Review of Goal Attainment Scaling as a Useful Outcome Measure in Psychogeriatric Patients with Cognitive Disorders

Sharon F.M. Bouwens a  Caroline M. van Heugten a, c  Frans R.J. Verhey a, b

a School for Mental Health and Neuroscience, Department of Psychiatry and Neuropsychology, Maastricht University, Maastricht, b University Hospital of Maastricht and Alzheimer Centre Limburg, Maastricht, c Centre of Excellence in Rehabilitation Medicine and the Rudolf Magnus Institute, Utrecht, The Netherlands

Key Words
Individualized approach  Clinimetrics  Feasibility  Ecological validity  Everyday functioning  Daily life activities

Abstract
Background: Since evidence-based interventions are the standard, there is an urgent need for more information concerning individual ways of measuring clinically relevant outcomes of interventions in cognitive disorders such as dementia. Goal Attainment Scaling (GAS) seems to offer a meaningful outcome measure. Aim: To examine the applicability of GAS in psychogeriatric patients with cognitive disorders. Methods: A systematic review was performed on the available literature on the clinimetric aspects and the feasibility of GAS when used as an outcome measure for psychogeriatric patients with cognitive disorders. Eight databases were screened. Two authors independently reviewed all the data. Ten studies were included. Results: Mixed results were found for responsiveness, content validity, inter-rater reliability and construct/convergent validity. The involvement of patient and/or caregiver in the goal-setting procedure is possible and multiple domains can be implemented. The possibility to set at least 3 realistic goals per patient in less than 30 min is unclear and the need for involvement of a blinded assessor is not well established. Conclusion: GAS proved to be useful on important aspects of an outcome measure for psychogeriatric patients with cognitive disorders. Since other relevant aspects showed mixed results and the number of studies investigating the use of GAS in psychogeriatric patients with cognitive disorders is small, the evidence is not strong enough yet to state that GAS is an applicable outcome measure in this population.

Introduction
Neurodegenerative disorders such as dementia cause serious cognitive impairments which lead to interference with daily life functioning [1]. There is an ongoing debate about the best way to measure clinically relevant outcomes of the treatment of dementia [2–4] such as with cholinesterase inhibitors [5, 6]. Outcome measures should not only be responsive, reliable and valid, but should also be tailored to the personal goals and needs of patients and their caregivers in relation to their daily life. However, current standardized outcome measures in the field of cognitive disorders do not always take account of these personal aspects. For instance, the widely used Alzheimer’s Disease Assessment Scale-Cognitive Subscale [7], which is a cognitive scale used in Alzheimer’s disease, measures effects on cognition, but does not take into account the individual situation of the patient [8].
Goal Attainment Scaling (GAS) is a global outcome measure that does take account of these individual needs. GAS is a measurement method first introduced by Kiresuk and Sherman [9] for program evaluation in mental health centres [10]. GAS allows for both individualization of patient goals according to the needs of each patient, and standardization of measurement by using a summary formula that calculates the extent to which a patient’s goals are met [11]. GAS reflects actual improvement in a patient’s functional ability [12] and can be adapted to any level or domain of the International Classification of Functioning, Disability, and Health [13], thereby covering aspects of life as a whole. Due to the increasing emphasis on a more client-centred approach, GAS received renewed attention, and has recently been used in patient groups suffering from various disorders, including dementia [14–16].

Despite the increasing use of GAS, its applicability is unclear in the case of psychogeriatric patients with cognitive disorders. Dementia is a multi-domain disorder, characterized by interference with daily life activities, which differs for each individual, as well as by a progressive nature. Hence, to be a useful measure, GAS must be responsive to both these aspects. In instances where the clinical relevance of interventions such as treatment with cholinesterase inhibitors is unclear [17], GAS could be used to reveal clinical information relevant to the patient, caregiver and clinician. However, as cognitively impaired patients often lack insight into their capabilities [10], it is to be expected that constructing GAS follow-up guides with these patients could be problematic and less feasible.

The aim here is to examine the applicability of GAS in people with cognitive disorders by reviewing the available literature on the clinimetric aspects and the feasibility of GAS when used as an outcome measure for psychogeriatric patients.

**Description of GAS**

The GAS method consists of a 6-step process. To select the goals that are relevant to the individual patient in step 1 the patient and/or caregiver is either interviewed by a clinician to identify problem areas and to determine goals for those areas in which intervention is planned, or team members set goals themselves after interviewing the patient and/or caregiver about problem areas. GAS reflects actual improvement in a patient’s functional ability [12] and can be adapted to any level or domain of the International Classification of Functioning, Disability, and Health [13], thereby covering aspects of life as a whole. Due to the increasing emphasis on a more client-centred approach, GAS received renewed attention, and has recently been used in patient groups suffering from various disorders, including dementia [14–16].

<table>
<thead>
<tr>
<th>Goal areas</th>
<th>goal 1 dealing with impaired memory</th>
<th>goal 2 social contacts</th>
<th>goal 3 coping with aggression</th>
</tr>
</thead>
<tbody>
<tr>
<td>Much less than expected (–2)</td>
<td>poor short-term memory, no use of memory aids</td>
<td>no interest in social contacts; sees nobody during the day except caregiver¹</td>
<td>becomes aggressive in all situations of misunderstanding</td>
</tr>
<tr>
<td>Somewhat less than expected (–1)</td>
<td>uses memory aids when another person cues to do so¹</td>
<td>ambivalent to visit others; caregiver needs to be present</td>
<td>becomes aggressive in almost all situations of misunderstanding¹</td>
</tr>
<tr>
<td>Program goal (0)</td>
<td>able to use memory aids with external cue-like watch²</td>
<td>enthusiastic to visit others; caregiver needs to be present²</td>
<td>learned how to stay calm, but does not succeed in all situations</td>
</tr>
<tr>
<td>Somewhat better than expected (+1)</td>
<td>able to use memory aids spontaneously</td>
<td>visits others on her own when asked to do so</td>
<td>seldom becomes aggressive²</td>
</tr>
<tr>
<td>Much better than expected (+2)</td>
<td>no need for memory aids</td>
<td>initiates visiting others on her own</td>
<td>never becomes aggressive anymore</td>
</tr>
</tbody>
</table>

¹ Baseline level. ² Follow-up level.

Applicability of Goal Attainment Scaling
Levels do not overlap. The expected outcome is scored with ‘0’. Step 5 consists of completion of the other scale levels, using possible outcomes including much less than expected (–2), somewhat less than expected (–1), somewhat better than expected (+1) and much better than expected (+2). In step 6, GAS baseline levels are noted on the follow-up guide. If a clinically relevant deterioration is plausible, the description of the current status can be scored as ‘–1’, with the worsened state scored as ‘–2’. If the problem is at its worst, the current level of functioning is scored as ‘–2’. Patients receive an intervention and at the defined follow-up time (e.g. end of intervention), the patient is scored at the attained level. The overall GAS score at baseline and follow-up is calculated with a standard formula (appendix 1). When goals are weighted, this weight is inserted in this formula. The hypothetical mean GAS score at follow-up is 50 (SD = 10); consistently high weight is inserted in this formula. The hypothetical mean GAS score at follow-up of 50 indicates that all predefined goals met the expected outcome at follow-up. Table 1 provides an example of a goal attainment scaling follow-up guide.

Methods

Literature Search and Study Selection

Figure 1 shows the process of literature search and study selection. Studies on ‘GAS’ and ‘cognitive disorders’ and ‘clinimetric aspects’ were identified by searching the following computerized databases until September 2007: CINAHL, PsychINFO, Pubmed, Medline, Cochrane Library, EMBASE, Rehabdata, and Amed. Both controlled vocabulary words and free text words in the title, abstract or key words were combined in our searches. Details of the search are shown in appendix 2. The initial search resulted in 823 studies. Two reviewers (S.F.M.B. and C.M.v.H.) independently assessed all studies for inclusion based on the title. In case of doubt, the first author obtained the abstracts or the full texts, if available, and both reviewers decided on inclusion. References of these studies were hand-searched to find additional potentially relevant studies. International requests were submitted when a study was not available in the Netherlands. After excluding 637 references due to non-relevant subject matter and 64 duplicates, 122 studies remained. These studies were further selected for this review based on the criteria that (1) the study included – at least partially – patients with cognitive disorders (dementia or mild cognitive disorders); (2) the study described clinimetric aspects of GAS; (3) the study used GAS as developed by Kiresuk and Sherman [9], or deviations from this original approach if the aforementioned authors were credited; (4) the study was a full report of a group study describing original data (i.e. not a case study, review, editorial); (5) the study was published in English. Careful reading of the 122 remaining studies resulted in the rejection of 112 studies because of different patient population [21], lack of description of clinimetric aspects [55], no full report (5), review (28) or non-English publication (3). Finally, a total of 10 studies were included in this review.

All included studies were judged on study characteristics (design, intervention, setting and sample); participant characteristics (mean age, severity level of disease) [20]; clinimetric aspects (responsiveness, reliability, validity), and feasibility (time required for setting the goals, number of goals, the extent to which goals were realistic, involvement of patient and/or caregiver in setting the goals, involvement of a blinded assessor in assessing the follow-up GAS scores and domains that were included in the goals).

Evaluation of the Clinimetric Aspects

Various definitions of responsiveness exist and responsiveness can be used for different purposes [21]. For our review, we defined responsiveness as the extent to which a measure is capable of detecting change when actual change occurred [22]. Furthermore, a variety of methods exist for calculating responsiveness [21]. In our study, responsiveness was interpreted as positive in the following instances: if the paired sample t test for analyzing the difference in assessment and the follow-up GAS scores were significant (p < 0.05); or if Norman’s [23] method involving repeated measures analysis of variance was >0.5; or if Cohen’s d estimate of effect size proposed by Kazis et al. [24] was >0.5; or if the standardized response mean [25] was >0.5; or if the relative efficiency [26] was >1. Responsiveness was interpreted as negative if the changes were non-significant [27]. Effect size is frequently used to measure the magnitude of an effect of an intervention, but it can be used for any other numerical comparison or contrast [28]. We used the effect size as an operationalization of the sensitivity...
to change of GAS as outcome measures compared to secondary measures [29].

Reliability was good when Pearson’s correlation coefficients (r) or intraclass correlation coefficients were >0.8, moderate between 0.7 and 0.8, and low when <0.7 [30].

Content validity was considered good when an expert panel reviewed the selected domains for GAS as good, or when domains correlated with position statements.

Construct/convergent validity was considered to be good when the correlation coefficient (Pearson or Spearman) between GAS and secondary outcome measures was >0.7, moderate if the coefficient was between 0.5 and 0.7, and low if the coefficient was <0.5 [31].

**Evaluation of Feasibility**

The time required for setting the goals was considered good when it took a maximum of 30 min to construct a GAS follow-up guide of at least 3 goals.

The number of goals was considered sufficient when at least 3 goals were set.

Goals were considered realistic when the hypothetical mean GAS score at follow-up was 50 (SD = 10, range 40–60).

Feasibility was considered good when the patient or the caregiver was involved in setting the goals.

Feasibility was considered good when a blinded assessor was involved in assessing the follow-up GAS scores.

Domains that were included in the goals had to cover at least 2 of the following domains: cognition, instrumental activities of daily life/self-care, mood, behaviour or mobility.

**Results**

Ten studies met the inclusion criteria, of which 9 studies were from Rockwood and co-workers [19, 32–40]. Two independent raters checked the titles and abstracts. The third rater was not consulted as there were no discrepancies between rater 1 and 2. Table 2 presents the main characteristics and the clinimetric aspects of GAS of the selected studies.

**Study Characteristics**

Four of the 10 studies were samples with dementia patients [32–34, 37]. In the study of Gordon et al. [19], 77% of the sample with nursing-home patients had dementia. The characteristics of the remaining 33% in this study were not described in further detail. In Rockwood et al. [36], 47% had dementia and the other patients in this study all showed at least a moderate degree of functional impairment. Stolee et al. [38] provided data of a sample in which 13% had a primary diagnosis of dementia, the remaining patients had stroke, depression or physical problems like hip fracture or congestive heart failure. The remaining 3 studies [35, 39, 40] did not provide details about numbers and specific diagnosis of the patients with cognitive disorders. The mean age of the participants was 79 years (n = 846, range = 51–96, 551 females). Most studies had a prospective, descriptive study design (n = 6). Other designs used were randomized controlled trials (n = 3). There was one retrospective, descriptive study. All studies used GAS as a primary outcome measure for an intervention. Seven studies described some form of specialized care and 3 studies were medical trials. Sample sizes ranged from 10 to 170 participants.

Table 3 shows a comparison of the clinimetric aspects of GAS.

**Clinimetric Aspects**

**Responsiveness**

Responsiveness was examined in 8 studies. Table 2 shows the details for every study. Six of these 8 studies showed a medium to large effect size (0.6–4.9), standardized response mean (0.2–1.7) or relative efficiency (0.5–100) [19, 32, 35, 36, 39, 40]; or a significant Norman’s analysis of variance (0.3 to 0.8) [32, 35, 39]; or a paired sample t test for analyzing the difference in assessment and the follow-up GAS scores (t = 2.9, p = 0.02) [32]. One study had mixed results, i.e. a small relative efficiency but a large effect size [37], and 1 study showed small standardized response means [33].

**Reliability**

Only 3 studies examined inter-rater reliability. The intraclass correlation coefficient was used by Rockwood et al. [36], Stolee et al. [38], and Stolee et al. [39] and was 0.9. Stolee et al. [38] also used Pearson’s correlation coefficient (0.9).

**Content Validity**

Five studies examined the content validity. Stolee et al. [38] and Yip et al. [40] assessed content validity by comparing identified goal areas with a list of assessment areas derived from position statements describing the essential components of geriatric assessment [41–43]. In Stolee et al. [38], 2 geriatricians independently identified 82% (71 out of 87) of the goal areas; the remainder were set by consensus. Both geriatricians identified the same chief goals in 93% (14 out of 15) of the patients and were able to agree on the relative importance of goals. In Yip et al. [40] all recommended domains were addressed. Stolee et al. [39] and Rockwood et al. [37] examined content validity through content analysis of identified goal areas. In Stolee et al. [39], clinicians from the geriatric rehabilitation unit reviewed the categorizations and sug-
Table 2. Study and participant characteristics and clinimetric aspects of GAS used in psychogeriatric patients

<table>
<thead>
<tr>
<th>Study</th>
<th>Design</th>
<th>Intervention</th>
<th>Sample and setting</th>
<th>Age, years</th>
<th>Responsiveness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gordon et al. [19]</td>
<td>prospective descriptive study</td>
<td>specialized geriatric medicine consultation</td>
<td>53 nursing-home patients of which 41 had dementia</td>
<td>81 ± 8</td>
<td>ES = 1.3 RE = 53.7</td>
</tr>
<tr>
<td>Hartman et al. [32]</td>
<td>prospective descriptive study</td>
<td>program of occupational therapy and therapeutic recreation</td>
<td>10 male persons with AD or another form of dementia</td>
<td>76</td>
<td>paired t test = 2.9, d.f. = 9, p = 0.02 ES by Kazis = 2.3 ANOVA by Norman = 0.4</td>
</tr>
<tr>
<td>Rockwood et al. [36]</td>
<td>prospective descriptive study</td>
<td>specialized geriatric inpatient wards</td>
<td>45 consecutive elderly patients of which 21 had dementia</td>
<td>81 ± 8</td>
<td>ES = 4.9 RE = 4.5</td>
</tr>
<tr>
<td>Rockwood et al. [34]</td>
<td>prospective study</td>
<td>Linopirdine, Aviva, DuPont Pharma</td>
<td>15 patients with AD of at least mild severity</td>
<td>73</td>
<td>ES = 0.6 RE = 0.5</td>
</tr>
<tr>
<td>Rockwood et al. [37]</td>
<td>RCT</td>
<td>a 12-month phase IV trial with donepezil</td>
<td>100 community-dwelling patients with mild to moderate AD</td>
<td>76 ± 8 range: 51–96</td>
<td>NA</td>
</tr>
<tr>
<td>Rockwood et al. [35]</td>
<td>RCT</td>
<td>interdisciplinary mobile geriatric assessment team</td>
<td>165 rural, community-dwelling, geriatric frail elderly patients</td>
<td>82 ± 7</td>
<td>ES = 2.9–4.4, Cohen’s d SRM = 0.8–1.1 RE = 29.9 to &gt;100 Norman’s responsiveness statistics = 0.3–0.6</td>
</tr>
<tr>
<td>Rockwood et al. [33]</td>
<td>RCT</td>
<td>Galantamine</td>
<td>130 patients with mild to moderate AD</td>
<td>78 ± 8 range: 51–94</td>
<td>SRM patient-caregiver = 0.4/0.2 for SRM clinician = 0.4/0.4</td>
</tr>
<tr>
<td>Stolee et al. [38]</td>
<td>prospective descriptive study</td>
<td>geriatric assessment unit</td>
<td>15 patients who were consecutively admitted to a geriatric service, of which 2 were demented</td>
<td>79 ± 10 range: 65–94</td>
<td>NA</td>
</tr>
<tr>
<td>Stolee et al. [39]</td>
<td>prospective descriptive study</td>
<td>geriatric rehabilitation unit</td>
<td>170 consecutively admitted geriatric elderly patients</td>
<td>81 ± 7 range: 61–96</td>
<td>ES = 3.5 SRM = 1.7 RE = 3.1 ANOVA (Norman) = 0.8</td>
</tr>
<tr>
<td>Yip et al. [40]</td>
<td>retrospective descriptive study</td>
<td>geriatric assessment and rehabilitation unit</td>
<td>143 geriatric patients</td>
<td>77 ± 8</td>
<td>SRM = 1.6 RE = 3.2</td>
</tr>
</tbody>
</table>

* p < 0.05. ADAS-Cog = Alzheimer’s Disease Assessment Scale-Cognitive Section [7]; BCRS = Brief Cognitive Rating Scale [53]; BI = Barthel Index [54]; CDS = Cornell Depression Scale [55]; CES-D = Centre for Epidemiologic Studies Depression Scale [56]; CGI = Clinical Global Impression; CIBIC-plus = Clinician’s Interview-Based Impression of Change-Plus caregiver input [49]; CIRS = Cumulative Illness Rating Scale [57]; FAQ = Functional Activities Questionnaire [58]; FIM = Functional Independence Measure [59]; GAS = Goal Attainment Scaling [9]; GCOR = Global Clinical Outcome Rating; GDS = Global Deterioration Scale [60]; HABAM = Hierarchical Assessment of Balance and Mobility [61]; IADL = Instrumental Activities of Daily Living [62]; KADL = Katz Activities of Daily Living Index [63]; MMSE = Mini-Mental State Examination [64]; NHP = Nottingham Health Profile [65]; OARS IADL = IADL subscale of the Older Americans Resources and Services [66]; PSMS = Physical Self-Maintenance Scale [62]; SMMSE = Standardized Mini-Mental State Examination [67]; SQLI = Spitzer Quality of Life Index [68].

ANOVA = Analysis of variance; ES = effect size; ICC = intraclass correlation coefficient; NA= not an aim of the study; RCT = randomized controlled trial; RE = relative efficiency; SRM = standardized response mean.
<table>
<thead>
<tr>
<th>Reliability</th>
<th>Content validity</th>
<th>Construct/convergent validity</th>
<th>Major findings</th>
</tr>
</thead>
</table>
| NA                      | inferred from its use in other geriatric settings | Spearman rank correlations of change scores  
BJ: $r = 0.2$; HABAM: $r = 0.2$  
CIRS: $r = -0.2$; Axis 8 of BCRS: $r = -0.2$ | GAS was the most responsive measure  
GAS did not correlate well with other measures |
| NA                      | NA                                | NA                                                                                           | GAS was responsive                                                 |
| inter-rater reliability (ICC = 0.9) | NA                                | Pearson correlations of change scores  
BJ: $r = 0.6$; FIM: $r = 0.5$  
MMSE: $r = 0.0$; KADL: $r = 0.5$  
PSMS: $r = -0.3$; SQLI: $r = 0.4$ | GAS appears to be somewhat more responsive to change than the other outcome measures and correlated moderately with the BI |
| NA                      | expert panel                      | Pearson correlations of change scores  
ADAS-Cog: $r = -0.5$; GDS: $r = -0.6$  
CGI: $r = -0.9$; MMSE: $r = 0.0$  
PSMS: $r = -0.3$; IADL: $r = -0.4$ | results on responsiveness and construct/convergent validity are mixed and content validity is unclear |
| NA                      | NA                                | MMSE; ADAS-Cog; PSMS; IADL; FAQ; CDS; CES-D; CIBIC-plus  
Patient/caregiver: Spearman $r = 0.0$ to 0.6  
Clinician: Spearman $r = 0.0$ to 0.8 | with the exception of the CIBIC-plus, most correlations between other measures and GAS were low to moderate  
These correlations increased slightly over time |
| NA                      | NA                                | NA                                                                                           | GAS was more responsive than any of the standard tests |
| NA                      | NA                                | NA                                                                                           | responsiveness of GAS was low for both clinicians and patients/caregivers |
| inter-rater reliability (ICC = 0.9 for change, 0.9 for discharge) | position statements               | Pearson correlations of change scores  
BJ: $r = 0.9$  
GCOR: $r = 0.8$ | GAS showed high content and construct/convergent validity and good inter-rater reliability |
| inter-rater reliability (ICC = 0.9 and 0.9) | expert panel                      | Pearson correlations of change and follow-up scores  
Change scores:  
BJ: $r = 0.6$; OARS IADL: $r = 0.5$  
MMSE: $r = 0.2$; NHP (overall): $r = 0.0$  
Follow-up scores:  
BJ: $r = 0.7$; OARS IADL: $r = 0.5$  
MMSE: $r = 0.3$; GCOR: $r = 0.7$  
NHP (overall): $r = -0.2$ | GAS was responsive and showed high inter-rater reliability  
Content validity and construct/convergent validity is unclear |
| NA                      | position statements               | SMMSE: Spearman $r = 0.1$  
BJ; KADL; OARS IADL  
Spearman $r$ between 0.4$^*$ and 0.5$^*$ | GAS was responsive and showed good content validity  
Correlations with other measures were low |

Applicability of Goal Attainment Scaling  
Dement Geriatr Cogn Disord 2008;26:528–540  
533
gested modifications. In Rockwood et al. [37], a panel with expertise in anti-dementia drug trials consisting of 3 neurologists, 2 geriatricians, an epidemiologist and a biostatistician reviewed the methodology and suggested modifications. Gordon et al. [19] inferred content validity from the use of GAS in other geriatric settings, but it was unclear whether the domains found in their study were related to the other settings as no results were shown.

Construct/Convergent Validity

Seven studies examined the construct/convergent validity of GAS using several secondary outcome measures (table 2). Spearman correlations were calculated by Gordon et al. [19], Rockwood et al. [34], Yip et al. [40] and ranged from 0.0 to 0.8. Pearson’s correlation coefficients were calculated by Rockwood et al. [36, 37] and Stolee et al. [38, 39], and ranged from 0.0 to 0.9.

Table 4 outlines the details about the feasibility of GAS of the selected studies and table 5 shows the comparison of the details about the feasibility of GAS.

Feasibility

Time Required for Setting the Goals

Five studies described the time required for setting the goals. Gordon et al. [19], Hartman et al. [32], Stolee et al. [38] and Yip et al. [40] stated that they were able to set goals in less than 30 min. The study by Rockwood et al. [37] was the only one in which it was not possible to set goals in less than half an hour.

Number of Goals

All 10 studies reported the number of goals set in their study. The exact number of goals for each study varied from 1 to 9. Further details are shown in table 4. In the study of Rockwood et al. [34], physicians set fewer goals than patients/caregivers (3 and 9, respectively). The same pattern is shown in Rockwood et al. [33]. Rockwood et al. [35] reported that the care-as-usual group and the intervention group set a similar number of goals (5 and 6, respectively). Two out of 10 studies had a mean number of goals set below 3. Eight studies showed a mean number of goals set above 3.

To What Extent Were Goals Realistic?

Whether goals were realistic was examined in 8 studies. Six studies had GAS follow-up scores between 40 and 60 [19, 32, 36, 38–40]. One study showed mixed results [35] and 1 study had GAS follow-up scores below 40, indicating that the goals were too difficult [37]. No study had a GAS follow-up score above 60, which would have indicated that the goals were too easy.

Patient and Caregiver Involved in Setting the Goals

All 10 studies described who was involved in the goal-setting process. Seven studies involved the patient and/or the caregiver in the goal-setting process [32–34, 37, 39, 40]. Two studies did not involve the patient and/or the caregiver because the clinician set the goals [19, 36]. In Stolee [38], 2 geriatricians independently assessed patients comprehensively. Then goals were set for each patient by consensus between the 2 geriatricians. However, it is not clear to what extent patients were involved in the goal-setting process.

A Blinded Assessor Involved in Assessing the Follow-Up GAS Scores

Six studies reported who assessed the follow-up GAS scores. Two studies used a blinded assessor [35, 36]. Four studies described that the persons who assessed the follow-up GAS scores were the persons who set the

Table 3. Comparison of the clinimetric aspects of GAS used in psychogeriatric patients

<table>
<thead>
<tr>
<th>Study</th>
<th>Responsiveness</th>
<th>Reliability</th>
<th>Content validity</th>
<th>Construct/convergent validity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gordon et al. [19]</td>
<td>+</td>
<td>NA</td>
<td>?</td>
<td>–</td>
</tr>
<tr>
<td>Hartman et al. [32]</td>
<td>+</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Rockwood et al. [36]</td>
<td>+</td>
<td>+</td>
<td>NA</td>
<td>+/–</td>
</tr>
<tr>
<td>Rockwood et al. [37]</td>
<td>+/–</td>
<td>NA</td>
<td>+</td>
<td>+/–</td>
</tr>
<tr>
<td>Rockwood et al. [34]</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>–, +/-</td>
</tr>
<tr>
<td>Rockwood et al. [35]</td>
<td>+</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Rockwood et al. [33]</td>
<td>–</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Stolee et al. [38]</td>
<td>NA</td>
<td>+</td>
<td>+</td>
<td>+/-</td>
</tr>
<tr>
<td>Stolee et al. [39]</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>–, +/-</td>
</tr>
<tr>
<td>Yip et al. [40]</td>
<td>+</td>
<td>NA</td>
<td>+</td>
<td>–</td>
</tr>
</tbody>
</table>

NA = Not an aim of the study.

a Changes between baseline and follow-up of GAS: – = NS; +/– = mixed results; + = S (p < 0.05, effect size >0.5, standardized response mean >0.5, relative efficiency >1); ? = unclear.
b Pearson correlation coefficient or intraclass correlation coefficient: – = <0.7; +/– = between 0.7 and 0.79; + = >0.8; ? = unclear.
c An expert’s panel reviewed the selected domains for GAS: – = reviewed as insufficient; + = reviewed as sufficient; ? = unclear.
d Correlation coefficient (Pearson or Spearman) between GAS and secondary outcome measures: – = low correlation (<0.5); +/- = moderate correlation (between 0.5–0.69); + = high correlation (>0.69); ? = unclear.
goals. Therefore, these assessors were not blinded [33, 34, 38, 40].

Domains That Were Included in the Goals

Nine studies described the domains that were covered by the goals. The number of domains varied from 2 to 12. Six studies reported 2 or more of the following 5 domains: cognition, instrumental activities of daily life/self-care, mood, behaviour or mobility [33–35, 37, 40, 44]. Three studies set goals in less than 2 predefined domains [19, 32, 38].

**Discussion**

We reviewed the literature on the clinimetric aspects and the feasibility of GAS when used as an outcome measure for psychogeriatric patients with cognitive disorders in order to study the applicability of GAS in this population. The literature search resulted in the identification of 10 studies of which most were performed by Rockwood and co-workers.

Although 1 study [19] argued that the presence of dementia did not make it feasible to involve the patients in...
setting their own goals, our review shows that the involvement of the demented patient and/or caregiver actually is feasible. Moreover, GAS can cover multiple domains relevant to the psychogeriatric population such as cognition, daily life activities and mood.

Six of the 10 studies did not provide full details of their sample. We could not rule out that the conclusions in these studies were based on non-cognitively impaired patients. Therefore, we were not able to make firm conclusions about those clinimetric aspects and feasibility when the majority of studies investigating these aspects were those with heterogeneous samples. This was the case for all the clinimetric aspects, time required for setting goals, realistic goals and the involvement of a blinded assessor in assessing the follow-up GAS score. Reliability was positive in all studies that investigated this clinimetric aspect, but since these were just 3 studies and all had a heterogeneous sample, results must be interpreted with caution.

**Strengths and Limitations of the Studies in the Review**

Nine of the 10 studies reviewed were performed by one centre (Rockwood and co-workers). Very few studies outside this research group investigated GAS in psychogeriatric patients with cognitive disorders [45]. None of them described clinimetric aspects and feasibility of GAS. The absence of independent studies raises questions about the use of GAS in cognitively impaired psychogeriatric patients and may be a reason why this potential novel and promising approach is not studied and used more in this population. Another explanation of the fact that we did not find studies from other research groups could be publication bias: these groups might have found negative results with regard to the use of GAS in psychogeriatric patients with cognitive disorders that they were not willing to publish.

With regard to the methodology of the reviewed studies, we noted first of all that the intraclass correlation coefficient ICC for measuring IRR is preferred over Pearson’s correlation when the sample size is small (n < 15). As Pearson’s correlation makes no assumptions about rater means, a t-test of the significance of the correlation is necessary to see whether inter-rater means differ [46]. All 3 studies assessing inter-rater reliability used the intraclass correlation coefficient. However, the range of sample size varied between n = 15 and n = 170, so the intraclass correlation coefficient was not the preferred statistic. Secondly, large sample sizes are generally preferred in order to find reliable results [31]. However, with a highly responsive instrument – as has turned out to be the case with GAS – smaller sample sizes are sufficient to demonstrate the efficacy of an intervention [36]. Therefore, the small sample sizes of some of the reviewed studies do not form a limitation.

### Table 5. Comparison of the feasibility of GAS used in psychogeriatric patients

<table>
<thead>
<tr>
<th>Study</th>
<th>Time required for setting goalsa</th>
<th>Number of goalsb</th>
<th>Realistic goalsc</th>
<th>Patient and/or caregiver involved in goal settingd</th>
<th>Blinded assessor involved in assessing the follow-up GAS scoresē</th>
<th>Main domainsf</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gordon et al. [19]</td>
<td>+</td>
<td>–</td>
<td>+</td>
<td>–</td>
<td>NR</td>
<td>–</td>
</tr>
<tr>
<td>Hartman et al. [32]</td>
<td>+</td>
<td>–</td>
<td>+</td>
<td>+</td>
<td>NR</td>
<td>–</td>
</tr>
<tr>
<td>Rockwood et al. [36]</td>
<td>NR</td>
<td>+</td>
<td>+</td>
<td>?</td>
<td>+</td>
<td>NR</td>
</tr>
<tr>
<td>Rockwood et al. [37]</td>
<td>–</td>
<td>+</td>
<td>–</td>
<td>+</td>
<td>NR</td>
<td>–</td>
</tr>
<tr>
<td>Rockwood et al. [34]</td>
<td>NR</td>
<td>+</td>
<td>NR</td>
<td>+</td>
<td>–</td>
<td>+</td>
</tr>
<tr>
<td>Rockwood et al. [35]</td>
<td>NR</td>
<td>+</td>
<td>+/-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Rockwood et al. [33]</td>
<td>NR</td>
<td>+</td>
<td>NR</td>
<td>+</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Stolee et al. [38]</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>?</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Stolee et al. [39]</td>
<td>NR</td>
<td>+</td>
<td>+</td>
<td>?</td>
<td>NR</td>
<td>+</td>
</tr>
<tr>
<td>Yip et al. [40]</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>?</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

NR = Not reported.

a – = >30 min; + = ≤30 min.
b – = <3; + = ≥3.
c – = Follow-up GAS score <40 or >60; + = follow-up GAS score between ≤40 and ≤60; +/- = mixed results.
d – = Absence of patient or caregiver; + = presence of patient or caregiver; ? = role of patient or caregiver unclear.

ē – = Absence of blinded assessor; + = presence of blinded assessor.
f – = ≤1 domain out of 5 (i.e. cognition, instrumental activity of daily life, mood, behaviour or mobility); + = 2 or more domains out of 5.
**Recommendations for Future Use of GAS**

We found that mixed results were obtained when evaluating the validity of GAS. This can be due to the fact that the secondary outcome measures used in the studies are fixed in terms of the scoring items and cannot vary across individuals while GAS can. Therefore, correlations between GAS and standardized tests are generally expected to be low [47]. Rockwood et al. [36] mentioned that there is no gold standard (criterion) for measuring clinically important change; Donnelly and Carswell [48] underscore the idiosyncratic nature of GAS, both making it hard to validate it. Although several authors have pointed to the fact that validation of GAS may be problematic because of its highly personalized approach, the claim that it is a highly ecologically valid instrument still needs to be proven. Using a blinded assessor for measuring the follow-up GAS scores can enhance cross-validity; this also ensures an unbiased outcome of interventions. In addition, comparative measures such as the Clinical Global Impression and the Clinician’s Interview-Based Impression of Change-Plus caregiver input [49] are also sensitive to clinical change and can thus be used to validate the GAS. Although these latter instruments do not take into account the specific needs of the patient and/or caregiver, their aim is to measure relevant clinical change after an intervention.

We have selected studies based on the criterion that the original GAS was used, and only small deviations were accepted. Two studies [34, 40] described small adaptations of which the adaptation made by Rockwood et al. [34] to set the baseline level on ‘0’ instead of ‘−2’ or ‘−1’ is recommended. According to them, this modification incorporates more levels of decline, which seems preferable, given that deterioration is likely in demented patients. When shifting the baseline status to ‘0’, the GAS formula must be adapted too since the score of 50 in the GAS formula is not the hypothetical mean for the follow-up score anymore. Yip et al. [40] developed a standard menu of goal areas developed by the geriatric assessment and rehabilitation unit team, from which relevant goals could be selected. However, the expected outcome was a non-numerical scoring which is qualitative and therefore difficult to interpret.

As most studies reviewed were performed by one centre (Rockwood and co-workers), a more widespread use by other independent research groups may produce more data on the applicability of GAS.

**Clinical Implications**

As dementia is characterized by a complex multidimensional nature, it is important that outcome measures reflect this multidimensionality. Our review showed that GAS is an outcome measure that can cover several relevant domains like cognition, behaviour and activities of daily life. Therefore, it is a relevant method for measuring outcomes in the psychogeriatric field. Outcome measures also need to take into account the progressive nature of dementia; as GAS can be adapted at any time, it is useful in dementia.

Furthermore, depending on the aim for which GAS is used, both process and outcome goals can be set. However, for intervention research it is important not to adapt the GAS follow-up guide because a baseline measure and predefined follow-up measure are needed.

Part of the definition of dementia is that cognitive impairments lead to interference with daily life activities. Traditional measures of dementia often do not reflect this interference [50]; GAS, however, does. Outcome measures seldom focus on individual needs relevant to the individual patient and the caregiver, focusing instead on general concepts, like mood and fatigue, with norms based on groups. As we showed in our review, many studies actively involved patients and/or caregivers in the goal-setting process, thereby ensuring that relevant goals were selected. In other words, the fact that GAS takes into account individual preferences is a safeguard for clinical relevance of this outcome measure. Patients with dementia may lack insight into their problems, in which case a caregiver can be asked to help with setting the goals. Caregiver bias can be present here, but since dementia also affects the caregiver [51], caregiver’s goals are highly relevant as well. Our conclusion is that caregivers should not only be asked to help setting goals for their demented partners, but should also set goals that are relevant to themselves.

GAS has a special position among regular outcome measures in the field of cognitive disorders and it is a useful complementary instrument [52] providing information that cannot be obtained by traditional standard outcome measures like the ADAS-cog. Although it can be complicated to involve a whole multidisciplinary team in setting goals for each patient, this approach provides a lot of relevant information that will remain hidden when a monodisciplinary approach is used. We realize that the GAS constructing time of 30 min is relatively long compared to a quick MMSE, but the extra information provided by GAS justifies the time involved in training and constructing the GAS follow-up guides.
Since GAS is developed to fit the patient’s individual needs and is constructed separately for each patient, it bridges possible language and culture barriers. As there is no uniform description of how to use GAS in a cognitively impaired population, training and practice to set goals in this particular population are recommended, and practice guidelines for standardizing GAS in a cognitively impaired population would be desirable.

**Conclusion**

GAS proved to be useful in evaluating important aspects of an outcome measure for psychogeriatric patients with cognitive disorders. Since other relevant aspects showed mixed results and the number of studies investigating the use of GAS in psychogeriatric patients with cognitive disorders is small, the evidence is not strong enough yet to state that GAS is an applicable outcome measure in this population. We do think that GAS has potential value and although the use of GAS requires training, which is time-consuming, we believe that this investment is worthwhile because GAS is a unique example of an instrument able to reflect the multidimensionality of dementia and other psychogeriatric conditions, including interference with daily life activities, for both patient and caregiver.

**Appendix 1**

Formula to calculate GAS scores:

\[
GAS \text{ score} = \frac{50 + 10 \sum (w_i x_i)}{\sqrt{0.7 \sum w_i^2 + 0.3 \left( \sum w_i \right)^2}}
\]

where \(x_i\) = the attainment level and \(w_i\) = the weight assigned to the goal area. Because the expected outcome/programme goal is assigned a score of 0, if all goals are met, the outcome score would be 50.

**Appendix 2**

Details of Literature Search

<table>
<thead>
<tr>
<th>Category</th>
<th>Terms</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Controlled vocabulary words (combinations of these terms)</strong></td>
<td></td>
</tr>
<tr>
<td>Goal attainment scaling</td>
<td>Alzheimer’s disease, delirium, dementia, amnestic cognitive disorders, dementia, multi-infarct dementia, vascular dementia, cognitive rehabilitation, cognition disorders psychometrics, psychometry, interrater reliability, reliability and validity, reliability, intrarater reliability, test-retest reliability, validity, face validity, qualitative validity, content validity, consensual validity predictive validity, criterion related validity, concurrent validity, external validity, discriminant validity, internal validity, construct validity, sensitivity to change, sensitivity and specificity, reproducibility of results</td>
</tr>
</tbody>
</table>

**References**

Applicability of Goal Attainment Scaling

41 Fox RA, Clarfield AM, Hogan DB: Competencies required for the practice of geriatric medicine as a consultant physician. CMAJ 1989; 141:1045–1048.

Dement Geriatr Cogn Disord 2008;26:528–540

539


