Chapter 7

Responsiveness of outcome measurements in rehabilitation of patients with Posterior Pelvic Pain since pregnancy

a clinical exploration

Jan MA Mens[1,3], Andry Vleeming[1], Chris J Snijders[2], Henk J Stam[3]

[1] The Spine & Joint Centre the Netherlands
The departments of
[2] Biomedical Physics and Technology,
[3] Rehabilitation Medicine,
of the Faculty of Medicine and Health Sciences, Erasmus University Rotterdam, the Netherlands

Acknowledgments
The authors thank Hugo Duivenvoorden for advice on data analysis.

CHAPTER 7

Abstract

Study Design. Cohort study
Objectives. To develop a test battery to evaluate the course of peripartum pelvic pain (PPPP).
Summary of Background Data. Properly validated scales to evaluate the course of peripartum pelvic pain are hardly available. The use of a large amount of tests is strenuous for the patient and has an unfavorable cost/benefit ratio.
Methods. Responsiveness of 48 effect measures was tested in 100 patients with PPPP; 35 measures were evaluated in 44 patients and 16 tests in 56 patients. The tests were performed at baseline and after 8 weeks treatment. Global impression of improvement scored by the patient (improved/not improved) was used as criterion standard for the course of the disease. Responsiveness was examined by calculating effect sizes and by comparing the mean change in patients who were improved and not improved by means of the Mann-Whitney test. The set of measurement instruments was further reduced by means of regression analysis.
Results. Twenty-one measures showed good properties. By means of regression analysis 7 tests were indicated as the most relevant tests: the Quebec Back Pain Disability Scale (QBPDS), pain in the evening, strength of adduction of the hips, the Active Straight Leg Raise (ASLR) test, pain provoked with mediorotation of the hips, range of trunk flexion/extension, and strength of trunk extension. Mobility of the lumbosacral junction (measured with the modified-modified Schober technique) and mobility of the pelvic joints, (measured radiographically), were unsound measures to assess clinical change in PPPP.
Conclusion. The usefulness of the QBPDS, hip adduction strength and the ASLR test as effect measures in PPPP is substantiated in the present study. The value of 19 other instruments is promising, but further studies are needed to confirm their usefulness.

Introduction

In general, a measurement instrument used to measure the course of a disease should be reliable, valid and responsive. Responsiveness of a measurement instrument is defined as the ability to detect clinical relevant changes over time. The most popular measurements in lumbopelvic pain research are self-reported scales.[6,9] The value of physical examination and radiography to assess disease severity in non-specific lumbopelvic pain is limited.[33] Several scales are based on decreased mobility of the spine and/or the strength of the trunk muscles. Although the reliability of these measurements is high, their relation with clinical parameters is questionable or weak.[12,19,20,24,27,28,30,32,34] Thus, there is still a need for objective tests with high reliability and validity.

Pregnancy is frequently complicated by the occurrence of lumbopelvic pain; the reported cumulative nine-month incidence ranges from 48-56%.[2,7,12,21] Posterior pelvic pain since pregnancy (PPPP) is often described as a distinct category.[10,21,22,31] However, it remains questionable whether PPPP is a specific syndrome, or just non-specific lumbopelvic pain with an onset during pregnancy or delivery. Regardless of the answer, detailed study on the characteristics of PPPP could provide better understanding of lumbopelvic pain in general.

In many cases improvement of the clinical situation is assessed by means of global impressions of both the patient and physician. These impressions are subjective and often influenced by irrelevant factors. A problem arises when their is a discrepancy between the impression of the patient and that of the physician. Moreover, there may be a need to measure the extent of the improvement. To measure the course of the disease with a large amount of tests is strenuous for the patient and has an unfavorable cost/benefit ratio.

The aim of the present study is to develop a test battery that gives appropriate information with a minimum of strain to the patient. In order to avoid inconveniencing a large group of patients with a large group of tests, at first an orientation was made conceirning the responsiveness of a large set of tests in two small populations.

Methods

Study populations

Patients were selected from the outpatient clinic of a rehabilitation center, specialized in treatment of lumbopelvic pain. Patients were treated with instructions with or without home exercises. Two study groups were formed in whom the selection criteria were:
Inclusion criteria:

1. Pain in the lumbopelvic region. Defined as pain experienced between the upper level of the iliac crests and the gluteal fold.
2. Pain beginning during pregnancy or within 3 weeks after delivery.
3. The patient was not pregnant.

Exclusion criteria:

1. A history of fracture, neoplasm or previous surgery of the lumbar spine, the pelvic girdle, the hip joint or the femur.
2. Signs indicating radiculopathy: asymmetric Achilles tendon reflex and/or (passive) Straight Leg Raising restricted by pain in the lower leg.
3. A systemic disease of the locomotor system.
4. Insufficient knowledge of the Dutch language to fill in forms, or any restriction to be tested.

The first group consisted of a 44 patients, mean age 31.7 ± 3.2 years and a mean period postpartum 4.1 ± 2.2 months. The second group comprised 56 patients, mean age 33.5 ± 4.9 years and a mean period postpartum 2.3 ± 3.0 years.

Measurement instruments

The studied tests covered 8 categories: 1) restrictions of activities of daily living (ADL), 2) pain, 3) fatigue, 4) mobility of the spine, 5) mobility of the pelvic girdle, 6) strength of the hip muscles, 7) strength of the trunk muscles, and 8) pain provocation tests. Test were selected on the basis of face validity and intra-examiner reliability. The majority were tests validated for use in non-specific low back pain, and/or sacro-iliac joint dysfunction.

To evaluate restrictions of ADL the Quebec Back Pain Disability Scale (QBDS) was used, 11 sub-scales of the Nottingham Health Profile (NHP), and the interval of time (in minutes, ranging from 0-60) without increase of pain when standing, walking, bicycling, sitting and lying down.[8,9,28]

The mean severity of pain and fatigue was scored on a 100 mm horizontal Visual Analogue Scale (VAS); 0 = 'no pain' to 100 = 'very severe pain'. Because of the large variation in pain and fatigue between morning and evening, both were scored. Pain and fatigue were was also measured on two subscales of the NHP (pain and energy). Radiation of pain was scored by means of the classification as proposed in the Quebec Task Force: no leg pain = 0, pain in the leg, but not beyond the knee = 1, pain in the leg beyond the knee = 2.[29] In case of radiation in both legs the highest score was used for analysis.

Mobility of the lumbosacral spine was assessed by the modified-modified Schober technique.[1] Mobility and muscle strength of the trunk were measured in a B200® isostation (Isotechnologies Inc., Hillsborough, NC, VS) in six directions. Mobility of the pelvic joints was measured radiographically according to Chamberlain.[5]

Strength of hip muscles was measured in three ways: isometric abduction and adduction and by the Active Straight Leg Raise test (ASLR test).[15,16,18] The method used to measure abduction and adduction strength was based on former investigations.[14,18] Both strengths were measured (in newton) with a handheld dynamometer (Microfet®, Hoggan Health Industries Inc., Draper, Utah, USA) in supine position, the knees 90% and the feet placed on the couch. The ASLR test was performed in the supine position

after the instruction: 'Try to raise your legs, one after the other, above the couch for 5 cm without bending the knee.' In the first part of the study the test was scored as 0 = no restriction or 1 = restriction. The sum score of both legs was used in the analysis. In a pilot study intertester reliability was 100%. In the second part a four-point scale was used as described in a previous study.[15]

Pain provocation tests were scored on a modification of the scale proposed by the American College of Rheumatology (ACR) to grade tender points in fibromyalgia: no pain = 0; complaint of pain without grimace, flinch, or withdrawal = 1 (mild); pain plus grimace or flinch = 2 (moderate); the examiner is not able to complete the test because of withdrawal = 3 (unbearable).[35] 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 ACR scale from zero to two. If tests were performed at both sides the sum score of both was used for analysis. Selection of the pain provocation tests was based on face validity and intra-examiner reliability. [3,11,13,23,25,]

In the first group of patients 35 measures were evaluated; 16 in the second group. Three measures were evaluated in both groups:QBPDS, hip adduction strength and the ASLR test. All measurements were performed twice by the same examiner (JM) at baseline and about 8 weeks later. The performer of the first physical examination was not blinded for outcome of the self-assessment scales at baseline. The performer of the second physical examination, the research assistant performing the trunk strength and mobility measurements and the radiologist who judged the X-rays were blinded for each other and for the outcome of the self-assessment scales.

Analysis

Global impression of improvement scored by the patient was used as criterion standard for the course of the disease. Responsiveness was examined by calculating the effect sizes (the mean change scores divided by the standard deviation of the change score) of the improved and not improved patients. Moreover a two-tailed Mann-Whitney nonparametric test was used to compare the improved and not improved patients.

Linear regression analysis (dependent variable global impression of improvement, method backward, at each step variables were included at 5% level and excluded at 10%) was used to reduce the amount of tests to an essential minimum. Only those measurement instruments were entered in the analysis for which the size of the effect (improved versus not improved) was larger than (arbitrarily) 0.5 or for which the Mann-Whitney test was significant at $p < 0.05$. The regression analysis was performed within each of the measurement instrument caregories of the two study groups.

Results

In 21 measures the effect size, improved versus not improved, increased 0.5. In 18 measures the mean difference between improved and not improved reached the

significance level of 0.05 (Tables 1 and 2). In 17 measures both criteria were fulfilled. In general the best scores were reached in the categories pain and hip strength. The highest scores were reached with adduction strength of the hips (effect size 1.67 in the first group and 1.10 in the second group) and QBDS (effect size 1.52 and 1.17 respectively; Tables 1 and 2). The measures in the category fatigue changed in the expected direction, but the effect sizes were relatively low. The change of the scores for mobility of the lumbosacral junction measured with the modified-modified Schober technique and mobility of the pelvic joints showed absolutely no relation with the global impression of improvement.

In each of the eight categories regression analysis was performed. In the category 'hip strength' both the adduction strength and the ASLR test were included in the regression model; in the other categories only one variable was sufficient to represent all the variables of that category.

Discussion

Responsiveness was studied for a large number of measurements in two small study groups. Global impression of improvement given by the patient was used as criterion standard. In general the global impression of the patient has two dimensions: the change in the disease itself and the change in the way the patient copes his problems. The use of patient's global impression to judge the change of the clinical situation is disputable. It has been shown in patients with arthritis, that the correlation between the global impression of the patient and objective clinical parameters is rather weak.[4,26] However in the absence of objective clinical parameters, the patient's global impression is useful to develop measurement instruments. It is noteworthy that many of the measurement instruments in the present study had a very low responsiveness; however, it is not necessary to abandon all these tests because many are performed for diagnostic or prognostic purposes. For example: X-rays of the pubic symphysis may be helpful for diagnosis, however, radiographs to follow the progression of the disease during rehabilitation may not be useful.

In the category ADL the QBPDS overlapped all other scales. The test is easy to perform; completing the form and calculating the score takes 2-5 minutes. A disadvantage of the NHP is that completing the forms takes the patient 10-15 minutes and a computer is generally needed to calculate the scores. An advantage of the sitting, walking and bicycling time is that these measures allow to evaluate clinical progress using simple questions and without the need for forms or a computer.

The effect sizes in the category fatigue were relatively low; this is in agreement with clinical experience that fatigue is one of the last symptoms in PPPP to improve.

The small effect size of the flexion mobility (measured with the modified-modified Schober technique) and the good properties of the 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 flexibility of the spine of the whole trunk is measured and also in extension direction.

Abduction and adduction strength of the hips and the ASLR test are relatively uncommon measurements in low back pain research. We assume that patients with PPPP score low on hip strength not because the involved hip muscles are weak, but because contraction of these muscles is painful.[16] A previous study demonstrated that weakness of ASLR could be (partly) restored instantly by fastening a belt around the pelvis.[15] In the present study both adduction strength and the ASLR test were included in the regression analysis model, indicating that they may not measure the same aspects of the disease. It could be that adduction strength is mainly impaired by symphysial pain and ASLR mainly by sacro-iliac pain.

The good properties of the mediorotation of the hips compared with other provocation tests was surprising. The superiority of this test needs confirmation in a future study. It was striking to observe that in the category pain provocation tests the effect size was acceptable in only three of the 12 tests; 6 of these 12 tests were scored bilateral so the patient was tested 18 times where only mediorotation seemed to be sufficient. Especially in this category the use of many tests is strenuous for the patient and offers little benefit with respect to evaluating the course of the disease.

The categories range of motion and strength of the spine scored surprisingly well; this might be because the majority of the patients were encouraged to do exercises to reinforce trunk muscle strength. The global impression of the patient could easily be influenced by the ability to perform these exercises.

Confirmation of the results of the present study is needed in a future validation study. The need for confirmation is less for the QBPDS, hip adduction strength and the ASLR test because their usefulness, indicated in the first study group, was confirmed in the second. These three latter instruments might be used as criterion standards in future research.

Conclusion

When global impression of improvement scored by the patient was used as criterion standard it seems possible to gain appropriate information about the course of peripartum pelvic pain with the use of a small test battery. The usefulness of QBPDS, hip adduction strength and the ASLR test is proven in the present study. The value of 19 other instruments is promising, but further studies are needed to confirm their usefulness.

Table 1. Responsiveness of outcome measurements in study group 1.

Outcome measurement improved or not improved based on patient's impression	Mean change in scores	SD of the change score	No. of patients	Effect sizes	Effect size improved versus not improved	Significance test
Activities of daily living						
QBPDS						
not improved	- 3.00	8.80	16	- 0.34	1.52	$p < 0.001$
improved	+ 13.04	11.02	28	+ 1.18		
NHP physical mobility						
not improved	- 3.82	20.30	16	- 0.19	0.89	$p < 0.01$
improved	+ 11.55	16.49	28	+ 0.70		
NHP sexual life						
not improved	- 0.06	0.68	16	- 0.09	0.53	$p < 0.05$
improved	+ 0.32	0.72	28	+ 0.44		
NHP emotional reactions						
not improved	- 3.55	13.41	16	- 0.26	0.48	NS
improved	+ 2.88	12.92	28	+ 0.22		
NHP holiday						
not improved	0.00	0.52	16	0.00	0.44	NS
improved	+ 0.29	0.66	28	+ 0.44		
NHP outdoor social life						
not improved	0.00	0.63	16	0.00	0.41	NS
improved	+ 0.29	0.71	28	+ 0.41		
NHP hobby						
not improved	- 0.06	0.77	16	- 0.08	0.35	NS
improved	+ 0.18	0.67	28	+ 0.27		
NHP sleep						
not improved	+ 1.31	22.15	16	+ 0.06	0.28	NS
improved	+ 4.67	13.57	28	+ 0.34		
NHP social isolation						
not improved	+ 0.09	26.03	16	+ 0.00	0.27	NS
improved	+ 4.59	17.22	28	+ 0.27		
NHP home social life						
not improved	+ 0.13	0.81	16	+ 0.16	0.14	NS
improved	+ 0.21	1.69	28	+ 0.30		
NHP domestic activities						
not improved	0.00	0.00	16	0.00	0.00	NS
improved	0.00	0.27	28	0.00		
NHP work						
not improved	0.00	0.37	16	0.00	-0.20	NS
improved	- 0.14	0.71	28	- 0.20		

Table 1 Continuation

Outcome measurement improved or not improved based on patient's impression	Mean change in scores	SD of the change score	No. of patients	Effect sizes	Effect size improved versus not improved	Significance test
pain						
VAS (pain evening)						
not improved	- 5.56	17.11	16	- 0.32	1.02	$p < 0.01$
improved	+ 12.04	17.08	28	+ 0.70		
VAS (pain morning)						
not improved	- 1.19	19.74	16	- 0.06	0.79	$p < 0.05$
improved	+ 12.82	17.51	28	+ 0.73		
NHP pain						
not improved	- 1.82	25.46	16	- 0.07	0.76	$p < 0.05$
improved	+ 17.62	25.62	28	+ 0.69		
Radiation						
not improved	- 0.38	0.96	16	- 0.40	0.70	$p < 0.05$
improved	+ 0.18	0.61	28	+ 0.30		
fatigue						
VAS (fatigue morning)						
not improved	- 3.21	24.00	16	- 0.13	0.46	NS
improved	+ 9.07	27.26	28	+ 0.33		
VAS (fatigue evening)						
not improved	+ 2.71	21.74	16	+ 0.12	0.45	NS
improved	+ 14.26	24.93	28	+ 0.57		
NHP energy						
not improved	- 2.58	34.80	16	- 0.07	0.39	NS
improved	+ 11.37	35.55	28	+ 0.32		
joint mobility (spine)						
Schober						
not improved	+ 0.39	1.38	14	+ 0.28	0.06	NS
improved	+ 0.15	0.67	25	+ 0.22		
joint mobility (pelvic ring)						
Chamberlain radiography†						
not improved	+ 0.17	0.58	12	+ 0.29	0.09	NS
improved	+ 0.24	1.25	17	+ 0.20		
strength (hip)						
adduction						
not improved	- 11.00	19.08	12	- 0.58	1.67	$p < 0.001$
improved	+ 24.48	22.40	23	+ 1.09		
ASLR† (2 point scale)						
not improved	- 0.27	0.46	15	- 0.59	0.99	$p < 0.01$
improved	+ 0.38	0.94	26	+ 0.40		

Table 1 Continuation

Outcome measurement improved or not improved based on patient's impression	Mean change in scores	SD of the change score	No. of patients	Effect sizes	Effect size improved versus not improved	Significance test
pain at provocation						
Hip mediorotation†						
not improved	- 0.50	0.76	14	- 0.66	1.13	$p < 0.01$
improved	+ 0.58	1.23	26	+ 0.47		
Pelvic torsion†						
not improved	- 0.71	1.86	14	- 0.38	0.76	$p < 0.05$
improved	+ 0.73	1.93	26	+ 0.38		
Posterior pelvic pain provocation test†						
not improved	+ 0.14	1.17	14	+ 0.12	0.55	NS
improved	+ 0.92	1.38	26	+ 0.67		
Active trunk flexion						
not improved	- 0.14	0.66	14	- 0.21	0.49	NS
improved	+ 0.15	0.54	26	+ 0.28		
Sacral thrust						
not improved	- 0.21	0.89	14	- 0.24	0.42	NS
improved	+ 0.15	0.83	26	+ 0.18		
Hip laterorotation†						
not improved	- 0.07	1.49	14	- 0.05	0.39	NS
improved	+ 0.38	1.13	26	+ 0.34		
Active trunk side flexion†						
not improved	- 0.14	0.66	14	- 0.21	0.37	NS
improved	+ 0.15	0.92	26	+ 0.16		
Active trunk extension						
not improved	- 0.21	0.89	14	- 0.24	0.34	NS
improved	+ 0.08	0.80		+ 0.10		
Gapping test						
not improved	- 0.07	1.33	14	- 0.05	0.08	NS
improved	+ 0.04	1.18	26	+ 0.03		
Pelvic compression†						
not improved	+ 0.14	1.61	14	+ 0.09	-0.02	NS
improved	+ 0.12	1.07	26	+ 0.11		
Lumbar pressure						
not improved	+ 0.17	1.03	12	+ 0.17	-0.03	NS
improved	+ 0.15	1.05	26	+ 0.14		
Active isometric hip adduction						
not improved	0.00	0.43	12	0.00	-0.14	NS
improved	- 0.08	0.57	25	- 0.14		

Variables are ordered within each category by size of the effect. By means of regression analysis the most important representatives of each category are indicated (printed bold).
† sum score of both sides
NS; non significant
QBPDS = Quebec Back Pain Disability Scale
NHP = Nottingham Health Profile
VAS = Visual Analogue Scale
ASLR = Active Straight Leg Raise test

Table 2. Responsiveness of outcome measurements in study group 2.

Outcome measurement improved or not improved based on patient's impression	Mean change in scores	SD of the change score	No. of patients	Effect sizes	Effect size improved versus not improved	Significance test
activities of daily living						
QBPDS						
not improved	- 0.60	10.60	10	- 0.06	1.18	p = 0.01
improved	+ 13.89	12.35	46	+ 1.12		
Sitting time						
not improved	- 2.80	16.19	10	- 0.17	0.97	p < 0.05
improved	+ 14.18	17.81	45	+ 0.80		
Walking time						
not improved	+ 2.40	10.81	10	+ 0.22	0.79	p = 0.01
improved	+ 17.14	17.02	44	+ 1.01		
Bicycling time						
not improved	+ 0.50	3.27	10	+ 0.15	0.50	p < 0.05
improved	+ 11.79	18.05	43	+ 0.65		
Standing time						
not improved	+ 0.80	3.19	10	+ 0.25	0.38	p < 0.05
improved	+ 10.43	16.58	44	+ 0.63		
Time laying down						
not improved	+ 8.80	16.26	10	+ 0.54	-0.04	NS
improved	+ 7.44	15.03	45	+ 0.50		
joint mobility (spine)						
extension + flexion						
not improved	+ 0.50	18.36	9	+ 0.03	1.25	p < 0.05
improved	+ 15.78	12.34	43	+ 1.28		
rotation†						
not improved	+ 1.74	7.48	9	+ 0.23	1.14	p < 0.05
improved	+ 9.53	6.96	43	+ 1.37		
side flexion†						
not improved	+ 2.69	5.17	9	+ 0.52	0.63	p = 0.05
improved	+ 6.45	5.63	43	+ 1.15		
strength (hip)						
adduction						
not improved	- 2.40	35.48	10	- 0.07	1.03	p = 0.01
improved	+ 41.76	43.54	46	+ 0.96		
ASLR† (4 point scale)						
not improved	- 0.30	1.66	10	- 0.18	0.95	p < 0.05
improved	+ 1.09	1.41	46	+ 0.77		
abduction						
not improved	- 3.60	38.47	9	- 0.09	0.83	p = 0.01
improved	+ 49.20	66.10	47	+ 0.74		

Table 2. Continuation

Outcome measurement improved or not improved based on patient's impression	Mean change in scores	SD of the change score	No. of patients	Effect sizes	Effect size improved versus not improved	Significance test
strength (spine)						
extension						
not improved	+ 2.82	22.85	9	+ 0.12	0.73	p = 0.01
improved	+ 24.90	29.27	43	+ 0.85		
rotation†						
not improved	+ 6.49	9.90	9	+ 0.66	0.55	NS
improved	+ 10.93	9.03	43	+ 1.21		
side flexion†						
not improved	+ 10.59	17.46	9	+ 0.61	0.44	NS
improved	+ 22.82	21.74	43	+ 1.05		
flexion						
not improved	+ 6.18	18.48	9	+ 0.33	0.40	NS
improved	+ 18.21	24.79	43	+ 0.73		

Variables are ordered within each category by size of the effect. By means of regression analysis the most important representatives of each category are indicated (printed bold).
† sum score of both sides
NS; non significant
QBPDS = Quebec Back Pain Disability Scale
NHP = Nottingham Health Profile
VAS = Visual Analogue Scale
ASLR = Active Straight Leg Raise test

References

1. Adrichem JAM van, Korst JK van der. Assessment of the flexibility of the lumbar spine: a pilot study in children and adolescents. Scan J Rheumatol 1973;2:87-91.
2 Berg G, Hammar M, Möller-Nielsen J, Lindén U, Thorblad J. Low back pain during pregnancy. Obstet Gynecol 1988;71:71-5.
3. Blower PW, Griffin AJ. Clinical sacroiliac tests in ankylosing spondylitis and other causes of low back pain - 2 studies. Ann Rheum Dis, 1984;43:192-5.
4. Buchbinder R, Bombardier C, Yeung M, Tugwell P. Which outcome measures should be used in rheumatoid arthritis clinical trials? Arthrtitis Rheum 1995;11:1568-80.
5. Chamberlain WE. The symphysis pubis in the roentgen examination of the sacro-iliac joint. Am J Roentgenol Radium Ther 1930;24;621-5.
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. Fast A, Shapiro D, Ducommun EJ, Friedmann LW, Bouklas T, Floman Y. Low back pain in pregnancy. Spine 1987;12:368-71.
8. Hunt SM, McKenna SP, McEwen J et al. The Nottingham Health Profile: subjective health status and medical consultations. Soc Sci Med 1981;15A:221-9.

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, Pynsent PB. Reproducibility of physical signs in low back pain. Spine 1989;14:908-17.
14. Meeteren J van, Mens JMA, Stam HJ. Reliability of strength measurement of the hip with a hand-held dynamometer in healthy women. Eur J Phys Med Rehabil 1997:7;17-20.
15. Mens JMA, Vleeming A, Snijders CJ, Stam HJ, Ginai AZ. The active straight leg raising test and mobility of the pelvic joints. Eur Spine J 1999;8:468-73.
16. Mens JMA, Vleeming A, Snijders CJ, Stam HJ. Diagonal trunk muscle exercises in peripartum pelvic pain, a randomized clinical trial. Accepted for publication in Physical Therapy.
17. Mens JMA, Vleeming A, Snijders CJ, Stam HJ. Reliability and validity of the active straight leg raise test in posterior pelvic pain since pregnancy. Accepted for publication in Spine.
18. Mens JMA, Vleeming A, Snijders CJ, Stam HJ. Reliability and validity of hip ad- and abduction weakness as measure in posterior pelvic pain since pregnancy. Submitted.
19. Michel A, Kohlmann T, Raspe H. The association between clinical findings and self-reported severity in back pain. Spine 1997;22:296-304.
20. Nattrass CL, Nitschke JE, Disler PB, Chou MJ, Ooi KT. Lumbar spine range of motion as a measure of physical and functional impairment: an investigation of validity. Clin Rehab 1999;13:211-8.
21. Östgaard HC, Andersson GBJ, Karlsson K. Prevalence of back pain in pregnancy. Spine 1991;16:549-52.
22. Östgaard HC, Zetherström GBJ, Roos-Hansson E. The posterior pelvic pain provocation test in pregnant women. Eur Spine J 1994;3:258-60.
23. Potter NA, Rothstein JM. Intertester reliability for selected clinical tests of the sacroiliac joint. Phys Ther 1985; 65:1671-5.
24. Rissanen A, Alaranta H, Saino P, Harkonen H. Isokinetic and non-dynamometric tests in low back pain patients related to pain and disability index. Spine;1994,19:1963-7.
25. Rudge SR, Swannell AJ, Rose DH, Todd JH. The clinical assessment of sacro- iliac joint involvement in ankylosing spondylitis. Rheumatol Rehabil 1982;21:15-20.
26. Ruta DA, Hurst NP, Kind P, Hunter M, Stubbings A. 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.
27. 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.
28. Schoppink LEM, Tulder MW van, Koes BW, Beurskens SAJHM, Bie RA de. Reliability and validity of the Dutch adaptation of the Quebec Back Pain Disability Scale. Physical Ther 1996;76:268-75.
29. 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.
30. Strender LE, Sjöblom A, Sundell K, Ludwig R, Taube A. Interexaminer reliability in physical examination of patients with low back pain. Spine 1997;15:814-20.
31. 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.

32. Thomas E, Silman AJ, Papageorgiou AC, Macfarlane GJ, Croft PR. Association between measures of spinal mobility and low back pain. An analysis of new attenders in primary care. Spine 1998;23:343-7.
33. Van Tulder MW, Assendelft WJ, Koes BW, Bouter LM. Spinal radiographic findings and nonspecific low back pain. A systematic review of observational studies. Spine 1997;22:427-34.
34. Waddell G, Somerville D, Henderson L, Newton M. Objective clinical evaluation of physical impairment in chronic low back pain. Spine 1992;17:617-28.
35. 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.