Concise report

Goal attainment scaling as a measure of treatment success after physiotherapy for chronic low back pain

Anne F. Mannion[1,2], Filomena Caporaso[2], Natascha Pulkovski[2] and Haiko Sprott[2]

Abstract

Objectives. In some chronic conditions, patient-specific tools with individualized items have proved to be more sensitive outcome instruments than fixed-item tools; their use has not yet been investigated in chronic low back pain (cLBP).

Methods. Eleven males and 21 females [mean age 44.0 (12.3) years] with cLBP, undergoing a spine-stabilization physiotherapy programme, completed the Roland Morris (RM) Disability Scale and a 0–10 pain scale pre- and post-therapy. Post-therapy, goal attainment scaling (GAS) scores were calculated regarding achievement of 2–6 priority GAS goals established pre-therapy; global outcome of therapy was assessed on a 5-point Likert scale.

Results. Approximately one-fifth of the individualized goals were not covered by items of the RM. Of the 121 individualized goals, 41 (34%) were achieved at the expected level, 42 (35%) were exceeded and 38 (31%) were not reached. GAS scores correlated with change scores for pain ($r = 0.61$, $P < 0.0001$) and RM ($r = 0.49$, $P = 0.006$). Sixty-five per cent of the patients had a successful outcome according to GAS (i.e. a score $\geqslant 50$); 55%, according to global outcome (therapy helped/helped a lot); 39%, according to the RM score change (score decrease $\geqslant 30\%$); and 44%, according to the pain score change (score decrease $\geqslant 30\%$).

Conclusions. GAS demonstrates the achievement of important goals undetected by fixed-item measures and is a valid and sensitive outcome measure for assessing the success of rehabilitation in patients with cLBP.

Key words: Low back pain, Outcome measures, Goal attainment scaling, Physiotherapy.

CLINICAL SCIENCE

Introduction

Consideration of the patient's perspective in the planning of therapy is associated with greater adherence [1] and increased satisfaction with care [2]. Similarly, outcome measures that are patient centred provide the best indication of the success of treatment. In the field of low back pain (LBP), fixed-item condition-specific instruments such as the Oswestry Disability Index [3] and the Roland Morris

(RM) Disability Questionnaire [4] are commonly used to this end. In some fields of rehabilitation medicine, it has been shown that wide inter-individual variability in the manifestations of a given disorder leads to reduced responsiveness of fixed-item instruments; in contrast, individualized outcome measures allow the achievement of important goals to be demonstrated [5]. Goal attainment scaling (GAS) [6] is one such instrument that overcomes the 'one size fits all' limitation of fixed-item instruments by measuring programme success using individualized patient goals. The instrument has not been investigated within the context of non-specific chronic LBP (cLBP), a condition with differing causes and associated disability. Assessed with traditional outcome measures, the treatment of cLBP is associated with only small to moderate effect sizes [7]; this is sometimes attributed to the failure

[1]Spine Center Division, Department of Research and Development, Schulthess Klinik and [2]Department of Rheumatology and Institute of Physical Medicine, University Hospital Zürich, Zürich, Switzerland.

Correspondence to: Anne F. Mannion, Department of Research and Development, Spine Center Division, Schulthess Klinik, Lengghalde 2, 8008 Zürich, Switzerland. E-mail: anne.mannion@kws.ch

of trials to consider the heterogeneous aetiology of LBP resulting in the group treatment effect being 'diluted down' by the inclusion of 'inappropriate' patients [8]. However, it is also possible that a lack of specificity in the outcome measures is partly responsible. This pilot study assessed the use of GAS in measuring treatment success as compared with the RM and a pain scale after a programme of physiotherapy exercises for patients with cLBP.

Patients and methods

Thirty-seven patients with cLBP participated in the study, which was part of a broader investigation of various aspects of deep trunk muscle function in cLBP [9]. Briefly, the patients were enrolled into the study before their commencement of a programme of 'spinal segmental stabilization exercise therapy' in tertiary care physiotherapy outpatient departments. They were all diagnosed by their referring physician as having non-specific cLBP according to the diagnostic triage reported in the current European guidelines [10]. The study was approved by the medical ethics committee of the Canton of Zürich. The patients gave their signed informed consent to participate after receiving verbal and written information about the study.

Therapy

Twelve specially trained physiotherapists administered the treatment once per week for 9 weeks; it was based on the methods described by Richardson *et al*. [11] and involved the preferential activation of the deep abdominal muscles, using the abdominal drawing-in manoeuvre. After mastering the basic exercise, patients integrated the deep trunk muscle contractions into their activities of daily living or sport. Patients performed home exercises comprising 10 repetitions, 10 times a day, documented in an exercise diary.

Questionnaires

Approximately 1–2 weeks before and 1–2 weeks after therapy, the patients completed a questionnaire containing (in addition to socio-demographic, medical and pain history questions):

- RM Disability Scale, which measures 24-activity limitations due to back pain (score 0–24: higher score, increased disability) [4]; and
- Pain Graphic Rating Scale (pain) [12]: a 0–10 scale to record average back pain intensity during the last week.

After therapy further questions inquired about the global outcome of treatment {'overall, how much did the treatment you received in the last few months help?', with a 5-item Likert scale dichotomized for describing the success of the treatment into 'good outcome' ('helped' or 'helped a lot') or 'poor' ('helped only little', 'didn't help', 'made things worse') [13]} and the changes that had taken place since before therapy in relation to back pain,

independence in everyday activities, ability to do sport, general physical capacity (at home or work), frequency and quality of social contacts and mental well-being [13] (response options: much better, better, unchanged, worse and much worse).

GAS

GAS was implemented in accordance with the original model described by Kiresuk and Sherman [6]. According to a recent systematic review, GAS delivers reliable and valid scores when employed as an outcome measure in working age and older people within a physical and neurological rehabilitation environment [14]. Before therapy, two to five goals were chosen through negotiation and consensus between the treating physiotherapist and the patient; for each goal, five levels of possible outcome were specified. The goals were to be specific, measurable, achievable, realistic and timed (SMART) [5, 6]. A score of −1 indicated where the patient saw himself at baseline, i.e. pretreatment. After treatment, each goal was examined by the corresponding treating therapist together with the patient and its relative achievement rated as follows: at the expected level (score of 0), less than expected (−1, no change from baseline; −2, much less than baseline) or more than expected (+1, more; +2, much more).

The scores were then converted to a GAS *T*-score, using the formula provided by Kiresuk *et al*. [15], and interpreted as follows: 50, expected level of achievement; <50, performance below the expected level; >50, performance above the expected level. A content analysis was carried out to investigate how many of the individual GAS goals were reflected in the items in the RM.

Statistical analysis

Descriptive data are presented as means (S.D.) or medians [interquartile range (IQR)], as appropriate. Responsiveness was given by the standardized response mean [SRM = (post-test mean − pre-test mean)/S.D. changes]. The relationships between GAS scores and the change scores for RM disability and pain were examined using partial correlation coefficients, holding baseline values for these variables constant, as recommended by Heavlin *et al.* [16]. Significance was accepted at the 5% level, with no corrections for multiple testing [17].

Results

Of the 37 patients, 32 (86%) completed the physiotherapy programme and questionnaires. Five patients were considered dropouts: two had never actually fulfilled the study's admission criteria (one language, one medical) and three chose not to continue due to the time commitment.

The baseline characteristics and main outcome results for the 32 completers have been reported elsewhere [9]. Briefly, there were 11 men and 21 women and their mean age was 44.0 (S.D. 12.3) years. They had had LBP on average for 92 (S.D. 129) months and had an RM disability score of 8.9 (S.D. 4.7), and an average pain intensity of

TABLE 1 Frequency of items in the RM disability questionnaire that featured as the area to be addressed in the individual's goals

RM item no.	Item content	Count, *n* (%)
2	I change position frequently to try and get my back comfortable	21 (15.7)
10	I only stand up for short periods of time because of my back	15 (11.2)
11	Because of my back I try not to bend or kneel down	15 (11.2)
4	Because of my back I am not doing any of the jobs that I usually do around the house	13 (9.7)
21	I avoid heavy jobs around the house because of my back	11 (8.2)
17	I only walk short distances because of my pain	9 (6.7)
13	My back is painful almost all the time	5 (3.7)
18	I sleep less well because of my back	5 (3.7)
14	I find it difficult to turn over in bed because of my back	3 (2.2)
23	Because of my back pain, I am more irritable and bad tempered with people than usual	3 (2.2)
12	I find it difficult to get out of a chair because of my back	2 (1.5)
16	I have trouble putting on my socks (or stockings) because of the pain in my back	1 (0.8)
5	Because of my back I use a handrail to get upstairs	1 (0.8)
3	I walk more slowly than usual because of my back	1 (0.7)
1	I stay at home most of the time because of my back	0 (0)
6	Because of my back I lie down to rest more often	0 (0)
7	Because of my back I have to hold on to something to get out of an easy chair	0 (0)
8	Because of my back I try to get other people to do things for me	0 (0)
9	I get dressed more slowly than usual because of my back	0 (0)
15	My appetite is not very good because of my back pain	0 (0)
19	Because of my back pain I get dressed with help from someone else	0 (0)
20	I sit down for most of the day because of my back	0 (0)
24	Because of my back I go upstairs more slowly than usual	0 (0)
-		29 (21.8)
Total		134 (100)

4.7 (S.D. 1.7). The main outcomes after therapy comprised a 2.3 (S.D. 4.2) point reduction in RM disability and a 1.1 (S.D. 2.1) point reduction in average pain (each $P < 0.01$); these constituted SRMs of 0.54 and 0.53, respectively [9].

In total, 134 individual goals were recorded for the whole group before the treatment began [median 4 (IQR 2; range 2–5) goals/patient]. With a score of −1 attributed to each GAS item pretreatment, the group mean baseline GAS score was 35.7 (S.D. 1.0).

The proportion of goals related to each of the RM items is shown in Table 1. Nine RM items did not feature in any of the GAS goals. Altogether, 105 (78%) of 134 goals could be matched to an RM item; those that could not tended to be related to sporting performance, work or psychological factors.

After therapy, it was possible to assess the level of achievement for 121 of the original 134 goal sets (goals from the dropouts, or goals related to seasonal sports, or influenced by subsequent musculoskeletal injuries, could not be assessed after therapy): 41 (34%) were achieved at the expected level; 38 (31%) were not reached and 42 (35%) were exceeded.

The mean GAS score after treatment was 51.0 (S.D. 13.7; range 24.6–80.2). This represented a mean change from pretreatment of 15.3 (S.D. 13.9), giving a SRM for the GAS of 1.10.

The GAS scores correlated with the change in pain (corrected for baseline pain; partial $r = 0.61$, $P < 0.0001$) and RM (corrected for baseline disability; partial $r = 0.49$,

$P = 0.006$), and with the post-therapy ratings of global treatment effectiveness (corrected $\rho = 0.39$, $P = 0.034$). They also showed low but significant correlations with the post-therapy ratings of improvement in the domains back pain, independence, ability to do sport and mental well-being (Table 2).

The proportions of patients achieving a minimum 30% reduction in RM and in pain (i.e. clinically relevant changes [18]) were 39 and 44%, respectively. The global outcome was rated as 'good' (therapy helped/helped a lot) in 55% of patients. The proportion of patients who achieved or exceeded their goals, as judged by a GAS score $\geqslant 50$, was 64.5%.

Discussion

A major challenge in the evaluation of treatment outcome is the identification of methods that adequately measure the success of a programme. Clinicians have long sought standardized tests against which an entire patient population can be measured; equally, however, such measures are criticized for being insensitive to the uniqueness of the goals of each patient within a given programme [19]. Overall, the findings of the present study supported those in other fields of rehabilitation, namely that there was a moderate correlation between the scores given by the fixed-item and patient-specific measures (suggesting adequate construct validity for the GAS), but that GAS

TABLE 2 Spearman's rank correlation coefficients between the global improvements in each domain measured post-therapy and the GAS scores, the change scores for pain and change scores for RM disability

Outcome domain	GAS score	Change score Average LBP	Change score RM
Global improvement in pain	**0.44**	**0.51**	**0.49**
Global improvement in independence	**0.40**	**0.59**	**0.54**
Global improvement in ability to do sport	**0.43**	**0.48**	**0.66**
Global improvement in physical activities in general	**0.33**	**0.46**	**0.50**
Global improvement in quality/quantity social contacts	0.01	0.13	−0.02
Global improvement in mental well-being	**−0.40**	−0.25	−0.08

Values in bold are $P < 0.05$.

was more responsive to change after treatment than were the fixed-item instruments.

Following the methods reported in previous studies on GAS (e.g. [5, 20]) and for the purposes of comparison, we assessed the instrument's responsiveness using the SRM (while mindful of the objections to doing so [6]). Encouragingly, the conclusions based on the SRMs (i.e. that GAS was the most responsive instrument) were wholly substantiated by those based on non-parametric methods of inferential analysis, such as frequency analysis of the proportion of patients with a successful outcome. It was established that almost two-thirds (65%) of the patients had a GAS score ⩾50, indicating that they had achieved or exceeded their goals, whereas only ∼40% patients achieved the minimum clinically important change of a 30% reduction in score for pain or disability, and 55% reported a 'good global outcome'. Hence, GAS does indeed appear to demonstrate the achievement of important goals that are less well detected by fixed-item measures.

We attempted to examine the content validity of GAS by performing a qualitative content analysis of the individual goals declared pretreatment. At face value, the goals appeared to be typical of those commonly identified in normal clinical practice for patients of this type. Also, most of the goals were approximated by the items of the RM, an instrument devised on the basis of typical complaints in a large group of LBP patients in primary care [4]. It was interesting, however, that 22% of the goals set using the GAS—most notably those to do with sports, work or psychological manifestations of the pain—could not be linked to any items of the RM, even in their broadest context of understanding. This highlights one of the benefits of GAS, i.e. it is able to obtain a comprehensive record of items that are important to the patients, regardless of how ambitious they may appear to be on an 'average patient' basis. Interestingly, many of the goals, although expressed in terms of 'function', were actually placed within a context of pain (e.g. 20 minutes standing without pain); this might explain why the GAS score correlated better with improvements in pain than with improvements in RM score. Despite what we might like to think or hope, it would seem that it is still the level of pain that is most important to these patients.

One of the potential disadvantages of GAS is that clinicians require sufficient knowledge, training and experience to carry out the procedure of goal setting. Nonetheless, the various steps in the process are all well documented [15]. The GAS formula is designed such that if goal achievement is predicted accurately and in an unbiased manner, the GAS at outcome should exceed and fall short of expectations in roughly equal proportions, and there should be an approximately normal distribution of GAS T-scores with a mean (S.D.) of 50 (10) [6]. This was the case in the present study [GAS score 51 (14), range 25–80] suggesting that there was no or little bias in estimating the potential for gain, and providing some support for the validity of the method applied. To the authors' knowledge, the impact of the actual number of goals set on the subsequent GAS scores has not been investigated in detail, and within the confines of the present study we could not shed any further light on this. The use of more than two goals is reported to deliver more reliable scores, as is the use of a 5-point rather than a 3-point scale for rating their achievement, but the number set must also be considered in relation to the length and scope of the programme and the needs of the patient [15]. We were also unable to assess the test–retest reliability of the GAS scores in the present study; however, a recent systematic review has testified to the reliability of GAS when employed as an outcome measure in other physical rehabilitation settings [14].

Another possible limitation of GAS as an outcome instrument is that it is designed for assessing the effectiveness of treatment after a designated period of time only. In many instances, however, it is desirable to follow up patients for longer to see how well any treatment effect is maintained. For this, the serial application of fixed-item instruments, which simply focus on current status at each assessment, are required. GAS scores also tell us nothing about the absolute level of adjustment or disability of the individual.

Important decisions regarding the quality of therapy and health policy are based on patient outcomes, and so it is essential to identify the most appropriate assessment tool. Our findings lead us to concur with previous authors [5, 20] that, in addition to documenting pretreatment expectations on change and sharpening the focus of treatment, GAS is able to capture subtle but important change

in client-centred functioning. It appears more suited to complement rather than replace established outcome measures—since both have their own unique advantages and disadvantages—and will perhaps have more relevance in one setting than in another. However, overall, GAS should be considered a useful adjunct to the present fixed-item options for assessing patient outcomes after rehabilitation for cLBP.

Rheumatology key messages

- Important decisions regarding therapy and health policy are based on patient outcomes.
- GAS is able to capture subtle but important changes in the functioning of patients with LBP.
- GAS appears more suited to complement rather than replace established fixed-item outcome measures.

References

1 Turk DC, Rudy TE. Neglected topics in the treatment of chronic pain patients–relapse, noncompliance, and adherence enhancement. Pain 1991;44:5–28.

2 Cleary PD, McNeil BJ. Patient satisfaction as an indicator of quality care. Inquiry 1988;25:25–36.

3 Fairbank JC, Couper J, Davies JB, O'Brien JP. The Oswestry low back pain questionnaire. Physiotherapy 1980;66:271–3.

4 Roland M, Morris R. A study of the natural history of back pain. Part 1: development of a reliable and sensitive measure of disability in low-back pain. Spine 1983;8: 141–4.

5 Khan F, Pallant JF, Turner-Stokes L. Use of goal attainment scaling in inpatient rehabilitation for persons with multiple sclerosis. Arch Phys Med Rehabil 2008;89:652–9.

6 Kiresuk TJ, Sherman R. Goal attainment scaling: general method for evaluating comprehensive community mental health programs. Community Ment Health J 1968;4: 443–53.

7 Keller A, Hayden J, Bombardier C, van Tulder M. Effect sizes of non-surgical treatments of non-specific low-back pain. Eur Spine J 2007;16:1776–88.

8 Turk DC. The potential of treatment matching for subgroups of patients with chronic pain: lumping versus splitting. Clin J Pain 2005;21:44–55.

9 Mannion AF, Helbling D, Pulkovski N, Sprott H. Spinal segmental stabilisation exercises for chronic low back pain: programme adherence and its influence on clinical outcome. Eur Spine J 2009;18:1881–91.

10 Airaksinen O, Brox JI, Cedraschi C *et al*. Chapter 4. European guidelines for the management of chronic non-specific low back pain. Eur Spine J 2006;15(Suppl. 2): S192–300.

11 Richardson C, Jull G, Hodges P, Hides J. Therapeutic exercise for spinal segmental stabilization in low back pain: scientific basis and clinical approach. Edinburgh: Churchill Livingstone, 1999.

12 Haefeli M, Elfering A. Pain assessment. Eur Spine J 2006; 15(Suppl. 1):S17–24.

13 Mannion AF, Junge A, Elfering A, Dvorak J, Porchet F, Grob D. Great expectations: really the novel predictor of outcome after spinal surgery? Spine 2009;34:1590–9.

14 Hurn J, Kneebone I, Cropley M. Goal setting as an outcome measure: a systematic review. Clin Rehabil 2006;20: 756–72.

15 Kiresuk TJ, Smith AG, Cardillo JE. Goal attainment scaling: applications, theory and measurement. Hillsdale, NJ: Lawrence Erlbaum, 1994.

16 Heavlin WD, Lee-Merrow SW, Lewis VM. The psychometric foundations of goal attainment scaling. Community Ment Health J 1982;18:230–41.

17 Perneger TV. What's wrong with Bonferroni adjustments. Br Med J 1998;316:1236–8.

18 Ostelo RW, Deyo RA, Stratford P *et al*. Interpreting change scores for pain and functional status in low back pain: towards international consensus regarding minimal important change. Spine 2008;33:90–4.

19 Williams RC, Stieg RL. Validity and therapeutic efficacy of individual patient goal attainment procedures in a chronic pain treatment center. Clin J Pain 1987;2:219–28.

20 Turner-Stokes L, Williams H, Johnson J. Goal attainment scaling: does it provide added value as a person-centred measure for evaluation of outcome in neurorehabilitation following acquired brain injury? J Rehabil Med 2009;41: 528–35.