UvA-DARE (Digital Academic Repository)

The impact of acute hospitalization on older persons
Experiences, outcomes and improvements
van Seben, R.

Link to publication

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (http://dare.uva.nl)
Chapter 7

Goal-setting instruments in geriatric rehabilitation: a systematic review

Rosanne van Seben
Lucienne A. Reichardt
Susanne M. Smorenburg
Bianca M. Buurman

J Frailty Aging 2017;6(1):37-45
Abstract
This systematic review summarizes the psychometric properties of goal-setting instruments that are applied within geriatric rehabilitation. PubMed Medline and Embase were systematically searched for eligible articles. Studies were included if they were conducted in a somatic or neurological rehabilitation setting, included patients aged ≥55 years and provided data on instruments’ psychometric properties (validity, reliability, responsiveness), utility and/or feasibility. Eleven studies were included. Seven studies, all conducted by one research group, evaluated Goal-Attainment Scaling (GAS); two studies assessed the Canadian Occupational Performance Measure (COPM) and one study the Self-Identified Goals Assessment (SIGA), which is based on the COPM. One study assessed a core set of the International Classification of Functioning, Disability and Health (ICF) framework. High concurrent, content and predictive validity and inter-rater reliability were found for GAS. Responsiveness appears to be excellent. Concurrent validity and inter-rater reliability of the COPM and content validity of both the COPM and SIGA appear to be good. Responsiveness of both instruments seems to be poor. Content validity of the ICF core set was found to be fair; responsiveness appears to be very poor. There is little published data on goal-setting instruments in geriatric rehabilitation. Evidence for its psychometric properties may support GAS as goal-setting instrument and additional outcome measure. However, more research is required in order to evaluate GAS, as research conducted in other health care settings may provide important additional findings. Before the COPM (or SIGA) can be recommended as goal-setting instrument, its psychometric properties require further research.
Introduction

Nearly a quarter of hospitalized patients aged 65 and over experience functional decline, defined as a loss of independence in Activities of Daily Living (ADL). ADLs reflect self-care activities that are crucial in order to live independently and a large number of older patients cannot directly return home after hospital discharge. Geriatric rehabilitation targets patients with new disabilities and aims to restore the patient’s pre-hospital level of functioning. Geriatric rehabilitation differs from general rehabilitation in many aspects, due to the complex health status of a geriatric patient. Older patients often suffer from co-morbid chronic medical conditions, are frequently subject to polypharmacy and may encounter premorbid disability in Instrumental ADLs and mobility. They often have low functional reserves and recovery to baseline functioning is arduous for this patient group.

Goal setting together with the patient is an essential aspect of geriatric rehabilitation. Based on their health status and individual lifestyle and environment, older individuals have their own desired outcomes of the rehabilitation process.6,12 Standardized instruments such as the Katz and Barthel ADL scale, which only assess functional status based on a standard set of activities, may overlook the individual wishes of the patient. In addition, goals help to establish an efficient rehabilitation program and a positive effect of goal setting on patient participation, patient satisfaction and the patient’s functional level has been reported.

Despite the positive effects described in the literature, goal setting is no daily practice in geriatric rehabilitation. Besides, goal setting can be challenging, in particular with geriatric patients. Therefore, in order to define and agree on realistic and patient-important goals at start of the rehabilitation process, a formal goal-setting instrument is required. Several approaches and instruments exist, ranging from simple to complex. Within the field of geriatric rehabilitation, goal-setting instruments are adopted from other clinical settings, including the Canadian Occupational Performance Measure (COPM) and Goal-Attainment Scaling (GAS).

Several reviews have been published that evaluated goal setting within a general rehabilitation context. For example, Rosewilliam and colleagues evaluate the application of goal setting in stroke rehabilitation. Although they found support within existing literature for the use of formalized instruments, they provide no further evaluation of the psychometric properties of these instruments. Other reviews supported the use of either GAS or the COPM but did not compare different instruments within the same review. In their review published more than a decade ago, Hurn and co-writers assessed goal-setting instruments and the psychometric properties within a physical or neurological rehabilitation setting for adults and older individuals (≥16) and recommend GAS in a physical rehabilitation setting.

To the best of our knowledge no systematic review exists that assesses more than one goal-setting instrument within the field of geriatric rehabilitation. Furthermore, as more emphasis has been placed on patient-centeredness and individualized approaches in the past five years, there is a need to update previous studies. The objective of this systematic review was to assess the psychometric properties (validity, reliability and responsiveness) of goal-setting instruments that
are applied within the field of geriatric rehabilitation.

**Methods**

**Search Strategy**

Databases PubMed Medline and Embase were systematically searched from January 1980 up to October 8th, 2015 to identify English articles on individualized goal-setting instruments within the field of geriatric rehabilitation. With the help from a clinical librarian a search was set up (Appendix 1). The search strategy was performed by one of the authors (R.S.). Based on title, a first selection was made by one investigator (R.S.). Potential eligible studies were included for the second phase of the screening procedure. After this first selection, two authors (R.S. and L.R.) screened titles and abstracts to decide whether full text review was required. Discrepancies between both investigators were resolved through discussion or by consulting a third reviewer (B.B.).

**Study Selection; inclusion and exclusion criteria**

We included studies that met the following inclusion criteria: 1) the study was conducted in a rehabilitation context according to the definition of the World Health Organization: “a process aimed at enabling people with disabilities to reach and maintain their optimal physical, sensory, intellectual, psychological and social functional levels. Rehabilitation provides disabled people with the tools they need to attain independence and self-determination”; 2) the study included patients aged 55 years and over, as these patients are eligible for geriatric rehabilitation; 3) the psychometric properties, utility and/or feasibility of the goal-setting instrument were assessed; 4) the study was conducted in a somatic or neurological rehabilitation settings; studies that focused on rehabilitation for patients suffering from mental diseases, cognitive disorders or chronic diseases were excluded; 5) All types of studies were included, except for reviews, case reports and practice-based guidelines.

**Data extraction**

Two researchers independently extracted data (R.S. and L.R.). A data extraction form was used to extract data on: study characteristics (authors, publication year, country, type of study, care setting, study population, and outcome measures); patient characteristics (mean age, rehabilitation setting); and description of the psychometric properties of the applied instrument. A quality assessment was performed, using the QUality Assessment of Diagnostic Accuracy Studies (QUADAS-2) that helps to assess the quality of test accuracy studies and assists in rating bias and applicability. Risk of bias consists of four areas including patient selection, index test, reference standard, and flow and timing. Applicability consists of four key areas: patient selection, index test, and reference standard. Both risk of bias and applicability are classified as unclear, low or high risk.

**Psychometric properties: definitions**

The goal-setting instruments were assessed in terms of their psychometric properties, including validity, reliability, and responsiveness. Validity is the extent to which an instrument measures what it is supposed to measure, which is further categorized into content, concurrent and predictive validity. Content
validity assesses the extent to which an instrument provides an adequate and representative sample of items. Concurrent validity demonstrates whether test scores correlate with validated measures of the same construct, i.e. whether the goal-setting instrument correlates well with instruments that measure, for example, functioning level. Predictive validity refers to the ability to predict what it theoretically should predict, i.e. whether feasible goals are set with the goal-setting instrument. Reliability assesses the stability of measures, i.e. consistency of measures. As geriatric rehabilitation aims to improve functioning of the older patient, responsiveness was also taken into account, reflecting the ability of an instrument to detect change over time.\textsuperscript{29} Feasibility was assessed, as instruments should be applicable in daily practice, within an appropriate timeframe.\textsuperscript{12} Also, as geriatric rehabilitation requires a multidisciplinary approach\textsuperscript{6}, it was taken into account whether the instrument is applicable in a multidisciplinary setting and can be performed by different professionals. Lastly, utility was considered to determine whether an instrument is useful, provides additional information and adds value in the geriatric rehabilitation process.

**Results**

**Search results**

After removing duplicates, the initial combined search retrieved 4711 journal articles, of which 4570 were screened as irrelevant by title (Figure 1). Two reviewers (R.S. and L.R.) independently screened 141 articles on title and abstract. 80 full text articles were excluded (N = 69): No somatic or neurological geriatric rehabilitation: 46 Goal-setting instruments’ psychometric properties, feasibility or utility not assessed: 20 Wrong language: 3

---

**Figure 1.** Flow diagram of search strategy and study selection
texts were then assessed, of which 69 articles were excluded: 46 studies were not conducted in a somatic or neurological geriatric rehabilitation setting; 20 studies did not assess psychometric properties, feasibility and/or utility of goal-setting instruments; and three studies were not written in English. Eleven articles met the selection criteria for this review.

**Study characteristics and risk of bias**
Table 1 provides an overview of the main characteristics of the eleven studies that were included for analysis. Four studies were prospective descriptive studies,30-33 one study was retrospective descriptive study,34 one study was a qualitative study35 and one study36 used a sub-analysis of a randomized control trial. Three studies were cohort studies37-39 and one study40 analyzed a subset of a larger cohort study. The number of included patients varied from 15 to 198 and weighted mean age across the studies was 80.33 (range 73.4-82.3). Seven studies were conducted in geriatric rehabilitation centers or geriatric restorative care services30,32-34,37,38,40; three studies were conducted in a hospital setting31,35,39 and in one study the goal-setting instrument was applied by a mobile geriatric assessment team36 in a community-dwelling older population. Except for the qualitative study,35 a quality assessment (QUADAS-2) was performed for all studies.28 In Appendix 2 a schematic overview is provided, displaying risk of bias for four domains, i.e. patient selection, index test, reference test and time and flow. The greatest risk of bias was most often associated with the reference standard, which was actually mostly due to absence of a reference standard. Two studies had issues with the index texts as no threshold scores were identified.

**Goal-setting instruments**
Seven studies evaluated GAS, which were all conducted by one research group, Rockwood and colleagues in Canada.30-36 Two studies examined the COPM.37,40 Based on the COPM, Melville and colleagues39 developed a less time-consuming, yet quite similar, alternative: the Self-Identified Goals Assessment (SIGA). Kus and colleagues38 assessed the use of a core set of the International Classification of Functioning, Disability and Health (ICF) framework for goal-setting in geriatric rehabilitation. A procedure description of the instruments is provided in Table 2.

**Validity**
Three studies assessed predictive validity of GAS and it was found that feasible goals were set as outcome scores were close to 50, which is the mean goal-attainment T-score31,32,34 (Table 3). There appears to be good evidence for content validity as appropriate goals were set, which reflect recommended problem areas for older rehabilitants.32-34 Yip et al.34 indicated that 40 percent of the goals consisted of items that were not addressed by standardized instruments such as the Katz and Barthel and GAS provided additional relevant outcome measures. Significant correlations were found between GAS scores and those of standardized instruments,32-34 which indicate good concurrent validity. With regard to the SIGA, Melville and colleagues39 found that meaningful goals were set of which basic self-care goals were most common. Two studies found that goals were commonly set within the areas self-care, leisure and productivity when the COPM is performed.37,40 Edwards et al40 found correlations between the COPM
<table>
<thead>
<tr>
<th>Study</th>
<th>Year</th>
<th>Country</th>
<th>Tool</th>
<th>Design</th>
<th>Care Setting</th>
<th>N</th>
<th>Mean age</th>
<th>Psychometric domains</th>
<th>Personnel</th>
<th>Outcome(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rockwood</td>
<td>1993</td>
<td>Canada</td>
<td>GAS</td>
<td>Prosp.</td>
<td>Geriatric restorative care</td>
<td>45</td>
<td>81</td>
<td>Reliability, Responsiveness</td>
<td>Multidisciplinary setting</td>
<td>Barthe l, Katz, FIM, PSMS, SQLI</td>
</tr>
<tr>
<td>Stolee</td>
<td>2012</td>
<td>Canada</td>
<td>GAS</td>
<td>Prosp.</td>
<td>Geriatric day hospital</td>
<td>90</td>
<td>76.2</td>
<td>Predictive validity, Responsiveness</td>
<td>NR</td>
<td></td>
</tr>
<tr>
<td>Stolee</td>
<td>1992</td>
<td>Canada</td>
<td>GAS</td>
<td>Prosp.</td>
<td>Geriatric assessment unit and restorative care</td>
<td>15</td>
<td>79</td>
<td>Content, concurrent, predictive validity Reliability</td>
<td>Multidisciplinary setting: geriatrician, nurse</td>
<td></td>
</tr>
<tr>
<td>Stolee</td>
<td>1999</td>
<td>Canada</td>
<td>GAS</td>
<td>Prosp.</td>
<td>Geriatric rehabilitation</td>
<td>173</td>
<td>81</td>
<td>Content, concurrent, predictive validity Reliability</td>
<td>Multidisciplinary setting</td>
<td>Barthe l, OARS-IADL, MMSE, NHP Self-rated health</td>
</tr>
<tr>
<td>Yip</td>
<td>1998</td>
<td>Canada</td>
<td>GAS</td>
<td>Retrosp.</td>
<td>Geriatric rehabilitation unit</td>
<td>143</td>
<td>77</td>
<td>Content, concurrent, predictive validity Reliability</td>
<td>Multidisciplinary setting</td>
<td>Barthe l, Katz, MMSE, OARS-IADL</td>
</tr>
<tr>
<td>Stolee</td>
<td>1999</td>
<td>Canada</td>
<td>GAS</td>
<td>Qualitative</td>
<td>Hospitals or community-based geriatric services</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Multidisciplinary setting</td>
<td>NA</td>
</tr>
<tr>
<td>Rockwood</td>
<td>2003</td>
<td>Canada</td>
<td>GAS</td>
<td>Sub-analysis of RCT</td>
<td>Interdisciplinary Mobile Geriatric Assessment Team</td>
<td>165</td>
<td>81.8</td>
<td>Responsiveness</td>
<td>NR</td>
<td>Barthe l ADL/ IADL, PSMS, MMSE, GSS, SQLI, GIC</td>
</tr>
</tbody>
</table>
### Table 1. Continued

<table>
<thead>
<tr>
<th>Study</th>
<th>Year</th>
<th>Country</th>
<th>Tool</th>
<th>Design</th>
<th>Care Setting</th>
<th>N</th>
<th>Mean age</th>
<th>Psychometric domains</th>
<th>Personnel</th>
<th>Outcome(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enemark Larsen(^37)</td>
<td>2012</td>
<td>Denmark</td>
<td>COPM</td>
<td>Cohort</td>
<td>Geriatric rehabilitation</td>
<td>185</td>
<td>Pre: 185 Post: 95</td>
<td>Content validity Reliability Responsiveness</td>
<td>Occupational therapist Physiotherapists</td>
<td>NA</td>
</tr>
<tr>
<td>Melville(^39)</td>
<td>2002</td>
<td>United States</td>
<td>SIGA</td>
<td>Cohort, Qualitative</td>
<td>Hospital based Sub-acute facility</td>
<td>30</td>
<td>73.4</td>
<td>Content validity Responsiveness</td>
<td>Occupational therapist</td>
<td>NA</td>
</tr>
<tr>
<td>Edwards(^40)</td>
<td>2007</td>
<td>Canada</td>
<td>COPM</td>
<td>Cohort of Cohort</td>
<td>In-patient rehabilitation (hip fracture)</td>
<td>198</td>
<td>82.3</td>
<td>Content, concurrent validity Responsiveness</td>
<td>Occupational Therapist</td>
<td>WOMAC, ADL/IADL</td>
</tr>
<tr>
<td>Kus(^38)</td>
<td>2011</td>
<td>Germany &amp; Austria</td>
<td>ICF</td>
<td>Cohort</td>
<td>Rehabilitation</td>
<td>209</td>
<td>80</td>
<td>Validity</td>
<td>Researchers</td>
<td>Length of stay Overall functioning</td>
</tr>
</tbody>
</table>

Note: FIM: Functional Independence Measure; PSMS: Physical Self-Maintenance Scale; SQLI: Spitzer Quality of Life Index; OARS: Older Americans Resource and Services Questionnaire; MMSE: Mini-Mental State Examination; NHP: Nottingham Health Profile; GSS: Geriatric Status Scale; GIC: global impression of change; WOMAC: Western Ontario and McMaster Universities Osteoarthritis Index; RCT: Randomized control trial; N.R.: not reported; N.A. not applicable; Prosp.: Prospective; Retrosp.: Retrospective
Table 2. Description of the goal-setting instruments included in the review

<table>
<thead>
<tr>
<th>Goal-setting instrument procedure</th>
<th>Description</th>
</tr>
</thead>
</table>
| **GAS** | Goal Attainment Scaling (GAS) was introduced in the 1960s by Kiresuk and Sherman, in order to assess outcomes in mental health care settings. GAS has been used in many other clinical settings since, including rehabilitation settings. (12) First, an interview is conducted with the patient in order to define the main problem areas and to establish three to five goals. In step two goals are weighted in terms of importance and difficulty, which are both rated from 0 (not important/difficult) to 3 (very important/difficult). These scores are multiplied to determine the weight. In step three the expected outcome is defined, using a five-point scale for each goal: -2 is the pre-treatment level, -1 indicates some progression towards the goal without achieving it, 0 is the expected level after treatment, +1 represents a better outcome than expected, and +2 is the best possible outcome for a certain goal. During step 5 goal attainment is scored. The outcome score for each goal is rated at the appointed review date, judging actual performance against the predefined levels. The goal-attainment T-score is then calculated by applying the formula: 

$$t = \frac{1}{10} \sum X_1$$

where: $X_1$ = attainment score, $n$ = number of scales, $p = 0.3$ expected inter-correlation among goals. A mean goal attainment T-score is approximately 50.12,23 |
| **COPM** | The Canadian Occupational Performance Measure (COPM) was developed as a tool in order to set occupational performance goals. These goals are based on the perception of the client and used in order to measure progress objectively in the problem areas defined by the client. In the first step, it is determined whether the client has any problems. Then the patient scores the problem in terms of importance on a 0-10 scale, which is continued by determining satisfaction with performance of the five most important problems, using a scoring-rate between 1 to 100. In the re-assessment step, the client is asked to score satisfaction with the performance of the activities and progress can be detected. At follow-up, the COPM is again applied in order to determine whether the problems are still present or new ones have developed.21 |
| **SIGA** | The Self-Identified Goals Assessment (SIGA) was developed for the same purpose as the COPM and has a quite similar goal-setting procedure but it was aimed to time-related problems with the COPM. After asking about prior functioning, home situation, and interests, the therapist asks the patient to identify tasks would like to improve. Then the patient is asked: how well can you do all of the things you want to do on a scale from 0 to 10? In contrast to the COPM, nt is thus asked to rate how well he or she is able to do all the things he or she wants to do. Optionally, the therapist asks the patient to rate each identified e same 0-10 scale. At follow-up the therapists shows the patient’s latest rating and asks whether the patient wants to add new goals or to change goals.39 |
| **ICF** | Kus, Muller conducted semi-structured interviews with the patient and patients’ goals were translated to the International Classification of Functioning, Disability and Health (ICF) and clustered into different categories. The ICF framework describes human functioning as a result of the dynamic interaction between several health conditions and contextual factors. The ICF identifies three health and health-related domains, including “body functions and structures”, “activities and participation” and “environmental factors”. In this study a comprehensive ICF core set for geriatric rehabilitation was developed that includes relevant domains and problems older patients encounter in functioning. The core set was used in order to code patients’ goals. |
### Table 3. Main Findings with regard to the psychometric properties of the goal-setting instruments

<table>
<thead>
<tr>
<th>Study</th>
<th>Main findings</th>
<th>Validity</th>
<th>Reliability</th>
<th>Responsiveness</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Content Concurrent Predictive</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rockwood30</td>
<td>Reliability: correlation between scores set by nurse and multidisciplinary team: 0.91. Responsiveness: compared to Barthel, FIM, MMSE, Katz, PSMS, Quality of Life index; GAS is most efficient in the detection of clinical important change.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stolee31</td>
<td>Predictive validity: the average GAS score for the patients were 32.0 (SD: 6.2) and 52.3 (SD: 8.7) at admission and discharge respectively. Responsiveness: GAS was able to detect clinical change (paired t-test t (89)= -17.48, p &lt; 0.001. Standardized Response Measure (SRM) = 1.85 (95% confidence interval 1.50-2.19, and ES =3.27).</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stolee32</td>
<td>Content validity: 12 of 13 goal areas were covered, after comparison literature. Concurrent validity: correlations with global clinical outcome rating and Barthel score: 0.82 and 0.86 respectively. Predictive validity: average discharge scores were 48.6 (SD=8.5) and 46.6 (SD=7.5). Reliability: two geriatricians identified similar outcome scores for 59% goal areas. There was initial agreement and consensus after discussion on 49% and 100% individual scale items respectively. Correlation for discharge and change scores between physician and nurse: 0.88 (discharge); 0.87 (change).</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stolee33</td>
<td>Content validity: appropriate goals were set according to geriatric rehabilitation professionals. Concurrent validity: except for MMSE and 6 items on NHP, strong correlations with standardized instruments for follow-up scores. Strong correlations with the two instruments of which domains are most similar to GAS goals: Barthel: 0.66 (0.54-0.77) and OARS IADL: 0.54 (0.42-0.67). Correlations with Global rating and MMSE: 0.67 (0.48-0.85) and 0.30 (0.14-0.46) respectively. Reliability: correlation between follow-up scores set by nurse and multidisciplinary team: 0.93 (intraclass correlation coefficient). Responsiveness: Effect size, relative efficiency and analysis of variance showed: GAS most responsive in comparison with Barthel, OARS IADL, MMSE, health question and NHP.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 3. Continued

<table>
<thead>
<tr>
<th>Study</th>
<th>Main findings</th>
<th>Validity</th>
<th>Reliability</th>
<th>Responsiveness</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Content</td>
<td>Concurrent</td>
<td>Predictive</td>
</tr>
<tr>
<td>Yip34</td>
<td>Content validity: all recommended areas were addressed. Concurrent validity: correlations between GAS and Barthel, Katz ADL/IADL, OARS were slightly significant. Predictive validity: mean GAS scores for outcome goals and overall goals were 50.6 (SD=10.6) and 50.5 (SD=10.0) respectively. Responsiveness: compared to the Barthel index, OARS-IADL and the (S) MMSE, GAS showed higher SRM. Relative efficiency (RE) for GAS was 3.16 (only one above 1.0, which indicates more efficient measurement of change than Barthel).</td>
<td>++ + +</td>
<td>+ + +</td>
<td></td>
</tr>
<tr>
<td>Stolee35</td>
<td>N.A.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rockwood36</td>
<td>Responsiveness: only GAS demonstrated a statistically significant change, in contrast to Barthel ADL/IADL, PSMS and SQUI. GAS remained the most responsive measure compared to the Barthel, CGA and PSMS for three common problem areas.</td>
<td></td>
<td></td>
<td>++</td>
</tr>
<tr>
<td>GAS Overall scoring</td>
<td>Content validity: goals were most commonly related to personal care, productivity, and leisure. Functional mobility was found to be the most important goal. Reliability: no statistically difference between data collected by occupational therapist and physiotherapist after six months of training. (x²=0.002, p=0.962). No significant difference in occupational performance issues (x²=0.464, p=0.793). No significant difference in score assessment (x²=0.453, p=0.501). Responsiveness: 84% of the participants showed a significant positive change in scores for performance and satisfaction for performance. However, scoring is difficult and baseline scores of 9 and 10 were documented, which are too high to detect improvement and affect to ability of the COPM to document change over time.</td>
<td>+/- + +</td>
<td>+ + +</td>
<td>+ + +</td>
</tr>
<tr>
<td>Enemark Larsen37</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study</td>
<td>Main findings</td>
<td>Validity</td>
<td>Reliability</td>
<td>Responsiveness</td>
</tr>
<tr>
<td>-------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>----------</td>
<td>-------------</td>
<td>----------------</td>
</tr>
<tr>
<td>MeVille39</td>
<td>Content validity: meaningful, specific, personal goals were described by the participants: transfers, bathing, walking to a particular place and toileting. Some goals were vague or concerned aspects such as pain and fatigue. Responsiveness: the mean change from admission to discharge was 3.3 (SD: 2.6). However, problems were reported in terms of quantifying performance.</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Edwards40</td>
<td>Content validity: goals were most commonly related to personal care, productivity, and leisure. Concurrent validity: significant correlation between COPM performance (-0.58 (-0.75, -0.35)) and satisfaction (-0.50 (-0.69, 0.24)) and WOMAC scores. Responsiveness: SRM: 1.77 (95% confidence interval: 1.38-2.35) and 1.89 (95% confidence interval: 1.46-2.63) for the COPM and WOMAC respectively. This is a not significant difference of 0.12 (95% confidence interval: -0.49-0.84), in favor of the WOMAC. The COPM detects change, but it cannot be concluded whether the COPM is more sensitive to change than the WOMAC.</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>COPM/SIGA</td>
<td>Overall scoring</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Kus38</td>
<td>Content validity: 72.7% linked to ICF (Mobility, walking and getting rid of pain) 27% not linked at all (autonomy, returning home, improving general health condition). Of those goals linked to ICF, 90% linked to core set (domestic life, recreation and leisure). Responsiveness: the ICF core set only has the potential to detect major changes: no scoring process is used and it can only be determined whether goals are attained</td>
<td>+/-</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

Rel. = reliability: the consistency of a measure; Content validity: the extent to which an instrument provides an adequate and representative sample of items; Concurr. = concurrent validity, demonstrates whether a test scores correlate well with validated measures of the same construct; Predict. = predictive validity: the ability of an instrument to predict what it theoretically predicts; Resp. = responsiveness: the ability to detect change over time. N.A. not applicable. – – Very Poor; – Poor; +/- Fair; + Good; ++ Excellent
and an instrument that measures functional status, indicating good concurrent validity. According to Kus et al.\textsuperscript{38} the ICF core set was helpful in categorizing and standardizing patients’ goals in the right problem areas. However, a part of the goals could not be linked to the core set.

Reliability
Regarding the stability of measures, inter-rater reliability was assessed for GAS and the COPM. Three studies assessed inter-rater reliability of GAS\textsuperscript{30,32,33} and high correlations were shown between the scores of different professionals or between one professional and the multidisciplinary team (Table 3). Regarding inter-rater reliability of the COPM, no statistically difference was found between data collected by physiotherapists and by occupational therapists in the study of Enemark Larsen and others.\textsuperscript{37}

Responsiveness
GAS scored excellent for responsiveness as it detects clinical changes better than several standardized instruments such as the Katz and Barthel index.\textsuperscript{30,31,33,34,36} Rockwood and colleagues\textsuperscript{36} found that, compared to four other measurement instruments, only GAS demonstrated a statistically significant difference after three months of treatment. The fact that more and/or different problem areas are assessed, that are not measured with standardized instruments was proposed as explanation in three studies.\textsuperscript{30,34,36} Moreover, GAS remained the most responsive instrument for three common problem areas\textsuperscript{36}, i.e. self-care, mobility and incontinence, which is most likely attributable to the ability of GAS to detect small changes in functioning. Although the COPM and SIGA appear to detect change,\textsuperscript{37,39,40} responsiveness of both instruments seems to be poor. In case high baseline scores are set the detection of any improvement is prevented.\textsuperscript{37} An instrument that assesses functional status, the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC), was found to detect change more accurate than the COPM.\textsuperscript{40} Kus and others stated\textsuperscript{38} that the ICF core set only has the potential to detect major changes: no scoring process is used and it can only be determined whether goals are attained.

Feasibility
The time required to perform GAS, approximately 20 minutes,\textsuperscript{31,32} was found to be feasible,\textsuperscript{31,32,34,35} particularly when performed by experienced professionals. Stolee et al\textsuperscript{35} indicated that GAS enhanced efficiency of the team meetings. In order to overcome potential time-related problems, Yip and co-writers\textsuperscript{34} developed a standardized menu with relevant goals from which the patient could choose. While remaining valid and responsive, this version took on average 10-15 minutes. Melville et al.\textsuperscript{39} developed the SIGA to overcome time-related problems with the COPM. The SIGA took approximately 13 minutes less than the COPM, but they do not mention how long these instruments actually take. Four studies\textsuperscript{30,32,34,35} supported GAS in a multidisciplinary setting. Three studies\textsuperscript{30,34,35} found that the focus during team meetings was enhanced and the balance between professionals in the goal-setting process was improved. Besides, GAS helps to spell out the multiple problems and acknowledges the importance of multidisciplinary interventions within a geriatric setting.\textsuperscript{32} The COPM was originally
developed for occupational therapy and only in one study performed by another professional. This was indicated at the most controversial part of this quite recent study and it took six months of training before high inter-rater reliability was achieved.37

Utility
Two studies31,35 indicated that GAS enhances communication with the patient and stimulates information sharing, which subsequently helps to provide a more structured and suitable rehabilitation. Three studies31,35,36 indicate that a measurement approach such as GAS is required when evaluating the outcomes of the rehabilitation process as it involves the patient’s desires, which remain unconsidered when standardized measures are used. It was found that the COPM37 and SIGA39 were useful as they provide an opportunity to gain insight into the rehabilitant’s perspective. However, difficulties were also encountered: identification of problems and categorization was found to be challenging and high baseline scores implicate difficulties with the scoring process.

Discussion
Main findings
The objective of this systematic review was to assess the psychometric properties of goal-setting instruments that are applied within the field of geriatric rehabilitation. With regard to GAS, excellent responsiveness and evidence for concurrent and predictive validity and inter-rater reliability were found. Concurrent validity and inter-rater reliability of the COPM and content validity of both the COPM and SIGA appear to be good, but these findings are based on only one or two studies. Responsiveness of both instruments appears to be poor. Based on one study, content validity of the ICF core set was found to be fair; responsiveness was found to be very poor.

Our review shows that little research has been conducted on goal-setting instruments in geriatric rehabilitation. Most studies evaluated GAS, which were, however, conducted by one research group: Rockwood and colleagues in Canada. A possible explanation for the fact that a small number of articles were included is the yet limited application of goal setting within geriatric rehabilitation. Rosewilliam and colleagues18 reviewed the literature on the extent of goal setting in rehabilitation and concluded that it is no daily practice. This is supported by two studies in which it was found that patients are not involved in the goal-setting process41 and professionals wrongly claim to apply a client-centered approach for goal-setting.42 In a more recent review43 these findings are confirmed: although most professionals aim for client participation within the goal-setting process, this is often not achieved.

Goal-attainment Scaling
Although one research group assessed GAS, the positive findings regarding predictive, content and concurrent validity and inter-rater reliability may support its use as goal-setting instrument and additional outcome measure within geriatric rehabilitation. GAS appears to be feasible in time, and can be performed by different professionals, e.g. physicians or physiotherapists. GAS provides an opportunity to establish relevant individual desired outcomes, which might be missed when
standardized instruments are used. Considering its excellent responsiveness, GAS accurately measures the outcomes of the rehabilitation process and has the ability to detect otherwise unnoticed, potentially clinically important, changes in functioning. GAS also acknowledges that independence is not necessarily affected by functional capacity but, for example, environmental factors may play a role. It should be noted that GAS is recommended as additional outcome measure rather than as replacement for standardized measures. GAS does not measure baseline and outcome functioning level of the patient; it provides the opportunity to measure change compared to expected outcome. Besides, although inter-rater reliability of GAS appears to be good, individualized measures remain less reliable compared to standardized instruments and, in contrast to the included studies, low agreement between GAS scores set by different assessors in a neurological rehabilitation setting has been reported. Also, as five articles were published in the 1990s, research is urgently required to evaluate GAS as routine goal-setting instrument and, besides, research from other researchers than the Rockwood group may provide additional or new information.

The COPM and the SIGA

The COPM and, based on the COPM, the SIGA were developed as tools to set occupational performance goals based on the perception of the client. Concurrent validity and inter-rater reliability of the COPM and content validity of both instruments appear to be good, but these findings are based on only one or two studies. Responsiveness appears to be poor both for the COPM and the SIGA. A standardized instrument that measures physical functioning detected change more accurately than the COPM and in case high baseline scores are set, the detection of any improvement is impossible. Besides, ratings of different goals are hard to compare, and group comparison is difficult due to the subjective nature of the scoring process of both instruments. The COPM was developed for the use in occupational therapy and is still mostly performed by occupational therapists, which does not support its use by different professionals. Within the literature, it remains a point of discussion whether the COPM, and thus the SIGA, can be applied as an outcome measure and it is proposed to apply the COPM purely as intervention rather than as outcome measure. Applied as such, communication between the patient and therapist may improve, providing information that cannot be obtained with standardized instruments. This is in line with the findings of the current review. However, before the COPM (or SIGA) can be recommended as goal-setting instrument, its psychometric properties require further research.

The ICF core set

Based on one study, content validity of the ICF core set appears to be fair; responsiveness was found to be very poor. For goal identification itself, an open-ended questionnaire was used in the study of Kus et al. The question arose whether open-ended questionnaires are feasible and effective in practice, when professionals instead of researchers conduct them. Hence, based on the current evidence and study within geriatric rehabilitation, it is not recommended to use the ICF framework as goal-setting instrument, neither as outcome measure. Nonetheless, the ICF framework appears to be useful for categorization and
standardization of patient’s goals. When used as such, the goal-setting process and evaluation of goal-attainment requires the use of additional instruments such as the FIM or COPM, which has been done in previous studies. However, in practice this method appears to be undesirable and too time demanding.

Strengths and Limitations
To our knowledge, this is the first systematic review that assessed the psychometric properties of goal-setting instruments that are applied within the field of geriatric rehabilitation. A thorough systematic literature search was conducted. However, this review has some potential limitations. First, the final literature strategy was conducted by one author only (R.S.). Second, a first selection of articles was also performed by R.S. Third, due to the nature of this present study, main limitations of this study arise from the included studies. Many of the included studies had issues with the reference standard and two studies with the index test. Once more, the Rockwood group conducted all articles that assessed GAS. Another limitation is related to the publication dates: the studies were published within a 20-year time frame; 1992-2012, of which more than half were published before 2004.

Conclusion
In conclusion, little research has been conducted in the area of goal setting in geriatric rehabilitation. The evidence for its psychometric properties, and for its excellent responsiveness in particular, may support GAS as goal-setting instrument and additional outcome measure within geriatric rehabilitation. However, more research is urgently required to evaluate GAS as goal-setting instrument and routine outcome measurement in geriatric rehabilitation, as research conducted in different health care settings and countries may add important findings. Before the COPM (or SIGA) can be recommended as goal-setting instrument, its psychometric properties require further research.
References


22. Kiresuk, TJ, Sherman, RE. Goal attainment scaling: A general method for evaluating comprehensive community mental health


Appendix 1. Search strategy performed and actual search Key

Search terms were: 'goals', 'rehabilitation' and 'geriatrics'.

Search terms for goals included: goal, 'goal attainment', 'goal attainment scaling', Canadian Occupational Performance Measure, 'life goals questionnaire' and 'improvement scaling'.

Search terms for rehabilitation included: rehabilitation, 'occupational therapy', 'restorative care service', 'skilled nursing facilities', 'skilled nursing facility' and 'geriatric care'.

Search terms for geriatric included: 'geriatric assessment', geriatrics, frail 'older people', 'older patients', 'older persons', 'middle aged', aged, 'aged, 80 and over'

Actual search Medline:

### Appendix 2 - QUAility Assessment of Diagnostic Accuracy Studies (Quadas-2)

<table>
<thead>
<tr>
<th>Study</th>
<th>PATIENT SELECTION</th>
<th>INDEX TEST</th>
<th>REFERENCE STANDARD</th>
<th>FLOW AND TIMING</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rockwood30</td>
<td>low</td>
<td>low</td>
<td>low</td>
<td>low</td>
</tr>
<tr>
<td>Stolee31</td>
<td>low</td>
<td>low</td>
<td>high</td>
<td>low</td>
</tr>
<tr>
<td>Stolee32</td>
<td>low</td>
<td>low</td>
<td>low</td>
<td>low</td>
</tr>
<tr>
<td>Stolee33</td>
<td>low</td>
<td>low</td>
<td>low</td>
<td>low</td>
</tr>
<tr>
<td>Yip34</td>
<td>low</td>
<td>Unclear</td>
<td>Unclear</td>
<td>low</td>
</tr>
<tr>
<td>Stolee35</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Rockwood36</td>
<td>low</td>
<td>low</td>
<td>low</td>
<td>low</td>
</tr>
<tr>
<td>Enemark - Larsen37</td>
<td>high</td>
<td>high</td>
<td>Unclear</td>
<td>low</td>
</tr>
<tr>
<td>Melville39</td>
<td>low</td>
<td>low</td>
<td>high</td>
<td>low</td>
</tr>
<tr>
<td>Edwards40</td>
<td>Unclear</td>
<td>low</td>
<td>high</td>
<td>high</td>
</tr>
<tr>
<td>Kus38</td>
<td>Unclear</td>
<td>high</td>
<td>high</td>
<td>low</td>
</tr>
</tbody>
</table>

Note, signaling questions for risk of bias of the four domains. Patient selection: could the selection of patients have introduced bias?; Index test: could the conduct or interpretation of the index test have introduced bias?; Reference standard: could the reference standard, its conduct, or its interpretation have introduced bias?; Flow and Timing: could the Patient Flow have Introduced Bias?