O. Hägg
P. Fritzell
K. Romberg
A. Nordwall

The General Function Score: a useful tool for measurement of physical disability. Validity and reliability

Received: 16 November 1999
Revised: 16 June 2000
Accepted: 5 September 2000
Published online: 2 February 2001
© Springer-Verlag 2001

O. Hägg (✉) · P. Fritzell · A. Nordwall
Department of Orthopedic Surgery,
Sahlgren University Hospital,
413 45 Gothenburg, Sweden
e-mail: ollehagg@hotmail.com,
Tel.: +46-31-3424435,
Fax: +46-31-825599

K. Romberg
Physiotherapy Unit,
Sahlgren University Hospital,
Gothenburg, Sweden

Abstract The General Function Score (GFS) is a disease-specific instrument consisting of nine items focusing on strict physical activities of daily living. It is intended as an alternative to the more complex scores of disability (such as the Oswestry Disability Index), serving as a complement to the quality of life instruments in the study of low back pain (LBP). It was developed from an original 17-item questionnaire, of which 11 of the items were tested for criterion validity in an observer-supervised performance test. Two items were excluded from further analysis because of too low a validity. The remaining nine-item GFS was tested for construct validity, reliability, feasibility and responsiveness in six different cohorts: 297 patients with chronic low back pain (cohort 1), an age- and sex-matched control group of 287 randomly allocated Swedish citizens (cohort 2), three separate groups of patients admitted for surgery due to low back disorders (cohorts 3–5) and outpatients with

spinal disorders (cohort 6). Correlations were tested with the Spearman Rank correlation coefficient, differences between groups with the Mann-Whitney test and the internal consistency with the Cronbach's coefficient α . The GFS total scores showed correlations of 0.78, 0.81 and 0.88 in the three aspects of the performance test. The response rate was 98.3–100%. The mean time to complete the questionnaire was 1.2 min. The internal consistency was 0.69 and 0.86 in cohorts 1 and 2 respectively. The test-retest correlation was 0.88. The GFS showed a high responsiveness to difference and change. The effect size was 0.82–0.96 in surgically treated disc herniation and 0.55–0.85 in spondylolisthesis. The GFS is a highly valid and reliable instrument with good responsiveness and feasibility, useful for evaluation of physical disability.

Keywords Low back pain · Physical disability · Questionnaire · Validity · Reliability

Introduction

The consequences of low back pain (LBP) for an individual can be generally described and measured on three different levels [6, 11, 25]:

1. Physical impairment (decreased range of motion, reduced muscle strength, neurologic deficiency, etc)
2. Physical disability (reduced walking ability, impaired lifting capacity, decreased ability to sit, stand and perform other physical functions) and

3. Social handicap (the resulting limitation in social, vocational and recreational life as well as life satisfaction)

During the last decade, several scales and questionnaires with acceptable validity and reliability have been developed to measure these aspects of human function [9, 20].

The Short Form 36 (SF36) [21, 29], Sickness Impact Profile (SIP) [1] and the EuroQoL [12] are recommended and widely used generic (non-disease specific) measures of social handicap, i.e. quality of life.

Several of these measures have lately been recommended for routine use in outcome studies of low back pain [10].

On the disease-specific level, the Oswestry Disability Index (ODI) [13], the Roland Morris Disability Score [23, 24] and the Million Scale (MS) [22] all describe disability as a combination of physical disability and social restriction, which, together with measurement of pain, are combined in a disability index. For example, the ODI includes questions about ability to sit, stand, lift and walk (physical disability) and questions about social life, sexual function, travelling and personal care (social handicap) as well as a measure of pain [19].

A disease-specific instrument strictly focused on physical activities of daily living, on the other hand, gives a picture of the purely physical restrictions inflicted by the disease. When used in combination with psychometric tests, it could serve to elucidate the relation between physical disability and psychosocial factors in the prognosis of low back pain.

Although the optimal scientific way of studying physical disability is to have each subject perform a set of specified tasks [28], in the clinical setting this is impractical and time consuming. A self-completed questionnaire validated against actual performance could serve as an acceptable substitute in clinical outcome studies.

The objective of the present study was to evaluate a new score specifically aimed at measuring the physical disability by a patient-completed questionnaire, the General Function Score.

Materials and methods

The questionnaire

The original General Function Score (GFS) consisted of 17 items concerning physical activities of daily living: walk a flight of stairs; sit more than 30 min; stand more than 30 min; walk more than 30 min; go by car more than 30 min; lift more than 10 kg; lean over a basin; carry a bag of groceries; carry a suitcase; carry a small child; make the bed; vacuum clean; wash the dishes; make food; dress; put on shoes; run.

The items were selected and aggregated from several sources, concentrating on strict physical disability rather than more complex social activities. Suggested items were discussed and evaluated theoretically by the author group, resulting in the present 17 items, which we considered describe daily physical requirements in an acceptable way.

Three response alternatives, "can perform", "can perform with difficulty" and "cannot perform", were considered appropriate for scaling of the disability. This type of three-level scaling is used, among others, in the EuroQoL [12]. The response alternatives were scored as 0, 1 and 2. The total score is calculated as a percentage, where 0 represents no physical disability and 100 represents maximal physical disability.

Subjects

Cohort 1

Cohort 1 comprised 297 patients with chronic low back pain, selected for surgery. There were 51% women, mean age 45 years, range 25–65 years. This group completed the GFS, and also filled in questionnaires with the ODI and the MS.

Cohort 2

A general population sample of Swedish citizens was collected by random allocation by the Swedish Central Board of Statistics (SCB). A mail questionnaire, including the GFS and several sociodemographic questions, yielded, with two reminders, 278 completed questionnaires from 400 subjects (70%). The control group was age and sex matched relative to cohort 1.

Cohort 3

Cohort 3 comprised 28 patients with different low back disorders admitted for surgery at our department. They completed the GFS form on an average of 17 days (SD 15, range 1–60) before admission and again on the day of admission.

Cohort 4

Cohort 4 comprised another group of 21 patients, with different low back disorders, admitted to our department for surgery. They completed the GFS form 2–3 weeks before admission to the ward. On the day of admission they carried out a performance test supervised by an observer (physiotherapist). The tasks in the performance test were those described in the GFS, and the observer supervising the performance assessed the patient's functional disability following strict criteria, and completed a GFS form for each patient. After the performance test the patients again completed a GFS.

Cohort 5

Cohort 5 comprised 26 consecutive patients operated for lumbar disc herniation, and 20 consecutive patients who underwent posterolateral fusion for isthmic lumbosacral spondylolisthesis. Both groups completed the GFS questionnaire before and 1 year after surgery.

Cohort 6

Cohort 6 comprised 35 outpatients with spinal disorders, who completed the GFS questionnaire while being timed by a nurse.

Evaluation

Evaluation of the GFS followed recommended methods [3, 25], outlined by Deyo et al. [7, 9, 10], covering: sensibility, validity, reliability, responsiveness and feasibility.

Sensibility

The *content validity*, i.e. the relevance of the items and the appropriateness of the questionnaire relative to the study purpose and the patient population, was considered in the construction of the GFS

Fig. 1 The triangular test model of the General Function Score (GFS) questionnaire

scale. The 17 items were considered to reflect a representative perspective of daily physical demands of the population in general as well as of patients with spinal problems.

Validity

The *criterion validity*, i.e. how well the questionnaire items mirror real life tasks, was evaluated in the performance test carried out in cohort 4. The performance of the tasks was supervised by an observer (physiotherapist) and assessed following strict criteria.

A triangular model of three sets of GFS questionnaires for each patient was created, as recently suggested by Salén et al. [26] (Fig. 1). The following correlations were calculated:

1. Pre-admission self-rating versus post-performance self-rating
2. Pre-admission self-rating versus observer rating
3. Observer rating versus post-performance self-rating

For practical reasons, some of the items could not be performed in the ward, resulting in 11 items to be included in the test (Table 1).

The *construct validity*, i.e. how the instrument behaves in comparison with other established/validated instruments, was also tested: as a correlation of the GFS with the ODI ($n=289$) and with the MS ($n=293$) in cohort 1.

Reliability

The reliability was tested in two ways.

1. The *internal consistency* of the GFS items was evaluated in cohorts 1 and 2. They were chosen because they both had large numbers of subjects.
2. A *test-retest* evaluation was performed in cohort 3, with a mean interval of about 2 weeks.

Responsiveness

The ability to detect clinical differences and changes after treatment was evaluated by testing four hypotheses:

1. Subjects with LBP have worse physical function than those without LBP. This hypothesis was tested in cohort 2, which was divided into two groups based on the presence of LBP (with LBP $n=98$, without LBP $n=176$, missing $n=4$).
2. Increasing age results in decreased physical function. This hypothesis was tested in cohort 2, which was divided into two age groups (≤ 45 years, $n=145$ and >45 years, $n=128$).
3. The presence of co-morbidity (any disease other than LBP) increases physical disability. The hypothesis was tested in cohort 2 (co-morbidity present $n=58$, no co-morbidity $n=214$).
4. Patients who have undergone discectomy for disc herniation and posterolateral fusion for isthmic spondylolisthesis (cohort 5) show an improved physical function at the 1-year follow-up.

Feasibility

As an indication of the usefulness of the GFS, the response-rate (the rate of missing values) was calculated in all patient cohorts.

Table 1 Items included in the performance test

Item	Performed tasks
1. Climb/descend staircase	Ascending and descending one storey
2. Sit more than 30 min	Sitting on an ordinary chair without armrest for at least 30 min
3. Walk more than 30 min	Walking back and forth in the corridor of the ward for at least 30 min
4. Stand more than 30 min	Standing in the corridor of the ward, allowed to read, to do hand work, or watch TV, for at least 30 min
5. Lift more than 10 kg	Lifting a crate of beer bottles (15 kg) from the floor to a desk 90 cm above the floor
6. Lean over a basin	Standing bent over a basin 90 cm above the floor, washing the face
7. Carry a shopping bag	Carrying a plastic shopping bag with 6 kg weight back and forth in the corridor (~100 m)
8. Make the bed	Making the hospital bed in the lowest position (~60 cm above the floor) with blankets, cover and cushion
9. Dress	Undressing and dressing in civilian clothes
10. Put on shoes	Putting on shoes, with or without laces, in the way the patient found it most comfortable
11. Run	Running, dressed in ordinary civilian clothes and shoes, at least half the length of the corridor (~25 m)

Additionally, the time needed to fill out the questionnaire was measured in cohort 6.

The test of the questionnaire was performed in two stages, in the following order:

- Step I: test of criterion validity (performance test of 11 items)
- Step II: only items with acceptable validity in the performance test were tested further in remaining aspects, as outlined above

Statistical analysis

Since the GFS scores are basically ordinal data, even though calculated as a percentage function, non-parametric tests were used in the statistical analysis.

The internal consistency of the GFS was calculated with Cronbach's coefficient α [4]. To be acceptable, the coefficient should be at least 0.6. On the other hand, a very high value, approaching 1, indicates that several items are more or less identical, and may be unnecessary [2, 14].

In the evaluation of construct validity, Spearman's rank correlation test was used to calculate the correlation between the GFS and the ODI and the MS. A correlation coefficient of at least 0.6 is considered acceptable [5, 15].

The reliability was calculated, item per item, with percentage agreement between test and retest. The total scores were compared with the intraclass correlation coefficient (ICC) [8]. An ICC of 1 denotes total agreement without any measurement error, 0 means total absence of agreement between test and retest results.

To calculate the sensitivity of the GFS to clinical differences, the Mann-Whitney test was used to compare age groups and groups with/without LBP in cohort 2. The responsiveness to change (surgical treatment) was tested with the Wilcoxon Signed Ranks test in cohort 5.

To estimate the effect size of the responsiveness, we used two accepted, but different, methods: those of Kazis et al. and of Cohen. According to Kazis et al. [18], the difference between pre-treatment and post-treatment scores is used as the numerator. The denominator is the standard deviation (inter-quartile range, in calculations of ordinal data) of the pre-treatment score. Cohen [3] uses the same numerator, but uses the standard deviation (inter-quartile range) of the score change as the denominator. An effect size of 0.2 is considered to be small, and 0.8 or greater to be large [3].

The statistical analyses were all performed using the Statistical Package for Social Sciences (SPSS), version 9.0.

Results

Step I

Criterion validity

Correlations between individual items in the same test procedures were all high, except for two items. The item "put on shoes" showed a very low correlation between the pre-admission and the post-performance self-ratings (0.023, $P=0.92$) and between the pre-admission self-ratings and the observer ratings (0.12, $P=0.62$). The other item, "run", showed a very low correlation in the comparison between the pre-admission and post-performance self-ratings (0.047, $P=0.84$). However, there was a significant, but not very high, correlation between the observer ratings and the post-performance self-ratings for both items (see Table 2).

Table 2 Correlations between the individual items as scored by the patient prior to admission (PA), by the patient after the performance test (PP) and during the performance test by an observer (OB). Calculations were performed with the Spearman Rank Correlation test

Item	Correlation coefficients		
	PA vs PP	PA vs OB	OB vs PP
Climb/descend staircase	0.59**	0.51*	0.62**
Sit >30 min	0.81***	0.65**	0.8***
Walk >30 min	0.41	0.7***	0.62**
Stand >30 min	0.53**	0.58**	0.54*
Lift >10 kg	0.58**	0.48*	0.79***
Lean over basin	0.551**	0.6**	0.76***
Carry a shopping bag	0.67**	0.65**	0.78***
Make the bed	0.61**	0.67**	0.82***
Dress	0.69**	0.59**	0.72***
Put on shoes	0.023	0.12	0.50*
Run	0.047	0.44*	0.60**

* $P<0.05$; ** $P\leq 0.01$; *** $P<0.001$

Table 3 General Function Score (GFS) total scores pre-admission (self-rated), in the performance test (observer) and post-performance test (self-rated). Correlations were calculated with the Spearman Rank Correlation Coefficient (*IQR* inter-quartile range)

GFS completed	GFS score		Correlation coefficient (<i>P</i> -value)
	Median	IQR	
Pre-admission	61	31–72	0.78 ^a (<0.001)
Observer	44	25–69	0.88 ^b (<0.001)
Post-performance test	44	22–61	0.81 ^c (<0.001)

^a Pre-admission vs observer

^b Post-performance test vs observer

^c Post-performance test vs pre-admission

Table 4 Percentage agreement of separate items in the test-retest analysis of the GFS

Item	Percentage agreement
Climb/descend staircase	86
Sit >30 min.	78
Stand >30 min.	68
Walk >30 min.	57
Lift >10 kg	71
Lean over a basin	71
Carry a bag of groceries	96
Make the bed	71
Dress	86

The two above-mentioned items with low correlation were excluded from further testing, leaving nine items to be tested.

Correlations between the nine-item GFS scores in the three aspects of the performance test were all high (see Table 3).

Table 5 Responsiveness of the GFS as illustrated by differences in GFS scores when comparing subgroups of the general population sample (cohort 2). The statistical significance of the difference was calculated with the Mann-Whitney Test (*LBP* low back pain)

Test sample	GFS score		Mann-Whitney test	
	Median	IQR	Z	P-value
LBP yes/no (<i>n</i> =98/176)	6/0	0–17/0	8.71	<0.001
Comorbidity yes/no (<i>n</i> =58/214)	6/0	0–11/0–6	4.68	<0.001
Age ≤45/>45 (<i>n</i> =145/128)	0/0	0–6/0–10	2.29	0.02

Step II

The results of the tests performed subsequently on the nine-item GFS were as follows.

Construct validity

The Spearman Rank correlation coefficient for the GFS with the ODI was 0.61 ($P<0.001$). The correlation with the MS was 0.54 ($P<0.001$).

Reliability

The internal consistency coefficients α of the items of the GFS in cohorts 1 and 2 were 0.69 and 0.86 respectively.

In the test-retest analysis, the percentage agreement of the separate items ranged from 57 to 96% (see Table 4). The total median scores (inter-quartile ranges) were 50 (39–61) and 56 (28–65). The intra-class correlation coefficient (ICC) of the total score was 0.87. We also tested for any systematic change of the GFS, as a function of time, by correlating the time interval between test and retest with the difference in GFS score at the two occasions. We found no systematic changes in the GFS score, (Spearman Rank correlation coefficient=0.09).

Responsiveness

The GFS showed significant differences in median values when cohort 2 was analysed in relation to the existence of LBP (hypothesis 1) with the Mann-Whitney test.

The median values of the GFS in the two age groups of cohort 2 were equal, but a difference in inter-quartile range and a significant difference according to the Mann-Whitney test (hypothesis 2) was found.

There was also a significant difference in the GFS score, comparing the subgroups with and without comorbidity (hypothesis 3) (see Table 5).

The GFS scores were significantly decreased at the 1-year follow-up after surgery for both isthmic spondylolisthesis and disc herniation (hypothesis 4). The effect size calculated with the two different methods in the spondylolisthesis group was 0.96 [3] and 0.82 [18]. In the

Fig. 2 Box plots of the GFS score before and 1 year after surgery for disc herniation and isthmic spondylolisthesis. Scores are presented as median values, including 25th/75th and 5th/95th percentiles

disc herniation group, the effect size was 0.55 [3] and 0.85 [18] (see Fig. 2).

Feasibility

The response rates of 98.3% in cohort 1, 99.3% in cohort 2, and 100% in cohorts 3–5 indicate that there were no major difficulties in understanding or answering the questions. The mean time needed to fill the questionnaire was 1.2 min, range 0.5–3.5 min.

Correlation of the nine-item GFS with the 17-item GFS

The correlation between the nine-item GFS and the original 17-item GFS was also tested in order to estimate potential loss of information resulting from reducing the number of items to nine. The correlation was 0.94 ($P<0.001$) in cohort 1 and 0.95 ($P<0.001$) in cohort 2, indicating that very little information is lost by reducing the number of items to nine (Fig. 3).

Fig. 3 Scatter plot of the nine-item GFS versus the 17-item GFS in cohorts 1 and 2

Discussion

The most important aspect of validity is the way the questionnaire mirrors real life. The gold standard is the actual performance of the physical activities described in a ques-

tionnaire. This was tested in 11 of the 17 original items. The nine items that showed high individual correlations were selected to form the final GFS (Appendix). The score of the final GFS showed a high correlation in all three aspects of the performance test, suggesting that the scale actually describes physical activities in a valid way.

However, the mean GFS score was higher (61%) before admission than both the patients' assessments after the performance test (44%) and the observer rating of the performance (44%).

This difference is reflected in the correlation coefficient, which was lower in the pre-admission self-rating versus observer rating (0.78), and the pre-admission versus post-performance self-rating (0.81), than in the post-performance self-rating versus observer rating (0.88). The same tendency was found in a recent study by Salén et al. [26], in which the corresponding correlation coefficients were reported as 0.48, 0.69, and 0.78, and also in a study by Simmonds and Claveau [27].

Patients appear to overestimate their disability in the self-report. It seems that, after trying the activity under controlled circumstances, they correct their assessment to such a degree that the correlation with the observer rating becomes very high. This hypothesis is, at least partly, supported by the fact that both the observer rating and the patients' own ratings are generally lower than the patients' self-rating before admission.

Over-estimation of self-reported physical disability might be diagnosis dependent. Hidding et al. [16] reported that discordance between self-reported disability and observed performance was high in fibromyalgia, moderate in rheumatoid arthritis and negligible in ankylosing spondylitis. Contrastingly, Jacobs et al. [17] found a very high agreement between rheumatoid patients' self-report and observed performance.

The consistency in patients' over-estimation of disability is also shown by the fact that the reliability test in cohort 3 showed a higher correlation and a difference in disability between test and retest of only six percentage points (56% vs 50%).

Correlation coefficients of the GFS scores with the performance test of between 0.78 (self-rated pre-admission vs observer) and 0.88 (self-rated post-performance vs observer) are well above what is generally considered as acceptable in clinical trials [5, 15]. The nine individual items showed correlation coefficients of between 0.41 and 0.82 in the three aspects of the performance test. This compares favourably with a recent performance versus self-report analysis [28], where the individual items showed correlation coefficients of 0.40–0.60. The nine-item GFS score thus seems to be a valid measure of physical disability.

The reliability, estimated with the Cronbach's coefficient α of internal consistency (0.69 and 0.86 in the two cohorts tested), was substantial. The internal consistency, or homogeneity, of the items of a score is a measure of how related the items are with one another. The present correlation coefficients confirm that the individual items are related on an adequate level [5, 15]. The measurement error was small, as indicated by the item percentage agreements, ranging from 57 to 96%, and by the high intra-class correlation coefficient (0.87).

The construct validity, tested as the correlations of the GFS with the ODI (0.61) and with the MS (0.54), were low. This was anticipated, since both the ODI and the MS measure several functions that are complex social functions rather than strict physical activities, e.g. sexual function, travelling, personal care, need of change at workplace, restriction of normal life.

The responsiveness of the GFS to change and the ability to discriminate between different clinical statuses were corroborated in the test of the four hypotheses, which were all confirmed. The effect size, calculated in the cohort of surgically treated disc herniations and isthmic spondylolisthesis was high, indicating a clinically relevant sensitivity to change in physical disability.

The mean time needed to complete the questionnaire was 1.2 min, which is shorter than the time needed for the ODI (5 min) [13], the MS (10 min) [26] and the Roland-Morris Disability Scale (5 min) [23].

Conclusion

The results of the present study indicate that the GFS is a useful disease-specific instrument of measuring physical disability, showing high validity, reliability, responsiveness and feasibility.

Acknowledgements The authors acknowledge the assistance of the statistical advisor, A. Odén, PhD, Romelanda, Sweden. Financial support was granted by the Greta and Einar Asker Foundation, Sweden.

Appendix

The General Function Score

	Can perform	Can perform with difficulty, due to low back pain	Can not perform, due to low back pain
Walk a staircase	☐	☐	☐
Sit >30 min	☐	☐	☐
Stand >30 min	☐	☐	☐
Walk >30 min	☐	☐	☐
Lift >10 kg	☐	☐	☐
Lean over a basin	☐	☐	☐
Carry a bag of groceries	☐	☐	☐
Dress	☐	☐	☐
Make the bed	☐	☐	☐

References

1. Bergner M, Bobbitt RA, Carter WB, Gilson BS (1981) The sickness impact profile: development and final revision of a health status measure. *Med Care* 19:787–805
2. Carmines EG, Zeller RA (1979) Reliability and validity assessment. Sage, Beverly Hills
3. Cohen J (1977) Statistical power: analysis for the behavioral sciences. Academic Press, New York, pp 1–27
4. Cronbach LJ (1951) Coefficient alpha and the internal structure of tests. *Psychometrika* 16:297–334
5. Cronbach LJ (1969) Essentials of psychological testing. Harper & Row, New York, pp 97–124, 348–355
6. Delitto A (1994) Are measures of function and disability important in low back care? *Phys Ther* 74:452–462
7. Deyo RA (1988) Measuring the functional status of patients with low back pain. *Arch Phys Med Rehabil* 69: 1044–1053
8. Deyo RA, Diehr P, Patrick DL (1991) Reproducibility and responsiveness of health status measures. Statistics and strategies for evaluation. *Control Clin Trials* 12:142S–158S
9. Deyo RA, Andersson G, Bombardier C, et al (1994) Outcome measures for studying patients with low back pain. *Spine* 19:2032S–2036S
10. Deyo RA, Battie M, Beurskens AJ, et al. (1998) Outcome measures for low back pain research. A proposal for standardized use. *Spine* 23:2003–2013
11. Dropsy R, Marty M (1994) Indices de Qualité de vie pour l'évaluation des lombalgies. *Rev Rhum Ed Fr* 61:44S–48S
12. EuroQoL Group (1990) EuroQoL – a new facility for the measurement of health-related quality of life. *Health Policy* 16:199–208
13. Fairbank JC, Couper J, Davies JB, O'Brien JP, et al (1980) The Oswestry Low Back Pain Disability Questionnaire. *Physiotherapy* 66:271–273
14. Fries JF, Spitz P, Kraines RG, Holman HR (1980) Measurement of patient outcome in arthritis. *Arthritis Rheum* 23:137–145
15. Guilford JP (1965) Fundamental statistics in psychology and education. McGraw-Hill, New York, pp 104
16. Hidding A, Santen M van, De Klerk E, et al (1994) Comparison between self-report measures and clinical observations of functional disability in ankylosing spondylitis, rheumatoid arthritis and fibromyalgia. *J Rheumatol* 21: 818–823
17. Jacobs JW, Oosterveld FG, Deuqbouts N, et al (1992) Opinions of patients with rheumatoid arthritis about their own functional capacity: how valid is it? *Ann Rheum Dis* 51:765–768
18. Kazis LE, Anderson JJ, Meenan RF (1989) Effect size for interpreting changes in health status. *Med Care* 27: 178S–189S
19. Kopec JA (1994) Conceptualization, development and evaluation of a new self-report measure of disability in people with back pain. Thesis, Department of Epidemiology and Biostatistics, McGill University
20. Kopec JA, Esdaile JM (1995) Functional disability scales for back pain. *Spine* 20:1943–1949
21. McHorney CA, Ware JE Jr, Raczek AE (1993) The MOS 36-item short-form health survey (SF-36). II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. *Med Care* 31:247–263
22. Million R, Hall W, Nilsen KH, Baker RD, Jayson MI (1982) Assessment of the progress of the back-pain patient. *Spine* 7:204–212
23. Roland M, Morris R (1983) A study of the natural history of back pain. I. Development of a reliable and sensitive measure of disability in low-back pain. *Spine* 8:141–144
24. Roland M, Morris R (1983) A study of the natural history of back pain. II. Development of guidelines for trials of treatment in primary care. *Spine* 8: 145–150
25. Roos H, Roos E, Ryd L (1997) On the art of measuring. *Acta Orthop Scand* 68:3–5
26. Salén BA, Spangfort EV, Nygren AL, Nordemar R (1994) The Disability Rating Index: an instrument for the assessment of disability in clinical settings. *J Clin Epidemiol* 47:1423–1435
27. Simmonds MJ, Claveau Y (1997) Measures of pain and physical function in patients with low back pain. *Physiother Theory Pract* 13:53–65
28. Simmonds MJ, Olson SL, Jones S, et al (1998) Psychometric characteristics and clinical usefulness of physical performance tests in patients with low back pain. *Spine* 23:2412–2421
29. Ware JE, Sherbourne CD (1992) The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care* 30:473–483
30. WHO (1980) International Classification of Impairments, Disabilities and Handicaps: a manual of classification relating to the consequences of disease. World Health Organization, Geneva