ORIGINAL REPORT

GOAL ATTAINMENT SCALING: DOES IT PROVIDE ADDED VALUE AS A PERSON-CENTRED MEASURE FOR EVALUATION OF OUTCOME IN NEUROREHABILITATION FOLLOWING ACQUIRED BRAIN INJURY?

Lynne Turner-Stokes, DM, FRCP[1,2], Heather Williams, MSc[2] and Jane Johnson, MSc[2]

From the [1]Department of Palliative Care, Policy and Rehabilitation, King's College London, School of Medicine and [2]Regional Rehabilitation Unit, Northwick Park Hospital, North West London Hospitals Trust, Middlesex, UK

Objective: To compare goal attainment scaling (GAS) and standardized measures in evaluation of person-centred outcomes in neurorehabilitation.

Design: A prospective cohort analysis from a tertiary inpatient neuro-rehabilitation service for younger adults with complex neurological disability.

Subjects/patients: Consecutive patients ($n = 164$) admitted for rehabilitation following acquired brain injury (any cause) over 3 years. Mean age 44.8 (standard deviation 14.4) years. Diagnosis: 66% strokes, 18% trauma, 16% other. Male:female ratio 102:62.

Methods: GAS-rated achievement of 1–6 patient-selected goals was compared with the Functional Assessment Measure (UK FIM+FAM), and Barthel Index (BI), rated on admission and discharge. Personal goals were mapped retrospectively to the FIM+FAM and International Classification of Functioning, Disability and Health (ICF).

Results: Median (interquartile range; IQR) GAS T-scores were 50.0 (44.2–51.8) and moderately correlated with changes in FIM+FAM and BI (both rho 0.38 ($p < 0.001$)). Standardized response means were 2.2, 1.6 and 1.4 for GAS, FIM+FAM and BI, respectively. Of 667 personal goals set, 495 (74%) were fully achieved. Although 413 (62%) goals were reflected by changes in FIM+FAM, over one-third of goals were set in other areas.

Conclusion: GAS appeared to be more responsive, and captured gains beyond the FIM+FAM, thus providing added value as an adjunct to outcome measurement in patients with complex disability.

Key words: brain injuries, rehabilitation, outcome assessment, goals, ICF.

INTRODUCTION

Patients undergoing neurorehabilitation following brain injury present with a diverse pattern of impairments and disabilities, and have varying potential for improvement; hence the goals for rehabilitation may vary widely (1). One of the key challenges for research is to identify outcome measures that capture the individual's own aims and aspirations for their rehabilitation programme and, at the same time, are comparable across different centres and patient populations.

Global disability measures such as the Barthel Index (BI) (2) or the Functional Independence Measure (FIM™) (3) provide a valid and reliable assessment of basic physical function and are widely used as standardized outcome measures across the world. However, they have recognized floor and ceiling effects in brain injury, in which cognitive and psychosocial factors are often the main factors limiting outcome (4). A variety of tools has been developed to extend the range to include psychosocial function (e.g. the Functional Assessment Measure (FIM+FAM) (5) or to address participation and community integration (6). However, these are often more time-consuming to administer, and still have significant ceiling effects (7). Small but important changes, affecting only 1 or 2 items, may easily be missed against the "noise" of a large number of unchanged items.

Goal-setting has become a standard part of practice in rehabilitation (8). Goal attainment scaling (GAS) is a method for assimilation of achievement in a number of individually-set goals into a single aggregated "goal attainment score", providing a person-centred outcome, focused on that individual's priorities. Originally described by Kirusek & Sherman in the 1960s (9) it has been applied in various areas of complex intervention (8, 10, 11) including brain injury rehabilitation (12).

GAS offers a number of potential advantages as an outcome measure for patients with complex disabilities. As well as providing a quantitative assessment of goal attainment, it also affords qualitative information about the patient's priority goals for treatment and their respective importance. The process of goal-setting and rating supports dialogue between the patient and their treating team, and offers an additional opportunity to negotiate mutually agreed expectations for outcome. However, clinicians require sufficient knowledge, training and experience to support patients to set realistic goals (13).

Nevertheless, the use of GAS is still somewhat controversial. Although large-scale studies on inter-rater reliability are lacking, the small studies published to date are generally favourable (14, 15). Some authors have been impressed by its responsiveness and sensitivity to patients' values (12, 16) and its flexibility across the domains of impairment, disability and

participation (11). However, others have raised concerns about the mathematical concepts underlying the tool – particularly its non-linearity (13, 17, 18), lack of uni-dimensionality (19) and the fact that it is excessively time-consuming (17). Contrary to the originators' assertions that the GAS formula produces interval quality data, Steenbeek et al. (13, 18) point out that, as GAS is based on a 5-point scale, the data can at best be only of ordinal quality and they recommend the use of medians and non-parametric statistics. To overcome some of the scaling properties of GAS, Tennant (19) and Yip et al. (20) have proposed the development of standardized goals or "item banks".

The World Health Organization's International Classification of Functioning Disability and Health (ICF) (21) has been developed as an international common language to facilitate classification and comparison of the impact of disease across health and health-related domains. The domains are classified from body, individual and societal perspectives by means of 2 principal lists: a list of "body functions and structure", and a list of domains of "activity and participation". If the development of standardized goal sets is to be a useful way forward, the ICF potentially offers a framework for goal classification.

Like it or hate it, GAS has a rapidly expanding literature (8) and there is growing interest from clinicians who, frustrated by the limitations of standardized scales, are starting to take a broader view of outcome assessment. It is pertinent therefore to gain a better understanding of the relationship between GAS and our traditional standardized measures, and also how the various goals relate to ICF domains.

The aim of this study was to examine the relationship between GAS and 3 commonly-used global disability measures (the BI, the FIM and FIM+FAM) in the assessment of outcome from an in-patient rehabilitation programme. We explored both the quantitative aspects of measurement, and the qualitative nature of the goals set, to determine whether GAS has the potential to offer added value as a person-centred outcome measure for rehabilitation following brain injury. Our specific research questions were:

- What is the relationship between individualized GAS and the standard global outcome measures, and is GAS more responsive?
- What types of goals are commonly chosen by individuals undergoing rehabilitation. In which ICF domains do they lie, and to what extent do they overlap with standardized measures?

METHODS

Design and setting

The study was undertaken in a tertiary regional specialist neuro-rehabilitation service for younger adult patients (predominantly aged 16–65 years) with complex neurological disabilities in the UK. In a prospective cohort analysis, routinely collected standardized outcome data (the BI and UK FIM+FAM) were compared with GAS scores for consecutive in-patients admitted to the unit for rehabilitation following acquired brain injury (of any cause) during a 3-year period between 1 March 2005 and 28 February 2008.

Ethics permission was obtained from the local research ethics committee for the analysis of clinical data for the purpose of research.

All patients admitted to the unit were advised that their data would be used for research and were given the opportunity to opt out if they wished. No-one opted out during this period, and data collection was therefore complete.

Participants

A total of 164 patients with brain injury were admitted for neurorehabilitation during this period, 102 (62%) males and 62 (38%) females. Their mean age was 44.8 (standard deviation (SD) 14.4) years, and the mean length of stay was 88 days (SD 51). The cause of brain injury was stroke in 108 (66%), trauma in 30 (18%), and the remaining 26 (16%) had other causes, such as anoxia, inflammation and tumour.

Measures

The following outcome measures are routinely scored on admission and discharge by the treating team for all patients admitted for rehabilitation.

Goal attainment scaling (GAS). GAS is applied by a method previously described (22) based on that of Kirusek & Sherman (9). Goal setting is part of routine practice on the unit. Every patient has a defined set of goals or objectives for their rehabilitation programme and progress towards these is reviewed at fortnightly intervals by the treating team, together with the patient wherever possible. Regular in-house training for all staff on the unit includes the setting of goals that are SMART (specific, measurable, achievable, realistic, and timed); negotiation of realistic goals; and the evaluation of goal attainment using the 5-point scale. GAS has been applied as a part of our inter-disciplinary outcome evaluation set since 2004. In the first year of application we set pre-defined criteria for all 5 possible outcome levels, using a follow-up guide as recommended (9). However, this was found to be too time-consuming, and since 2005 we have used an abbreviated method in which criteria for the "zero" attainment score are clearly pre-defined at the outset, but any over- or under-achievement is rated by agreement between the patient and team at the end of the programme. Applied in this manner, GAS takes no more than 3–5 min over and above the process of goal-setting itself. The method is detailed in a separate publication (23), but is broadly as follows:

- Out of the defined set of individual objectives for their programme, 1–6 priority "personal goals" are identified and agreed between the patient (and/or family carer) and their treating team during the initial goal-planning meeting.
- Every effort is made to maximize patient involvement in identifying their personal goals, which are then made "SMART" through a process of negotiation with the patient and/or family.
- The chosen goals are then weighted by "importance" (determined by the patient/family) and the degree of "difficulty" (judged by the treating team). These are each graded on a scale of 0–3 ranging from 0 = "not at all" to 3 = "very" important or difficult (22), and goal weighting is the multiplicand of "importance × difficulty".
- In order to allow for deterioration, and in accordance with previous applications (22, 24, 25), baseline scores for each goal are allocated on admission as "–1" unless no clinically plausible worse outcome is possible, in which case a score of "–2" is given (e.g., for a patient whose goal is to be able to walk independently indoors with a walking aid: if at baseline they were starting to take some steps with the assistance of 2 people, this would score –1. However, if they were unable even to stand, let alone take steps, this would score –2.)
- At discharge from the programme, goal attainment is reviewed together with the patient and/or family, and rated on a 5-point scale, where:
  - "0" denotes the expected level of achievement;
  - "+1" and "+2" are respectively "a little" and "a lot" better than expected;
  - "–1" and "–2" are correspondingly "a little" and "a lot" less than the expected level.

Rating is based on their actual performance in relation to the expected level of achievement for each goal (i.e. what they do, not what they could do). In the case of disagreement, the lower level is scored.

- The attainment levels for the chosen personal goals are then combined in a single aggregated "T-score" by applying the formula recommended by Kiresuk & Sherman (9) which accounts for variable numbers of goals, inter-correlation of goal areas and variable weighting: Total score $= 50 + \{[10\Sigma(w_i x_i)]/[0.7\Sigma w_i^2 + 0.3(\Sigma w_i)^2]^{1/2}\}$, where $w_i =$ weight assigned to the ith goal and $x_i =$ the score of the ith goal.

If goals are set in an unbiased fashion so that results exceed and fall short of expectations in roughly equal proportions, over a sufficiently large number of patients, one would expect a normal distribution of GAS T-scores with a mean of 50 and SD of ±10.

UK Functional Assessment Measure (5). This is a 30-item global measure of disability, each item being scored on 7 levels, to give a range of 30–210. The UK FIM+FAM includes Version 4 of the FIM™ [1] (motor (13-items) and cognitive scales (5 items)) and adds a further 12 "FAM" items relating mainly to psychosocial and cognitive function. It has been shown to be reliable (5) and is increasingly widely used as an outcome measure for brain injury rehabilitation in the UK. Psychometric evaluation shows that the FIM+FAM can be analysed in 2 subscales: motor (16 items) and cognitive (14 items) (26). The UK FIM+FAM software also calculates a BI from the FIM data (27), scored on a range of 0–20 according to the manual of Collin et al. (28).

Our unit is the national UK FIM+FAM training centre and all staff are trained users. The UK FIM+FAM is scored for all patients on the unit, by the treating multidisciplinary team, within 10 days of admission. In addition to these baseline scores, the team also routinely notes the intended "target" score that they hope to reach by discharge from the programme. This target score is set in the recognition that full independence is unlikely to be achieved in all items, and indeed some may not be expected to change at all. The discharge score is then rated by the treating team (without reference to the baseline or target scores) during the last 7 days before discharge.

Data handling and analysis

Data were collated in a spreadsheet (Microsoft Excel) and transferred to SPSS version 11.5 for statistical handling. In this analysis, the majority of instruments yielded ordinal data. Although the GAS formula should theoretically deliver normally distributed data, 1-sample Kolmogorov-Smirnov tests revealed that the data in this set did not conform to normality. Non-parametric statistical tests were therefore used wherever possible. Wilcoxon signed-rank tests were used to evaluate changes in GAS and standardized measures from baseline.

Current techniques for evaluation of responsiveness rely on the estimation of means and SD. Responsiveness was compared using the effect size (mean change from baseline/SD baseline) and the standardized response mean (SRM) (mean change from baseline/SD change). Effect sizes have been reported to over-estimate response if the distribution of baseline scores is narrow, as they tend to be with GAS (29), and the SRM avoids this problem to a certain extent. As effect size and SRM rely on parametric assumptions, Wilcoxon z values were also given. Spearman's rank correlations were used to examine associations between the various measures.

The routine rating of target (or "goal") scores for the UK FIM+FAM provided an opportunity to apply the GAS formula to FIM+FAM scores and facilitates direct comparison of the tools as a measure of the achievement of both personal and FIM+FAM goals. Therefore, in addition to analysis of raw sum-scores for the FIM+FAM, we applied the same principles of goal attainment scaling to the analysis of FIM+FAM data. For each patient, each item of the FIM+FAM was retrospectively allocated a score on the 5-point scale of –2 to +2 (according to the rules below), rated both on admission (baseline) and at discharge (achieved).

- At baseline: if the FIM+FAM item level was not expected to change (i.e. the admission and target levels were the same) a baseline score of 0 was allocated. If the target level was higher than baseline, a score of –1 was allocated, unless the baseline FIM+FAM score was 1 (i.e. worst possible), in which case –2 was allocated.
- At discharge: if the target level was achieved, a GAS rating of 0 was allocated. If the target level was not reached, –1 was allocated, unless it had deteriorated from baseline, in which case –2 was allocated. Similarly if the discharge level exceeded the target level, +1 was allocated unless level 7 (complete independence) was achieved, in which case +2 was allocated.

The scores for the 30 items were then combined using the GAS formula to derive GAS-transformed FIM and FIM+FAM scores (baseline and T-scores for Motor and Cognitive sub-scores, in each case). (This transformation could not be applied to the BI, as its 3-level structure does not allow translation to 5 levels).

Because no specific information was available to weight the FIM+FAM goals, goal weights were all set at 1, for GAS transformation. Therefore, to facilitate comparison, both weighted and un-weighted GAS scores were computed for the personal goals.

We also undertook a qualitative analysis of the personal goals that were chosen, in order to identify the common goal areas. These were mapped retrospectively onto the FIM and/or FAM items and also onto domains of the WHO ICF (21) with reference to the linking rules published by Cieza et al. (30) and with the assistance of the ICF illustration library online (www.icfillustration.com). Second-level categories (3-digit codes) were used as they are considered to provide the best trade-off between breadth and depth of coding (31, 32). Two investigators (LTS and HW) coded the goal descriptions independently, then pooled and discussed their results to produce an agreed ICF code (or set of codes) for each personal goal. The principal codes were then assembled for each of the common goal areas.

RESULTS

Descriptive statistics for the admission (baseline) and discharge (achieved) scores are shown in Table I. Weighting made very little difference to the personal GAS scores, as the weighted and un-weighted (achieved) GAS T-scores were highly correlated (Spearman's rho 0.9, $p < 0.001$) with no systematic bias between them (Wilcoxon $z = -1.3$, $p = 0.19$).

All measures changed significantly from baseline to discharge (Wilcoxon $z = -7.9$ to -11.0; $p < 0.001$). The effect sizes were all "large", interpreted according to Cohen (33) (see foot note to Table I), with the exception of the FIM and FIM+FAM cognitive scales. Median GAS (achieved) T-scores were close to 50, both for personal goals and for GAS-transformed FIM ± FAM data, indicating that goals exceeded and fell short of expectation in roughly equal proportions.

The effect sizes were substantially higher for both GAS and GAS-transformed scores than for the raw standardized measure scores but, for the reasons previously noted (29), were probably over-estimated due to the small baseline SD. The SRMs were therefore thought to provide a better basis for comparison in this context. At 2.23, the SRM was substantially higher for personal GAS (un-weighted) than for the raw standardized measures, which ranged from 1.37 for the BI to 1.61 for the FIM+FAM. GAS transformation of the FIM+FAM scores improved the SRM only modestly to 1.73.

GAS (achieved) T-scores were closely correlated with the change in GAS from baseline (rho 0.77, $p < 0.0001$). For comparison with other measures, the aggregated GAS T-score was used

Table I. *Summary of change from admission to discharge within the various different scores*

| Measure | Admission | | Discharge | | Change score | Wilcoxon | Effect | |
	Median (IQR)	Range	Median (IQR)	Range	Median (IQR)	z*	size†	SRM
Standardized measures: raw scores								
Barthel Index	9 (4–13)	0–20	16 (11–20)	0–20	6 (2–9)	–10.4	1.0	1.37
FIM Motor	44 (27–65)	13–91	73 (53–86)	13–91	20 (9–32)	–10.6	0.93	1.44
FIM Cognitive	22 (16–28)	5–35	27 (22–31)	5–35	3 (0–6)	–9.3	0.47	0.94
FIM Total	67 (45–87)	18–122	101 (77–115)	18–126	24 (12–38)	–10.8	0.89	1.53
FIM+FAM Motor	53 (35–75)	16–111	87 (66–103)	16–112	26 (14–38)	–10.8	1.0	1.53
FIM+FAM Cognitive	61 (44–76)	14–98	71 (58–80)	13–91	6 (–1 – 15)	–7.9	0.36	0.75
FIM+FAM Total	118 (84–141)	30–205	163 (129–184)	30–209	36 (21–55)	–10.8	0.88	1.61
Goal attainment scores, including GAS transformed measures								
Personal GAS (weighted)	35.0 (30.6–35.7)	20–40	50.0 (44.2–51.8)	24–64	14.4 (11.0–19.4)	–11.0	3.54	2.29
Personal GAS (unweighted)	35.0 (31.9 35.5)	21–46	50.0 (46.4 – 50)	25–65	14.5 (10.9–18.1)	–11.0	3.16	2.23
FIM GAS	35.7 (31.9–40.5)	16–50	50.9 (47.1–54.8)	27.1–73.9	15.3 (7.6–21.0)	–10.8	2.20	1.63
FIM+FAM GAS	36.2 (33.0–40.4)	19–50	51.2 (47.8–54.7)	28.3–75.2	14.1 (8.8–19.9)	–10.8	2.37	1.73

*Significance all $p<0.001$.
†Effect sizes may be interpreted according to Cohen (0.2=small, 0.5=moderate, 0.8=large).
For comparability with the un-weighted transformed scores, both weighted and un-weighted personal GAS scores are recorded.
FIM: Functional Independence Measure; FIM+FAM: UK Functional Assessment Measure; GAS: Goal Attainment Scaling; SRM: standardized response mean (mean change from baseline/standard deviation change); effect size: mean change from baseline/standard deviation baseline; IQR: interquartile range.

in preference to the change in GAS score from baseline, as it is inherently a measure of change (see discussion). The relationship between change in the standardized measures between admission and discharge and GAS T-scores is shown in Table II.

As expected there was strong correlation between changes in the FIM+FAM and BI (rho 0.84, $p<0.001$). In contrast, only moderate correlations were seen between personal GAS T-scores and the standardized measures (0.36–0.43 for the raw change scores, and 0.41–0.49 for the GAS-transformed FIM ± FAM scores), suggesting that GAS may indeed encompass areas of change not included in the FIM+FAM or BI. We therefore undertook an analysis of the actual individual goals that were set, mapping these on to domains of the WHO's ICF (21) (see Table III).

Of a total of 667 individual goals set, 531 (79.6%) were rated at –1 at baseline, and 136 (20.4%) were rated at –2. In

Table II. *Spearman's correlations between personal goal attainment scaling (GAS) T-scores and change from baseline in other measures and their GAS-transformed counterparts*

| | GAS achieved T-scores | |
Measure	Personal GAS (unweighted) rho	Personal GAS (weighted) rho
Barthel Index	0.360*	0.377*
FIM Motor	0.363*	0.388*
FIM Cognitive	0.102	0.087
FIM Total	0.350*	0.364*
FIM+FAM motor	0.406*	0.431*
FIM+FAM Cognitive	0.123	0.087
FIM+FAM total	0.370*	0.379*
GAS transformed measures (achieved T-scores)		
FIM+FAM GAS	0.462*	0.494*
FIM GAS	0.411*	0.434*

*Correlations were significant at $p<0.001$.
FIM: Functional Independence Measure; FAM: Functional Assessment Measure.

all, 495 (74%) were fully achieved (scoring 0 or above); 85 (17.1%) of these scored "+1" and 12 (2.4%) scored "+2". Of the 172 goals that were not achieved, 149 (87%) goals scored "–1" and 23 (13%) scored "–2". A total of 638 goals (95.6%) were rated as either "moderately" (23.1%) or "very" important (72.6%); 356 (53.4%) were rated as "moderately" difficult and 242 (36.3%) as "very" difficult. Although there was a slight trend towards lower levels of achievement for more difficult goals, this did not reach significance.

Whilst the philosophy of the unit is to encourage goals to be set as far as possible in areas relating to activities and participation, a small proportion (approximately 10%) were necessarily process goals concerned with areas such as discharge planning and setting up care, for example in patients with low awareness states. A further 5% of goals were set in domains of "body functions" (i.e. impairment-related goals), but the remaining 85% addressed activities and participation. The most popular areas for goal setting were mobility, self-care and communication, which accounted for approximately 65% of the total goals set. Approximately 20% addressed extended or community-based activities such as domestic tasks or recreation/leisure activities. Goal attainment was generally quite consistent across all these goal areas, with 67–76% of goals being achieved or over-achieved.

Mapping of personal goals onto the FIM+FAM items could not be precise, as the SMART goal description often did not coincide with FIM+FAM level descriptors, and some goals were reflected in multiple items of the standardized scale (e.g., 6 FIM+FAM items address different aspects of toileting and incontinence). In some instances, a goal might lie within the area of a FIM+FAM item (e.g. walking) but may lie outside the range (e.g. walking over distances longer than 50 m). Goals were therefore considered individually to assess whether they were likely to have been reflected by changes in the FIM and/or FAM scores. In all, 315 (47%) goals addressed areas that could feasibly have been reflected by changes in the FIM, and 413 (62%) by the FIM+FAM.

Table III. *Analysis of goal areas including coverage by the FIM+FAM: UK Functional Assessment Measure (FIM±FAM) and International Classification of Functioning, Disability and Health (ICF) domains*

Goal category	n set	n or % achieved	FIM	FIM* +FAM	FIM+FAM Items†	Principal ICF domain codes – second level
Body functions	37	67%				
Mental function						
Low level: Awareness/interaction	10	7	±	±	18, 29	b110, b164 (d335)
Higher level: Memory/orientation	11	7	–	±	27, 28, 30	b114, b144, b140, b180
Emotion/behaviour	5	3	–	±	23, 24	b152, b147
Senses: Vision/hearing	7	5	–	–	–	b210, b230, b250
Pain	4	2	–	–	–	b280
Activities						
Motor function/coordination e.g.	41	76%				
Improving control of upper/lower limb	41	31	–	–	–	d440, d445, b710, b735
Mobility	168	76%				
Low level – standing/transfers/postural	47	30	±	±	10–12	d410, d415, d420, d465 (d450)
Medium level – walking/stairs	97	79	+	+	14, 15	d450, d455, d460, d465
High level – outdoors/running	24	19	–	–	–	d450, d455, d460, d465 (d920)
Selfcare	134	75%				
General independence	13	9	+	+	1–6	d599
Eating/drinking/nutrition	27	22	+	+	1,7	d550, d560, b510, b530
Toileting/continence	33	23	+	+	6, 8, 9, 11	d530, b525, b620 (d420, d450)
Washing/dressing/grooming	61	46	+	+	2–5	d510, d520, d540
Communication	93	69%				
Total communication (with aids, etc)	29	18	+	+	17, 22	d330, d335, d360, d399, d730, d760
Speech – talking/understanding	43	32	+	+	17, 18, 21	d330, d350, d360, d730, b310, b320, b126
Reading/writing	21	14	–	+	19,20	d140, d145, d166, d170, d325, d345
Participation						
Extended activities of daily living*	47	74%				
Cooking/meal preparation	37	28	–	–	(EADL)	d630, d620
Household/gardening	9	6	–	–	(EADL)	d620, d640, d650, d920
Finance	1	1	–	–	(EADL)	d860
Community access	21	71%				
General	8	6	–	+	13, 16	d460, d470, d499 (d620)
Using public transport	10	6	–	+	16	d460, d470, (b164)
Driving	3	3	–	–	–	d475
Recreation/leisure	31	68%				
Computers/using EAT	11	9	–	±	25	d360, d920, d825, b140, b144
Active	5	2	–	±	25	d920, d455
Arts/crafts	3	2	–	±	25	d920, d650
Social and general	12	8	–	±	22, 25	d910, d920, d760
Work/education/responsibility	28	68%				
Work	16	10	–	–	–	d825, d845, d840
Education	4	3	–	–	–	d820, d830
Parenting	8	6	–	–	–	d660, d760
Other goals						
Process e.g.	67	86%				
Discharge/care	61	54	–	–	–	e115,e120, e125, e310, e340, e575
Spending time at home	6	4	–	–	–	
Total	667	495				

*"+": most goals reflected by changes in FIM/FAM, "±" some goals reflected (less then half), "–": not reflected.

†Numbers refer to the item no in the UK FIM+FAM. The EADL items form a separate module not included in the original 30 FIM+FAM items. These are available from the authors on request.

ICF domain codes in brackets indicate codes from different constructs that were commonly included in the goal statement (e.g. community access goals frequently specified the purpose of shopping (d620)).

EAT: Electronic Assistive Technology; FIM: Functional Independence Measure; FIM+FAM: UK Functional Assessment Measure; EADL: extended activities of daily living.

DISCUSSION

In this analysis we explored the relationship between personal GAS and standard global outcome measures (the BI, the FIM and the UK FIM+FAM) in the assessment of outcome from in-patient rehabilitation in young adults with complex neurological disabilities following acquired brain injury. We found GAS to be more responsive than the standard instruments. Although there was considerable overlap between the personal goals chosen and those represented in the standard instruments,

GAS also recorded gains in other important areas not addressed by those tools.

Our results tally with those of previous authors who have also found GAS to be more responsive than the BI (25, 29) and the FIM (10, 16) when compared in terms of effect size or SRM. However, the comparison of different outcome measures presents a number of methodological problems that merit further discussion.

Firstly, subjecting ordinal data to mathematical manipulation, such as the computation of difference scores (change from baseline), can be unreliable, especially where baseline and outcome scores are closely correlated. Previous applications of GAS have variously used the GAS achieved T-score (16, 29) or change in GAS score from baseline (24, 25) for comparison as the indicator of outcome. As Kirusek points out (34), because change over time is built into the way in which GAS scores are derived, the T-score is in itself a measure of change and avoids some of the problems of computing change scores. Therefore, GAS-achieved T-scores were used for comparison in this analysis.

Secondly, the calculation of effect size and SRM is a parametric technique that necessarily involves the computation of mean change and SD, but the very different nature of the data-sets may limit comparison. Because the large majority of baseline GAS scores are –1, the baseline variation in GAS score is small and this may lead to over-estimation of the effect size (29). For this reason we used the SRM. Even so, the possible range of GAS (mainly 40–60) is very different from the FIM+FAM (30–210) so that differences in SRM could simply reflect the range of data, rather than the responsiveness of the instrument *per se*. In this study we took the novel approach of applying the GAS formula to transform the FIM+FAM data to a similar range, so that personal and standardized scores could be compared "on a level playing field". The difference between them is then more likely to reflect the achievement of gain itself, rather than the way it is calculated. Applying this method, the SRM was approximately 50% greater for the individual GAS than the FIM+FAM. However, we would once again stress that any parametric analysis of ordinal data must be interpreted with caution.

We recognize that there is still considerable debate about the validity of GAS. A full rehearsal of the various arguments for and against it is beyond the scope of this article, but the findings reported here may go some way towards clarification. Mackay & Somerville (17) have raised concern about the validity of goal weighting. Whilst weighting may provide useful qualitative information for clinical interpretation of variance in goal achievement, it can sometimes have a perverse effect (23). In this analysis, the inclusion of goal weighting made little difference in numerical terms. For research purposes, it would therefore appear reasonable to exclude it from the formula and so eliminate a possible source of bias. Another concern has been the time take to rate GAS, and a balance must be found between timeliness and rigour. In this study we have used an abbreviated method which was practical to apply in routine practice, but could potentially have introduced bias. Our median achieved GAS T-score for personal goals was 50, and showed a slight skew towards under-achievement of goals in the inter-quartile range, which suggests that at least the team was not overly generous in its allocation of retrospective scores.

Whilst previous studies have compared GAS with the BI and FIM, this is the first published comparison with the FIM+FAM. Our analysis of goals confirmed that the FIM+FAM items covered a larger proportion of personal goals (approximately two-thirds) than the FIM items (approximately half). This suggests that, even though the additional FAM items confer little benefit in terms of the measurement properties of the scale (35, 36), they do appear to capture additional qualitative outcome information in areas of importance to patients. Even so, the personal goals covered a much wider area of personal experience than any of the standardized scales.

To overcome the problems of variability, whilst still maintaining the recognized benefits of GAS, some authors have proposed the development of standardized goals or "item banks" (19, 20) so that these may be calibrated onto a uni-dimensional scale, thus satisfying the metric requirements of a linear or interval measure. It could in fact be argued that the FIM+FAM items already represent item banks, as they were originally selected because they represent common goal areas for rehabilitation (37). These instruments have been carefully standardized and validated over more than a decade, and therefore might represent a reasonable starting point for item banks in some of the common goal areas. However, the disadvantage of that approach is that goal setting is constrained to fit the descriptors, and the essential individualized nature of GAS is lost.

As this study demonstrates, the range of personal goals extends well beyond the confines of these measures. The ICF is already being used to compare the content of different tools (38), interventions (39) and rehabilitation goals (32). Mapping of goals onto the ICF may potentially provide a common framework for goal classification in the future that would support development of standardized goal set for GAS. However, this mapping process presents its own challenges. In this study, ICF codes were allocated retrospectively and this was not always straightforward. Even though the SMART goal wording provided considerable definition, the rationale behind the goal, was often complex, for example, a goal to achieve independent use of computers could be for any of all of the following: (*i*) therapeutic purposes (e.g. cognitive training of attention (b140) and memory (b144), hand-eye co-ordination (b760) or fine hand use (d440)); (*ii*) for recreational purposes (d920); (*iii*) for vocational training (d825); or (*iv*) for communication with the outside world (d360). However, codes could only be allocated if these were specified in the goal statement. Ideally, ICF codes should be allocated prospectively, as the rationale for each goal is more likely to be apparent at the time of goal setting. Unfortunately, given the complexity of the ICF coding system and the considerable pressures on clinicians' time, prospective code allocation has not yet proven practical in the context of routine clinical practice in our setting.

Strengths and weaknesses

The strengths of this study are that it represents a sizeable cohort of data collected in the course of real-life clinical practice. This is the first published study to compare GAS with the

FIM+FAM, and to map goals across both measures and to the ICF. The inclusion of patients with brain injuries of any cause extends the perspective and experience of the study group more widely than if selection was limited to one diagnostic group. The systematic process of goal-setting and negotiation on our unit ensured that the priority goals chosen were important to the patients, and the expectation for outcome was mutually agreed between staff and patients and/or their family.

We also recognize a number of weaknesses, over and above the challenges for analysis that have been discussed above. This was a single-centre study, which may limit the generalizability of our findings. Moreover, it was conducted in a tertiary centre selecting young adults with more complex disabilities than typically present in a district neuro-rehabilitation service. We hope, however, that it will encourage other centres to undertake similar analyses in different settings. The mapping of goals onto the FIM+FAM and ICF was undertaken retrospectively. Although ICF linking rules were applied as carefully as possible, in the absence of a clearly stated rationale for all the goals, coding accuracy cannot be fully guaranteed. Similarly, it was not possible to be 100% certain whether goal attainment was actually reflected in FIM+FAM levels, especially when the goal crossed multiple items. As noted above, prospective coding would overcome these problems, but in its current form the ICF is unwieldy for most busy clinicians to use. However, as common goals linked to ICF core-sets start to emerge in the future, it is possible that the production of tools such as localized decision trees embedded in electronic records may assist with coding, making it a more feasible option for use in the context of clinical practice (40).

In conclusion, this study has provided further evidence that GAS is a responsive measure that identifies the achievement of person-centred goals for rehabilitation, which may not be detected by commonly-used standardized measures. In addition to quantitative assessment, it provides useful qualitative information about the patient's priority goals. Whilst it cannot replace standardized measures, GAS provides useful added value as a measure of outcome from rehabilitation. Applied by our simplified method within the context of a goal-orientated rehabilitation programme, it provides a valuable adjunct to routine measurement for very little further investment of effort. Mapping of goals onto the ICF may provide a useful framework for describing common goal areas, and further work with prospective mapping is now required.

REFERENCES

1. Wade DT. Evidence relating to goal-planning in rehabilitation. Clin Rehabil 1998; 12: 273–275.
2. Mahoney FI, Barthel DW. Functional evaluation: the Barthel Index. Md State Med J 1965; 14: 61–65.
3. Heinemann AW, Linacre JM, Wright BD, Hamilton BB, Granger C. Relationships between impairment and physical disability as measured by the Functional Independence Measure. Arch Phys Med Rehabil 1993; 74: 566–573.
4. Turner-Stokes L. Outcome measurement in brain injury rehabilitation – towards a common language. Clin Rehabil 1999; 13: 273–275.
5. Turner-Stokes L, Nyein K, Turner-Stokes T, Gatehouse C. The UK FIM+FAM: development and evaluation. Functional Assessment Measure. Clin Rehabil 1999; 13: 277–287.
6. Willer B, Button J, Rempel R. Residential and home-based postacute rehabilitation of individuals with traumatic brain injury: a case control study. Arch Phys Med Rehabil 1999; 80: 399–406.
7. Hall KM, Mann N, High WMJ, Wright J, Kreutzer JS, Wood D. Functional measures after traumatic brain injury: Ceiling effects of FIM, FIM+FAM, DRS, and CIQ. J Head Trauma Rehabil 1996; 11: 27–39.
8. Hurn J, Kneebone I, Cropley M. Goal setting as an outcome measure: a systematic review. Clin Rehabil 2006; 20: 756–772.
9. Kiresuk T, Sherman R. Goal attainment scaling: a general method of evaluating comprehensive mental health programmes. Com Mental Health J 1968; 4: 443–453.
10. Rockwood K, Stolee P, Fox RA. Use of goal attainment scaling in measuring clinically important change in the frail elderly. J Clin Epidemiol 1993; 46: 1113–1118.
11. Schlosser RW. Goal attainment scaling as a clinical measurement technique in communication disorders: a critical review. J Commun Disord 2004; 37: 217–239.
12. Malec JF, Smigielski JS, DePompolo RW. Goal attainment scaling and outcome measurement in postacute brain injury rehabilitation. Arch Phys Med Rehabil 1991; 72: 138–143.
13. Steenbeek D, Ketelaar M, Galama K, Gorter JW. Goal attainment scaling in paediatric rehabilitation: a critical review of the literature. Dev Med Child Neurol 2007; 49: 550–556.
14. Shefler G, Canetti L, Wiseman H. Psychometric properties of goal-attainment scaling in the assessment of Mann's time-limited psychotherapy. J Clin Psychol 2001; 57: 971–979.
15. Palisano RJ. Validity of goal attainment scaling in infants with motor delays. Phys Ther 1993; 73: 651–658; discussion 658–660.
16. Rockwood K, Howlett S, Stadnyk K, Carver D, Powell C, Stolee P. Responsiveness of goal attainment scaling in a randomized controlled trial of comprehensive geriatric assessment. J Clin Epidemiol 2003; 56: 736–743.
17. Mackay G, Somerville W. Reflections on goal attainment scaling. Educational Res 1996; 2: 161–171.
18. Steenbeek D, Meester-Delver A, Becher JG, Lankhorst GJ. The effect of botulinum toxin type A treatment of the lower extremity on the level of functional abilities in children with cerebral palsy: evaluation with goal attainment scaling. Clin Rehabil 2005; 19: 274–282.
19. Tennant A. Goal attainment scaling: current methodological challenges. Disabil Rehabil 2007; 29: 1583–1588.
20. Yip AM, Gorman MC, Stadnyk K, Mills WG, MacPherson KM, Rockwood K. A standardized menu for Goal Attainment Scaling in the care of frail elders. Gerontologist 1998; 38: 735–742.
21. World Health Organization. International classification of func-

tioning, disability and health. Geneva: World Health Organization; 2002. Report no.: ISBN 91 4 154542 9.

22. Ashford S, Turner-Stokes L. Goal attainment for spasticity management using botulinum toxin. Physiother Res Int 2006; 11: 24–34.

23. Turner-Stokes L. Goal attainment scaling in rehabilitation; a practical guide. Clin Rehabil 2009; 23: 362–370.

24. Rockwood K, Joyce B, Stolee P. Use of goal attainment scaling in measuring clinically important change in cognitive rehabilitation patients. J Clin Epidemiol 1997; 50: 581–588.

25. Rushton PW, Miller WC. Goal attainment scaling in the rehabilitation of patients with lower-extremity amputations: a pilot study. Arch Phys Med Rehabil 2002; 83: 771–775.

26. Hawley CA, Taylor R, Hellawell DJ, Pentland B. Use of the functional assessment measure (FIM+FAM) in head injury rehabilitation: a psychometric analysis. J Neurol Neurosurg Psychiatr 1999; 67: 749–754.

27. Nyein K, McMichael L, Turner-Stokes L. Can a FIM score be converted to a Barthel Index? Clin Rehabil 1999; 13: 56–63.

28. Collin C, Wade DT, Davis S, Horne V. The Barthel ADL Index: a reliability study. Int Disabil Stud 1988; 10: 61–63.

29. Gordon JE, Powell C, Rockwood K. Goal attainment scaling as a measure of clinically important change in nursing-home patients. Age Ageing 1999; 28: 275–281.

30. Cieza A, Brockow T, Ewert T, Amman E, Kollerits B, Chatterji S, et al. Linking health-status measurements to the international classification of functioning, disability and health. J Rehabil Med 2002; 34: 205–210.

31. Stucki G, Grimby G. Forward: Applying the ICF in medicine. J Rehabil Med 2004; 36: 5–6.

32. Soberg HL, Finset A, Roise O, Bautz-Holter E. Identification and comparison of rehabilitation goals after multiple injuries: an ICF analysis of the patients', physiotherapists' and other allied professionals' reported goals. J Rehabil Med 2008; 40: 340–346.

33. Cohen J, editor. Statistical power analysis for the behavioural sciences. 2nd edn. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.

34. Kiresuk T, Smith A, Cardillo J, editors. Goal attainment scaling: application, theory and measurement. New York: Lawrence Erlbaum Associates; 1994.

35. Hobart JC, Lamping DL, Freeman JA, Langdon DW, McLellan DL, Greenwood RJ, et al. Evidence-based measurement: which disability scale for neurologic rehabilitation? Neurol 2001; 57: 639–644.

36. Linn RT, Blair RS, Granger CV, Harper DW, O'Hara PA, Maciura E. Does the functional assessment measure (FAM) extend the functional independence measure (FIM) instrument? A Rasch analysis of stroke inpatients. J Outcome Meas 1999; 3: 339–359.

37. Granger CV. A conceptual model for functional assessment. In: Granger CV & Gresham GE, editors. Functional assessment in rehabilitation medicine. Baltimore: Williams and Wilkins; 1984.

38. Schepers VP, Ketelaar M, van de Port IG, Visser-Meily JM, Lindeman E. Comparing contents of functional outcome measures in stroke rehabilitation using the International Classification of Functioning, Disability and Health. Disabil Rehabil 2007; 29: 221–230.

39. Boldt C, Brach M, Grill E, Berthou A, Meister K, Scheuringer M, et al. The ICF categories identified in nursing interventions administered to neurological patients with post-acute rehabilitation needs. Disabil Rehabil 2005; 27: 431–436.

40. Rauch A, Cieza A, Stucki G. How to apply the International Classification of Functioning, Disability and Health (ICF) for rehabilitation management in clinical practice. Eur J Phys Rehabil Med 2008; 44: 329–342.