HHS Public Access

Author manuscript

Gerontology. Author manuscript; available in PMC 2024 June 29.

Published in final edited form as:

Gerontology. 2023 ; 69(8): 1002–1013. doi:10.1159/000529984.

Suitability of Goal Attainment Scaling in Older Adult Populations with Neurodegenerative Disease Experiencing Cognitive Impairment: A Systematic Review and Meta-Analysis

Ollie Fegter[a,b], **Haylie Santos**[c], **Alfred W. Rademaker**[a], **Angela C. Roberts**[a,c,d,e], **Emily Rogalski**[a,b]

[a]Mesulam Center for Cognitive Neurology and Alzheimer's Disease, Northwestern University Feinberg School of Medicine, Chicago, IL, USA

[b]Department of Psychiatry and Behavioral Sciences, Northwestern University Feinberg School of Medicine, Chicago, IL, USA

[c]School of Communication Sciences and Disorders, Western University, London, ON, Canada

[d]Department of Computer Science, Western University, London, ON, Canada

[e]Canadian Centre for Activity and Aging, Western University, London, ON, Canada

Abstract

Introduction: Identifying responsive outcome measures for assessing functional change related to cognition, communication, and quality of life for individuals with neurodegenerative disease is important for intervention design and clinical care. Goal Attainment Scaling (GAS) has been used as an outcome measure to formally develop and systematically measure incremental progress toward functional, patient-centered goals in clinical settings. Evidence suggests that GAS is reliable and feasible for use in older adult populations and in adult populations with cognitive impairment, but no review has assessed the suitability of GAS in older adults with neurodegenerative disease experiencing dementia or cognitive impairment, based on responsiveness. This study conducted a systematic review to evaluate the suitability of GAS as an outcome measure for older adult populations with neurodegenerative disease experiencing dementia or cognitive impairment, based on responsiveness.

Methods: The review was registered with PROSPERO and performed by searching ten electronic scientific databases (PubMed, Medline OVID, CINAHL, Cochrane, Embase, Web of Science, PsycINFO, Scopus, OTSeeker, REHABDATA) and four registries (Clinicaltrials.gov, Grey Literature Report, Mednar, OpenGrey). A summary measure of responsiveness (post-intervention

Correspondence to: Ollie Fegter, olivia.fegter@northwestern.edu.

Author Contributions

O.F., E.R., and A.C.R. designed the study. O.F. performed the literature search. O.F. and H.S. screened articles, extracted data, and conducted risk of bias assessments, with input from E.R. and A.C.R. as needed. A.W.R. performed statistical analyses. O.F. wrote the article with input from all authors.

Conflict of Interest Statement

The authors have no conflicts of interest to declare.

Statement of Ethics

An ethics statement is not applicable because this study is based exclusively on published literature.

minus pre-intervention mean GAS T-score) was compared across eligible studies using a random-effects meta-analysis. Risk of bias in included studies was assessed using the NIH Quality Assessment Tool for Before-After (Pre-Post) Studies with No Control Group.

Results: 882 eligible articles were identified and screened by two independent reviewers. Ten studies met inclusion criteria for the final analysis. Of the ten included reports, 3 focus on all-cause dementia, 3 on multiple sclerosis, 1 on Parkinson's disease, 1 on mild cognitive impairment, 1 on Alzheimer's disease, and 1 on primary progressive aphasia. Responsiveness analyses showed pre- and post-intervention GAS goals were significantly different from zero ($Z = 7.48$, $p < 0.001$), with post-intervention GAS scores being higher than pre-intervention GAS scores. Three included studies showed a high risk of bias, 3 showed a moderate risk of bias, and 4 showed a low risk of bias. Overall risk of bias of included studies was rated as moderate.

Conclusion: GAS showed an improvement in goal attainment across different dementia patient populations and intervention types. The overall moderate risk of bias suggests that while bias is present across included studies (e.g., small sample size, unblinded assessors), the observed effect likely represents the true effect. This suggests that GAS is responsive to functional change and may be suitable for use in older adult populations with neurodegenerative disease experiencing dementia or cognitive impairment.

Keywords

Goal Attainment Scaling; Dementia; Neurodegenerative disease; Systematic review; Meta-analysis

Introduction

Goal Attainment Scaling (GAS) was initially developed as a program evaluation instrument for community mental health programs [1]. With time, GAS was adopted as a method to standardize the measurement of change in individualized goals in response to intervention [2].

GAS provides a systematic approach to defining goals and evaluating goal achievement. For each goal, an ordinal scale of possible treatment outcomes is developed [1]. Each achievement level is anchored by a descriptive statement that takes into account the personal relevance and level of difficulty of the goal. Each response rating can then be transformed into a T-score which reflects level of achievement for that goal [1]. Transforming response ratings into T-scores allows for standardizing scores across goals and across participants. These scales are specific to the goals of each patient and should be clear enough that a follow-up assessor unfamiliar with the goals can evaluate goal attainment [3]. An example of a GAS goal format from the Communication Bridge-2 randomized controlled trial [4] is provided below.

A participant with language and memory difficulties may want to improve their recall of items on a grocery list. The starting point (i.e., patient's current status) may be that the patient forgets 3–4 items on the grocery list (assigned a scale value of 0). Forgetting to buy all items on the grocery list may be significantly worse than the expected treatment outcome

(assigned a scale value of −2), and remembering to buy every item on the grocery list may be significantly better than the expected treatment outcome (assigned a scale value of +2).

It is important to note that while some studies assign a scale value of “0” to the expected treatment outcome, other studies assign a scale value of “0” to the patient’s current level of functioning. When a patient’s current level of functioning is anchored at “0,” scores above “0” represent functional improvement and scores below “0” represent functional decline. For an example of this approach, see the study by Roberts et al. [4]. These approaches are mathematically equivalent [2]. The number of scale values and treatment goals can be adapted to the needs of each patient [3]. When this process is followed, GAS allows for the measurement of incremental progress toward patient-centered goals [2, 5]. Additionally, Kiresuk et al. [2] suggested that patient participation in goal setting may improve treatment compliance and outcomes.

GAS has been used in both clinical and research settings to measure individual goal attainment for individuals with mild cognitive impairment (MCI) and dementia [6–8], communication disorders [9], intellectual disabilities [10], motor delays [11], traumatic brain injury [12], and Alzheimer’s disease dementia [13, 14]. GAS has been used to compare treatment outcomes for pharmaceutical interventions [15–17], occupational therapy to restore activities of daily living (ADLs) [18], geriatric rehabilitation [19, 20], memory and aging programs [21], and geriatric day hospitals [22].

Existing reviews have evaluated the psychometric properties (i.e., reliability, validity) and feasibility of GAS in geriatric rehabilitation and in adults with cognitive impairment. One systematic review found high responsiveness related to patient-centered goals in geriatric rehabilitation [20]. Another systematic review evaluated the suitability of GAS in individuals with cognitive impairment based on psychometric properties (e.g., validity, reliability) and feasibility and concluded that GAS is feasible and useful for measuring treatment outcomes in adults with cognitive impairment (e.g., MCI, depression, dementia, or subjective cognitive complaints after stroke or congestive heart failure) [23]. However, they cautioned that more research is needed to draw conclusions related to the reliability and validity of the GAS instrument in this population [23].

While these reviews provide support for the use of GAS in geriatric rehabilitation and with adults with cognitive impairment (e.g., MCI, depression, subjective cognitive complaints after stroke) based on psychometric properties and feasibility, no review has evaluated the responsiveness of GAS in older adult populations with neurodegenerative disease experiencing dementia or cognitive impairment. Cognitive impairment is characterized by a deficit in any number of cognitive functions (e.g., memory, language, executive functioning) and may be the result of another disease or disorder [24]. This is different from MCI, which is a syndrome characterized by cognitive decline that does not interfere with daily activities but is greater than expected for one’s age [25]. MCI is distinct from dementia, where cognitive decline interferes more substantially with daily functioning. However, amnestic MCI shows a high rate of progression to dementia [25]. Neurodegenerative disease is the progressive loss of neuron populations and can be classified by clinical features, neuroanatomical changes, or neuropathological changes [26–

28]. Common neurodegenerative diseases include Alzheimer's disease, Parkinson's disease, Huntington's disease, and motor neuron disease. Because cognition declines over time with neurodegenerative dementia, understanding GAS' ability to measure functional change over time is essential to understanding its utility in this population.

The current review evaluates GAS responsiveness to assess its suitability for use as an intervention outcome measurement in older adult populations with neurodegenerative disease experiencing dementia or cognitive impairment. Responsiveness is defined as a measure's ability to detect clinically meaningful change over time [29, 30]. Clinically meaningful change suggests functional improvement or change in the prognosis or treatment of disease [31]. While there is no widely accepted measure of clinically meaningful change, it has been suggested that statistical significance may be used to infer clinical significance [31]. For this review, responsiveness is operationalized as a measure's ability to detect statistically significant pre- and post-intervention scores across studies [32]. This definition is consistent with best practices in measuring responsiveness and has been used in other studies examining the suitability of GAS in older adults [22, 23, 29, 30]. In older adults generally and in adults with cognitive impairment, GAS has been highly responsive in measuring functional change related to cognition, communication, quality of life (QOL), and ADLs [8, 18, 21–23, 33]. Additionally, GAS has shown greater responsiveness to functional gains in older adults when compared to other commonly used instruments, such as the Comprehensive Geriatric Assessment (CGA), Barthel Index (BI) for ADLs and the Physical Self-Maintenance Scale (PSMS) [12, 15, 16, 34].

This suggests that GAS is appropriate for measuring functional change toward predetermined goals related to cognition, communication, QOL, and ADLs. However, GAS may be differentially responsive in different populations, and understanding its responsiveness in specific populations is essential to understanding its clinical utility. Therefore, the purpose of this systematic review was to answer the following question: is GAS suitable for use in older adult populations with neurodegenerative disease experiencing dementia or cognitive impairment, based on responsiveness? While existing reviews have examined the suitability of GAS in older adults generally [20] and for use with adults with cognitive impairment [23], no study has specifically reviewed the suitability of GAS for use with older adults with neurodegenerative disease experiencing dementia or cognitive impairment, based on responsiveness.

Method

This systematic review was conducted in adherence to the PRISMA guidelines [35]. Consistent with PRISMA guidelines, details of the protocol for this systematic review were registered on PROSPERO prior to the initial literature search and can be accessed online [36].

Eligibility Criteria

An article was considered eligible for inclusion if the report (i) used an adult sample, (ii) included adults with neurodegenerative disease experiencing any level of dementia or cognitive impairment, (iii) used GAS as originally described [1], (iv) used GAS in the

context of an intervention, (v) included pre- and post-intervention GAS scores, and (vi) was available in English. Published randomized control trials, non-randomized control studies, and cohort studies were included in this review. Because the objective of this review was to address the suitability of GAS and not to evaluate the efficacy of a specific intervention, all intervention types were eligible for inclusion in this review.

After the initial screening by title and abstract, a significant number of articles were included that did not use GAS in the context of an intervention that targeted cognition, communication, ADLs, or QOL. Most of these articles used GAS in the context of a motor intervention. To better address the research question in context of cognition, communication, ADLs, and QOL, two additional inclusion criteria were added. With the additional criteria, reports were only eligible for inclusion if (i) the study met all original inclusion criteria, (ii) the outcome measures used in the study assessed cognition, communication, QOL, or ADLs, and (iii) the study intervention targeted cognition, communication, QOL, or ADLs. Figure 1 illustrates the inclusion decision-making process. Reports were compared using cohort names, author names, study locations, study dates, and intervention description to determine which reports came from the same study. When multiple reports from the same study were identified, the earliest published report meeting inclusion criteria was used to avoid duplicating results in the analysis.

Literature Search and Study Selection

The following databases were searched: PubMed, MEDLINE (Ovid), CINAHL (EBSCOhost), Cochrane Reviews, Embase (Elsevier), Web of Science, PsycINFO (Ovid), Scopus, OTSeeker, and REHABDATA. The following registers and gray literature databases were also searched: ClinicalTrials.gov, Grey Literature Report, Mednar, and OpenGrey. Search terms for each database were created in consultation with a librarian and following the methodology proposed by Bramer et al. [37]. Search terms included both controlled vocabulary terms (for databases and registers with thesauri) and free-text terms. The methodology proposed by Bramer et al. [37] includes macros for Microsoft Word that allow for a single-line query in Embase to be translated into other major databases. These macros were used to translate search query syntax. Thesaurus terms were changed manually for MEDLINE (Ovid), PsycINFO (Ovid), and CINAHL (EBSCOhost), as these databases have custom controlled-text thesauri. Documentation of search queries used can be found in online supplementary Material 1 (for all online suppl. material, see https://doi.org/10.1159/000529984). Databases and registers were searched from inception through October 11, 2021. Identified citations were screened, and an updated search was performed using the same search strategy to identify eligible reports from October 11, 2021, through March 3, 2022. No additional eligible reports were identified.

Citations for each identified report were downloaded to EndNote version 20.2. Duplicates were removed, and citations were uploaded to Rayyan for screening. All reviewers were blinded to other reviewers' decisions during screening, and blinding was removed only to compare inclusion decisions. Two independent reviewers (O.F. and H.S.) screened all reports alphabetically by title and abstract with the initial inclusion criteria in groups of 50–75. After each group of reports was screened, reviewers met to compare inclusion decisions

and resolve inconsistencies. Two additional reviewers (A.C.R. and E.R.) were consulted in cases of disagreement. After the additional inclusion criteria were added, reports were again screened by two independent reviewers (O.F. and H.S.) by title and abstract in groups of 50. After screening by title and abstract was complete, two reviewers (O.F. and H.S.) independently screened full-text reports in groups of 3–5 for inclusion. Reviewers met to compare and resolve inconsistencies, and two additional reviewers (A.C.R. and E.R.) were consulted in cases of disagreement. Disagreement between the two initial reviewers was at or below 10% for each group of articles screened. Citations that did not meet inclusion criteria were excluded, and the reason for exclusion was recorded at both the title/abstract screening stage and the full-text screening stage.

Data Collection and Analysis

Data Extraction—A modified version of the Cochrane data collection form (2013) was developed in Microsoft Excel and used for data collection. Two independent reviewers (O.F. and H.S.) extracted data from each report. They conferred after collecting data from 1–3 reports to compare extracted data and discuss any inconsistencies. No automation tools were used to extract data. For eligible articles that did not report pre- and post-intervention GAS scores, authors were contacted directly via email to obtain GAS scores. If authors could not provide GAS scores, the report was not included in analysis.

Risk of Bias Assessment—In systematic reviews, risk of bias is an assessment of the systematic error in a study's methodology (e.g., design, implementation, analysis, reporting) that can lead to underestimation or overestimation of the intervention effect [38]. Different risk of bias measures are recommended for assessing methodological quality based on study type [39]. For this review, risk of bias for pre-post studies with no control group was assessed using the NIH Quality Assessment Tool for Before-After (Pre-Post) Studies with No Control Group [40]. The NIH Quality Assessment Tool assesses risk of bias for 12 domains, such as sample size, intervention fidelity, inclusion criteria, attrition rate, and reliability and validity of outcome measures. An overall quality rating of "good," "fair," or "poor" is then assigned to each report. A full list of risk of bias domains for the NIH Quality Assessment Tool can be found in online supplementary Material 2. The Grading of Recommendations, Assessment, Development, and Evaluation (GRADE) framework was used to evaluate the overall quality of evidence for this review [41]. The GRADE framework allows researchers to evaluate the quality of a body of evidence based on risk of bias, inconsistency, imprecision, and publication bias across reports. Quality of evidence is then rated as being very low, low, moderate, or high. These descriptors correlate to the level of confidence the authors have that the true effect is similar to the estimated effect.

Included reports were assessed for risk of bias by two independent reviewers (O.F. and H.S.). The reviewers met to discuss risk of bias decisions and resolve discrepancies after independently rating 1–3 articles. No automation tools were used in evaluating risk of bias with the NIH Quality Assessment Tool. Disagreement between reviewers was at or below 10% for each group of reports assessed.

Synthesis Methods—As studies were only included if they reported pre- and post-intervention GAS scores, all included studies were eligible for synthesis and included in the summary measure of responsiveness. Responsiveness is defined as a measure's ability to detect clinically meaningful change over time [29, 30]. Clinically meaningful change has generally been defined as the extent to which an intervention makes a difference in the everyday life of an individual [42]. While there is no widely accepted measure of clinically meaningful change, Crosby et al. [31] suggested that clinically meaningful change will result in functional improvement, symptom reduction, or change in the treatment or prognosis of disease. For one-group repeated measures designs, paired *t* tests may be used to determine statistical significance and infer clinical significance [31]. For this review, responsiveness is defined as a measure's ability to detect statistically significant pre- and post-intervention scores across studies [32].

The measure of responsiveness for each included study was the post-intervention minus pre-intervention mean difference in the GAS T-score. Paired *t* tests were conducted to determine if mean study differences were significantly different from zero. These mean differences did not account for time between pre- and post-intervention GAS goals. The results of the individual studies were combined into a summary measure using random-effects meta-analysis, with an assumed correlation of zero between pre- and post-intervention GAS scores [32]. Kiresuk and Sherman [1] suggested an assumed correlation of 0.3, but this was for correlation among multiple goals measured at one time rather than for the correlation between pre- and post-intervention scores. A conservatively assumed correlation of zero between pre- and post-intervention scores within each study was used for this analysis as this maximized the within-study variability. However, the zero-correlation assumption likely has little effect on the results since the between-study variance overwhelmed the within-study variance (the percent of variance due to between-study variability, I-squared = 97%). Review Manager [43] software was used to perform the meta-analysis calculation.

Results

Study Selection and Characteristics

A total of 882 records were identified through database and register searching. Of the initial search, 326 duplicates were removed, and 556 articles were screened by title and abstract with the original inclusion criteria. Based on the eligibility criteria, 459 articles were excluded. Four identified reports included only a title and not an abstract. The authors of these reports were contacted, and we were unable to retrieve three of these reports. As such, the reports were excluded. 94 articles were screened again by title and abstract with the additional inclusion criteria. Full-text documents for 23 articles were screened. Ten reports met all eligibility criteria and were included in the analysis. No additional reports were added after the updated search. Figure 2 illustrates the full study selection process. Of the ten included reports, 3 focus on all-cause dementia, 3 on multiple sclerosis, 1 on Parkinson's disease, 1 on MCI, 1 on Alzheimer's disease, and 1 on primary progressive aphasia. In these studies, GAS was used to measure functional improvement in cognitive rehabilitation, drug trials, speech-language therapy, and memory training. Table 1 provides a full summary of

study characteristics. Average time between pre- and post-intervention GAS goals was 17.5 weeks (SD = 14.25, range = 4–52).

Risk of Bias Assessment

Risk of bias as assessed with the NIH Quality Assessment Tool can be seen in Figure 3. Of the ten included studies, four showed an overall low risk of bias, three showed a moderate risk of bias, and three showed a high risk of bias. Across domains, there was a low risk of bias for defining the study's objective and eligibility criteria, intervention's description and fidelity, reliability and validity of outcome measures, attrition rate, and administering outcome measures pre- and post-intervention. There was a moderate risk of bias for representativeness of the sample, blinded assessors, and reporting p values for pre-to-post-intervention changes. Four studies did not use blinded assessors at GAS follow-up [44–47], and five studies did not report p values for pre-to-post changes in GAS scores [46–50]. There was a moderate-high risk of bias for sample size. Four studies [45, 46, 49, 50] used a small sample size (i.e., less than 15 participants). Risk of bias for one domain remained unclear, as no study reported if all eligible participants were enrolled. According to the guidelines given in the NIH Quality Assessment Tool, overall risk of bias was rated as moderate. Using the GRADE tool, the quality of evidence of this review was rated as moderate [41].

Responsiveness

The summary measure of responsiveness was the post-intervention minus pre-intervention GAS T-score, and the mean difference was calculated as a weighted average using sample size. The difference between post- and pre-intervention GAS scores was significantly different from zero ($z = 7.48$, $p < 0.001$), with post-intervention GAS scores being higher than pre-intervention GAS scores (M = 18.11, 95% CI [13.36, 22.85]; Figure 4).

Discussion

This systematic review was conducted to evaluate the suitability of GAS as an outcome measure for older adult populations with neurodegenerative disease experiencing dementia or cognitive impairment, based on responsiveness. The results show that the difference between post-intervention and pre-intervention GAS scores is significantly different from zero in the positive direction. This suggests that across older adult populations with neurodegenerative disease experiencing dementia or cognitive impairment, GAS may be able to detect clinically meaningful change and may be responsive to measuring progress toward functional patient-centered goals related to communication, cognition, QOL, or ADLs. These findings are consistent with other systematic reviews suggesting that GAS is responsive to functional change in geriatric rehabilitation [8, 20] and in adults with cognitive impairment [23]. While these reviews did not specifically evaluate GAS in older adults with neurodegenerative disease experiencing dementia or cognitive impairment, responsiveness trends are similar between this study and the studies mentioned above (i.e., statistically significant pre- and post-intervention GAS scores).

The NIH Quality Assessment Tool was used to rate limitations in methodology of the ten studies included in this review and found an overall moderate risk of bias rating (see Fig. 3). For example, goal setting, scaling, and evaluation were not thoroughly described in all articles. Four reports (40%) did not describe how many goals each participant set [14, 45, 47, 50], and one report (10%) did not describe who was involved in goal setting and scaling [45]. Four studies (40%) did not use blinded assessors at GAS follow-up [44–47], and two studies (20%) did not report if GAS assessors were blinded at follow-up [48, 49]. This increases risk of bias, as assessors may be influenced by knowledge of the participants' baseline GAS scores. A moderate-high risk of bias was introduced due to small sample sizes in four studies (40%) [45, 46, 49, 50]. While these were appropriately weighted in the meta-analysis, these studies inevitably had a higher risk of bias than studies with a larger sample size [51]. All studies used for analysis failed to report if all eligible participants were included. Despite these limitations, each report clearly described the study's objective and eligibility criteria, used reliable and valid outcome measures (e.g., California Verbal Learning Test, Patient-Reported Outcomes Measurement Information System, Rivermead Behavioural Memory Test, Mini-Mental Status Examination) in addition to using GAS, and assessed goal attainment pre- and post-intervention. Additionally, a majority of reports (90%) clearly defined the study intervention, and only one study had an attrition rate above 20% [14]. According to the GRADE tool for assessing quality of evidence in a systematic review, there is an overall moderate risk of bias and moderate level of certainty in the findings of this review [41]. This suggests that while the included studies and overall result are susceptible to some bias, the problems are not severe enough to invalidate the results of the meta-analysis [52]. As such, the observed effect in this review is likely close to the true effect in this population [41, 52].

Evaluating the psychometric properties and feasibility of GAS in specific settings and populations is critical to understanding its suitability for assessing individual treatment outcomes and the quality of treatment programs [5, 53]. Only one report in this review included information on the psychometric properties of GAS use in this population (see Table 1). O'Sullivan et al. [49] reported inter-rater reliability for follow-up assessors (kappa = 0.73). No other studies reported metrics for reliability or validity. This is consistent with other systematic and scoping reviews that have found sparse reporting on the validity and reliability of GAS use in rehabilitation research [33] and in adults with communication disorders [9], MCI, or dementia [8]. One systematic review of GAS use in geriatric rehabilitation [20] found that GAS showed high concurrent, content, and predictive validity, as well as high inter-rater reliability. Some studies not meeting eligibility criteria for this review have shown high reliability and validity using GAS in older adult populations [3, 15, 20], while others have suggested that GAS has low reliability, specifically inter-rater reliability [9, 16]. More research is needed to understand the possible factors (e.g., rater training, goal target, cognition) that may contribute to the reliability of GAS scoring in older adult populations with neurodegenerative disease experiencing dementia or cognitive impairment.

It is important to note that the structure of GAS scales may affect researchers' ability to measure change and compare goal attainment across interventions. The structure of the GAS scale, including where 0 is anchored (i.e., as the baseline performance, as the expected

outcome with all scores being positive, or as the expected outcome with negative values signifying decline and positive values signifying gains), may change the scale range and limit the ability to compare scores across scales. For example, in three reports included in this study, a scale value of "0" was assigned to the participants' baseline performance [13, 14, 47]. In these cases, had "2" been used as the most likely treatment outcome rather than "0," the magnitude of the mean change in GAS score may have been even greater. Therefore, while the results of this review suggest that GAS is highly responsive in this population, these results may be a conservative estimate of actual change.

Limitations

There were some limitations to the methodology of this review. There was one deviation from the published protocol: additional inclusion criteria were added after the initial screening by title and abstract. These additional criteria were added to narrow the scope of this review and focus the suitability of GAS in this population on interventions targeting cognition, communication, QOL, and ADLs. Adding these criteria allowed for interventions targeting motor functioning to be excluded from this review. Of the 71 articles excluded at the additional title and abstract screening, 64 were excluded because they related to motor functioning. The other seven articles excluded at this stage were not in the context of an intervention and would have been excluded at the full-text screening even if the additional criteria had not been added. There is also some potential bias from the use of the NIH Quality Assessment Tool. The NIH Quality Assessment Tool relies on reviewer judgment to make a global risk of bias judgment. While there was less than 10% disagreement between reviewers when assessing risk of bias, there is potentially more bias in using this tool than in using other algorithm-driven tools to assess risk of bias. Additionally, there may have been language, cultural, and publication biases in the articles included in this analysis. One relevant article was excluded because no English translation was available, and no efforts were made to contact authors for unpublished results. Although we planned to conduct subgroup analyses by patient population and intervention type, this was not possible due to small sample sizes. While GAS may be differentially responsive based on disease type and intervention type, this review was unable to explore these potential differences.

Future Directions

While the results of this review suggest that GAS may be suitable for use in an older adult population with neurodegenerative disease experiencing dementia or cognitive impairment, GAS may be differentially responsive depending on disease type or dementia severity. As stated above, subgroup analyses based on disease type or dementia severity could not be conducted for this review due to small sample sizes. Similarly, GAS may be differentially responsive based on intervention type (e.g., speech-language therapy, occupational therapy, cognitive rehabilitation) or intervention target (e.g., cognition, communication, ADLs, QOL). Future studies could use less specific inclusion criteria and perform subgroup analyses to explore differential responsiveness to dementia severity, disease type, or intervention modality.

GAS may have therapeutic effects independent of intervention effects [21, 54]. For example, Herdman et al. [21] found that clinically meaningful gains were found in both intervention

and non-intervention groups using GAS in a memory and aging program. The act of systematically setting and scaling goals may provide therapeutic benefits in the absence of another intervention. As such, GAS may by nature have a therapeutic effect and be suitable for measuring functional change for prespecified goals. Studies using GAS to evaluate treatment outcomes could include control groups that do not receive intervention but still complete GAS to parse out the therapeutic effects attributable to GAS and the therapeutic effects attributable to the intervention. The studies included in this analysis did not all include GAS scores for control groups; as such, no comparison between GAS scores in intervention and control groups could be made. Understanding the extent to which goal attainment is achieved through goal setting without specific treatment is essential to evaluating the efficacy of interventions.

Conclusion

Overall, results of this review suggest that GAS may be suitable for use in older adult populations with neurodegenerative disease experiencing dementia or cognitive impairment, based on responsiveness. Risk of bias and certainty of evidence for this review were rated as moderate, suggesting that while there is some bias in the included studies, the observed effect is likely representative of the true effect in this population. Specifically, this review provides evidence that GAS can measure functional change related to cognition, communication, ADLs, and QOL in this population. Further research is needed to understand psychometric properties of GAS in this population and to understand if GAS is differentially responsive based on disease type, dementia severity, or intervention modality. Lastly, further research is needed to better understand the therapeutic effects of GAS in the absence of intervention.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We would like to thank Annie Wescott for her expertise and guidance in developing search terms for this review. Additionally, we would like to thank Daniel Gutstein for his assistance in making figures for this review.

Funding Sources

This project was supported, in part, by the National Institute on Aging of the National Institute of Health under grant numbers R01AG055425 and R56AG055425 and by the National Institutes of Health's National Center for Advancing Translational Sciences, grant number UL1TR001422. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Data Availability Statement

All data generated or analyzed during this study are included in this article and its supplementary material files. Further inquiries can be directed to the corresponding author. The protocol for this systematic review was preregistered on PROSPERO and can be accessed at: https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42021276371.

References

1. Kiresuk TJ, Sherman RE. Goal attainment scaling: a general method for evaluating comprehensive community mental health programs. Community Ment Health J. 1968;4(6):443–53. [PubMed: 24185570]
2. Kiresuk TJ, Lund SH, Larsen NE. Measurement of goal attainment in clinical and health care programs. Drug Intell Clin Pharm. 1982;16(2):145–53. [PubMed: 7075465]
3. Kiresuk TJ. Goal attainment scaling at a county mental health service. Eval Int J Theor Res Pract. 1973;Mo 1:12–8.
4. Roberts AC, Rademaker AW, Salley EA, Mooney A, Morhardt D, Fried-Oken M, et al. Communication Bridge™−2 (CB2): an NIH Stage 2 randomized control trial of a speech-language intervention for communication impairments in individuals with mild to moderate primary progressive aphasia. Trials. 2022 Jun 13;23(1):487. [PubMed: 35698099]
5. Kiresuk T, Lund S. Goal attainment scaling: a medical-correctional application. Med L. 1982;1:227.
6. Shabbir SH, Sanders AE. Clinical significance in dementia research: a review of the literature. Am J Alzheimers Dis Other Demen. 2014;29(6):492–7. [PubMed: 24526758]
7. Rockwood K, Stanley J, Dunn T, Howlett SE. Goal attainment scaling scores, without defined attainment levels, were associated with standardized measures in people with vascular and mixed dementia. J Prev Alzheimer's Dis. 2018;5(1):S158–59.
8. Jogie P, Rahja M, van den Berg M, Cations M, Brown S, Laver K. Goal setting for people with mild cognitive impairment or dementia in rehabilitation: a scoping review. Aust Occup Ther J. 2021;68(6):563–92. [PubMed: 34346077]
9. Schlosser RW. Goal attainment scaling as a clinical measurement technique in communication disorders: a critical review. J Commun Disord. 2004;37(3):217–39. [PubMed: 15063144]
10. Shogren KA, Burke KM, Antosh A, Wehmeyer ML, LaPlante T, Shaw LA, et al. Impact of the Self-Determined Learning Model of Instruction on self-determination and goal attainment in adolescents with intellectual disability. J Disabil Pol Stud. 2019;30(1):22–34.
11. Palisano RJ. Validity of goal attainment scaling in infants with motor delays. Phys Ther. 1993;73(10):651–8; discussion 658–60. [PubMed: 7690976]
12. Grant M, Ponsford J. Goal Attainment Scaling in brain injury rehabilitation: strengths, limitations and recommendations for future applications. Neuropsychol Rehabil. 2014;24(5):661–77. [PubMed: 24787703]
13. Rockwood K, Graham JE, Fay S; ACADIE Investigators. Goal setting and attainment in Alzheimer's disease patients treated with donepezil. J Neurol Neurosurg Psychiatry. 2002;73(5):500–7. [PubMed: 12397141]
14. Rockwood K, Fay S, Song X, MacKnight C, Gorman M; Video-Imaging Synthesis of Treating Alzheimer's Disease VISTA Investigators. Attainment of treatment goals by people with Alzheimer's disease receiving galantamine: a randomized controlled trial. CMAJ. 2006;174(8):1099–105. [PubMed: 16554498]
15. Vu M, Law AV. Goal-attainment scaling: a review and applications to pharmacy practice. Res Soc Adm Pharm. 2012;8(2):102–21.
16. Gaasterland CMW, Jansen-van der Weide MC, Weinreich SS, van der Lee JH. A systematic review to investigate the measurement properties of goal attainment scaling, towards use in drug trials. BMC Med Res Methodol. 2016;16(1):99–. [PubMed: 27534620]
17. McGarrigle L, Rockwood K. The responsiveness of goal attainment scaling using just one goal in controlled clinical trials: an exploratory analysis. J Patient Rep Outcomes. 2020;4(1):35–10. [PubMed: 32399731]
18. Weidenbohm K, Parsons M, Dixon R. Goal attainment scaling: a tool for use in evaluation of services for older people. N Z J Occup Ther. 2005;52(2):9–14.
19. Stolee P, Stadnyk K, Myers AM, Rockwood K. An individualized approach to outcome measurement in geriatric rehabilitation. J Gerontol A Biol Sci Med Sci. 1999 Dec;54(12):M641–7. [PubMed: 10647971]
20. van Seben R, Reichardt L, Smorenburg S, Buurman B. Goal-setting instruments in geriatric rehabilitation: a systematic review. J Frailty Aging. 2017;6(1):37–45. [PubMed: 28244557]

21. Herdman KA, Vandermorris S, Davidson S, Au A, Troyer AK. Comparable achievement of client-identified, self-rated goals in intervention and no-intervention groups: reevaluating the use of Goal Attainment Scaling as an outcome measure. Neuropsychol Rehabil. 2019;29(10):1600–10. [PubMed: 29430998]
22. Stolee P, Awad M, Byrne K, DeForge R, Clements S, Glenny C, et al. A multi-site study of the feasibility and clinical utility of Goal Attainment Scaling in geriatric day hospitals. Disabil Rehabil. 2012;34(20):1716–26. [PubMed: 22397694]
23. Bouwens SFM, Van Heugten CM, Verhey FRJ. Review of goal attainment scaling as a useful outcome measure in psychogeriatric patients with cognitive disorders. Dement Geriatr Cogn Disord. 2008;26(6):528–40. [PubMed: 19052453]
24. Folstein M, Anthony JC, Parhad I, Duffy B, Gruenberg EM. The meaning of cognitive impairment in the elderly. J Am Geriatr Soc. 1985;33(4):228–35. [PubMed: 3989183]
25. Gauthier S, Reisberg B, Zaudig M, Petersen RC, Ritchie K, Broich K, et al. Mild cognitive impairment. Lancet. 2006;367(9518):1262–70. [PubMed: 16631882]
26. Hardy J, Gwinn-Hardy K. Genetic classification of primary neurodegenerative disease. Science. 1998;282(5391):1075–9. [PubMed: 9804538]
27. Bertram L, Tanzi RE. The genetic epidemiology of neurodegenerative disease. J Clin Invest. 2005;115(6):1449–57. [PubMed: 15931380]
28. Dugger BN, Dickson DW. Pathology of neurodegenerative diseases. Cold Spring Harb Perspect Biol. 2017;9(7):a028035. [PubMed: 28062563]
29. Husted JA, Cook RJ, Farewell VT, Gladman DD. Methods for assessing responsiveness: a critical review and recommendations. J Clin Epidemiol. 2000 May;53(5):459–68. [PubMed: 10812317]
30. Terwee CB, Dekker FW, Wiersinga WM, Prummel MF, Bossuyt PMM. On assessing responsiveness of health-related quality of life instruments: guidelines for instrument evaluation. Qual Life Res. 2003 Jun;12(4):349–62. [PubMed: 12797708]
31. Crosby RD, Kolotkin RL, Williams GR. Defining clinically meaningful change in health-related quality of life. J Clin Epidemiol. 2003;56(5):395–407. [PubMed: 12812812]
32. Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. A basic introduction to fixed-effect and random-effects models for meta-analysis. Res Synth Methods. 2010;1(2):97–111. [PubMed: 26061376]
33. Krasny-Pacini A, Evans J, Sohlberg MM, Chevignard M. Proposed criteria for appraising goal attainment scales used as outcome measures in rehabilitation research. Arch Phys Med Rehabil. 2016;97(1):157–70. [PubMed: 26343173]
34. 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 Aug;56(8):736–43. [PubMed: 12954465]
35. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372:n71. [PubMed: 33782057]
36. Fegter O, Rogalski E, Roberts A. Responsiveness of goal attainment scaling in persons with neurodegenerative diseases with dementia or cognitive impairment. CRD42021276371 ed. PROSPERO2021.
37. Bramer WM, de Jonge GB, Rethlefsen ML, Mast F, Kleijnen J. A systematic approach to searching: an efficient and complete method to develop literature searches. J Med Libr Assoc. 2018;106(4):531–41. [PubMed: 30271302]
38. Boutron I, Page MJ, Higgins JP, Altman DG, Lundh A, Hróbjartsson A. Considering bias and conflicts of interest among the included studies. Cochrane Handb Syst Rev Interv. 2019:177–204.
39. Ma LL, Wang YY, Yang ZH, Huang D, Weng H, Zeng XT. Methodological quality (risk of bias) assessment tools for primary and secondary medical studies: what are they and which is better? Mil Med Res. 2020;7(1):7–11. [PubMed: 32111253]
40. National Heart L, Institute B. Study quality assessment tools. 2019. https://www.nhlbi.nih.gov/health-topics/study-quality-assessment-tools (accessed 2019).

41. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ. 2008;336(7650):924–6. [PubMed: 18436948]
42. Kazdin AE. The meanings and measurement of clinical significance. 1999.
43. Review Manager (RevMan). 5.4 ed. The Cochrane Collaboration; 2020.
44. Chew J, Chong MS, Fong YL, Tay L. Outcomes of a multimodal cognitive and physical rehabilitation program for persons with mild dementia and their caregivers: a goal-oriented approach. Clin Interv Aging. 2015;10:1687–94. [PubMed: 26543358]
45. Reilly S, Hynes SM. A cognitive occupation-based programme for people with multiple sclerosis: a study to test feasibility and clinical outcomes. Occup Ther Int. 2018;2018:1614901. [PubMed: 29853813]
46. Spencer KA, Paul J, Brown KA, Ellerbrock T, Sohlberg MM. Cognitive rehabilitation for individuals with Parkinson's disease: developing and piloting an external aids treatment program. Am J Speech Lang Pathol. 2020;29(1):1–19. [PubMed: 31697899]
47. Chester H, Beresford R, Clarkson P, Entwistle C, Gillan V, Hughes J, et al. The Dementia Early Stage Cognitive Aids New Trial (DESCANT) intervention: a goal attainment scaling approach to promote self-management. Int J Geriatr Psychiatry. 2021;36(5):784–93. [PubMed: 33271639]
48. Hanssen KT, Saltytė Benth J, Beiske AG, Landrø NI, Hessen E. Goal attainment in cognitive rehabilitation in MS patients. Neuropsychol Rehabil. 2015;25(1):137–54. [PubMed: 25356835]
49. O'Sullivan M, Coen R, O'Hora D, Shiel A. Cognitive rehabilitation for mild cognitive impairment: developing and piloting an intervention. Neuropsychol Dev Cogn B Aging Neuropsychol Cogn. 2015;22(3):280–300. [PubMed: 24955493]
50. Volkmer A Better Conversations with Primary Progressive Aphasia (BCPPA): developing and piloting a communication partner training intervention for people with primary progressive aphasia and their significant others. UCL (University College London); 2020.
51. Rannisto M, Rosti-Otajärvi E, Mäntynen A, Koivisto K, Huhtala H, Hämäläinen P. The use of goal attainment scaling in neuropsychological rehabilitation in multiple sclerosis. Disabil Rehabil. 2015;37(21):1984–91. [PubMed: 25495288]
52. Viswanathan M, Patnode CD, Berkman ND, Bass EB, Chang S, Hartling L, et al. Assessing the risk of bias in systematic reviews of health care interventions. Methods guide for effectiveness and comparative effectiveness reviews [Internet]. 2017.
53. Kiresuk TJ, Stelmachers ZT, Schultz SK. Quality assurance and goal attainment scaling. Prof Psychol. 1982;13(1):145–52.
54. Haladay D, Swisher L, Hardwick D. Goal attainment scaling for patients with low back pain in rehabilitation: a systematic review. Health Sci Rep. 2021 Sep;4(3):e378. [PubMed: 34589616]

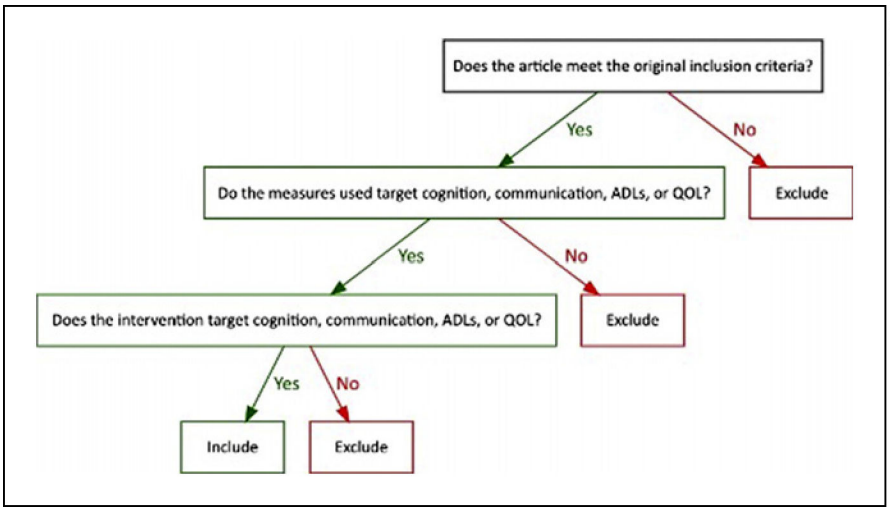

Fig. 1.
Additional inclusion criteria decision tree.

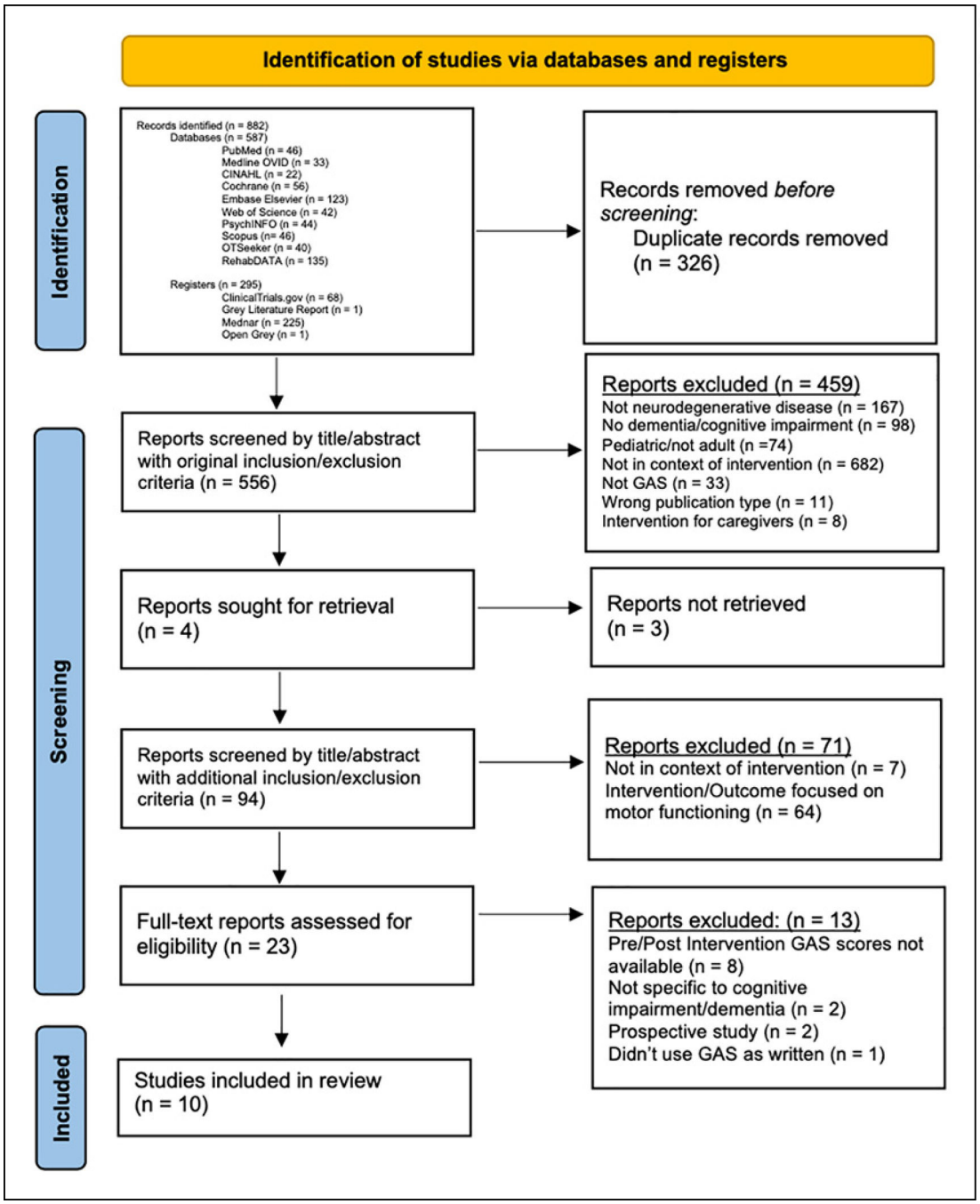

Fig. 2.
PRISMA 2020 flow diagram for new systematic reviews.

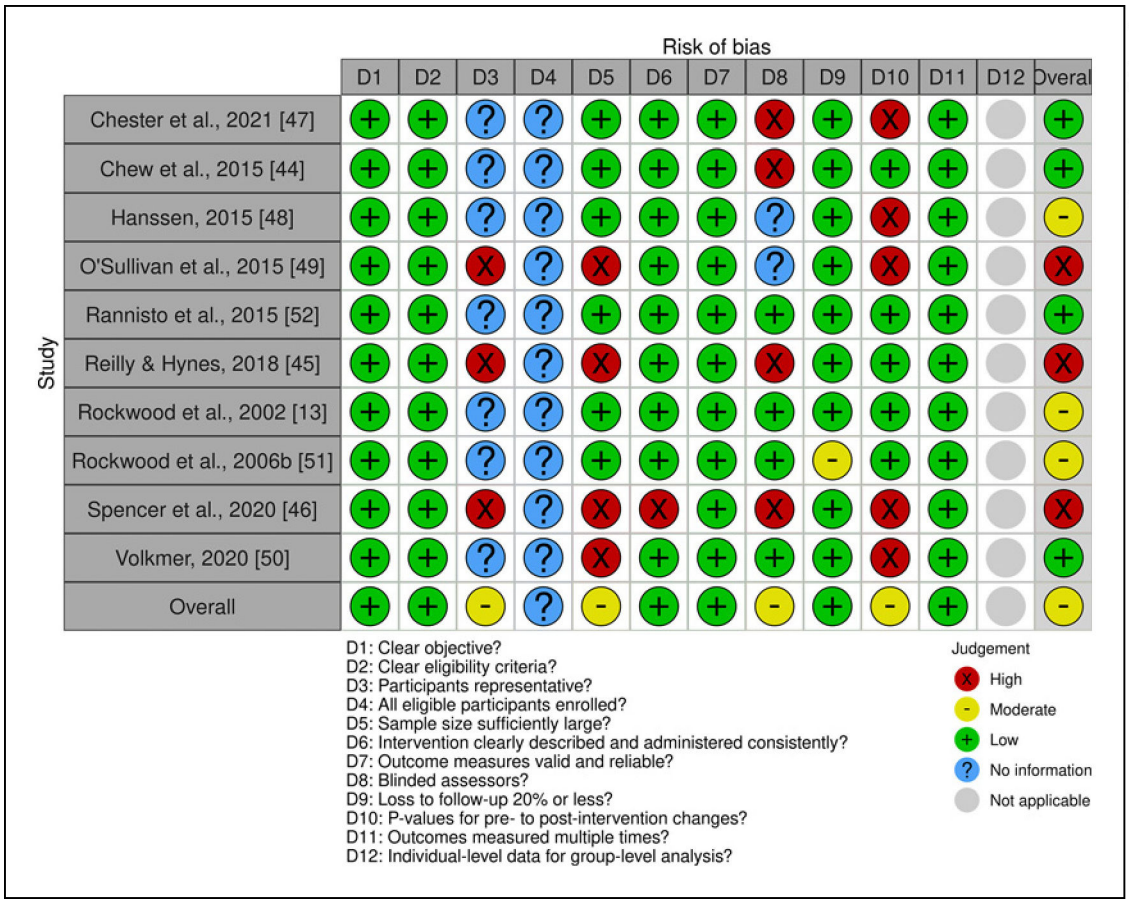

Fig. 3.

Stoplight plot of risk of bias for the 10 pre-post intervention studies included using the NIH Quality Assessment Tool. D1–D10 indicate each of the questions in the NIH Quality Assessment Tool. The final column represents overall risk of bias for each included study. The final row represents overall risk of bias for each domain. Domain 12 is related to group-level analyses and was not applicable to any of the included studies.

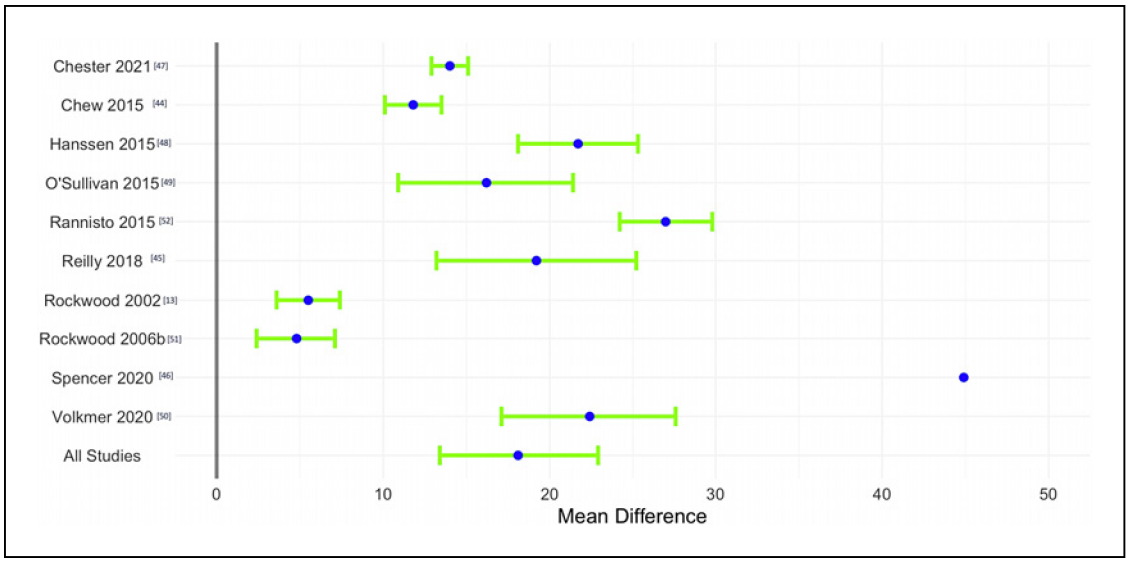

Fig. 4.

Random-effects meta-analysis indicates responsiveness for pre-post intervention GAS T-scores. Blue dots represent the average change in GAS scores for each study. The blue diamond represents the summary responsiveness measure or average change in GAS scores across studies. Values to the right of zero represent better than expected levels of goal attainment.

Table 1.

Study characteristics

Study	Sample and setting	Intervention	Intervention target	Intervention length	Time between pre-post GAS scores	Number of goals	People involved in goal setting	Psychometric properties
Chester et al. (2021) [47]	111 patients with mild-moderate dementia in healthcare centers	Using external memory aids	QOL, cognition	4 weeks	4 weeks	293 total	Participant, partner, and researcher	Not provided
Chew et al. (2015) [44]	44 patients with mild dementia in an outpatient geriatric clinic	Rehabilitation program through group therapy	Cognition	8 weeks	8 weeks	M = 2.6 per participant	Participant and caregiver	Not provided
Hanssen et al. (2015) [48]	57 patients with multiple sclerosis in an inpatient rehabilitation program	Inpatient cognitive rehabilitation	ADLs, communication, QOL	16 weeks	28 weeks	148 total (1–4 per participant)	Participant	Not provided
O'Sullivan et al. (2015) [49]	5 patients with MCI in an outpatient memory clinic	Cognitive rehabilitation (psychoeducation, habit training, deep breathing, external aids)	ADLs, cognition	6–8 weeks	8 weeks	3–4 per participant	Participant and caregiver	Inter-rater reliability (kappa = 0.73)
Rannisto et al. (2015) [52]	58 patients with multiple sclerosis in an outpatient unit of a university hospital	Neuropsychological rehabilitation through memory retraining, compensatory strategies, and psychoeducation	Cognition	12 weeks	12 weeks	116 total (1–3 per participant)	Participant and clinician	Not provided
Reilly and Hynes (2018) [45]	12 patients with multiple sclerosis in a university research center	Learning compensatory strategies to manage demands of daily life	ADLs, cognition	9 weeks	18 weeks	Not provided	Not provided	Not provided
Rockwood et al. (2002) [13]	88 patients with mild-moderate dementia in healthcare clinics	Donepezil hydrochloride	ADLs, cognition	52 weeks	52 weeks	855 total (M = 9, SD = 3)	Participant, caregiver, and clinician	Not provided
Rockwood et al. (2006) [51]	64 patients with probable AD in research facilities	Galantamine	Cognition	16 weeks	16 weeks	377 total	Participant, caregiver, and clinician	Not provided
Spencer et al. (2020) [46]	3 patients with Parkinson's disease	Individualized external aids	ADLs, cognition	8 weeks	22 weeks	3–6 per participant	Participant and researcher	Not provided
Volkmer (2020) [50]	9 patients with mild-moderate PPA in healthcare centers	Speech-language therapy	Communication	Not provided	7 weeks	30 total	Participant, partner, and researcher	Not provided