

Responsiveness of Outcome Measurements in Rehabilitation of Patients With Posterior Pelvic Pain Since Pregnancy

Jan M. A. Mens, MD, PhD,*‡ Andry Vleeming, PhD,* Chris J. Snijders, PhD,†
Inge Ronchetti, MSc,* Abida Z. Ginai, MD, PhD,§ and Henk J. Stam, MD, PhD‡

Study Design. A cohort study was conducted.

Objective. To develop a test battery for evaluating the course of posterior pelvic pain since pregnancy.

Summary of Background Data. Properly validated scales to evaluate the course of posterior pelvic pain since pregnancy are scarce. Moreover, the use of many tests would be too strenuous for the patient and has an unfavorable cost–benefit ratio.

Methods. The ability of 48 effect measures to detect clinically relevant changes over time (responsiveness) was tested in patients with posterior pelvic pain since pregnancy. In this test, 35 measures were evaluated in a group of 44 patients, and 16 measures in a group of 56 patients (three measures were evaluated in both groups). All the tests were performed at baseline and after 8 weeks treatment. A global impression of improvement (improved or not improved) scored by the patient was used as the standard for assessing the course of the disease. Responsiveness was examined by calculating the standardized response mean of the improved patients and by using a two-tailed Mann–Whitney nonparametric test to compare the patients who had improved and those who had not improved.

Results. Of the 48 effect measures, 26 measures of five categories (activities of daily living, pain, hip muscle strength, spine mobility, and spine muscle strength) showed good correlation with the patient's global impression of improvement. The measures in the "mobility of the pelvic joints" category were insufficient for assessing clinical change in posterior pelvic pain since pregnancy. The measures in the "fatigue" and "pain provocation tests" categories correlated only moderately with clinical change.

Conclusions. It seems possible to gain appropriate information about the course of posterior pelvic pain since pregnancy with a small test battery. The usefulness of the Québec Back Pain Disability Scale, the hip adduction strength assessment, and the active straight-leg-raise test was proved by the current study. The value of 23 other instruments was substantiated, but further studies are needed to confirm their usefulness. The correlation of 22 evaluated measures with the patient's global improvement was too weak for them to be recommended as measures of clinical changes over time in posterior pelvic pain since pregnancy. It is recommended that clinicians

and investigators compile a small test battery by selecting the best representatives of the five measurement categories that have good correlation with the patient's global impression of improvement. [Key words: diagnostic tests, low back pain, outcome measure, pregnancy, sacroiliac joint] **Spine 2002;27:1110–1115**

In general, an instrument used to measure the course of a disease should be reliable, valid, and responsive. Responsiveness of an instrument is defined as its ability to detect clinically relevant changes over time. In many situations, changes in the disease status are assessed by global impressions of both the patient and physician. These impressions are subjective and often influenced by irrelevant factors. A problem arises when there is a discrepancy between the impression of the patient and that of the physician. Moreover, there is a need to measure the extent of the change. Following the course of the disease with a large number of tests is too strenuous for the patient, and this approach has an unfavorable cost–benefit ratio.

In studies of lumbopelvic pain, the most popular measures are self-reported scales.^{6,9} The value of pain provocation tests, mobility and strength measurements, and radiography is limited. Although the reliability of these measurements is high, their relation to clinical parameters is questionable or weak.^{12,18,19,23,26,27,29,31,34,35} Thus, there still is a need for objective tests with high reliability and validity.

Pregnancy is complicated frequently by the occurrence of lumbopelvic pain. The reported cumulative 9-month incidence ranges from 48% to 56%.^{1,8,12,20} Posterior pelvic pain since pregnancy (PPPP) often is described as a distinct pain type.^{10,20,21,30} However, it remains questionable whether PPPP is a specific syndrome or just nonspecific lumbopelvic pain with an onset during pregnancy or delivery. Regardless, detailed study on the characteristics of PPPP could provide better understanding of lumbopelvic pain in general.

The aim of the current study was to develop a test battery that gives appropriate information with minimum strain to the patient. To avoid burdening a sizable number of patients with a large group of tests, an orientation was first carried out in two small groups of patients to evaluate the responsiveness of a large set of tests.

From the *Spine and Joint Centre, the Netherlands, and the Departments of †Biomedical Physics and Technology, ‡Rehabilitation Medicine, and §Radiology, Faculty of Medicine and Health Sciences, Erasmus University Rotterdam, the Netherlands.

Acknowledgment date: December 19, 2000.

First revision date: May 1, 2001.

Second revision date: August 17, 2001.

Acceptance date: November 28, 2001.

Device status category: 1.

Conflict of interest category: 12.

Table 1. Baseline Characteristics of the Two Study Groups

	First group (n = 44)	Second group (n = 56)
Age (mean ± SD)	31.7 ± 3.2 y	33.5 ± 4.9 y
Duration of the period postpartum (mean ± SD)	4.1 ± 2.2 mo	2.3 ± 3.0 y
Parity (median, range)	2, 1–4	2, 1–5
Score on QBPDS (mean ± SD)	44.5 ± 14.8	54.7 ± 15.2

SD = standard deviation; QBPDS = Québec Back Pain Disability Scale.

■ Methods

Study Populations. Patients were selected from the outpatient clinic of a rehabilitation center that specialized in the treatment of lumbopelvic pain. All the patients were receiving treatment based on instructions, with or without home exercises. Two study groups were formed. The following selection criteria were used for both groups.

Inclusion Criteria. Patients were included in the study if they could report pain in the lumbopelvic region, defined as pain experienced between the upper level of the iliac crests and the gluteal fold; pain beginning during pregnancy or within 3 weeks after delivery; and freedom from pregnancy.

Exclusion Criteria. Patients were excluded from the study if they had a history of fracture, neoplasm, or previous surgery of the lumbar spine, pelvic girdle, hip joint, or femur; signs indicating radiculopathy (asymmetric Achilles tendon reflex or passive straight-leg-raising restricted by pain in the lower leg); a systemic disease of the locomotor system; insufficient knowledge of the Dutch language to fill in forms; or any restriction of testing. The first study group consisted of 44 patients, and the second group involved 56 patients. Data on these 100 patients are presented in Table 1.

Measurements. The 48 tests investigated in this study covered eight main measurement categories:

1. restrictions of activities of daily living
2. pain
3. pain provocation tests
4. fatigue
5. mobility of the pelvic girdle
6. mobility of the spine
7. strength of the hip muscles
8. strength of the trunk muscles.

Tests were selected on the basis of face validity (*i.e.*, validity based on the assumption of an expert that an instrument measures the things it should measure) and good intraexaminer reliability. Most of the tests were validated for use in assessing nonspecific low back pain or sacroiliac joint dysfunction.

To evaluate restrictions of activities of daily living, the Québec Back Pain Disability Scale (QBPDS) and 11 subscales of the Nottingham Health Profile (NHP) were used.^{7,9,27} Moreover, the patients were asked how long, using a range of 0 to 60 minutes, they could stand, walk, bicycle, sit, and lie without any increase of pain. For these questions, a pilot study showed the test–retest reliability of two measurements 1 week apart to be acceptable to good. The intraclass correlation coefficients were 0.94, 0.81, 0.77, 0.85, and 0.74, respectively.

The mean severity of pain and fatigue was scored on a 100-mm horizontal visual analog scale with choices ranging from 0 (no pain) to 100 (very severe pain). Because of the large variation in pain and fatigue between morning and evening,¹⁴ the responsiveness of both measures was evaluated. Pain and fatigue also were measured on two subscales of the NHP (pain and energy). Radiation of pain was scored using a modification of the classification proposed by the Québec Task Force: 0 (no leg pain), 1 (pain in the leg, but not beyond the knee), 2 (pain in the leg beyond the knee), and 3 (pain in the foot).²⁸ In cases of radiation in both legs, the highest score was used for the analysis.

Pain provocation tests were scored using a modification of the scale proposed by the American College of Rheumatology to grade tender points in fibromyalgia: 0 (no pain), 1 (mild: report of pain without grimace, flinch, or withdrawal), 2 (moderate: pain plus grimace or flinch), 3 (unbearable: the examiner is not able to complete the test because of withdrawal).³⁶ Pain generated at the place where the body of the patient was in contact with the hand of the examiner or with the couch was ignored. Pain provoked by active tests was scored on the American College of Rheumatology scale from 0 to 2. If tests were performed on both sides, the summed score of both was used for the analysis.

The selection of the pain provocation tests was based on face validity and intraexaminer reliability.^{2,11,13,21,22,24} The performance of these tests was conducted as described in the literature.^{11,21} The following tests were selected: posterior pelvic pain provocation test, hip internal and external rotation, pelvic torsion, active trunk movements in four directions, pelvic compression in the supine position (“gapping test”) and the side-lying position, sacral thrust, lumbar pressure, and pain at isometric adduction of the hips.

Mobility of the pelvic joints was measured radiographically according to Chamberlain.⁴ Mobility of the lumbosacral spine was assessed by the modified–modified Schober technique.³² Mobility and muscle strength of the trunk were measured on the B200 Isostation (Isotechnologies, Hillsborough, NC) in six directions.

Strength of the hip muscles was measured in three ways: by isometric abduction and adduction, and by the active straight-leg-raise (ASLR) test.^{15–17} The method used to measure abduction and adduction strength was based on a former investigation.³³ Both strengths were measured in newtons using a handheld dynamometer (Microfet; Hoggan Health Industries, Draper, UT) with the subject in the supine position, knees at 90°, and feet placed on the couch. In a pilot study, the intratester reliability of two measurements 1 week apart was acceptable. The intraclass correlation coefficients were 0.79 for abduction and 0.76 for adduction measurement (unpublished data). The ASLR test was performed with the subject in the supine position after the instruction: “Try to raise your legs, one after the other, above the couch 5 cm without bending the knee.” In the first patient group of the current study, the test was scored as 0 (no restriction) or 1 (restriction). The summed score for both legs was used in the analysis. In a pilot study, intertester reliability was 100%.¹⁶ In the second patient group, a 4-point scale was used, as described in a previous study.¹⁶ In a pilot study the intertester reliability of the score was high (Kendall’s Tb = 0.81).¹⁶

Using 48 tests in the same patient would be too strenuous for the patient. Therefore, it was decided to split the battery of tests and perform them in two patient groups: 35 measures

Table 2. Responsiveness of Outcome Measurements in the First Study Group

Outcome Measurement Improved or Not Improved Based on Patient's Global Impression	Standardized Response Mean of Improved Patients	Standardized Response Mean of Not Improved Patients	Significance, Mann-Whitney Test
<i>Activities of daily living</i>			
QBPDS	+1.18	-0.34	<i>P</i> < 0.001
NHP-physical mobility	+0.70	-0.19	<i>P</i> < 0.01
NHP-sexual life	+0.44	-0.09	<i>P</i> < 0.05
<i>Pain</i>			
Morning (VAS)	+0.73	-0.06	<i>P</i> < 0.05
Evening (VAS)	+0.70	-0.32	<i>P</i> < 0.01
NHP	+0.69	-0.07	<i>P</i> < 0.05
Radiation	+0.30	-0.4	<i>P</i> < 0.05
<i>Pain provocation tests</i>			
Posterior pelvic pain provocation test*	+0.67	+0.12	NS
Hip internal rotation*	+0.47	-0.66	<i>P</i> < 0.01
Pelvic torsion*	+0.38	-0.38	<i>P</i> < 0.05
<i>Fatigue</i>			
Evening (VAS)	+0.57	+0.12	NS
<i>Muscle strength (hip)</i>			
Adduction	+1.09	-0.58	<i>P</i> < 0.001
ASLR (2-point scale)*	+0.40	-0.59	<i>P</i> < 0.01

Positive values indicate improvement. Outcome measurements are not listed in case the standardized response mean of the improved patients was below 0.50 and the Mann-Whitney test did not reach the level of significance ($P < 0.05$). Outcome measurements which fulfilled both criteria are printed bold. Variables are ordered within each category by standardized response mean of improved patients.

* Sum score of both sides.

NS = not significant; QBPDS = Québec Back Pain Disability Scale; NHP = Nottingham Health Profile; VAS = Visual Analogue Scale; ASLR = Active Straight Leg Raise Test.

evaluated in the first group and 16 in the second group. Three measures were evaluated in both study groups: the QBPDS, the hip adduction strength assessment, and the ASLR test. The rationale for allocating the tests to the first or second patient group was arbitrary, but based, in part, on logistical reasons.

All the measurements were performed at baseline, then approximately 8 weeks later. Physical examination and hip muscle strength measurements were performed by the main investigator (J.M.). A research assistant (I.R.) performed the trunk muscle strength and mobility measurements, and a radiologist (A.G.) judged the radiographs. The three investigators were blinded to the results of each other, to the results of the first examination, and to the outcome of the self-assessment scales.

The global impression of improvement scored by the patient was used as the standard for assessing the course of the disease. The score was given on a 5-point Likert scale: 1 (much worse), 2 (slightly worse), 3 (unchanged), 4 (slightly improved), and 5 (much improved). The distribution of the answers among the 100 patients of both study groups was as follows: much worse ($n = 1$), slightly worse ($n = 5$), unchanged ($n = 19$), slightly better ($n = 51$), and much better ($n = 24$). Because the much worse and slightly worse groups were very small, and because it would simplify the statistical analysis, the categories 1 to 3 were pooled into one category designated not improved, and categories 4 and 5 into one category designated improved. In fact, the largest part of the not improved group consisted of patients who scored "unchanged."

Statistical Analysis. The global impression of improvement scored by the patient was used as the standard for assessing the course of the disease. Responsiveness was examined by calculating the standardized response mean (the mean change scores divided by the standard deviation of the change score) of the patients who improved and those who did not improve. Cohen⁵ suggested that standardized response means of 0.5 be considered "moderate." Arbitrarily, this value was used in the

current study as cutoff for consideration of a measure as possibly useful clinically. Moreover, a two-tailed Mann-Whitney nonparametric test was used to compare the patients who improved and those who did not improve. A P value less than 0.05 was considered significant.

■ Results

The standardized response mean of improved patients increased 0.5 in 21 of the 48 effect measures. In 20 measures, the mean difference between the patients who improved and those who did not improve reached the significance level of 0.05 (Tables 2 and 3). Both criteria were fulfilled in 15 measures and one of the criteria in 11 measures.

In general, the best scores were reached in the "activities of daily living," "pain," and "hip muscle strength" categories. The highest scores for standardized response mean were reached with adduction strength of the hips (1.09 in the first patient group and 0.96 in the second group) and the QBPDS (1.18 and 1.12, respectively) (Tables 2 and 3).

Only 3 of the 13 NHP subscales were able to assess clinical change in PPPP (Table 2). In the fatigue category, the visual analog scale in the morning and NHP energy did not meet any of the criteria for responsiveness. Only the visual analog scale for fatigue in the evening showed moderate responsiveness for one of the two criteria. Only 3 of the 12 pain provocation tests showed a (weak) relation to clinical improvement.

The change in the scores for mobility of the lumbosacral junction, measured with the modified-modified Schober technique and the radiographically assessed mo-

Table 3. Responsiveness of Outcome Measurements in the Second Group

Outcome Measurement Improved or Not Improved Based on Patient's Global Impression	Standardized Response Mean of Improved patients	Standardized Response Mean of Not Improved Patients	Significance, Mann-Whitney Test
<i>Activities of daily living</i>			
QBPDS	+1.12	-0.06	<i>P</i> = 0.01
Walking time	+1.01	+0.22	<i>P</i> = 0.01
Sitting time	+0.80	-0.17	<i>P</i> < 0.05
Bicycling time	+0.65	+0.15	<i>P</i> < 0.05
Standing time	+0.63	+0.25	<i>P</i> < 0.05
Time laying down	+0.50	+0.54	NS
<i>Joint mobility (spine)</i>			
Rotation*	+1.37	+0.23	<i>P</i> < 0.05
Extension + flexion	+1.28	+0.03	<i>P</i> < 0.05
Side flexion*	+1.15	+0.52	<i>P</i> = 0.05
<i>Muscle strength (hip)</i>			
Adduction	+0.96	-0.07	<i>P</i> = 0.01
ASLR (4-point scale)*	+0.77	-0.18	<i>P</i> < 0.05
Abduction	+0.74	-0.09	<i>P</i> = 0.01
<i>Muscle strength (spine)</i>			
Rotation*	+1.21	+0.66	NS
Side flexion*	+1.05	+0.61	NS
Extension	+0.85	+0.12	<i>P</i> = 0.01
Flexion	+0.73	+0.33	NS

Positive values indicate improvement. Outcome measurements which fulfilled both criteria are printed bold. Variables are ordered within each category by standardized response mean of improved patients.

* Sum score of both sides.

NS = not significant; QBPDS = Québec Back Pain Disability Scale; ASLR = Active Straight Leg Raise test.

bility of the pelvic joints, showed no relation to the patient's global impression of improvement.

■ Discussion

In the two study groups (44 and 56 patients, respectively), responsiveness was studied for a large number of effect measurements. Global impression of improvement reported by the patient was used as the standard. The use of the patient's global impression to judge any change in the clinical situation is debatable. The use of other criteria (*e.g.*, a change in the score on the QBPDS) could yield different results. In general, the global impression of the patient has two dimensions: a change in the status of the disease itself and a change in the way the patient copes with and judges the problems. For example, in patients with arthritis, the correlation between the global impression of the patient and objective clinical parameters was shown to be weak.^{3,25} However, in the absence of objective clinical parameters, the patient's global impression is useful for developing measurement instruments. In the current study, it is noteworthy that many of the effect measurements demonstrated very low responsiveness. However, it is not necessary to abandon all these tests because many are performed for diagnostic or prognostic purposes. For example, radiographs of the pubic symphysis may be helpful for diagnosis, but not for following the progression of the disease during rehabilitation.

In the current study, use of the QBPDS proved to be best for the activities of daily living category. The test is easy to perform. Completing the form and calculating

the score takes 2 to 5 minutes. The NHP has certain disadvantages. Completing the forms takes the patient 10 to 15 minutes, and a computer generally is needed to calculate the scores. Thus, the authors recommend abandoning the use of the NHP to measure clinical changes in PPPP. An advantage of the questions about sitting, walking, and bicycling time is that they allow evaluation of clinical progress by means of simple questions, without the need for forms or a computer.

The relative good responsiveness of the internal rotation of the hips measure, as compared with other pain provocation tests, was surprising. The superiority of this test needs confirmation in a future study. It was striking to observe that in the pain provocation tests category, the responsiveness was acceptable in only 3 of the 12 tests. Of these 12 tests, 6 were scored bilaterally, so the patient was tested 18 times. It seemed that 6 times was more than sufficient. The use of many tests is strenuous for the patient and offers little benefit in terms of evaluating the course of the disease.

The responsiveness of the measures in the fatigue category was relatively low. This is consistent with clinical experience, which shows that fatigue is one of the last symptoms in PPPP to improve.

The mobility of the pelvic joints, as measured radiographically, appears to be insufficient for assessment of clinical change in PPPP. This seems to contradict the theory that PPPP is related to enlarged mobility of the pelvic joints.^{15,16} However, the observation supports the idea that, besides enlarged joint mobility, other factors must play a role in the development of PPPP. Achieving a decrease in the mobility of the pelvic joints by

means of a rehabilitation program probably is impossible, but the current study showed that a decrease in the mobility of the pelvic joints is not necessary for clinical improvement.

The small standardized response mean of flexion mobility (measured with the modified–modified Schober technique) and the large standardized response mean of flexion–extension mobility (measured with the B200 Isostation) seem to be contradictory. A possible explanation is that the modified–modified Schober technique measures only mobility of the lumbosacral junction, and only in flexion direction, whereas with the B200, spine flexibility of the whole trunk is measured, and also in the direction of extension.

The abduction and adduction strength of the hips measurement and the ASLR test are relatively uncommon measurements in low back pain research. It is assumed that patients with PPPP score low on hip muscle strength not only because the involved hip muscles are weak, but also because contraction of these muscles is impaired by pain.¹⁴ A previous study demonstrated that weakness of ASLR could be reduced instantly by fastening a belt around the pelvis.¹⁶

The study patients scored surprisingly well in the range of motion and muscle strength of the spine categories, perhaps because most of the patients were encouraged to do exercises to improve trunk mobility and to reinforce trunk muscle strength. The global impression of the patients could be influenced easily by their ability to perform these exercises.

Most of the evaluated measures seemed to have low power for detecting deterioration in the disease status, possibly because the largest part of the group that was not improved consisted of patients who scored unchanged. The absence of a large group of patients who felt their disease had deteriorated is a limitation of the current study.

The results of the current study need to be confirmed in a validation study. However, such confirmation is less applicable for the QBPDS, the hip adduction strength assessment, and the ASLR test because their usefulness, indicated in the first study group, was confirmed in the second group. These three instruments might be used as standards in future research.

■ Conclusion

When global impression of improvement scored by the patient is used as a standard, it seems possible to gain appropriate information about the course of PPPP with the use of a small test battery. The usefulness of the QBPDS, the hip adduction strength assessment, and the ASLR test was proved in the current study. The value of 23 other instruments was substantiated, but further studies are needed to confirm their usefulness. With 22 of the evaluated measures, the correlation with patient's global improvement was too weak

for them to be recommended as measures of clinical changes over time in PPPP.

■ Key Points

- The Québec Back Pain Disability Scale, the hip adduction strength assessment, and the active straight-leg-raise test are recommended for the measurement of clinical changes over time in patients with posterior pelvic pain since pregnancy.
- In rehabilitation of posterior pelvic pain since pregnancy, a decrease in mobility of the pelvic joints, as assessed radiographically, is not necessary for clinical improvement.

References

1. Berg G, Hammar M, Möller-Nielsen J, et al. Low back pain during pregnancy. *Obstet Gynecol* 1988;71:71–5.
2. Blower PW, Griffin AJ. Clinical sacroiliac tests in ankylosing spondylitis and other causes of low back pain: Two studies. *Ann Rheum Dis* 1984;43:192–5.
3. Buchbinder R, Bombardier C, Yeung M, et al. Which outcome measures should be used in rheumatoid arthritis clinical trials? *Arthritis Rheum* 1999; 42:1568–80.
4. Chamberlain WE. The symphysis pubis in the roentgen examination of the sacroiliac joint. *Am J Roentgenol Radium Ther* 1930;24:621–5.
5. Cohen J. *Statistical Power Analysis for the Behavioural Sciences*. New York: Lawrence Erlbaum Associates, 1988.
6. Deyo RA, Battie M, Beurskens AJ, et al. Outcome measures for low back pain research: A proposal for standardized use. *Spine* 1998;23:2003–13.
7. Erdman RAM, Passchier J, Kooijman M, et al. The Dutch version of the Nottingham Health Profile: Investigations of psychometric aspects. *Psychol Rep* 1993;72:1027–35.
8. Fast A, Shapiro D, Ducommun EJ, et al. Low back pain in pregnancy. *Spine* 1987;12:368–71.
9. Kopec JA, Esdaile JM, Abrahamowicz M, et al. The Québec back pain disability scale. *Spine* 1995;20:341–52.
10. Kristiansson P, Svärdsudd K. Discriminatory power of tests applied in back pain during pregnancy. *Spine* 1996;21:2337–44.
11. Laslett M, Williams M. The reliability of selected pain provocation tests for sacroiliac joint pathology. *Spine* 1994;19:1243–9.
12. Mantle MJ, Greenwood RM, Currey HLF. Backache in pregnancy. *Rheumatol Rehabil* 1977;16:95–101.
13. McCombe PF, Fairbank JCT, Cockersole BC, et al. Reproducibility of physical signs in low back pain. *Spine* 1989;14:908–17.
14. Mens JMA, Snijders CJ, Stam HJ. Diagonal trunk muscle exercises in peripartum pelvic pain: A randomized clinical trial. *Phys Ther* 2000;80:1164–73.
15. Mens JMA, Vleeming A, Snijders CJ, et al. Reliability and validity of the active straight-leg-raise test in posterior pelvic pain since pregnancy. *Spine* 2001;26:1167–71.
16. Mens JMA, Vleeming A, Snijders CJ, et al. The active straight-leg-raising test and mobility of the pelvic joints. *Eur Spine J* 1999;8:468–73.
17. Mens JMA, Vleeming A, Snijders CJ, et al. Validity and reliability of the active straight leg raise test for measuring disease severity in patients with posterior pelvic pain after pregnancy. *Spine* 2002;27:196–200.
18. Michel A, Kohlmann T, Raspe H. The association between clinical findings and self-reported severity in back pain. *Spine* 1997;22:296–304.
19. Nattress CL, Nitschke JE, Disler PB, et al. Lumbar spine range of motion as a measure of physical and functional impairment: An investigation of validity. *Clin Rehabil* 1999;13:211–18.
20. Östgaard HC, Andersson GBJ, Karlsson K. Prevalence of back pain in pregnancy. *Spine* 1991;16:549–52.
21. Östgaard HC, Zetherström GBJ, Roos-Hansson E. The posterior pelvic pain provocation test in pregnant women. *Eur Spine J* 1994;3:258–60.
22. Potter NA, Rothstein JM. Interrater reliability for selected clinical tests of the sacroiliac joint. *Phys Ther* 1985;65:1671–5.
23. Rissanen A, Alaranta H, Saino P, et al. Isokinetic and nondynamometric tests in low back pain patients related to pain and disability index. *Spine* 1994;19: 1963–7.
24. Rudge SR, Swannell AJ, Rose DH, et al. The clinical assessment of sacroiliac

- joint involvement in ankylosing spondylitis. *Rheumatol Rehabil* 1982;21:15-20.
25. Ruta DA, Hurst NP, Kind P, et al. Measuring health status in British patients with rheumatoid arthritis: Reliability, validity, and responsiveness of the Short Form 36-Item Health Survey (SF-36). *Br J Rheumatol* 1998;37:425-6.
 26. Sachs BL, Achmad SS, LaCroix M, et al. Objective assessment for exercise treatment on the B-200 isostation as part of work tolerance rehabilitation: A random prospective blind evaluation with comparison control population. *Spine* 1994;19:49-52.
 27. Schoppink LEM, Tulder MW van, Koes BW, et al. Reliability and validity of the Dutch adaptation of the Québec Back Pain Disability Scale. *Phys Ther* 1996;76:268-75.
 28. Spitzer WO, LeBlanc FE, Dupuis M, et al. Scientific approach to the assessment and management of activity-related spinal disorders: A monograph for clinicians. Report of the Québec Task Force on Spinal Disorders. *Spine* 1987;12:S1-59.
 29. Strender LE, Sjöblom A, Sundell K, et al. Interexaminer reliability in physical examination of patients with low back pain. *Spine* 1997;15:814-20.
 30. Sturesson B, Uden G, Uden A. Pain pattern in pregnancy and "catching" of the leg in pregnant women with posterior pelvic pain. *Spine* 1997;15:1880-3.
 31. Thomas E, Silman AJ, Papageorgiou AC, et al. Association between measures of spinal mobility and low back pain: An analysis of new attenders in primary care. *Spine* 1998;23:343-7.
 32. van Adrichem JAM, van der Korst JK. Assessment of the flexibility of the lumbar spine: A pilot study in children and adolescents. *Scan J Rheumatol* 1973;2:87-91.
 33. van Meeteren J, Mens JMA, Stam HJ. Reliability of strength measurement of the hip with a handheld dynamometer in healthy women. *Eur J Phys Med Rehabil* 1997;7:17-20.
 34. van Tulder MW, Assendelft WJ, Koes BW, et al. Spinal radiographic findings and nonspecific low back pain: A systematic review of observational studies. *Spine* 1997;22:427-34.
 35. Waddell G, Somerville D, Henderson L, et al. Objective clinical evaluation of physical impairment in chronic low back pain. *Spine* 1992;17:617-28.
 36. Wolfe F, Smythe HA, Yunus MB, et al. The American College of Rheumatology 1990 criteria for the classification of fibromyalgia. *Arthritis Rheum* 1990;33:160-72.

Address reprint requests to

Jan M. A. Mens, MD, PhD
Spine & Joint Centre
Westerlaan 10
3016 CK Rotterdam
The Netherlands
E-mail: sjceco@wxs.nl