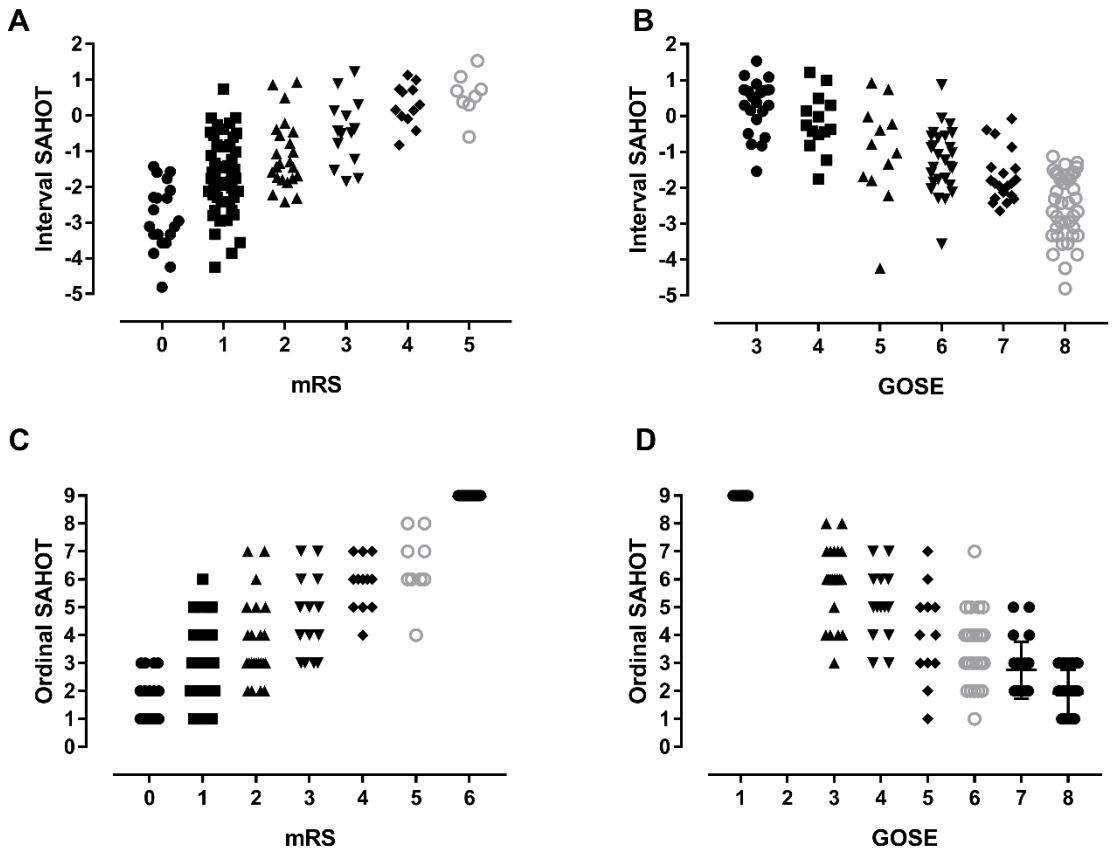
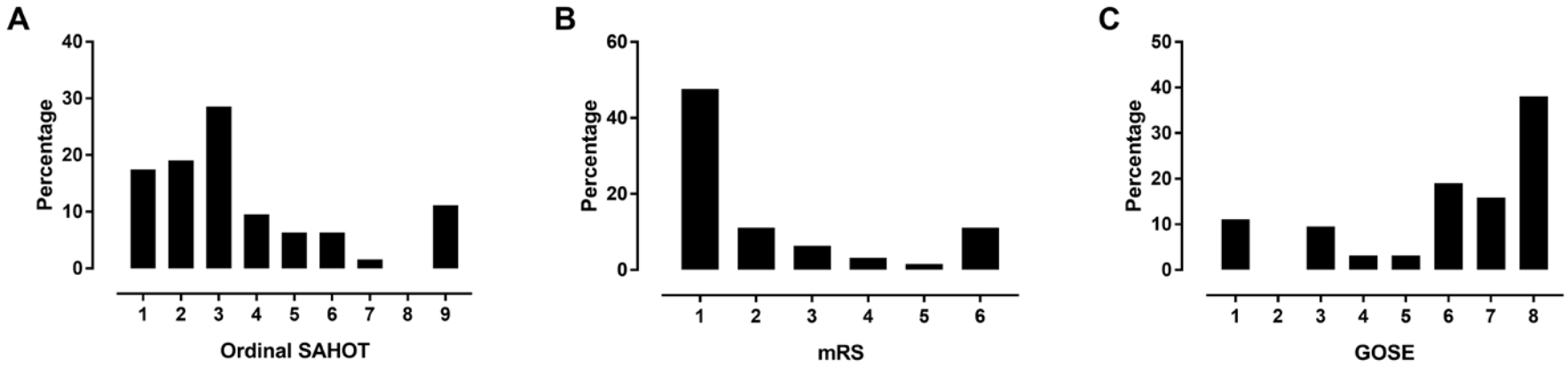


Figure 2. Frequency distribution of scores for the ordinal SAHOT (A), mRS (B) and GOSE (C)



References

Al-Khindi T, Macdonald RL, Schweizer TA. Cognitive and functional outcome after aneurysmal subarachnoid hemorrhage. Stroke 2010; 41(8): e519-36.

Andrich D. An index of person separation in latent trait theory, the traditional KR. 20 index, and the Guttman scale response pattern. Education Research and Perspectives 1982; 9(1): 95-104.

Andrich DS, B.E.; Luo, G. RUMM2030 - Professional Edition (Rasch Unidimensional Measurement Model, version 5.2). Perth, Western Australia: RUMM Laboratory; 2012.

Baisch SB, Schenk T, Noble AJ. What is the cause of post-traumatic stress disorder following subarachnoid haemorrhage? Post-ictal events are key. Acta Neurochir (Wien) 2011; 153(4): 913-22.

Berry E. Post-traumatic stress disorder after subarachnoid haemorrhage. Br J Clin Psychol 1998; 37 (Pt 3): 365-7.

Boosman H, Passier PE, Visser-Meily JM, Rinkel GJ, Post MW. Validation of the Stroke Specific Quality of Life scale in patients with aneurysmal subarachnoid haemorrhage. J Neurol Neurosurg Psychiatry 2010; 81(5): 485-9.

Cohen J. A Coefficient of Agreement for Nominal Scales. Educational and Psychological Measurement 1960; 20(1): 37-46.

Cronbach LJ, Meehl PE. Construct validity in psychological tests. Psychol Bull 1955; 52(4): 281-302.

Dromerick AW, Edwards DF, Diringer MN. Sensitivity to changes in disability after stroke: a comparison of four scales useful in clinical trials. J Rehabil Res Dev 2003; 40(1): 1-8.

Farrell B, Godwin J, Richards S, Warlow C. The United Kingdom transient ischaemic attack (UK-TIA) aspirin trial: final results. J Neurol Neurosurg Psychiatry 1991; 54(12): 1044-54.

Fisher CM, Kistler JP, Davis JM. Relation of cerebral vasospasm to subarachnoid hemorrhage visualized by computerized tomographic scanning. Neurosurgery 1980; 6(1): 1-9.

Hedlund M, Zetterling M, Ronne-Engstrom E, Carlsson M, Ekselius L. Depression and post-traumatic stress disorder after aneurysmal subarachnoid haemorrhage in relation to lifetime psychiatric morbidity. Br J Neurosurg 2011; 25(6): 693-700.

Huang J, van Gelder JM. The probability of sudden death from rupture of intracranial aneurysms: a meta-analysis. Neurosurgery 2002; 51(5): 1101-5; discussion 5-7.

Hutter BO, Gilsbach JM. Which neuropsychological deficits are hidden behind a good outcome (Glasgow = I) after aneurysmal subarachnoid hemorrhage? Neurosurgery 1993; 33(6): 999-1005; discussion -6.

Jennett B, Bond M. Assessment of outcome after severe brain damage. Lancet 1975; 1(7905): 480-4.

Jennett B, Snoek J, Bond MR, Brooks N. Disability after severe head injury: observations on the use of the Glasgow Outcome Scale. J Neurol Neurosurg Psychiatry 1981; 44(4): 285-93.

Kazis LE, Anderson JJ, Meenan RF. Effect sizes for interpreting changes in health status. Med Care 1989; 27(3 Suppl): S178-89.

Kim DH, Haney CL, Van Ginhoven G. Utility of outcome measures after treatment for intracranial aneurysms: a prospective trial involving 520 patients. Stroke 2005; 36(4): 792-6.

Kreiter KT, Copeland D, Bernardini GL, Bates JE, Peery S, Claassen J, *et al.* Predictors of cognitive dysfunction after subarachnoid hemorrhage. Stroke 2002; 33(1): 200-8.

Linacre J. Category disordering vs. step (threshold) disordering. Rasch Measurement Transactions 1999; 13(1): 675.

Macdonald RL, Cusimano MD, Etminan N, Hanggi D, Hasan D, Ilodigwe D, *et al.* Subarachnoid Hemorrhage International Trialists data repository (SAHIT). World Neurosurg 2013a; 79(3-4): 418-22.

Macdonald RL, Jaja B, Cusimano MD, Etminan N, Hanggi D, Hasan D, *et al.* SAHIT Investigators--on the outcome of some subarachnoid hemorrhage clinical trials. Transl Stroke Res 2013b; 4(3): 286-96.

Meyer B, Ringel F, Winter Y, Spottke A, Gharevi N, Dams J, *et al.* Health-related quality of life in patients with subarachnoid haemorrhage. Cerebrovasc Dis 2010; 30(4): 423-31.

Mills RJ, Calabresi M, Tennant A, Young CA. Perceived changes and minimum clinically important difference of the Neurological Fatigue Index for multiple sclerosis (NFI-MS). Mult Scler 2013; 19(4): 502-5.

Noble AJ, Baisch S, Mendelow AD, Allen L, Kane P, Schenk T. Posttraumatic stress disorder explains reduced quality of life in subarachnoid hemorrhage patients in both the short and long term. Neurosurgery 2008; 63(6): 1095-104; discussion 04-5.

Passier PE, Visser-Meily JM, van Zandvoort MJ, Post MW, Rinkel GJ, van Heugten C. Prevalence and determinants of cognitive complaints after aneurysmal subarachnoid hemorrhage. Cerebrovasc Dis 2010; 29(6): 557-63.

Powell J, Kitchen N, Heslin J, Greenwood R. Psychosocial outcomes at three and nine months after good neurological recovery from aneurysmal subarachnoid haemorrhage: predictors and prognosis. J Neurol Neurosurg Psychiatry 2002; 72(6): 772-81.

Powell J, Kitchen N, Heslin J, Greenwood R. Psychosocial outcomes at 18 months after good neurological recovery from aneurysmal subarachnoid haemorrhage. J Neurol Neurosurg Psychiatry 2004; 75(8): 1119-24.

Quinn AC, Bhargava D, Al-Tamimi YZ, Clark MJ, Ross SA, Tennant A. Self-perceived health status following aneurysmal subarachnoid haemorrhage: a cohort study. BMJ Open 2014; 4(4): e003932.

Quinn TJ, Dawson J, Walters MR, Lees KR. Reliability of the modified Rankin Scale: a systematic review. Stroke 2009; 40(10): 3393-5.

Rankin J. Cerebral vascular accidents in patients over the age of 60. II. Prognosis. Scott Med J 1957; 2(5): 200-15.

Rasch G. Probabilistic Models for Some Intelligence and Attainment Tests: University of Chicago Press; 1980.

Rivero-Arias O, Gray A, Wolstenholme J. Burden of disease and costs of aneurysmal subarachnoid haemorrhage (aSAH) in the United Kingdom. Cost Eff Resour Alloc 2010; 8: 6.

Schweizer TA, Macdonald RL. Brain hemorrhage: assessing outcome after subarachnoid hemorrhage. Nat Rev Neurol 2010; 6(8): 427-8.

Sembi S, Tarrier N, O'Neill P, Burns A, Faragher B. Does post-traumatic stress disorder occur after stroke: a preliminary study. Int J Geriatr Psychiatry 1998; 13(5): 315-22.

Shalev AY, Freedman S, Peri T, Brandes D, Sahar T, Orr SP, *et al.* Prospective study of posttraumatic stress disorder and depression following trauma. Am J Psychiatry 1998; 155(5): 630-7.

Taylor TN, Davis PH, Torner JC, Holmes J, Meyer JW, Jacobson MF. Lifetime cost of stroke in the United States. Stroke 1996; 27(9): 1459-66.

Teasdale GM, Drake CG, Hunt W, Kassell N, Sano K, Pertuiset B, *et al.* A universal subarachnoid hemorrhage scale: report of a committee of the World Federation of Neurosurgical Societies. J Neurol Neurosurg Psychiatry 1988; 51(11): 1457.

Tennant A, Conaghan PG. The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper? Arthritis Rheum 2007; 57(8): 1358-62.

Tennant A, McKenna SP, Hagell P. Application of Rasch analysis in the development and application of quality of life instruments. Value Health 2004; 7 Suppl 1: S22-6.

Tesio L. Measuring behaviours and perceptions: Rasch analysis as a tool for rehabilitation research. J Rehabil Med 2003; 35(3): 105-15.

van Heugten C, Rasquin S, Winkens I, Beusmans G, Verhey F. Checklist for cognitive and emotional consequences following stroke (CLCE-24): development, usability and quality of the self-report version. Clin Neurol Neurosurg 2007; 109(3): 257-62.

Visser-Meily JM, Rinkel GJ, Vergouwen MD, Passier PE, van Zandvoort MJ, Post MW. Post-traumatic stress disorder in patients 3 years after aneurysmal subarachnoid haemorrhage. Cerebrovasc Dis 2013; 36(2): 126-30.

Wilson JT, Pettigrew LE, Teasdale GM. Structured interviews for the Glasgow Outcome Scale and the extended Glasgow Outcome Scale: guidelines for their use. J Neurotrauma 1998; 15(8): 573-85.

Witteveen AB, Bramsen I, Hovens JE, van der Ploeg HM. Utility of the impact of event scale in screening for posttraumatic stress disorder. Psychol Rep 2005; 97(1): 297-308.

Wong GK, Lam SW, Ngai K, Wong A, Mok V, Poon WS. Quality of Life after Brain Injury (QOLIBRI) Overall Scale for patients after aneurysmal subarachnoid hemorrhage. J Clin Neurosci 2014; 21(6): 954-6.

Wright BD. Estimating Rasch measures for extreme scores. Rasch Measurement Transactions 1998; 12(2): 2.

Wright BD. Separation, Reliability and Skewed Distributions: Statistically Different Levels of Performance. Rasch Measurement Transactions 2001; 14(4).

SUPPLEMENTARY MATERIAL

A Subarachnoid Haemorrhage-Specific Outcome Tool

Adrian Pace[1,*], Sophie Mitchell[2,*], Lizzie Casselden[2*], Ardalan Zolnourian[3], James Glazier[2], Lesley Foulkes[3], Diederik Bulters[3,#], Ian Galea[2,3,#]

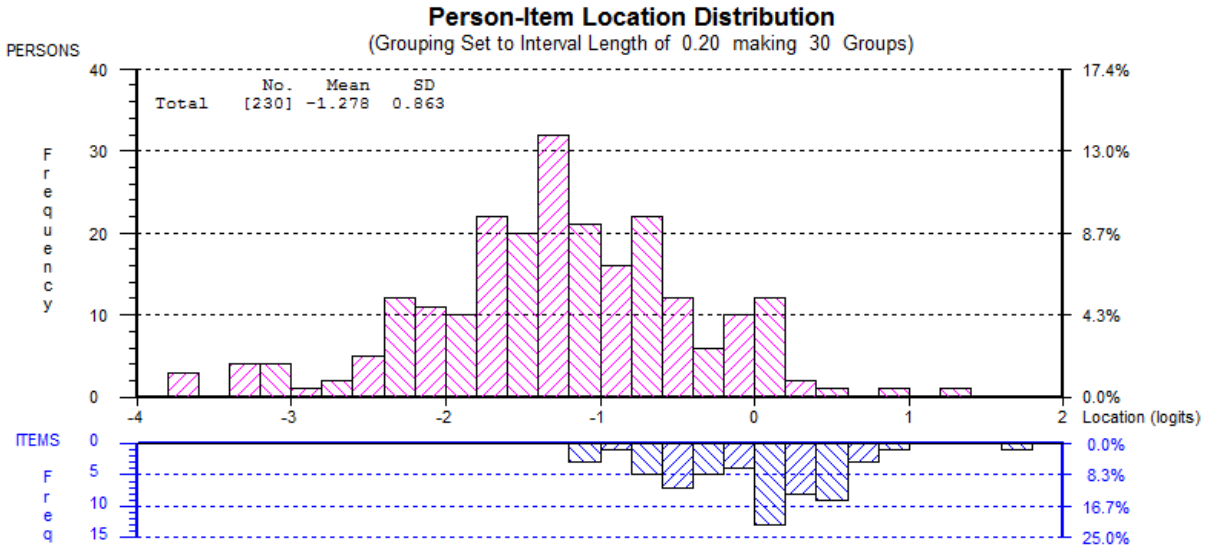
From:

1. Greater Manchester Neurosciences Centre, Salford Royal NHS Foundation Trust, UK
2. Clinical Neurosciences, Clinical & Experimental Sciences, Faculty of Medicine, University of Southampton
3. Wessex Neurosciences Centre, University Hospital Southampton NHS Foundation Trust

* These authors contributed equally to the work

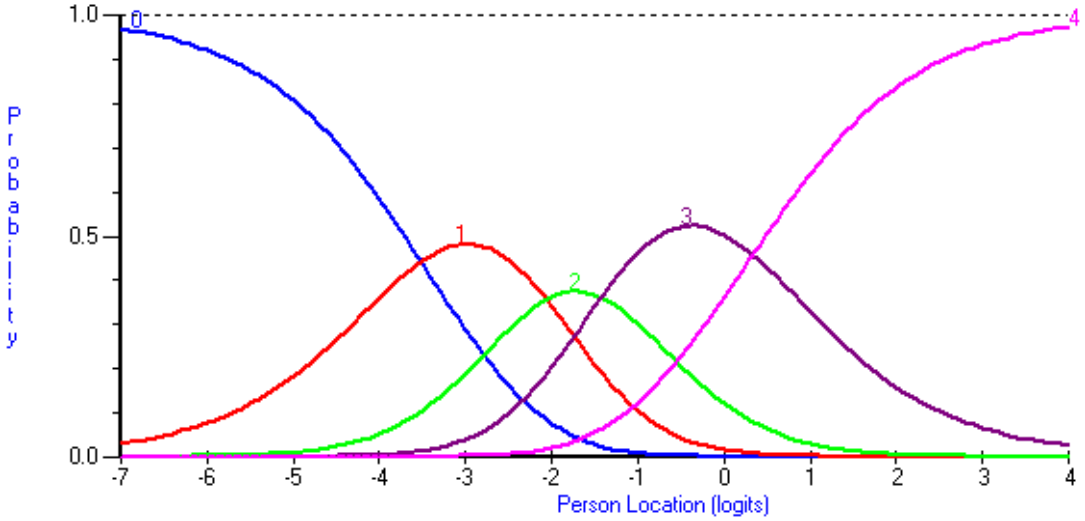
[#] Joint senior authors

Supplementary Figure 1. Person-Item Location Distribution Map

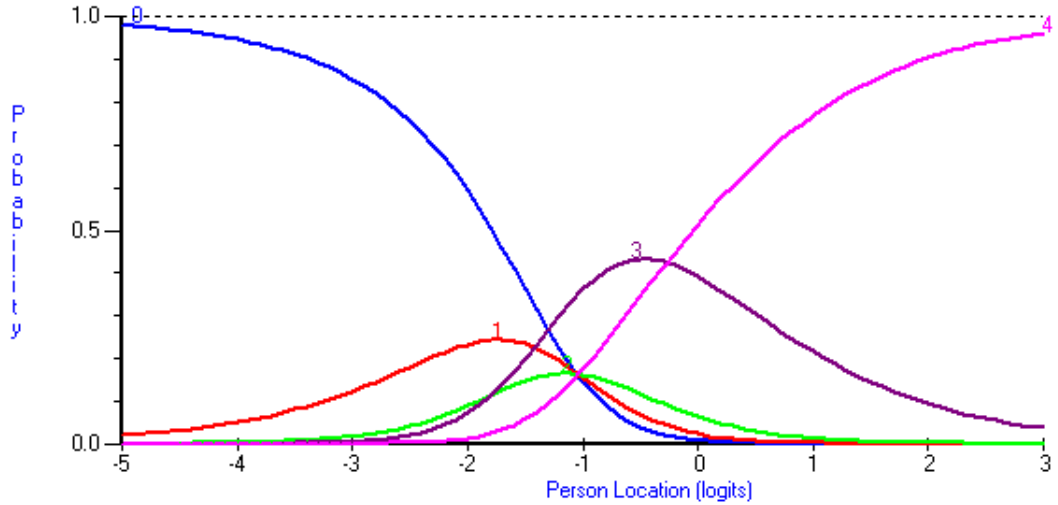


Relative distribution of location estimates for participants with non-extreme scores (upper red histogram) and the 60 SAHOT items (lower blue histogram) mapped on the same continuum of SAH impact. The height of each histogram bar reflects the number of persons or items whose location estimates fall in the same region of the continuum. In this graph, the grouping function for such regions is set to intervals of 0.20 logits.

Supplementary Figure 2A. Expected ordering of HASH-OT response categories

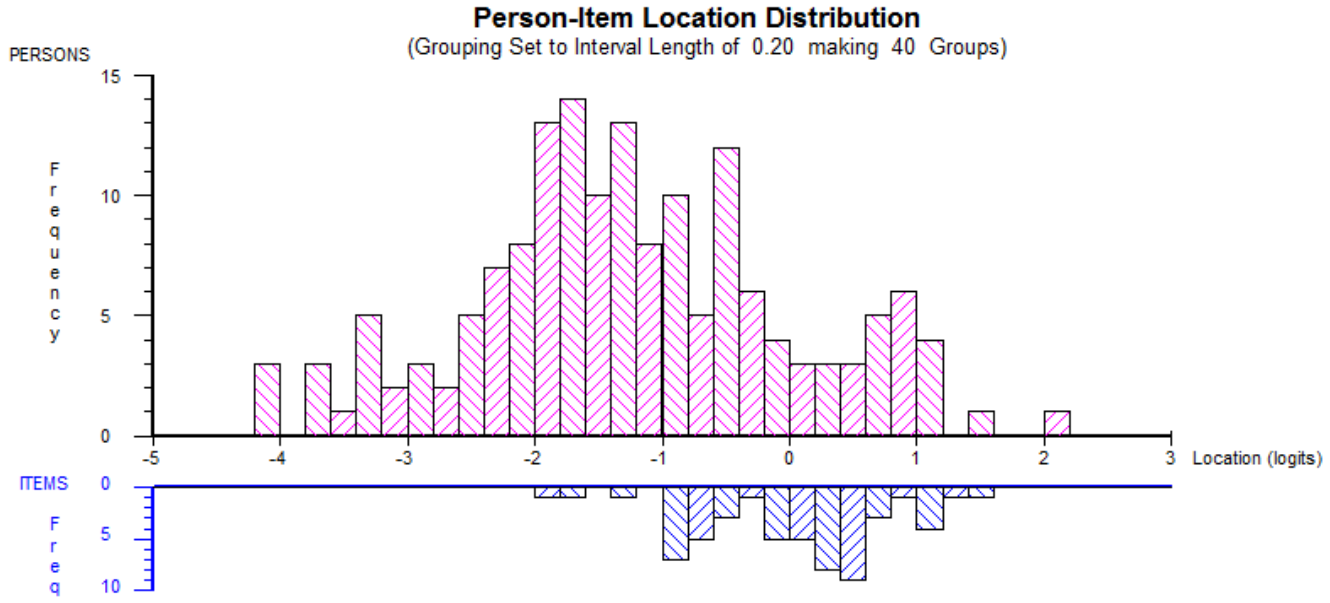


Supplementary Figure 2B. Observed ordering of HASH-OT response categories



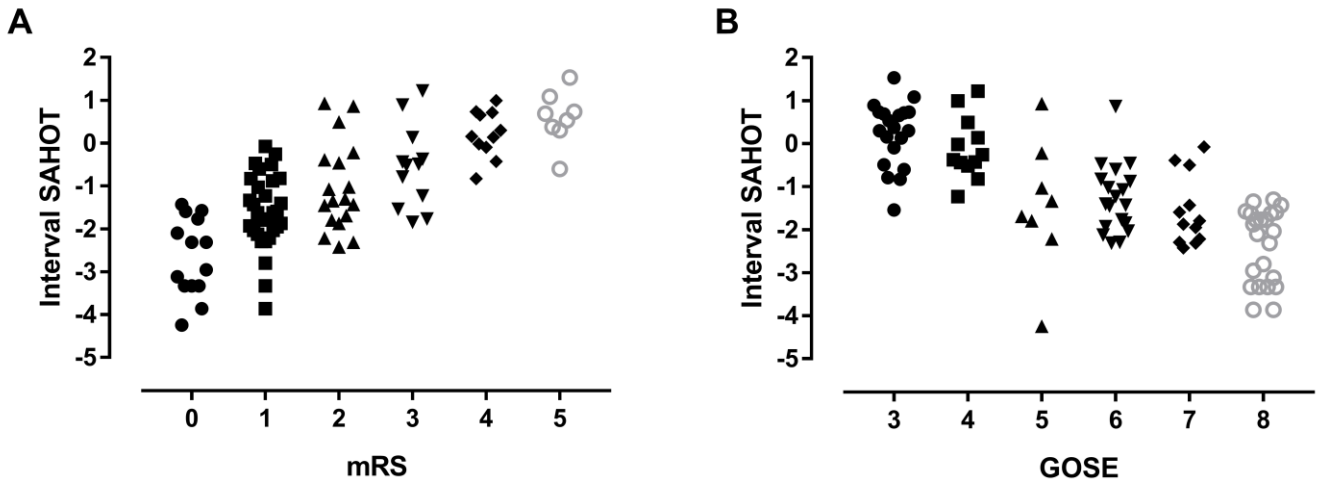
Category Probability Curves (CPCs) depict the probability (y-axis) of being graded in consecutive response categories (coloured lines) with increasing strength (x-axis). Figure 2a shows how the probability of being graded in each of the 5 HASH-OT categories should follow a logical ordered sequence from less to more. Figure 2b shows that in this item, response category 1 (red curve) and 2 (green curve) are never the most likely category for a patient to be graded in, at any level of SAH impact. Thus, the sequential ordering of the 5 categories did not work as intended.

Supplementary Figure 3. Person-Item Location Distribution Map (Aneurysmal-only SAH)



Relative distribution of location estimates for participants with non-extreme scores (upper red histogram) and the 56 SAHOT items (lower blue histogram) mapped on the same continuum of SAH impact. The height of each histogram bar reflects the number of persons or items whose location estimates fall in the same region of the continuum. In this graph, the grouping function for such regions is set to intervals of 0.20 logits.

Supplementary Figure 4. Discrimination between individuals (aneurysmal cases only). SAHOT *versus* mRS (A) and GOSE (B).



Supplementary Table 1. Baseline characteristics of patients – aneurysmal cases only. Mean and range[a], median, number and %[b]

Number	82
Age (years)[a]	58, 21-84
Fisher grade[b]	
1	2, 3%
2	5, 6%
3	25, 30%
4	50, 61%
Median, range	4, 1-4
WFNS[b]	
1	47, 57%
2	16, 19%
3	3, 4%
4	12, 15%
5	4, 5%
Median, range	1, 1-5

Gender	
male	16, 20%
female	66, 80%

Table 2. Rasch-based scale properties of the 56-item SAHOT (3 item-scoring categories) in aneurysmal-only cases

Scoring range	0-112
Person separation index	0.945
Mean person location (SD)	-1.25 (1.25)
Item category thresholds	46/56 ordered (82%)
Item fit	Fit residuals >±2.5 in 10.7%
	X^2 probability <Bonferroni adjustment in 6/56 items
Excess correlation between item residuals (>0.30)	1.8%

Supplementary Table 3. Convergent and discriminant validity of the SAHOT – aneurysmal cases only

	Correlation coefficient	**p**
A. Convergent validity: outcome		
mRS	0.743	$<10^{-4}$ **
GOSE	-0.749	$<10^{-4}$ **
CLCE24	0.640	0.003 **
A. Convergent validity: prognosis		
WFNS	0.444	0.026 *
Fisher score	0.217	0.296
Blood clot volume: total	0.419	0.074
Blood clot volume: subarachnoid	0.339	0.156
B. Discriminant validity		
BICRO39	0.321	0.180
Age	0.187	0.371

Supplementary Table 4. Responsiveness of the SAHOT – aneurysmal cases only

Tool	3 month score median (range)	6 month score median (range)	p	Effect size
Interval SAHOT	-1.1 (-2.2 to 1.2)	-1.3 (-3.3 to 0.9)	0.03*	0.28
mRS	1.5 (0 to 4)	1 (0 to 5)	0.89	0.05
GOSE	5 (3 to 8)	5.5 (3 to 8)	0.19	0.36

The 60-item development questionnaire follows next

This tool, upon which the study was based, is the authors' creation under a CC BY-SA licence.

Subarachnoid Haemorrhage Outcome Assessment

This form is designed to assess recovery following subarachnoid hemorrhage at this moment in time. The patient and their next of kin should fill in separate forms without consulting each other.

Please think back to how things were before the bleed, and compare this to how the bleed has IMPACTED on the following aspects of daily life NOW (i.e. this week). Please circle the correct response. If a question is not relevant, please circle "N/A'' (Not Applicable).

1. General Aspects of Daily Life

OVERALL FUNCTION	No Change	Small Change	Moderate Change	Large Change	Complete Change	N/A	Better / Worse
Physical activities of daily life *(e.g. walking, climbing stairs)*	No Change	Small Change	Moderate Change	Large Change	Complete Change	N/A	Better / Worse
Work *(i.e. number of working hours and how much one can do at work)*	No Change	Small Change	Moderate Change	Large Change	Complete Change	N/A	Better / Worse
Income *(gross income)*	No Change	Small Change	Moderate Change	Large Change	Complete Change	N/A	Better / Worse

Driving	No Change	Small Change	Moderate Change	Large Change	Complete Change	N/A	Better / Worse
Socializing	No Change	Small Change	Moderate Change	Large Change	Complete Change	N/A	Better / Worse
Pursuing previous hobbies	No Change	Small Change	Moderate Change	Large Change	Complete Change	N/A	Better / Worse
Household chores	No Change	Small Change	Moderate Change	Large Change	Complete Change	N/A	Better / Worse
Days / evenings out	No Change	Small Change	Moderate Change	Large Change	Complete Change	N/A	Better / Worse
Quality of relationship with those closest	No Change	Small Change	Moderate Change	Large Change	Complete Change	N/A	Better / Worse
Quality of relationships with others	No Change	Small Change	Moderate Change	Large Change	Complete Change	N/A	Better / Worse
Doing things on one's own *(e.g. shopping, going out)*	No Change	Small Change	Moderate Change	Large Change	Complete Change	N/A	Better / Worse
Tolerance of crowded, busy or noisy places	No Change	Small Change	Moderate Change	Large Change	Complete Change	N/A	Better / Worse
Sleep pattern	No	Small	Moderate	Large	Complete	N/A	Better

(day or night)	Change	Change	Change	Change	Change		Worse
Sex life	No Change	Small Change	Moderate Change	Large Change	Complete Change	N/A	Better / Worse
Basic self care *(e.g. ability to wash, dress)*	No Change	Small Change	Moderate Change	Large Change	Complete Change	N/A	Better / Worse
Exercise *(e.g. sporting activities)*	No Change	Small Change	Moderate Change	Large Change	Complete Change	N/A	Better / Worse

2. Physical Aspects

Physical fatigue / tiredness *(i.e. how much one can do before needing to stop to rest)*	No Change	Small Change	Moderate Change	Large Change	Complete Change	N/A	Better / Worse
Balance when walking	No Change	Small Change	Moderate Change	Large Change	Complete Change	N/A	Better / Worse
Clumsiness *(change in handwriting, difficulty with cutlery, knocking things over)*	No Change	Small Change	Moderate Change	Large Change	Complete Change	N/A	Better / Worse
Falls *(including trips / stumbling)*	No Change	Small Change	Moderate Change	Large Change	Complete Change	N/A	Better / Worse
Strength / coordination in arms and hands	No	Small	Moderate	Large	Complete	N/A	Better

	Change	Change	Change	Change	Change		Worse
Strength / coordination in legs	No Change	Small Change	Moderate Change	Large Change	Complete Change	N/A	Better Worse
Pain	No Change	Small Change	Moderate Change	Large Change	Complete Change	N/A	Better Worse
Headaches	No Change	Small Change	Moderate Change	Large Change	Complete Change	N/A	Better Worse
Urinary continence	No Change	Small Change	Moderate Change	Large Change	Complete Change	N/A	Better Worse
Eyesight	No Change	Small Change	Moderate Change	Large Change	Complete Change	N/A	Better Worse
Hearing	No Change	Small Change	Moderate Change	Large Change	Complete Change	N/A	Better Worse
Smell / taste	No Change	Small Change	Moderate Change	Large Change	Complete Change	N/A	Better Worse
Swallowing food or water	No Change	Small Change	Moderate Change	Large Change	Complete Change	N/A	Better Worse
Word finding when speaking	No Change	Small Change	Moderate Change	Large Change	Complete Change	N/A	Better Worse

Recognising faces	No Change	Small Change	Moderate Change	Large Change	Complete Change	N/A	Better / Worse
Ability to get a point across in conversation	No Change	Small Change	Moderate Change	Large Change	Complete Change	N/A	Better / Worse
Ability to compromise in discussion with others	No Change	Small Change	Moderate Change	Large Change	Complete Change	N/A	Better / Worse
Ability to recognise danger	No Change	Small Change	Moderate Change	Large Change	Complete Change	N/A	Better / Worse
Navigational skills *(i.e. getting lost)*	No Change	Small Change	Moderate Change	Large Change	Complete Change	N/A	Better / Worse

4. Behavioural / Psychological Aspects

Low mood	No Change	Small Change	Moderate Change	Large Change	Complete Change	N/A	Better / Worse
Mood swings	No Change	Small Change	Moderate Change	Large Change	Complete Change	N/A	Better / Worse
Strength of emotions	No Change	Small Change	Moderate Change	Large Change	Complete Change	N/A	Better / Worse

Easily moved to tearfulness or laughter	No Change	Small Change	Moderate Change	Large Change	Complete Change	N/A	Better / Worse
Ability to control one's reactions	No Change	Small Change	Moderate Change	Large Change	Complete Change	N/A	Better / Worse
Irritability	No Change	Small Change	Moderate Change	Large Change	Complete Change	N/A	Better / Worse
Anxiety	No Change	Small Change	Moderate Change	Large Change	Complete Change	N/A	Better / Worse
Feelings of fear	No Change	Small Change	Moderate Change	Large Change	Complete Change	N/A	Better / Worse
Feelings of paranoia	No Change	Small Change	Moderate Change	Large Change	Complete Change	N/A	Better / Worse
Agitation	No Change	Small Change	Moderate Change	Large Change	Complete Change	N/A	Better / Worse
Restlessness *(Inability to stand still)*	No Change	Small Change	Moderate Change	Large Change	Complete Change	N/A	Better / Worse
Self-confidence	No Change	Small Change	Moderate Change	Large Change	Complete Change	N/A	Better / Worse

Awareness of others' thoughts, feelings and/or needs	No Change	Small Change	Moderate Change	Large Change	Complete Change	N/A	Better / Worse

Motivation	No Change	Small Change	Moderate Change	Large Change	Complete Change	N/A	Better / Worse

Placidity *(e.g. remaining calm in situations when previously would have not)*	No Change	Small Change	Moderate Change	Large Change	Complete Change	N/A	Better / Worse

Feeling comfortable in new environments	No Change	Small Change	Moderate Change	Large Change	Complete Change	N/A	Better / Worse

The final 56-item questionnaire follows next

This tool, upon which the study was based, is the authors' creation under a CC BY-SA licence.

SAHOT

SubArachnoid Haemorrhage Outcome Tool

This form is designed to assess recovery following subarachnoid hemorrhage <u>at this moment in time</u>. The patient and their next of kin should fill in separate forms without consulting each other.

- **Please think back to how things were <u>BEFORE</u> the bleed, and compare this to how the following aspects of daily life are <u>NOW</u> (i.e. this week).**
- **Please circle the response that best describes this <u>CHANGE</u> for each aspect.**
- **If you have not yet tried an activity, or are unsure if you would be able to undertake a task, circle "large/severe change" for the purposes of this questionnaire.**
- **If you did not do an activity before the bleed, please select "N/A".**

1. General Aspects of Daily Life

OVERALL FUNCTION	No change	Some change	Large or severe change	N/A
Physical activities of daily life *(e.g. walking, climbing stairs)*	No change	Some change	Large or severe change	N/A
Socializing *(with people other than colleagues/family)*	No change	Some change	Large or severe change	N/A
Pursuing previous hobbies	No change	Some change	Large or severe change	N/A
Household chores	No change	Some change	Large or severe change	N/A
Days / evenings out	No change	Some change	Large or severe change	N/A
Quality of relationship with those closest	No change	Some change	Large or severe change	N/A
Tick if relationship is now better [ ] or worse [ ]				
Quality of relationships with others	No change	Some change	Large or severe change	N/A
Doing things on one's own *(e.g. shopping, going out)*	No change	Some change	Large or severe change	N/A
Coping in crowded, busy or noisy places	No change	Some change	Large or severe change	N/A
Sleep pattern *(day or night)*	No change	Some change	Large or severe change	N/A

Sex life	No change	Some change	Large or severe change	N/A
Basic self care *(e.g. ability to wash, dress)*	No change	Some change	Large or severe change	N/A
Recreational exercise	No change	Some change	Large or severe change	N/A

2. Physical Aspects

Physical fatigue / tiredness *(i.e. how much one can do before needing to stop to rest)*	No change	Some change	Large or severe change	N/A
Balance when walking	No change	Some change	Large or severe change	N/A
Clumsiness *(change in handwriting, difficulty with cutlery, knocking things over)*	No change	Some change	Large or severe change	N/A
Falls *(including trips / stumbling)*	No change	Some change	Large or severe change	N/A
Strength / coordination in arms and hands	No change	Some change	Large or severe change	N/A
Strength / coordination in legs	No change	Some change	Large or severe change	N/A
Pain	No change	Some change	Large or severe change	N/A
Urinary continence	No change	Some change	Large or severe change	N/A

Vision (excluding changes in prescription of glasses)	No change	Some change	Large or severe change	N/A
Hearing	No change	Some change	Large or severe change	N/A
Smell / taste	No change	Some change	Large or severe change	N/A
Swallowing food or water	No change	Some change	Large or severe change	N/A
Word finding when speaking	No change	Some change	Large or severe change	N/A

3. Cognitive Aspects

Mental fatigue *(i.e. tiredness with mental tasks)*	No change	Some change	Large or severe change	N/A
Short-term memory	No change	Some change	Large or severe change	N/A
Long-term memory *(i.e. remembering things that happened years ago)*	No change	Some change	Large or severe change	N/A
Learning a new skill	No change	Some change	Large or severe change	N/A
Concentration	No change	Some change	Large or severe change	N/A
Distractibility	No change	Some change	Large or severe change	N/A

Multitasking *(i.e. doing two or more things at the same time)*	No change	Some change	Large or severe change	N/A
Remembering names of familiar people	No change	Some change	Large or severe change	N/A
Recognising faces	No change	Some change	Large or severe change	N/A
Ability to get a point across in conversation	No change	Some change	Large or severe change	N/A
Ability to compromise in discussion with others	No change	Some change	Large or severe change	N/A
Ability to recognise danger	No change	Some change	Large or severe change	N/A
Navigational skills *(i.e. getting lost)*	No change	Some change	Large or severe change	N/A

4. Behavioural / Psychological Aspects

Low mood	No change	Some change	Large or severe change	N/A
Mood swings	No change	Some change	Large or severe change	N/A
Strength of emotions	No change	Some change	Large or severe change	N/A
Easily moved to tearfulness or laughter	No change	Some change	Large or severe change	N/A

Ability to control one's reactions	No change	Some change	Large or severe change	N/A
Irritability	No change	Some change	Large or severe change	N/A
Anxiety	No change	Some change	Large or severe change	N/A
Feelings of fear	No change	Some change	Large or severe change	N/A
Feelings of paranoia	No change	Some change	Large or severe change	N/A
Agitation	No change	Some change	Large or severe change	N/A
Restlessness *(inability to stand still)*	No change	Some change	Large or severe change	N/A
Self-confidence	No change	Some change	Large or severe change	N/A
Awareness of others' thoughts, feelings and/or needs	No change	Some change	Large or severe change	N/A
Motivation	No change	Some change	Large or severe change	N/A
Feeling comfortable in new environments	No change	Some change	Large or severe change	N/A
Apathy	No change	Some change	Large or severe change	N/A