Chapter 6

Validity and Reliability of reduced hip abduction and adduction strength as diagnostic instrument in Posterior Pelvic Pain since Pregnancy

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 Inge Ronchetti, research assistant, for assessing the test-retest and intertester reliability; and Ronald TM van Kalmthout, MD for examining a large part of the control subjects.

Submitted for publication

CHAPTER 6

Abstract

Study Design. A cross-sectional analysis was performed in a group of women meeting strict criteria for posterior pelvic pain since pregnancy (PPPP). The strength of abduction and adduction of the hips was measured and common severity measurement scales of lumbopelvic pain were scored. The results were compared with the scores of healthy controls.

Objectives. To develop new diagnostic instruments for use in patients with PPPP. The objectives of the present study were to assess the validity and reliability of hip abduction and adduction strength as diagnostic instruments.

Summary of Background Data. Various diagnostic tools are used to categorize patients with PPPP, but there is still a need for simple tests with high reliability, sensitivity and specificity.

Methods. Intra- and intertester reliability of hip abduction and adduction strength measurement were assessed in two groups of 50 women with lumbopelvic pain of various etiology and various degrees of severity. Sensitivity was assessed in 200 patients with PPPP and specificity in 100 healthy women. Validity of abduction and adduction strength as severity scale was investigated by comparing the test scores with the medical history, scores on self-reported scales on disability, pain and tiredness, and pain provocation tests. Sensitivity and specificity of abduction and adduction strength was compared with the Posterior Pelvic Pain Provocation test (PPPP test) and the Active Straight Leg Raise Test (ASLR test); the usefulness as severity scale was compared with the Québec Back Pain Disability Scale (QBPDS) and the score of the ASLR test. The influence of several demographic and anthropometric variables on abduction and adduction strength was investigated.

Results. The test-retest reliability measured with Pearson's correlation coefficient between two abduction and adduction measurements one week apart was 0.82 and 0.75, respectively. The intraclass correlation coefficients were 0.79 and 0.76; intertester reliability measured with Pearson's correlation coefficient for abduction and adduction strength measurement was 0.85 and 0.81, respectively. The intraclass correlation coefficients were 0.67 and 0.82. In the patient group hip abduction and adduction strength averaged 192 ± 73 and 123 ± 61 newton, respectively; in the control group these values were 296 ± 64 and 222 ± 52 newton, respectively. At a specificity level of 0.95 the abduction strength in the control group was 196 and adduction 129 newton. With these cut-off points sensitivity was 0.51 for abduction and 0.61 for adduction. The sensitivity of weakness of hip abduction and adduction strength was lower than the sensitivity of the PPPP test and the ASLR test; advantage of the hip abduction and adduction weakness is that this measurement is more objective. Hip abduction and adduction strength correlated as expected with almost all severity scales. Abduction and adduction strength correlated well with the scores on the QBPDS and the ASLR test (Pearson's correlation coefficients 0.48 - 0.56).

Conclusion. Hip abduction and adduction weakness are suitable phenomena to be used as diagnostic instruments to discriminate between patients who are disabled by PPPP and healthy subjects; the strengths can be recommended as disease severity scales in patients with PPPP. The measurements are easy to perform; reliability, sensitivity and specificity are acceptable. It seems that decreased hip abduction and adduction strength is caused by the inability to use the muscles rather than by weakness of the muscles.

CHAPTER 6

Introduction

The need for diagnostic instruments in lumbopelvic pain to categorize patients in groups with different prognoses and to measure disease severity is extensively described in a former publication.[9] The Posterior Pelvic Pain Provocation test (PPPP test) and the Active Straight Leg Raise Test (ASLR test) seem to be suitable to discriminate between patients who are disabled by posterior pelvic pain since pregnancy (PPPP) and healthy subjects.[4,9,12] The ASLR test is validated as a reliable scale to measure disease severity in patients with PPPP.[9] However, even with the use of thee tests, doubt about diagnosis and severity could persist. In order to increase objectivity a need remains for simple tests with high validity, sensitivity and specificity.

Many patients with PPPP experience pain when the hip muscles are tensed; even in the absence of pain the muscles around the hips are often improperly used or weak. Pain at resisted abduction has been recommended as a reliable test to assess sacroiliac joint dysfunction.[1,6] It has been suggested that weakness of abduction and adduction is caused by fear for pain and/or disturbed proprioception and/or fatigue.[8,15,16] It was decided to investigate the feasibility of using the strength of hip abduction and adduction as diagnostic tools in patients with lumbopelvic pain.

The aim of the present study is to establish the usefulness of reduced abduction and adduction strength as diagnostic instruments in women with PPPP. More specifically, the reliability and the validity of abduction and adduction strength to qualify and quantify PPPP were investigated.

Methods

Subjects

Patients were selected from the outpatient clinic of a rehabilitation center, specialized in treatment of lumbopelvic pain.

Intra- and intertester reliability

Intra- and intertester reliability were assessed in two groups of 50 women with lumbopelvic pain of various etiology and various degrees of severity.

Sensitivity

Patients were selected from those who visited the rehabilitation center for the first time; 200 consecutive patients with PPPP were included. The population is described extensively in a former publication.[9]

Specificity

Control subjects were 100 women who consulted a primary care unit because of local problems of the locomotor system, e.g. tennis elbow, trigger finger, frozen shoulder.

Only those women were included who were without previous medical consultations or time lost from work due to lumbopelvic pain, and scored zero on the Québec Back Pain Disability Scale (QBPDS) at the moment of examination.[3,13] The length of the post partum period, parity and age were not restricted.

Measurements

Abduction and adduction strength

The method to measure abduction and adduction strength was based on a former study.[7] Both strengths were measured (in newton) with a handheld dynamometer (Microfet®, Hoggan Health Industries Inc., Draper, Utah, USA) in supine position, the knees at 90% and the feet placed on the couch. When measuring abduction, the examiner places the dynamometer with his right hand against the lateral aspect of the left knee and holds the right knee by means of his left hand placed against the lateral aspect of the right knee. The patient is asked to spread the legs during 5-7 seconds as forcefully as possible; the examiner holds the knees in position. After 5-7 seconds rest the measurement was repeated two times. When the score of the last measurement was the highest an extra measurement was performed, etc. The highest value of all measurements was used for analysis. In the same position, the device was placed against the medial aspect of the left knee to measure adduction strength. The patient is asked to squeeze the device (and the right hand of the examiner) between the knees.

Intratester reliability

Abduction and adduction strength were measured on two separate occasions by the same assessor with a one-week period in between. The assessor was an experienced examiner (IR) and was blinded for the scores from the previous week, for the results of all other measurements and the patients' medical history. The results were not communicated to the patient.

Intertester reliability

Abduction and adduction strength were measured one hour apart by two experienced assessors (IR and JM). The order of the two sessions was at random. Both assessors were blinded for each other's scores and one of them (IR) was also blinded for the results of all other measurements and the patients' medical history.

Confounders

To assess the presence of variables having an influence on hip abduction and adduction strength (confounders), the influence of the most obvious variables was analyzed in the control group: the length of the postpartum period, age, parity, height and weight.

Sensitivity and specificity

In clinical use a high specificity (with a moderate high sensitivity) is of more value than a high sensitivity (with a moderate high specificity) to judge an individual patient. Therefore the specificity was fixed at 0.95. The cut-off values of abduction and adduction strength and the specificity were calculated at that level.

The sensitivity of both the 50% of the patients with the highest disability and the 50% with the lowest disability was computed. Disability was measured on the QBPDS; this scale was developed to measure the grade of disability in non-specific low back pain and appeared also suitable in patients with PPPP.[3,9,13]

Comparison with the PPPP test and the ASLR test

Because the PPPP test and the ASLR test are well-documented, reliable, sensitive and specific instruments to assess PPPP,[4,9,12] hip abduction and adduction strength was compared with these tests. The PPPP test was performed on both sides; the test was scored positive when pain was provoked on at least one side. The ASLR test was measured on a scale ranging from 0-10.[9] To score the PPPP test, the ASLR test and hip abduction and adduction strength as independently as possible from each other, the sequence was: at first the PPPP test was scored, then the ASLR test was scored by the patient without any help from the examiner; finally hip abduction and adduction strength were measured.

Validity as instrument to measure severity

The criterion-based assessment of validity of disease severity measurement is hampered by the lack of a "gold standard". To test external validity the associations between the scores of the abduction and adduction strength and four types of variables were investigated: 1) medical history variables, 2) self-reported scales on disability, pain and fatigue, 3) pain provocation tests, 4) the ASLR test.

The following medical history variables were expected to correlate with disease severity: radiation of pain to one or both legs, use of pain medication during the previous week, previous consultation of a medical specialist, sleep disturbance because of PPPP, absence of work because of PPPP at the moment of examination.[3,14]

Disability was measured by the QBPDS, the interval of time (in minutes, ranging from 0-60) without increase of pain when standing, walking, bicycling, sitting and lying down. Pain and fatigue were measured on visual analogue scales.

Pain was also assessed by provocation tests. If the test was performed at both sides, the test was scored as positive when pain was evoked at least on one of the two sides. The selection of the pain provocation tests was based on good intertester reliability, specificity and selectivity.[4,5,11,12] The following tests were performed: posterior pelvic pain provocation test, pelvic torsion, tenderness of the long dorsal sacro-iliac ligament, sacral thrust and lumbar pressure.

Statistical analysis

SPSS statistical software was used for data analysis. Intra- and intertester reliability of the measurements were determined by calculating Pearson's correlation coefficient and intraclass correlation coefficient (ICC). The estimates of variance for the ICC were obtained from a one-way random effect model. In the control group a multiple linear regression model was used to examine the presence of confounders. A Chi-square test and Pearson's correlation coefficient were used to investigate the correlation between abduction and adduction strength and the ASLR test and the PPPP test. Associations between abduction and adduction strength and severity scales were determined by calculating Pearson's correlation coefficient. For dichotomous variables the significance of the difference of the means of the scores in both groups were compared using the independent samples t-test. A p-value < 0.05 was considered significant.

Results

It took 15 months to include the 200 patients. Mean age of the patients was 32.7 ± 3.5 years. Parity ranged from 1-6 with a median of 2. Postpartum period ranged from 0.5 - 4.8 years with a median of 1.7 years.

Mean age of the 100 control subjects was 46.5 ± 14.3 years. Parity ranged from 0-11 with a median of 2. Postpartum period ranged from 0.27 - 60.9 years with a median of 19.5 years. No significant difference between the control group and the patient group was found with respect to parity. Age and duration of the postpartum period of the control group were higher than those of the patient group (independent samples t-test $p < 0.001$).

Intratester reliability

Pearson's correlation coefficients between the two measurements made one week apart by the same investigator for abduction and adduction strength were 0.82 and 0.75, respectively; the ICCs were 0.79 and 0.76, respectively.

Intertester reliability

Pearson's correlation coefficients between the two measurements made one hour apart by two investigators for abduction en adduction strength were 0.85 and 0.81, respectively; the ICCs were 0.67 and 0.81, respectively.

Possible confounders

In a multiple linear regression analysis in the control group the multiple correlation coefficient of the combination age, parity, duration of the postpartum period, height, weight with hip abduction and adduction strength was $R = 0.319$ and 0.249, respectively; the adjusted R^2 was 0.039 and -0.004, respectively (Tables 1a and 1b).

Table 1a. Results of multiple linear regression with scores on abduction strength as dependent variable in control subjects.

Predictor Variable	Higher Score Category	Standardized Beta Coefficient	t-Test	P-Value
Age	Older	0.417	1.271	0.208
Parity	Lower	0.178	1.375	0.173
Duration	Shorter	0.482	1.546	0.127
Height	Taller	0.077	0.671	0.505
Weight	Heavier	0.209	1.839	0.070

Multiple correlation coefficient R =0.319. $R^2 = 0.102$. Adjusted $R^2 = 0.039$.

Table 1b Results of multiple linear regression with scores on adduction strength as dependent variable in control subjects.

Predictor Variable	Higher Score Category	Standardized Beta Coefficient	t-Test	P-Value
Age	Older	0.565	1.687	0.096
Parity	Lower	0.162	1.226	0.224
Duration	Shorter	0.534	1.675	0.098
Height	Taller	0.142	1.207	0.232
Weight	Heavier	0.018	0.152	0.880

Multiple correlation coefficient R = 0.249. $R^2 = 0.062$. Adjusted $R^2 = -0.004$.

Specificity / cut-off levels

Mean values for hip abduction and adduction strength in the control group were 296 ± 64 and 222 ± 52 newton, respectively (Figure 1a and 2a). The hip abduction strength ranged from 168 - 606 newton and the adduction strength from 82 - 349 newton. The 5th percentiles were 196 and 129 newton, respectively. These latter values were chosen as cut-off levels to differentiate between patients with PPPP and healthy subjects.

Sensitivity

Mean values for hip abduction and adduction strength in the patient group were 192 ± 73 and 123 ± 61 newton, respectively (Figure 1b and 2b). In 51% of the 200 patients the abduction strength was less than the cut-off point (196 newton); in 61% the adduction strength was less than the cut-off point (129 newton). Thus at these levels sensitivity was 0.51 for abduction weakness and 0.61 for adduction weakness.

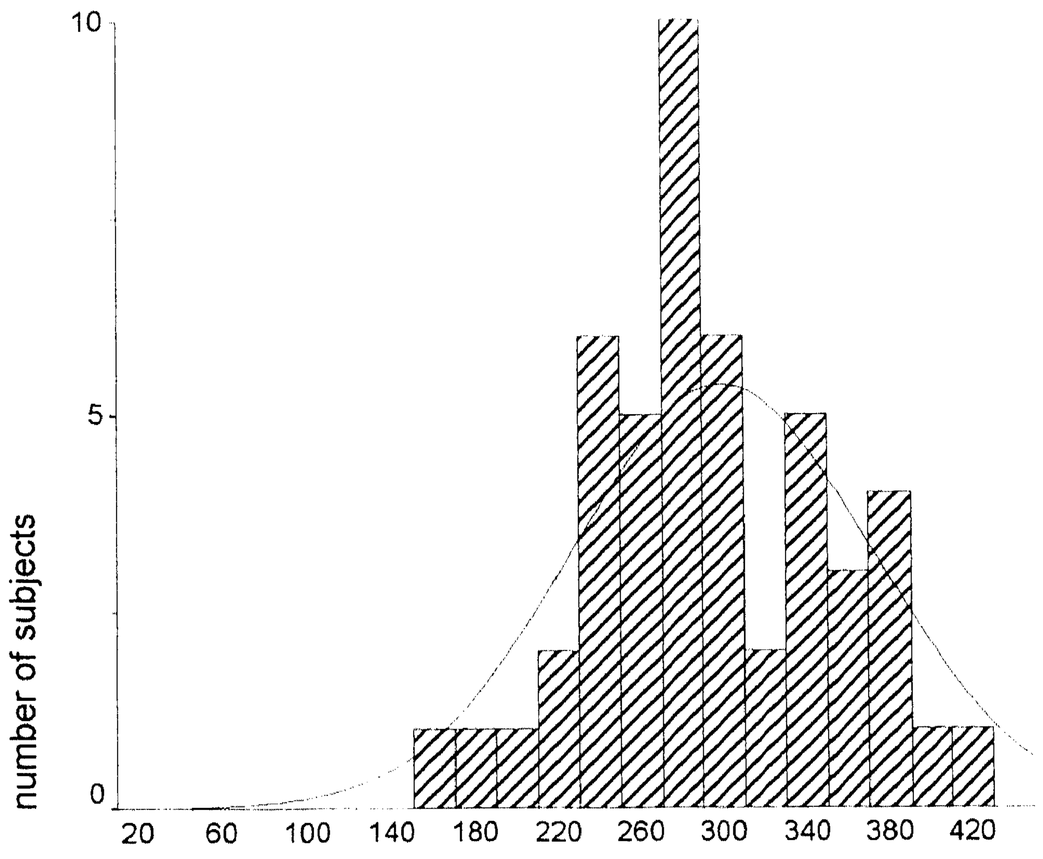

Figure 1a Abduction strength (in newton) of 100 healthy women.

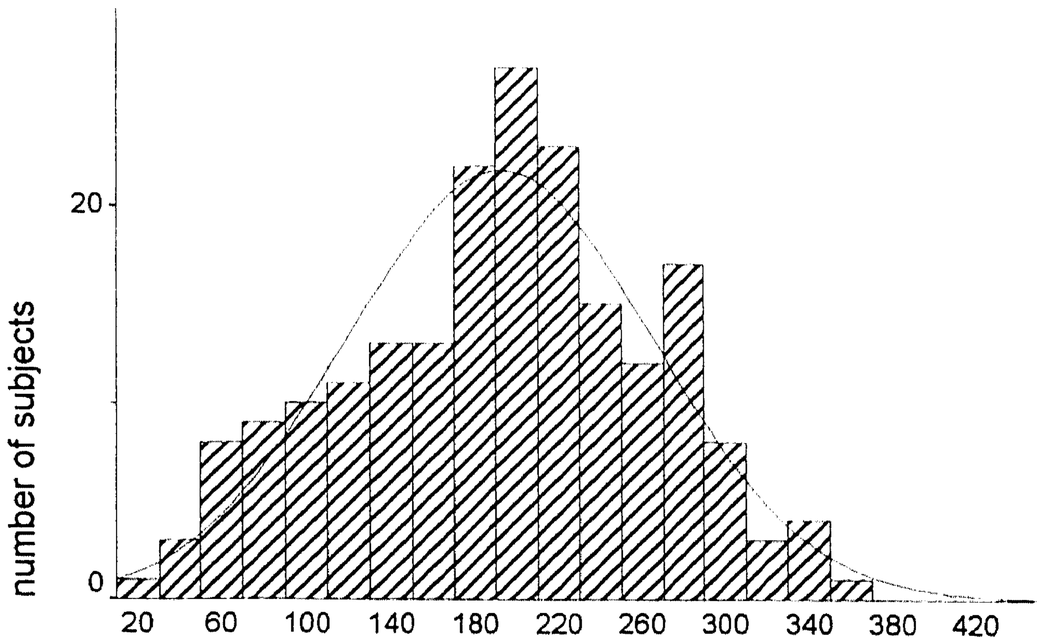

Figure 1b. Abduction strength (in newtons) of 200 patients with posterior pelvic pain since pregnancy.

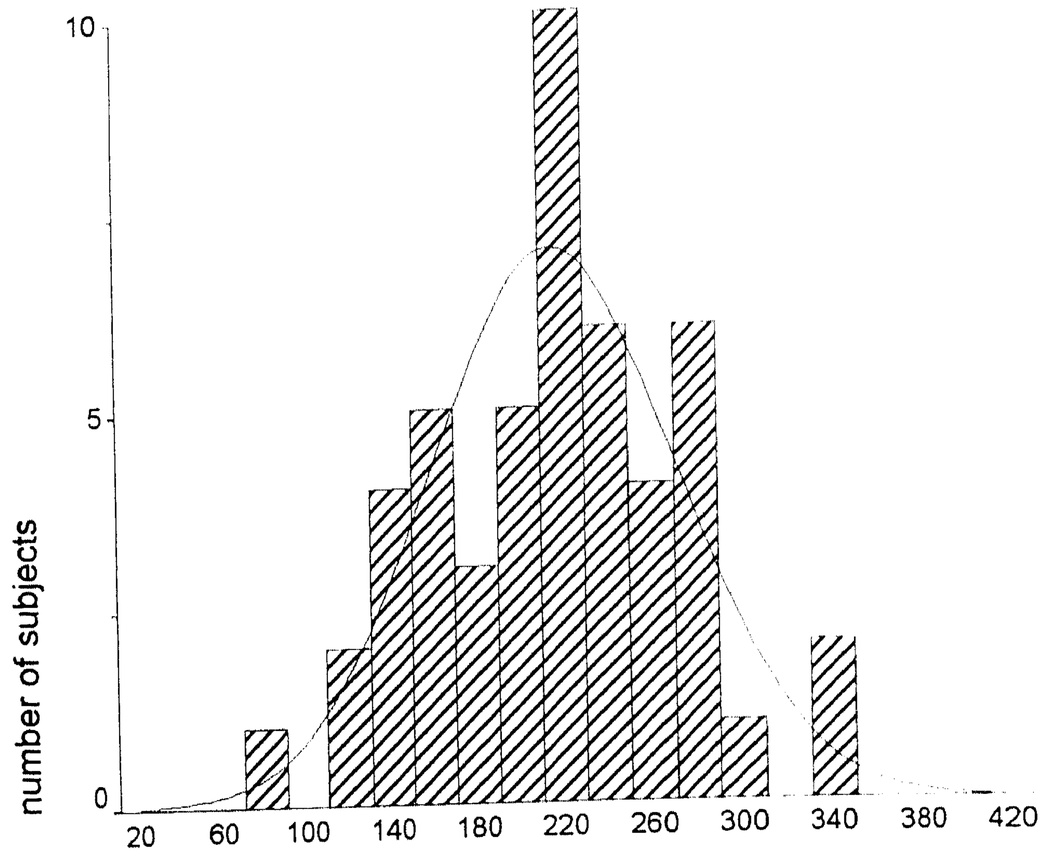

Figure 2a Adduction strength (in newton) of 100 healthy women.

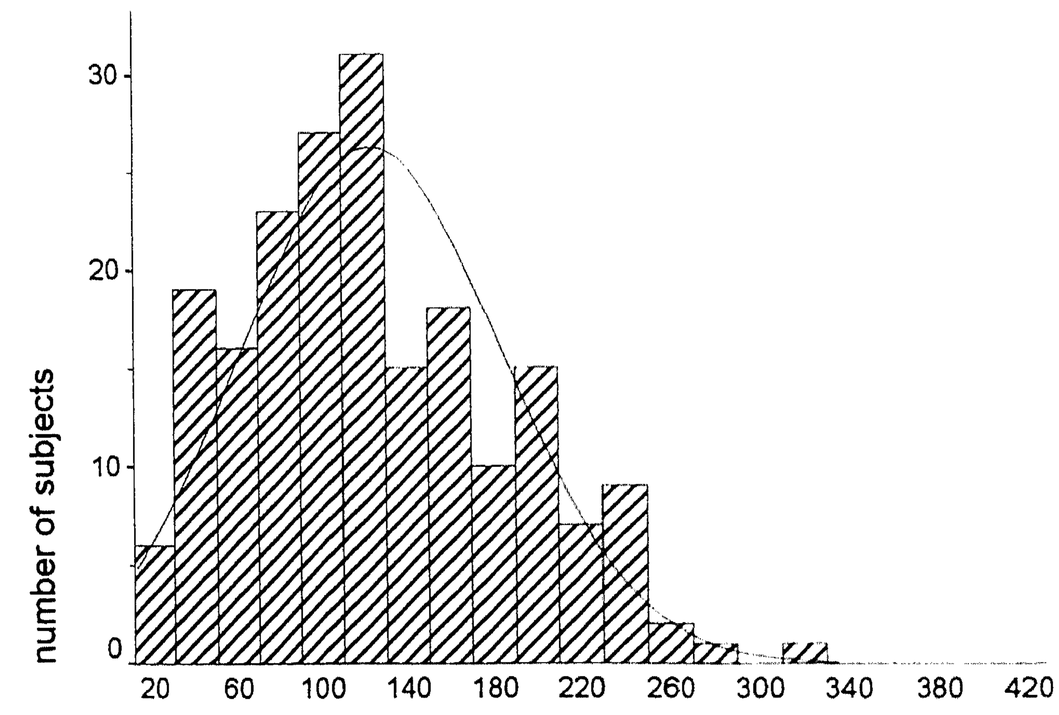

Figure 2b. Adduction strength (in newton) of 200 patients with posterior pelvic pain since pregnancy.

Median value of QBPDS was between 44 and 45. In the 100 patients with a QBPDS score of 45 or higher, sensitivity of abduction and adduction weakness was 0.70 and 0.73, respectively. In the 100 patients with a QBPDS score below 45, sensitivity was 0.31 and 0.48, respectively.

Comparison with the ASLR test and the PPPP test

In 173 patients (87%) the ASLR test was positive (Tables 2a and 2b). Correlations between hip abduction and adduction weakness and the ASLR test were significant. Significance of Pearson's Chi square was $p < 0.001$ and $p < 0.01$, respectively; Pearson's correlation coefficients were 0.31 and 0.22 ($p < 0.001$), respectively.

Tables 2a and 2b. Abduction and adduction weakness versus PPPP test in patients with PPPP.

	PPPP test negative	PPPP test positive	Total
Hip abduction strength normal	44	55	99
Hip abduction strength weak	18	83	101
Total	62	138	200

Pearson's correlation coefficient 0.29 ($p < 0.001$). Pearson's Chi square $p < 0.001$

	PPPP test negative	PPPP test positive	Total
Hip adduction strength normal	32	46	78
Hip adduction strength weak	30	92	122
Total	62	138	200

Pearson's correlation coefficient 0.17 ($p < 0.05$). Pearson's Chi square $p < 0.05$
Data are numbers of patients.

In 138 patients (69%) the PPPP test was positive (Tables 3a and 3b). Correlations between hip abduction and adduction weakness and the PPPP test were significant. Significance of Pearson's Chi square was $p < 0.001$ and $p < 0.05$, respectively; Pearson's correlation coefficients were 0.29 ($p < 0.001$) and 0.17 ($p < 0.05$), respectively. In 13 of the 200 patients (6.5%) all four tests (ASLR, PPPP and weakness of hip abduction and adduction) were negative.

Validity as instrument to measure severity

The values of the abduction and adduction strength correlated as expected with all severity scales (Tables 4 - 7). Scores were significantly higher when pain radiated to one or both legs, when medication to treat pain was used, when a medical specialist was consulted before the visit to the rehabilitation clinic, when sleep was disturbed by pain,