From the Neurotec Department, Division of Physiotherapy,
Karolinska Institutet, Stockholm, Sweden

SACROILIAC PAIN-PROVOCATION TESTING IN PHYSIOTHERAPY

TIME AND FORCE RECORDING

Ulla Levin

Stockholm 2004

ISBN 91-7349-975-7

Printed at **Repro Print AB**, Stockholm

To my family

Abstract

Introduction Some of the tests used for examining pain arising from the sacroiliac region are intended to provoke pain by using manually applied forces. Previous reliability and validity studies of sacroiliac pain-provocation tests show conflicting results, which may be due to lack of standardized force exposure.

Objectives The overall aim of the present work was to investigate force and time interval of force exposure and their influence on the intra- and inter-reliability, validity, sensitivity and specificity of sacroiliac pain-provocation tests.

Methods Two examination tables were used, one with digital scales measuring applied force in a vertical direction and one with two force plates, each capable of recording three orthogonal forces. In the first study three sacroiliac pain-provocation tests were performed by 18 physiotherapists on three occasions, and in a second study two tests were performed by 15 physiotherapists on two occasions, on the same healthy subject. In a third and fourth study one test was performed once by each of three physiotherapists on 11 subjects with verified sacroiliitis. In the third study 11 healthy subjects were also tested. Whether the subject had sacroiliitis or was healthy was unknown to the physiotherapist. The physiotherapists' experience in musculoskeletal evaluation and therapy varied. In the third and fourth study the subjects indicated change in pain by pressing a button causing an audible signal and simultaneously a mark in the data collection, whereupon the test was immediately discontinued.

Results The results of performing the compression test and the distraction test, and of applying pressure on apex sacralis using the digital scales, showed that the intra-examiner reliability was acceptable concerning applied force, ICC [1.1] 0.63, 0.71, 0.74. Inter-examiner reliability was unsatisfactory. In the distraction test and during pressure on apex sacralis, recorded force from the force plate closer to the physiotherapist was significantly less ($P \leq 0.05$) than that from the force plate further away. In the distraction test the lateral and vertical force components differed significantly ($P \leq 0.05$) between occasions. The recorded vertical force component dominated in both tests (subject supine/prone). For the distraction test, the recorded force was significantly smaller ($P \leq 0.05$) and the time interval significantly shorter ($P \leq 0.05$) in the sacroiliitis group than in the healthy group. The time interval of force exposure until pain was provoked varied between 1.8 and 19.5 s. The sensitivity was 0.55 calculated for all three physiotherapists, and varied between them, range 0.55-0.82. Recorded force and time interval of force exposure varied within and between physiotherapists in the sacroiliitis group. Significant differences between physiotherapists were found in the magnitude of the impulse (force · time) for the vertical and lateral force components ($P \leq 0.05$). A significant decrease was found in the amplitude of the vertical force component during the time interval of force exposure ($P \leq 0.05$) for two of the three physiotherapists.

Conclusion The accumulated results of the present work indicated that variation in examination technique within and between physiotherapists, irrespective of experience, explain varying outcomes of sacroiliac pain-provocation tests and, accordingly, varying sensitivity of the tests. Negative sacroiliac pain-provocation testing should be interpreted with caution in clinical situations, unless simultaneously monitoring of applied force is available.

Key words Biomechanics; Clinical assessment; Force, Pain-provocation tests; Physical examination; Reliability; Reproducibility; Sacroiliac joint; Sensitivity; Specificity; Validity

Ulla Levin, PT, Neurotec Department, Division of Physiotherapy, 23100, Karolinska Institutet, SE-141 83 Huddinge, Sweden. E-mail: Ulla.Levin@neurotec.ki.se

ISBN 91-7349-975-7

Sammanfattning

Introduktion Några av de tester som används vid undersökning av sacroiliacaledsregionen avser att provocera smärta via manuellt applicerad kraft. Tidigare reliabilitets- and validitetsstudier av smärtprovokationstester för sacroiliacaleder visar motsägelsefulla resultat, vilket skulle kunna bero på brist på standardiserad kraftapplikation.

Syfte Det övergripande syftet med avhandlingsarbetet var att utforska krafter och tidsintervall samt dess inverkan på intra- och interreliabilitet, validitet, sensitivitet och specificitet vid smärtprovokationstester för sacroiliacaleder.

Metod Två undersökningsbänkar användes, en med en digital våg, som mätte kraft i vertikal riktning och en med två kraftplattor, där var och en mätte tre vinkelräta krafter. I den första studien utfördes tre smärtprovokationstester för sacroiliacaleder av 18 sjukgymnaster vid tre tillfällen och i en andra studie utfördes två tester av 15 sjukgymnaster vid två tillfällen på samma friska person. I en tredje och fjärde studie utfördes ett test en gång av var och en av tre sjukgymnaster på 11 personer med verifierad sacroiliit. I den tredje studien testades också 11 friska personer. Huruvida den undersökta personen var frisk eller hade sacroiliilit var okänt för sjukgymnasten. Sjukgymnasternas erfarenhet av undersökning och åtgärder för led- och muskelrelaterade funktionshinder varierade. I den tredje och fjärde studien indikerades förändring i smärta genom tryck på en knapp, som gav en hörbar signal och samtidigt en markering i datainsamlingen varpå testet omedelbart avbröts.

Resultat Då kraften mättes av den digitala vågen vid kompressionstest, separationstest och vid tryck på apex sacralis visade resultaten att intra-test reliabiliteten var acceptabel för uppmätt kraft, ICC [1.1] 0.63, 0.71, 0.74. Inter-test reliabiliteten var otillfredsställande. Vid separationstest och vid tryck på apex sacralis, uppmätte kraftplattan närmast sjukgymnasten signifikant mindre kraft ($P \leq 0.05$) än kraftplattan längst ifrån. Vid separationstestet uppmättes signifikanta skillnader ($P \leq 0.05$) för den laterala och vertikala kraftkomponenten mellan tillfällen. Den uppmätta vertikala kraftkomponenten dominerade vid båda testerna (personen ryggliggande/magliggande). Vid separationstest var den uppmätta kraften signifikant mindre ($P \leq 0.05$) och tidsintervallet signifikant kortare ($P \leq 0.05$) i gruppen med personer med sacroiliit jämfört med gruppen med friska personer. Tidsintervallet för kraftapplikation till dess smärta provocerades varierade mellan 1.8 och 19.5 s. Sensitiviteten var 0.55 beräknad för samtliga tre sjukgymnaster och varierade mellan dem, 0.55-0.82. Uppmätt kraft och tidsintervall för kraftapplikation varierade inom och mellan sjukgymnasterna i gruppen med patienter med sacroiliit. Det var signifikanta skillnader mellan sjukgymnasterna för magnituden för impulsen (kraft · tid) för den vertikala och den laterala kraftkomponenten ($P \leq 0.05$). Resultaten visade en signifikant minskning i amplituden för den vertikala kraftkomponenten under tiden för kraftapplikation ($P \leq 0.05$) för två av de tre sjukgymnasterna.

Konklusion Det sammanlagda resultatet av avhandlingsarbetet indikerade att variation i examinationsteknik inom och mellan sjukgymnaster, oavsett erfarenhet, kan förklara varierande resultat av smärtprovokationstester för sacroiliacaleder och därmed varierande sensitivitet för testen. Negativa smärtprovokationstester bör tolkas med försiktighet i klinisk verksamhet, så länge den applicerade kraften inte kan avläsas under testets gång.

Nyckelord Biomekanik; Klinisk underökning; Kraft, Smärtprovokationstest; Sjukgymnastik; Reliabilitet; Reproducerbarhet; Sacroiliacaled; Sensitivitet; Specificitet; Validitet

Ulla Levin, leg sjukgymnast, Institutionen Neurotec, Sektionen för sjukgymnastik, 23100, Karolinska Institutet, SE-141 83 Huddinge. E-post: Ulla.Levin@neurotec.ki.se

ISBN 91-7349-975-7

LIST OF PUBLICATIONS

This thesis is based on the following publications, which are referred to in the text by their Roman numerals.

I **Levin U**, Nilsson-Wikmar L, Stenström CH, Lundeberg T.
Reproducibility of manual pressure force on provocation of the sacroiliac joint.
Physiotherapy Research International 1998; 3: 1-14.

II **Levin U**, Nilsson-Wikmar L, Harms-Ringdahl K, Stenström CH.
Variability of forces applied by experienced physiotherapists during provocation of the sacroiliac joint.
Clinical Biomechanics 2001; 16: 300-306.

III **Levin U**, Stenström CH.
Force and time recording for validating the sacroiliac distraction test.
Clinical Biomechanics 2003; 18: 821-826.

IV **Levin U**, Nilsson-Wikmar L, Stenström CH.
Variability within and between physiotherapists' evaluations of sacroiliac pain using distraction test.
Submitted for publication.

Additional data, not previously published, are added.

Please note that Study IV may not be the final version.

Errata and clarifications page 48

CONTENTS

ABBREVIATIONS

AMTI	Advanced Mechanical Technology Incorporated
ANOVA	Analysis of variance
ASCII	American Standard Code for Information Interchange
ASIS	Anterior superior iliac spine
BASDAI	Bath Ankylosing Spondylitis Disease Activity Index
BASFI	Bath Ankylosing Spondylitis Functional Index
BMS	Between-subject variance
CI	Confidence interval
CV	Coefficient of variation
Fx	Caudal/cranial force component
Fy	Lateral/medial force component
Fz	Vertical force component
F_{total}	Total force - $\sqrt{(Fx^2 + Fy^2 + Fz^2)}$
I	Impulse (force · time)
IASP	International Association for the Study of Pain
ICC	Intraclass correlation coefficient
ICF	Classification of Functioning, Disability and Health
kg	kilogram (1kg corresponds to 2.205 pounds)
m	metre (1 m corresponds to 39.37 inches)
LBP	Low back pain
md	median
mm	millimetre (1 mm corresponds to 0.039 inch)
M-VAS	Mechanical visual analogue scale
N	Newton (1 kg corresponds to 9.807 N)
OMT	Orthopaedic Manipulative Therapy
PA	Posteroanterior
PT	Physiotherapist
s	second
s^2	Variance
SD	Standard deviation
SEM	Standard error of mean
SEM/S_W	Standard error of measurement
SI	Sacroiliac
VAS	Visual analogue scale
VPS	Verbal pain scale
WHO	World Health Organization
WMS	Within-subject variance

1 INTRODUCTION

Pain arising from the sacroiliac (SI) and lumbosacral region is a common complaint for which physiotherapists (PTs) are consulted. Some of the examination tests used in assessment procedures are intended to provoke pain by using manually applied forces. In the present work applied forces and time interval of force exposure in SI pain-provocation tests have been systematically evaluated. It is of essential importance to find reliable and valid assessment methods as a prerequisite for adequate interventions. The International Classification of Functioning, Disability and Health (ICF), with its overall aim of providing a unified and standardized language and a framework for the description of health and health-related states, requires not only reliable and valid assessment methods but also the ability to denote the level of severity of the disability.

1.1 International Classification of Functioning, Disability and Health

According to the ICF all health conditions may concern body functions and body structures, activities and participation and the interacting environmental factors (WHO, 2001). The parts and components of the ICF are shown in Figure 1. Personal factors are not classified but are included as they may influence the results obtained in assessments as well as in interventions.

Pain is classified as an impairment of body functions and might influence all components in the ICF.

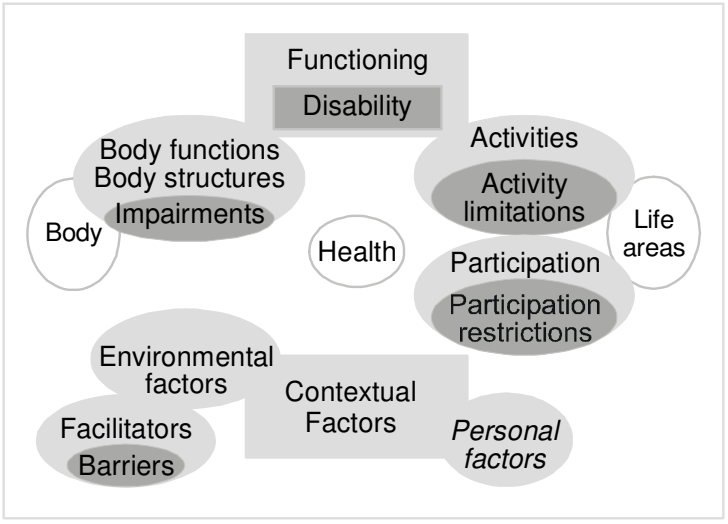


Figure 1 The parts and components of the ICF (WHO, 2001), negative terms (dark).

1.2 Sacroiliac joint pain

Pain has been defined by the International Association for the Study of Pain (IASP) as 'an unpleasant sensory and emotional experience associated with actual or potential tissue damage' (Merskey et al., 1994). Pain caused by inflammation or biomechanically is nociceptive – a result of the activation of primary afferent nociceptive neurons by noxious stimuli in an intact nervous system (Kandel, 2000).

Schwarzer et al. (1995) demonstrated that the SI joint is a likely source of pain in at least 13% and sometimes as many as 30% of patients who consult a clinic specialising in spinal pain. Known causes of SI joint pain include spondyloarthropathy, crystal and pyogenic arthropathies, fractures of the sacrum and pelvis, and diastasis resulting from pregnancy and childbirth (Dreyfuss et al., 1996).

Spondyloarthropathy is the generic term for several disorders which may cause inflammation in the sacroiliac joints, sacroiliitis, which leads to typical inflammatory back pain of varying intensity (Dougados et al., 1991; Braun et al., 2000). Spondyloarthropathy includes ankylosing spondylitis, arthritis of chronic inflammatory bowel disease, psoriatic arthritis, reactive arthritis and undifferentiated spondyloarthropathy (Dougados et al., 1991; Braun et al., 2000; Muche et al., 2003). Symmetrical sacroiliitis is found in more than 80% of patients with ankylosing spondylitis. Sacroiliitis is more often asymmetrical and less severe in early stages in other arthritis forms but can all progress to ankylosing spondylitis (Braun et al., 2000; Muche et al., 2003). Spondyloarthropathy may include asymmetric arthritis in the joints of the lower limbs (Braun et al., 2000).

When pain from the SI joint has no demonstrable lesion, it is presumed to have a biomechanical origin and is termed SI joint dysfunction, synonymous with SI joint syndrome (Dreyfuss et al., 1996). There is some confusion about the theories of SI joint dysfunction and its existence remains controversial. The term is often used to describe pain presumed to be due to asymmetry within the pelvic ring because of stress on associated soft tissues or low-back area caused by e.g. fixation, hypomobility, misalignment or subluxation of the SI joint (Levangie, 1999; Freburger et al., 2001). According to Daum (1995) SI joint dysfunction is acquired with no history of major trauma. The posterior zygapophyseal joints and nerve roots may refer pain to the SI region and might offer another explanation of the pain (Freburger et al., 2001).

1.3 Anatomy

The pelvic girdle is a closed osteoarticular ring composed of the two inominates, the sacrum and the coccyx. It provides a dynamic link between the vertebral column and the lower limbs (Lee, 1999).

There has long been controversy whether the SI joint is a true diarthrodial (synovial) joint (Sashin, 1930; Lee, 1999; Adams et al., 2003) or partly a syndesmosis and partly a synovial joint (Walker, 1992; Harrison et al., 1997; Gray, 2000). Using magnetic resonance imaging Puhakka et al. (2004) showed that the SI joint had the anatomical characteristics of a cartilaginous articulation consistent with a syndesmosis, but with some characteristics of a synovial joint confined to its distal portion. The joint has roughened articular surfaces with great individual variability in configuration according to right and left sides, age and gender (Harrison et al., 1997; Lee, 1999; Gray, 2000). There is also great variability in the orientation of the joint from dorsomedial to ventrolateral (Walker, 1992). The iliac cartilage is partly hyaline and partly fibrocartilage, while the sacral cartilage is hyaline and at least twice as thick as the iliac (McLauchlan et al., 2002; Puhakka et al., 2004). The most superficial aspect of the SI joint is approximately 2 cm below the skin surface, allowing only indirect manual examination (Russell et al., 1981). The joint has resistant and well-developed ligaments, of which the interosseous sacroiliac ligament – the strongest ligament of the body – is intimately blended with the capsule dorsally. It contributes to the stability of the joint dorsally as does the dorsal sacroiliac ligament. The relatively thin and weak ventral sacroiliac ligament

reinforces joint stability ventrally. In addition, the SI joint is indirectly stabilized through the iliolumbar, sacrotuberous and sacrospinous ligaments (Grieve, 1983; Lee, 1999; Pool-Goudzwaard et al., 2003). The connective-tissue systems of pelvis are the same for both sexes (Gerlach et al., 1992). The SI joint allows very restricted translation and rotation (Jacob et al., 1995; Sturesson et al., 2000; Adams et al., 2003).

The innervation of the SI joint is complex, with an abundant intraarticular nerve supply, although there is no clear consensus regarding this. The segmental origin of the nerve fibres differs between individuals and is not always the same bilaterally. Nerve fibres from the plexus, comprising dorsal rami from L5 and S1-S3, supply the ligaments and the skin of the lateral and distal gluteal areas. Ventrally, the joint is innervated by ventral rami from the L5–S2 and by the superior gluteal nerve (Solonen, 1957; Bradley, 1974; Grieve, 1983; Fortin et al., 1999; Yin et al., 2003). The lumbosacral cord lies directly on the lower third of the SI joint and only its thin capsule separates the nerve trunk and the joint space (Sashin, 1930; Grieve, 1983). The superior gluteal nerve and the obturator nerve cross just in front of the capsule, providing an anatomical explanation of pain referred to the abdomen, the hip, the groin, lateral thighs and other parts of the lower extremities (Norman, 1968). The SI joint ligaments contain mechanoreceptors, and both unmyelinated and myelinated axons are present in the capsule (Fortin et al., 1999; Vilensky et al., 2002). The SI joint may be sensitive to abnormal loading, excessive movement and inflammation, and may be a source of pain mediated by the nerves in the SI joint tissue (Fortin et al., 1999). Thus conditions obviously exist for provoking SI pain with manually applied forces.

1.4 Measurement properties

Measurement properties include reliability, validity, and the ability of an assessment method to detect changes (Finch et al., 2002).

A *reliable* measure must provide consistent values with only small errors, but also be able to differentiate among subjects (Finch et al., 2002). Terms used synonymously with *reliability* include *agreement, consistency, precision*, *repeatability* and *reproducibility*. *Agreement* is defined as the extent to which a measurement or a test procedure yields the same results among individuals over time: the term is equivalent to *inter-examiner reliability* (Johnston et al., 1992). Harms et al. (1997) define *repeatability* as the *consistency* with which a therapist can repeat a given technique under the same experimental conditions, and *reproducibility* as the ability of the therapist to repeat a procedure either at a different time, or under different experimental conditions. The terms *repeatability* and *reproducibility* are equivalent to *intra-examiner reliability*. Sources of error, random (chance) or systematic, may be due to subject, examiner, occasion and/or instrument variability. Variance is the statistical term used to describe the variability in data, and analysis of variance (ANOVA) is often used to calculate the different sources of variance (Finch et al., 2002).

Validity is commonly regarded as the degree to which a test actually measures what it is intended to measure (Johnston et al., 1992; Finch et al., 2002). The aspects of *validity* focused in the present work are *face validity*, *known-groups construct validity* and, to some extent, *clinical validity*. *Face validity* simply determines whether a measure appears to measure what it is intended to measure. *Known-groups construct validity* refers to validation of two or more groups that represent different levels of the functioning or disability of interest. *Clinical validity* is defined as the degree to which a measure is useful for specific clinical procedures for individuals. *Criterion validity* has been evaluated in some earlier studies comparing the

results of anaesthetic SI joint injections with the results of SI pain-provocation tests in patients presumed to have SI joint dysfunction (Dreyfuss et al., 1996; Maigne et al., 1996; Broadhurst et al., 1998; Slipman et al., 1998; Ferrante et al., 2001; Laslett et al., 2003; Young et al., 2003). *Criterion validity* is the degree to which a test correlates with an external criterion at approximately the same time (Finch et al., 2002).

Sensitivity describes the proportion of subjects with a disability who have a positive test result for the disability, and *specificity* is the proportion without a disability who have a negative result for the disability. A measure that can correctly identify every subject who has the actual disability has a sensitivity of 1.0 and a measure that can correctly identify every subject who does not have the actual disability has a specificity of 1.0. *Positive predictive value* is the proportion of individuals correctly identified by the test as actually having the disability, and *negative predictive value* is the proportion of those subjects testing negative who actually do not have the disability (Sackett et al., 1991; Fletcher et al., 1996; Altman, 1991; Finch et al., 2002).

1.5 Assessment of sacroiliac joint pain

A physiotherapeutic assessment of SI joint pain includes, in addition to the patient's history, a physical examination and self-reports of pain and related disabilities.

Manual tests for SI joint pain used in physical examination can be divided into pain-provocation tests aimed to stress the structures and thus reproduce the pain, and palpation tests for assessing position or movement (Laslett et al., 1994). The present work focused on pain-provocation tests, which have greater positive reliability than palpation tests (van der Wurff et al., 2000a).

Pain is always a subjective experience, and the primary source of its assessment should if possible be self-reports (American Pain Society, 2002). Methods for assessing different dimensions of pain include visual analogue scales, numeric scales or verbal scales for rating intensity, duration and/or affect (Huskisson, 1974; Borg, 1982; Price et al., 1994), body drawings on which to mark pain location (Margolis et al., 1988), and questionnaires with scales for scoring the effect of pain on activities (Calin et al., 1994). Pain may also be assessed with multidimensional questionnaires including several of the above-mentioned aspects (Melzack, 1975; Kerns et al., 1985).

1.6 Reliability and validity studies of sacroiliac pain-provocation tests

The SI pain-provocation tests most common in reliability and validity studies are shown in Table I. Each has at least three names; the names used in the following text are in bold in the table. How to perform the tests is not always described and the magnitude of the applied force is stated only in one study (Kokmeyer et al., 2002). The directions of applied force differ e.g. in the descriptions of the distraction test and so does the hand placement for applying pressure on apex sacralis. Consequently examiners probably apply force in different directions when performing the same SI joint test and also have different opinions on how to perform the tests (Figure 2). Further, there are many names for the tests such as 'compression test', 'separation test' and 'distraction test' (Table I). This circumstance is misleading as applied force in all SI pain-provocation tests undoubtedly causes stress in the structures by compression, tension and shear in different parts of the SI joint, the lumbosacral junction and surrounding soft tissues ventrally and posteriorly. Many of the tests certainly also affect the lumbar spine and the hips (Table I). In a few studies training sessions were conducted to ensure standardisation of

technique, but how these sessions were conducted or how the technique was judged to be standardized is not mentioned (Laslett et al., 1994; Albert et al., 2000; Laslett et al., 2003; Young et al., 2003). The applied force was recommended by Lee (1999) to be maintained for 20 s. Quick pressure was applied by Toussaint et al. (1999) while Kokmeyer et al. (2002) maintained the applied force for about 5 s. No further time intervals of force exposure were reported.

The inclusion criteria varied in the reliability and validity studies as did the examiners' professions and experience; and numbers are not always mentioned (Table II). In addition the SI pain-provocation tests evaluated in the studies varied in number from one to six (Table III). In the validity studies the SI pain-provocation tests have mainly been used for evaluating SI joint pain. Only in a few studies have the tests been used in an effort to make the diagnosis (Table III). Pain location was assessed only in four studies (Laslett et al., 1994; Dreyfuss et al., 1996; Ferrante et al., 2001; Laslett et al., 2003) and pain responses using rating scales were assessed only in the studies with anaesthetic SI joint injections (Table III). In the remaining studies the subjects answered yes or no to indicate a positive or negative SI pain-provocation test. No further self-reports were obtained. There also appeared to be confusion whether studies were designed to test the reliability or the validity of SI pain-provocation tests (van der Wurff et al., 2000a; van der Wurff et al., 2000b). All the factors mentioned above point to large differences in the conditions for the studies. Accordingly, the conflicting results of previous reliability and validity studies are not surprising (Table III).

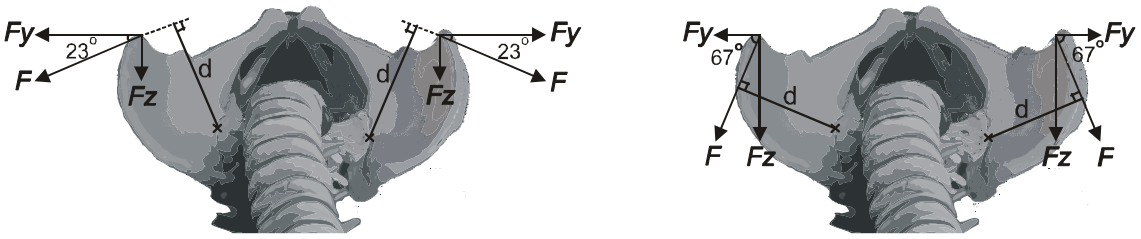


Figure 2 Schematic illustrations of different performances of the distraction test; lateral force vector (Fy), vertical force vector (Fz), resultant force vector (F), lever arm (d).

Table I Definitions of the most common sacroiliac (SI) pain-provocation tests in reliability and validity studies. The names used in the text are in bold. * Index letters in left column explained in right column.

Pain-provocation SI joint tests	Position and performance
Compression test (Laslett et al., 1994; Maigne et al., 1996; Strender et al., 1997 [a] *; Kokmeyer et al., 2002; Laslett et al., 2003; Young et al., 2003) • Approximation (Magee, 2002) • Iliac compression test (Potter et al., 1985) • Lateral pelvic compression (Russell et al., 1981; Blower et al., 1984) • Separation test (Albert et al., 2000) • Transverse posterior distraction-anterior compression (Lee, 1999)	**Lying on one side** Downward pressure on the uppermost iliac crest. [a] Prone
Distraction test (Laslett et al., 1994 [b,c]; Kokmeyer et al., 2002 [b,c]; Laslett et al., 2003 [b,d]; Young et al., 2003 [b,c]) • Anteroposterior pelvic pressure (Russell et al., 1981 [b]) • Gapping test (Potter et al., 1985 [b,c]; Albert et al., 2000 [c]; Magee, 2002 [b,c]) • Iliac compression test (Toussaint et al., 1999 [b,c]; Albert et al., 2000 [c]) • Pressure on anterior superior iliac spines (Blower et al., 1984 [e]) • Transverse anterior distraction-posterior compression (Lee, 1999 [b,c])	**Supine** Cross-armed pressure on anterior superior iliac spines posteriorly [b] and laterally [c]. [d] Pressure without arms crossed. [e] No position or performance defined.
Gaenslen's test (Dreyfuss et al., 1996; Maigne et al., 1996; Ferrante et al., 2001; Kokmeyer et al., 2002; Laslett et al., 2003; Young et al., 2003) • Stress test (Russell et al., 1981) • Pelvic torsion test (Laslett et al., 1994)	**Supine** Maximum flexion of the hip on one side; opposite hip extended with a light downward pressure.
Patrick's test (Dreyfuss et al., 1996; Maigne et al., 1996; Strender et al., 1997; Ferrante et al., 2001; Kokmeyer et al., 2002) • FABER test (Broadhurst et al., 1998) • Patrick's fabere test (Albert et al., 2000)	**Supine** Knee flexed; flexion, abduction and external rotation of the hip.
Pressure on apex sacralis (Russell et al., 1981 [f]) • Long dorsal sacroiliac ligament test (Lee, 1999 [f]) • Midline sacral thrust (Dreyfuss et al., 1996 [g]) • Pressure on lower half of sacrum (Blower et al., 1984) • Sacral apex pressure test (Magee, 2002) • Sacral pressure test (Maigne et al., 1996 [h]) • Sacral thrust (Laslett et al., 1994 [f]; Young et al., 2003 [f])	**Prone** Hand placement differs on sacrum. [f] Anterior force with heel of hand over sacrum. [g] Posteroanterior force. [h] Rapid short-amplitude vertical thrust.
Thigh thrust (Dreyfuss et al., 1996; Kokmeyer et al., 2002; Laslett et al., 2003; Young et al., 2003) • Posterior pelvic pain provocation test (Östgaard et al., 1994; Albert et al., 2000) • Posterior shear test (Laslett et al., 1994; Broadhurst et al., 1998)	**Supine, hip 90° flexion and knee relaxed** Downward force through femur.

Table II Criteria and examiners in reliability and validity studies of SI pain-provocation tests.

Studies	Inclusion criteria for subjects	Examiners
Russell et al., 1981	LPB > 6 months	1 of 3 authors
Blower et al., 1984	Half ankylosing spondylitis, half low back pain (LBP) (between 12th rib and gluteal fold) of mechanical/degenerative origin.	2 authors - physicians?
Potter et al., 1985	Unilateral buttock pain < 1 year, paralumbar pain or pain into posterolateral lower extremity to ankle.	4 PTs
McCombe et al., 1989	A group from standard orthopaedic referral practice and one group with LBP.	2 orthopaedic surgeons and 1 PT
Laslett et al., 1994	Unilateral LBP or buttock pain with or without radiation below knee.	1 PT for all and 1 blinded PT selected from a pool of 5
Östgaard et al., 1994	Pregnancy with LBP with or without radiation into calf or foot, tenderness of back muscles, pain before first pregnancy. Also women without pain.	2 PTs, blinded
Dreyfuss et al., 1996	Pain below L5 consistent with characteristic of sacroiliac (SI) joint pain in previous studies.	Examiners
Maigne et al., 1996	Chronic unilateral LBP with or without radiation to the posterior thigh, compatible to SI joint pain.	Physicians?
Strender et al., 1997	'Suitable patients' from a clinic specializing in back pain.	2 PTs and 2 physicians
Broadhurst et al., 1998	Pain below the lumbosacral junction, worse going down hills or inclines, LBP associated with groin pain, absence of lumbar symptoms.	Physicians?
Slipman et al., 1998	LBP or buttock pain regardless of associated hip or leg symptoms.	2 physicians?
Toussaint et al., 1999	Not reported	Physicians?
Albert et al., 2000	33rd week of gestation, diseased/healthy.	8 PTs
Ferrante et al., 2001	SI syndrome	Physicians?
Kokmeyer et al., 2002	LBP for < 2 weeks to >6 months or asymptomatic for LBP.	2 final-year PT students, blinded
Laslett et al., 2003	Buttock pain with or without lumbar or lower extremity symptoms.	PTs with McKenzie credits
Young et al., 2003	Chronic lumbar or lumbopelvic pain.	PTs with McKenzie credits, blinded

Table III Positive (+) and negative (-) results of sacroiliac (SI) pain-provocation tests used for evaluating SI joint pain or making diagnosis in reliability and validity studies. Anaesthetic SI joint injections were used as an independent, criterion standard.

	Pain	Diagn	Using anaesth inject	Compr test	Distr test	Gaenslen´s test	Patrick´s test sacralis	Pressure apex	Thigh thrust
Reliability									
Blower et al., 1984		•		+	+			+	
Potter et al., 1985	•			+	+				
McCombe et al., 1989	•			-	-				
Laslett et al., 1994	•			+	+	+		-	+
Dreyfuss et al., 1996	•		•			+	+	-	+
Strender et al., 1997	•			-			-		
Albert et al., 2000	•			+	+		+		+
Kokmeyer et al., 2002	•[a]			+	+	+	+	+	+
Validity									
Russell et al., 1981		•		-	-	-		-	
Blower et al., 1984		•		-	+			+	
Östgaard et al., 1994	•								+
Dreyfuss et al., 1996	•		•			-	-	-	-
Maigne et al., 1996	•		•	-	-	-	-	-	
Broadhurst et al., 1998	•	•	•				+		+
Slipman et al., 1998	•		•			-	-		
Toussaint et al., 1999	•				-				
Albert et al., 2000	•	•		-	-		+		+
Ferrante et al., 2001	•		•			+	+		
Laslett et al., 2003	•[a]		•	+	+	+		+	+
Young et al., 2003	•[a]		•	+	+	+		+	+

[a] + presupposes combinations of 3 to 5 tests

1.7 Relevance of the present work

One reason for the conflicting results in reliability and validity studies of SI pain-provocation tests might be different etiologies of SI joint pain. Another might be differing examination techniques, including the directions and magnitudes of applied force and the time interval of force exposure. The lack of data concerning applied force and its time interval in previous studies prevents comparisons of, and generalisations from, the present results. However, there is an increasing demand for reliable and valid physiotherapeutic assessment methods including the ability to denote the presence and severity of the disability. This and the need for an easily-performed manual examination of SI joint pain represented a challenge to record the magnitude of applied force and the time interval of force exposure as a step towards standardized testing, thus improving reliability and validity.

2 AIMS

The general aim was

- to investigate forces and time interval of force exposure and their influence on intra- and inter-examiner reliability, validity, sensitivity and specificity during SI pain-provocation tests in healthy subjects and subjects with sacroiliitis.

Specific aims were

- to evaluate PTs' ability to maintain a constant pressure force and the intra- and inter-examiner reliability of applied vertical force during distraction testing, compression testing and pressure on the apex sacralis.
  (Study I)

- to record applied forces intended for the right and left SI joints, respectively; to evaluate the consistency of distributed forces, whether PTs distributed forces in a similar way, and to compare the recorded force distribution with the PTs' perceptions of the directions in which they applied forces during distraction testing and pressure on the apex sacralis.
  (Study II)

- to investigate how long (seconds) applied force has to be maintained to provoke pain, and whether applied force and time interval of force exposure during the distraction test could discriminate subjects with pain caused by sacroiliitis from healthy subjects and be used as a technique for evaluating pain.
  (Study III)

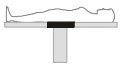
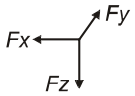
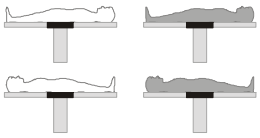
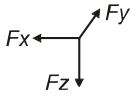
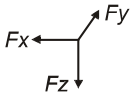
- to determine whether applied force and time interval of force exposure vary within and between PTs when testing subjects with sacroiliitis and whether reported pain varies during sacroiliac distraction testing by different PTs.
  (Study IV)

3 METHODS

3.1 Overview of Studies I-IV

An overview of PTs, SI pain-provocation tests, persons examined, instrumentation and recorded directions of the tests involved in all studies is given in Table IV.

Table IV Overview of physiotherapists (PTs), sacroiliac (SI) pain-provocation tests, examinees, instrumentation (force plate) and recorded directions of applied forces in Studies I-IV.

Study	PT	SI test	Examinee	Force plate	Direction
Study I	Performing SI tests often (n=7) Performing SI tests seldom or never (n=11)	Compression Distraction Pressure on apex sacralis	 Healthy (n=1)		
Study II	OMT [a] certificate (n=8) Without OMT certificate (n=7)	Distraction Pressure on apex sacralis	 Healthy (n=1)		
Study III	Experienced (n=3)	Distraction performed from both sides of the examinee	 Healthy (n=11) Sacroiliitis (n=11)		
Study IV	Experienced (n=3)	Distraction	 Sacroiliitis (n=11)		

[a] OMT – Certificate of Orthopaedic Manipulative Therapy

3.2 Participants

Eighteen PTs in Study I and 15 PTs in Study II examined one healthy woman. Three PTs in Study III examined 11 healthy subjects and 11 subjects with sacroiliitis, verified from radiography or magnetic resonance imaging in Studies III-IV. The characteristics for the PTs and examinees are shown in Table V. All PTs were right-handed. Mean duration of SI joint pain in the sacroiliitis group (Studies III-IV) was 13.8 years (SD 6.1) and medication use was unchanged during the test period. Ten subjects had bilateral SI joint pain and one unilateral. One male had bilateral hip involvement and one unilateral; one had bilateral knee involve-

ment and one unilateral. Median pain on mechanical visual analogue scale (M-VAS) was 18 mm (range 0-73). The median scores on all questions in the Bath Ankylosing Spondylitis Disease Activity Index (BASDAI) and the Bath Ankylosing Spondylitis Functional Index (BASFI) in Studies III-IV are shown in Figure 3.

Table V Characteristics of physiotherapists (PTs) and examinees in Studies I-IV, showing mean (SD)/median (range) for clinical experience, age, weight and height.

Participants	**Clin. experience years**	**Age years**	**Weight kg**	**Height m**
Study I				
PTs	mean (SD)	mean (SD)	mean (SD)	mean (SD)
7 PTs, SI tests often (♀)	18.9 (5.4)	41.1 (5.6)	64.7 (11.2)	1.70 (0.1)
11 PTs SI tests seldom or never [a] (10 ♀)	10.2 (11.6)	39.6 (9.8)	60.3 (4.4)	1.68 (0.1)
Examinee				
1 healthy subject (♀)		42	70	1.75
Study II				
PTs				
8 PTs with OMT [b] (7 ♀)	20.6 (11.2)	48.9 (10.9)	65.2 (10.2)	1.70 (0.1)
7 PTs without OMT (♀)	17.9 (12.5)	41.7 (11.3)	60.9 (6.1)	1.67 (0.1)
Examinee				
1 healthy subject (♀)		46	72	1.75
Study III	range	range	range	range
3 PTs (♀)	24-30	46-61	53-84	1.62-1.71
Examinees		md (range)	mean (SD)	mean (SD)
11 healthy subjects (1 ♀)		30 (27-51)	86.0 (12.5)	1.84 (0.1)
11 subjects with sacroiliitis (1 ♀)		36 (29-69)	82.5 (12.8)	1.80 (0.1)
Study IV	range	range	range	range
3 PTs (♀)	24-30	46-61	53-84	1.62-1.71
Examinees		md (range)	mean (SD)	mean (SD)
11 subjects with sacroiliitis (1 ♀)		36 (29-69)	82.5 (12.8)	1.80 (0.1)

[a] One PT was excluded as she was unable to participate on the second test occasion
[b] OMT – Certificate of Orthopaedic Manipulative Therapy

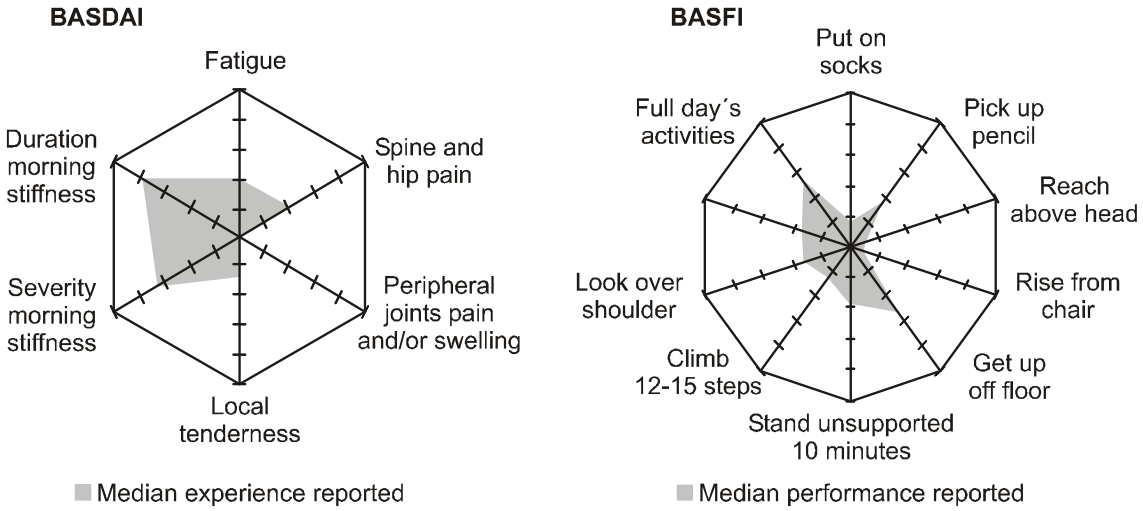


Figure 3 Median score for each question in Bath Ankylosing Spondylitis Disease Activity Index (BASDAI) and Bath Ankylosing Spondylitis Functional Index (BASFI), respectively, of all subjects with sacroiliitis (0 at centre, 10 at periphery). Studies III-IV

3.3 Sacroiliac pain-provocation tests

The following three tests were performed.

The compression test (Figure 4). With the examinee side-lying, a vertical force was applied on the uppermost margin of the iliac crest. The test is presumed to distract the posterior structures of the SI joints and to compress the anterior parts. (Lee, 1999; Magee, 2002).

The distraction test (Figure 5). With the examinee lying supine, a posterolateral cross-armed force was applied on both anterior superior iliac spines (ASIS). The test is presumed to distract the anterior ligaments and capsules of the SI joints and to compress the posterior parts (Lee, 1999; Magee, 2002).

Pressure on the apex sacralis (Figure 6). With the examinee lying prone an anterior force was applied on the apex sacralis thus forcing the sacrum to counternutate. The test is presumed to stress the structures of the SI joints and of the lumbosacral junction. (Lee, 1999; Magee, 2002).

The examinee was instructed to ignore any discomfort from the applied force of the PT's hands on her body. The tests were performed from one side or from both sides of the examinee (Table IV, VI).

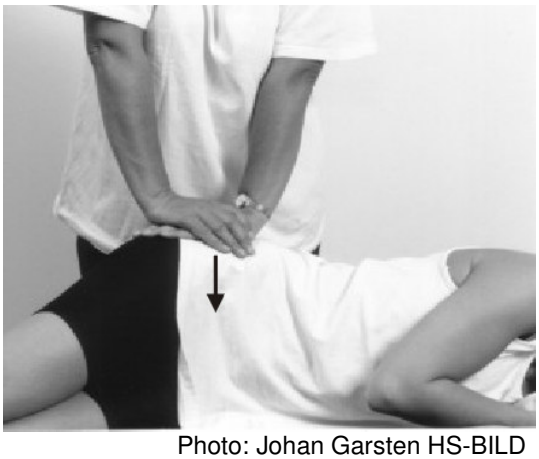


Photo: Johan Garsten HS-BILD
(Modified from Levin, 2003)

Figure 4 Compression test.
Study I

Photo: Johan Garsten HS-BILD
(Modified from Levin, 2003)

Figure 5 Distraction test.
Studies I, II, III, IV

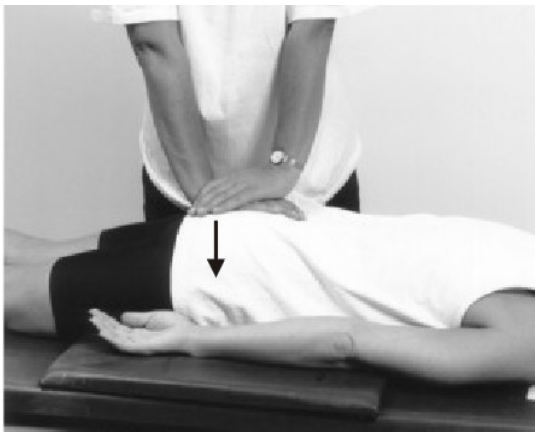


Photo: Johan Garsten HS-BILD
(Modified from Levin, 2003)

Figure 6 Pressure on apex sacralis.
Studies I, II

3.4 Instrumentation

In Study I an examination table with digital scales (EKS International AB, Smålandsstenar, Sweden) was used (Levin, 1989) (Figure 7) and in Studies II – IV an examination table with two AMTI force plates (Advanced Mechanical Technology, Inc., Watertown, MA, USA; model MC818-6-1000, size 457 x 203 mm^2), one for each SI joint with a distance of 40 mm (Figure 8).

The scales were connected to an electric ink recorder (Brush 220 Recorder, Gould., Inc., Valley View, Ohio, USA) equipped with graph paper. One square on the y-axis of the graph paper corresponded to 2 kg (2 · 9.81 N). One square on the x-axis corresponded to 1 s at a speed of 1 mm/s. During the trial the scales were calibrated before each individual PT's trial with standard weights, total 20 kg, in seven steps.

Each force plate, providing an accuracy of 0.25 N, was capable of recording three orthogonal forces, *Fx* (caudal/cranial), *Fy* (lateral/medial) and *Fz* (vertical). Raw signals from the force plates were recorded as action forces with a sampling frequency of 100 Hz during

the test periods (Table VI). Differences between the magnitudes of force recorded from the two force plates made it possible to present indications of unevenly applied forces (Studies II-IV).

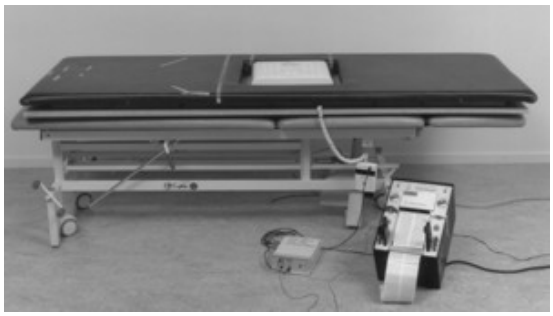


Photo: Johan Garsten HS-BILD

Figure 7 The height-adjustable examination table with digital scales and the ink recorder.

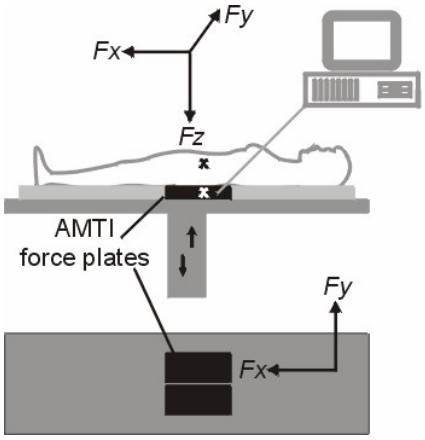


Figure 8 The height-adjustable examination table with two AMTI force plates and the computer. The anterior superior iliac spine was aligned with a mark alongside one force plate.

Table VI Details of data collection for each PT and each examinee (Studies I-IV).

	Occasions n	Test periods	Repetition of each test, n	Testing from one side	Testing from both sides	Data for each SI joint
Study I						
Applied force in vertical direction	3	20	1	•		
Study II						
Applied force in three orthogonal directions	2	15	6	•		•
Studies III, IV						
Applied force in three orthogonal directions	1	until pain or ≥ 20	3		•	•

3.5 Self-reports

The self-reports assessed for impairments and activity limitations before and after the distraction testing in Studies III-IV are shown in Table VII. The BASDAI (Garret et al., 1994) and the BASFI (Calin et al., 1994) consist of questions answered on 100 mm VAS, min-max score 0-10. Pain drawings, reliable as a measure of location of pain (Margolis et al., 1988), were used to mark the location of usual or present pain. The M-VAS (Pharmacia & Upjohn Sweden AB) was used for rating pain intensity. The M-VAS has a ruler with a movable marker on a horizontal, non-graded 100-mm line. 'No pain' and 'worst imaginable pain' are marked on the left and right ends, respectively, and the line is graded on the back. The pencil-and-paper VAS is valid and reliable (Carlsson, 1983) and the M-VAS is comparable to the pencil-and-paper VAS (Price et al., 1994). When the subject pointed to the region of pain, the area was marked. Pain was regarded as emanating from the SI joint if the mark was on the posterior superior iliac spine, immediately medially, laterally and/or caudally along the SI joint.

Table VII Self-reported assessments for the examinees in Studies III-IV.

	Impairments	Activity limitations
Once before all test sessions		
BASDAI [a]	•	
BASFI [b]		•
Marked usual or present pain on a pain drawing	•	
Before test sessions from each side		
Rated present pain on M-VAS [c]	•	
Pointed twice to the region of usual or present pain	•	
After test sessions from each side		
Rated provoked pain on M-VAS	•	
Pointed twice to region of provoked pain	•	

[a] Bath Ankylosing Spondylitis Disease Activity Index
[b] Bath Ankylosing Spondylitis Functional Index
[c] Mechanical visual analogue scale

3.6 Procedure

Each PT independently performed the SI pain-provocation tests at the same time of day on each examinee. Details of the data collection are shown in Table VI. In Study I the examinee was tested at one- or two-week intervals and in Study II at two-week intervals. In Studies III-IV the examinee was tested on three consecutive days by a different PT each day.

The order of the tests was randomised by lot for each PT in Studies I-II and the order of the sides in Studies III-IV. The height of the tables was adjusted for each PT according to individual preference. The PTs were instructed to perform the tests in their habitual manner (Studies I-IV), and to perform the tests on healthy subjects as if examining a patient by provoking pain (Studies I-II).

The examinee lay in a standardized and marked position with the pelvis and lower lumbar spine on the digital scales/force plates. The pelvis and the lumbar spine had no contact with the additional padded examination board (Figure 7, 8). In Study I the PTs indicated vocally

when the applied force they judged appropriate was attained and this moment was marked on the graph paper. The applied force and its duration were recorded on the ink recorder. In Studies II-IV an auditory signal cued the PT to start the test. The examinee was instructed to report any pain, i.e. any change from the status before testing, at once: vocally in Studies I-II, and by pressing a button giving an audible signal and simultaneously a mark in the data collection in Studies III-IV, whereupon the PT immediately discontinued the test. The subject rated, on the M-VAS, the pain intensity experienced when pressing the button. The subject also pointed to the region where pain was felt. In Studies III-IV the procedure was repeated with the PT standing on the other side of the subject.

After the tests in Study I the PT was asked whether the applied force had been the same as when examining a patient. The PTs rated their opinions on a 100 mm VAS, the ends defined as 'Not at all' and 'Absolutely' for each test.

After the test sessions in Study II, the PTs graded their perceptions of applied directions from the greatest direction to the least on a form with each force component drawn on a picture of a body and with written explanations. The PTs were also given a verbal explanation of the force components on the pictures.

3.7 Data analysis

In Study I the applied force was calculated by manually counting the squares every second on the graph paper and multiplying the number of squares by 19.62 (2 · 9.81) to yield N. For each PT and each SI joint test, the applied force per second, the mean pressure force for each occasion, and for three occasions, were calculated. The same variables were calculated for the entire group of PTs, for all the PTs who performed the SI joint test often and for those who performed the tests seldom or never.

In Studies II-IV the ASCII files of the force plate signals were imported to Axograph (Axon Instruments, Inc., Union City, CA, USA), a Macintosh-based program. The amplitudes of *Fx*, *Fy* and *Fz* and time events were manually defined in the graphics so generated. The analysis periods selected from the graphics in Studies II-IV, are shown in Figure 9. The mean amplitude of each analysis period is expressed as the magnitude of force in the following text. Onset of force in Studies III-IV was defined as the point where the curve definitely turned upwards and the start of the plateau in Study IV as where the curve became horizontal. The end of the plateau was defined as the point where the subject indicated pain or the end of force exposure (subjects without pain indication).

In Studies II-IV the means and the standard deviations (SD) were calculated across all repetitions of each SI joint test. The variables of *Fx*, *Fy* and *Fz* (Studies II-IV) and time interval (Studies III-IV) were calculated for each PT, each subject, each trial, each force plate and each occasion, from one side or from both (Table VI). In Study II the variables were also calculated for the entire group of PTs with and without OMT certificate, respectively. In Studies II-III the magnitude of total force (F_{total}) was calculated as $\sqrt{(Fx^2 + Fy^2 + Fz^2)}$ (Harms et al., 1997; Enoka, 2002).

In Study IV the integration of force exposure over time: impulse (*I*) = force (*F*) · time (*s*) for *Fy* and *Fz*, respectively, were calculated for each subject, each force plate and each PT for each repetition and also across three repetitions (Table VI). The magnitude of *Fz* for the last 100 ms before pain signal/end of force exposure was subtracted from *Fz* for the first 100 ms of the plateau for evaluating the ability to keep the same amplitude of force exposure during the plateau (Figure 9).

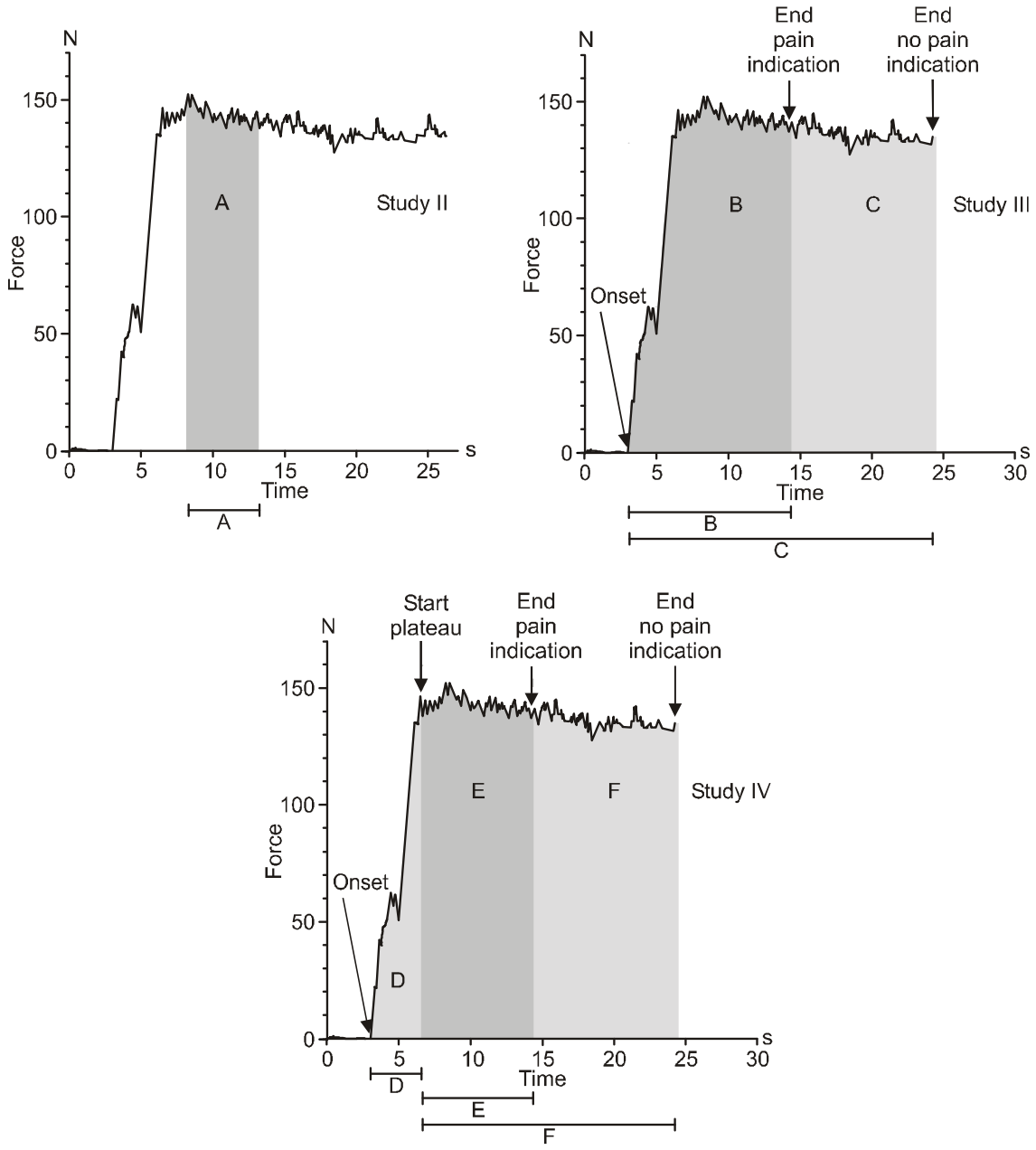


Figure 9 The analysis periods of the forces recorded from the AMTI force plates in Studies II-IV. The mean amplitudes of A, B, C, D, E and F are expressed as magnitudes of force.

Study II: A - 5 s.

Study III: B - from onset of force to pain indication (subjects with pain) or
C - from onset of force to end of force exposure (subjects without pain).

Study IV: D - from onset of force to start of plateau.
E - from start of plateau to pain indication (subjects with pain) or
F - from start of plateau to end of force exposure (subjects without pain).
Impulse = force (mean amplitude) · time (s) = total area under the curves of E and F.

3.8 Statistics

An overview of the statistical methods used in this work is given in Table VIII. In Study I StatView (data analysis software system) for MacIntosh was used for statistical analyses, except for the intraclass correlation coefficient (ICC). The ICC was calculated in Microsoft Excel. In Studies II - IV StatSoft, Inc. STATISTICA version 6 was used. The significance level was set at $P \leq 0.05$.

Table VIII Survey of statistical methods used in Studies I-IV.

	Study I	Study II	Study III	Study IV
Descriptive statistics				
Coefficient of variation (CV)	•	•		
Mean, ±SD	•	•	•	•
Median, range	•	•	•	•
Percent (%)				•
95% confidence interval (CI)	•		•	•
Sensitivity			•	•
Specificity			•	
Standard error of mean (SEM)	•			
Variance (s^2)	•			
Analytical statistics				
One-way ANOVA	•			
Two-way ANOVA		•	•	•
Three-way ANOVA		•	•	•
ICC [1.1] [a]	•			
Regression coefficient	•			
Standard error of measurement (SEM/S_W)	•	•		•

[a] Intraclass correlation coefficient: Between-subject mean squares (BMS) – within-subject mean square (WMS)/BMS+(k–1) WMS, k = number of occasions

In Study I the maintenance of a constant force for each occasion and each test was estimated from the slope of the regression line, calculated from the recorded force per second for each PT. The mean slope and its 95 % confidence interval (CI) were then estimated. One-way analysis of variance (ANOVA) for repeated measures was used to analyse the differences in slope on three occasions. The force exerted was reported as the mean of each occasion and the mean of three occasions. Intra-examiner reliability was assessed by calculating the variance (s^2) of three occasions for each individual PT, the mean s^2 for all PTs and from this the standard deviation (SD), and the coefficient of variance (CV). The intraclass correlation coefficient (ICC) [1,1] (Shrout et al., 1979; Lahey et al., 1983) was also calculated. The minimally acceptable ICC is suggested by Chinn (1991) to be at least 0.60 while Fleiss (1986) recommended that values between 0.40-0.75 represent fair to good reliability and values above 0.75 represent excellent reliability. Inter-examiner reliability was assessed by calculating SD, standard error of mean (SEM) and CV from the mean of three occasions for all PTs.

A three-way ANOVA for repeated measures was performed for the magnitudes of F_{total} (Studies II-III), *Fx*, *Fy* and *Fz* (Study II), *Fy* impulse and *Fz* impulse (Study IV). Where there was significant interaction simple main effect tests were performed (Kirk, 1995). A two-way

ANOVA for repeated measures was performed for the time interval (Study III) and for *Fz* for the first and last 100 ms (Study IV). The consistency of F_{total}, *Fx*, *Fy* and *Fz* between occasions (Study II), and of *Fy, Fz* and the time interval between repetitions (Study IV), recorded from each force plate, was assessed by calculating standard error of measurement (SEM/S_W) derived from the square root of the mean square within subjects obtained from a two-way ANOVA, and with CV calculated from S_W (Study II).

The mean, SD and 95% CI were calculated for the magnitudes of F_{total} recorded from each force plate, and for the time interval of force exposure: for healthy subjects, for all subjects with sacroiliitis and for the subjects with sacroiliitis who indicated pain (Study III). In Study IV the same variables were calculated for *Fz*, *Fy* and for the time interval of force exposure recorded from each force plate: for all repetitions, for all subjects and for each PT.

Sensitivity and specificity were calculated as the proportions of individuals, with pain indication or without, correctly identified by the distraction test (Sackett et al., 1991). The test was graded positive if the subject indicated provoked pain by pressing the button and in addition pointed out pain experienced in the region of the SI joints. At least two positive repetitions out of three were required (Studies III-IV).

3.9 Ethics approval

All participants were fully informed verbally and in writing of the experimental procedure and were free to withdraw at any time. All signed an informed consent statement.

The designs of the studies were approved by the Human Ethics Committee, Huddinge University Hospital, Dnr: 208/95 (Study I) and Dnr: 277/97 (Studies II-IV).

4 RESULTS

4.1 Study I

No systematic change over time was demonstrated in the applied force for the individual PT, or for the entire group of PTs on the three occasions in any of the three tests.

The mean force for each individual PT on the first, second and third test occasions are shown in Figure 10. The ICCs [1.1] on three occasions were 0.63, 0.71 and 0.74, respectively, for the compression test, the distraction test and pressure on the apex sacralis for all PTs. The S_W on the three occasions were 46.4, 34.4 and 36.6 N respectively, for the compression test, the distraction test and pressure on the apex sacralis. The mean force for all PTs was 251 N, 95% CI 218.3-285.6, CV 26.2 in the compression test; 232 N, 95% CI 202.8-261.6, CV 24.6 in the distraction test; 222 N, 95% CI 188.8-255.2, CV 29.1 in pressure on the apex sacralis.

Eleven of 17 PTs rated that they had definitely (100 mm VAS) applied force as on a

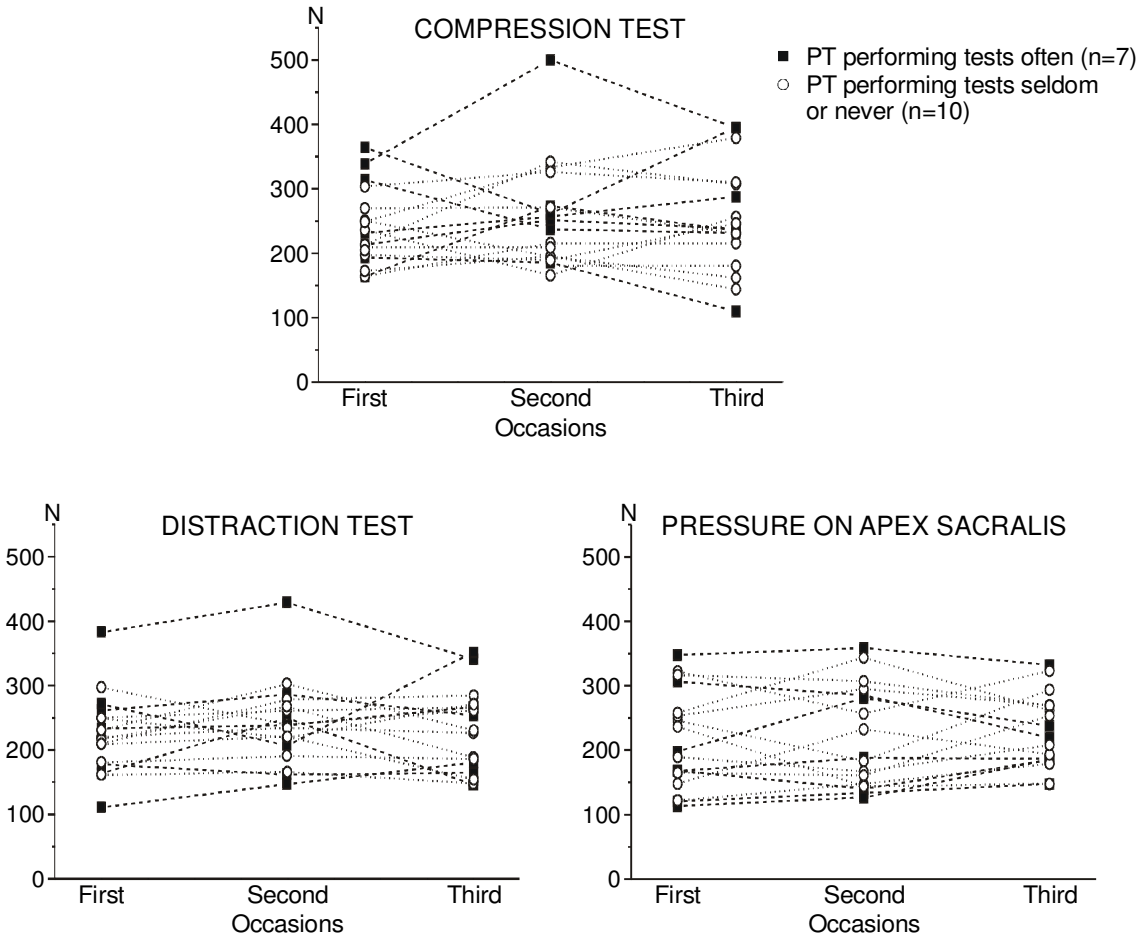


Figure 10 Mean force during 20 s for each physiotherapist (PT) on first, second and third occasions in the compression test, the distraction test and pressure on the apex sacralis.

patient in all tests. For the remaining 6 PTs data was missing for one, while the median for the remaining five was 37 mm, 78 mm, 81 mm for the compression test, the distraction test and pressure on apex sacralis, respectively.

4.2 Study II

No difference was found in the ANOVA regarding the level of the magnitudes of F_{total}, *Fx, Fy, Fz* in the profile of the two groups PTs with OMT certificate and PTs without, as there was no significant interaction between groups and force plates in the tests. Nor was any difference found in the tests between the occasions, except for *Fz* in the distraction test, as there was no significant interaction between occasions and force plates.

The magnitude of F_{total} recorded from the force plate closer to the PT was significantly less than for that from the force plate further away, for both groups, on both occasions and in both tests. The distribution of force components varied between and within PTs in both tests (Figures 11 and 12). In the distraction test, significant differences were found for the magnitudes of F_{total} and *Fz* independently of group, between the force plates and between occasions; and for *Fy* between occasions. In pressure on the apex sacralis, significant differences between the force plates were found for the magnitudes of F_{total}, *Fy* and *Fz* independently of group and occasion. In both tests, the variance regarding F_{total} between PTs was higher than, or about equal to, the variance within PTs.

In the distraction test four PTs considered that they had applied the most force vertically and 11 PTs laterally. Two PTs considered that no force was applied vertically. In pressure on apex sacralis 14 PTs considered that they had applied the most force vertically and one PT cranially. Nine PTs considered that no force was applied cranially.

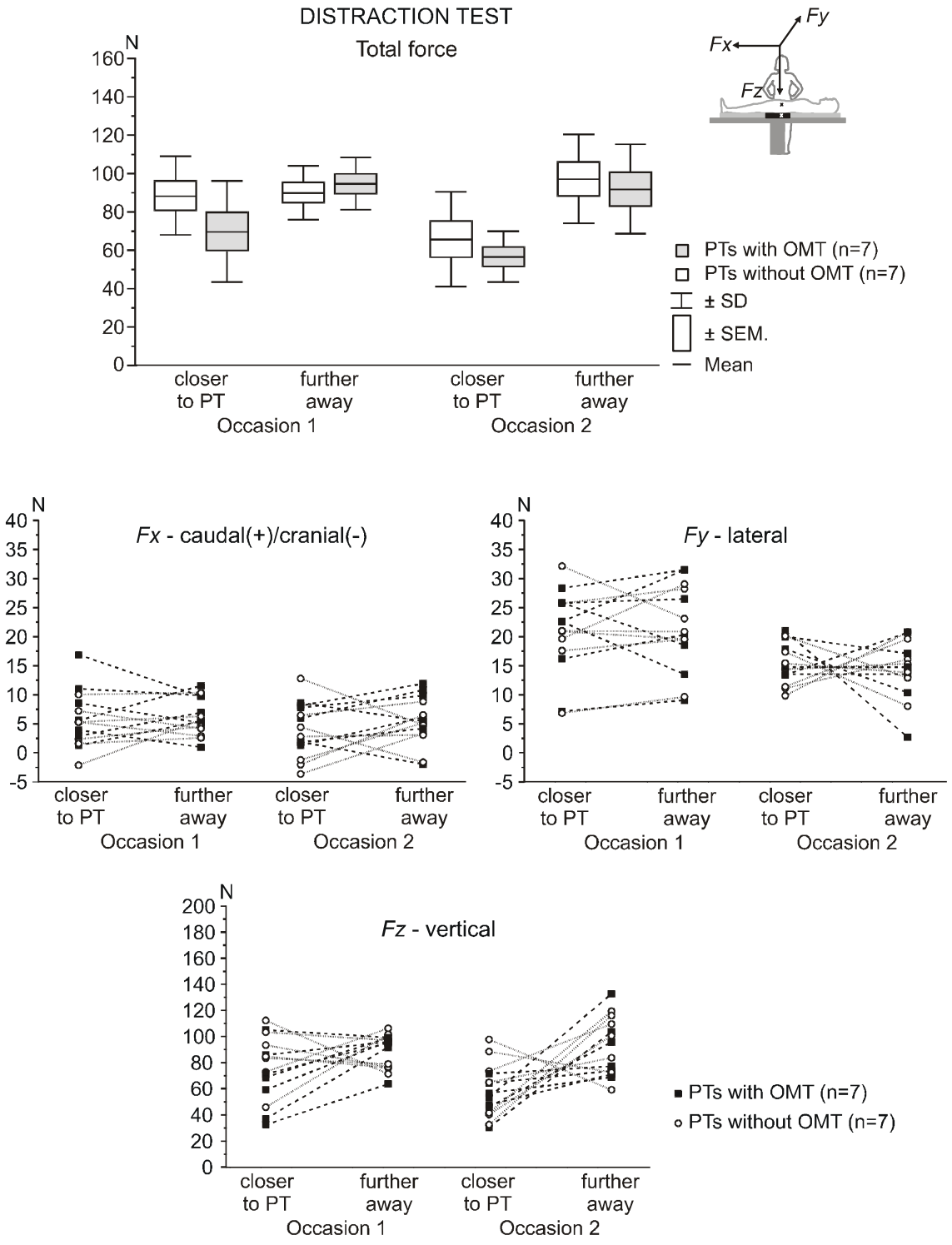


Figure 11 Magnitudes of total force vector and force components (N) on the force plate closer to the physiotherapist (PT) and the force plate further away on first and second occasion for PTs with and without Certificate in Orthopaedic Manipulative Therapy (OMT), in distraction test.

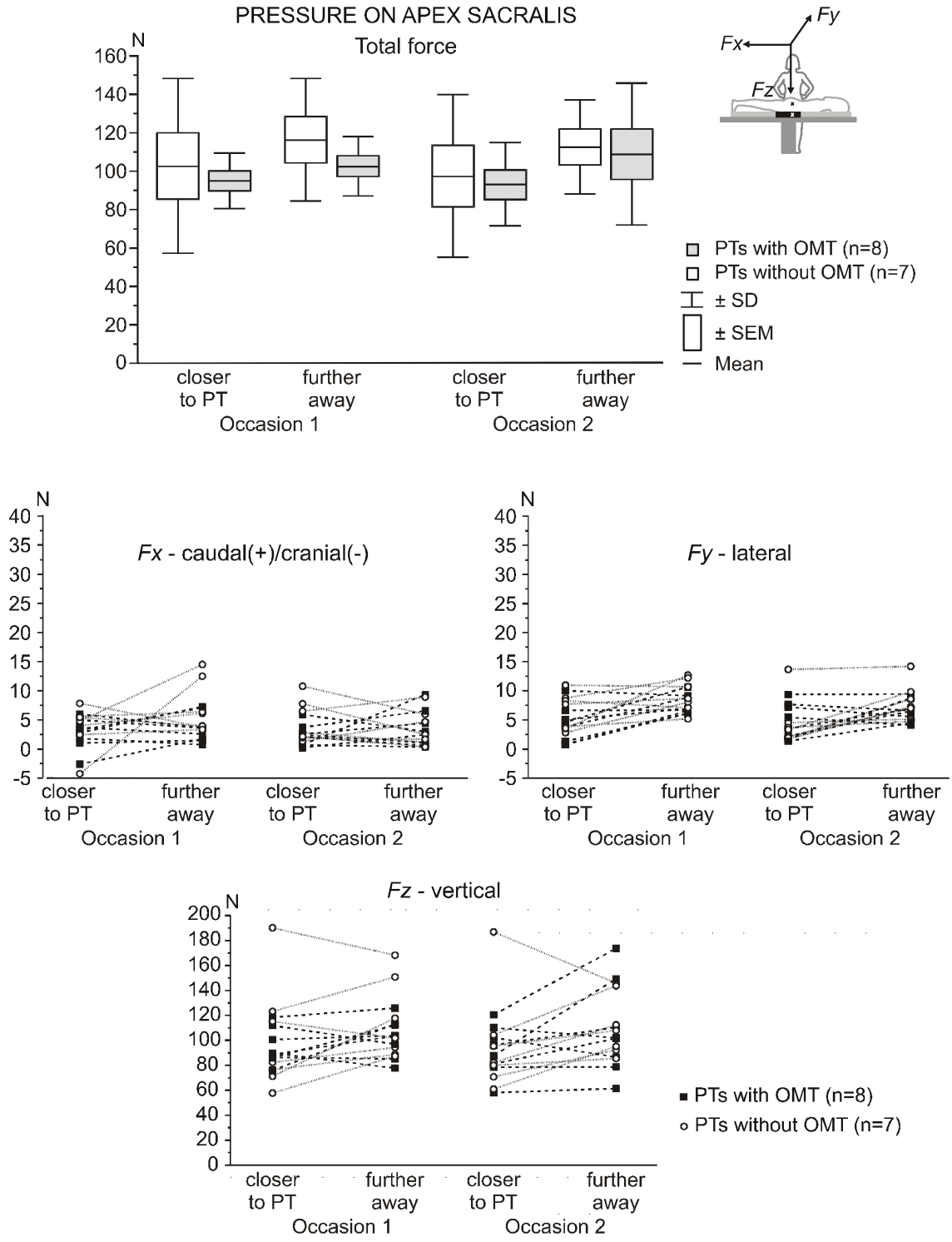


Figure 12 Magnitudes of total force vector and force components (N) on the force plate closer to physiotherapist (PT) and on the force plate further away on first and second occasion for PTs with and without Certificate in Orthopaedic Manipulative Therapy (OMT), in pressure on apex sacralis.

4.3 Study III

The time interval of force exposure until pain was provoked varied in subjects with sacroiliitis, range 1.8 – 19.5 s. Two subjects with sacroiliitis did not press the button although they experienced pain while tested by two of the PTs. No pain was provoked in healthy subjects. Of the subjects who gave a pain signal, the level of pain on M-VAS after the test sessions compared to before was rated higher (≥ 5 mm) by 46% of the subjects, equal or no pain by 39.5% and lower (≥ -5 mm) by 14.5%. The pain level on M-VAS after the test sessions from the side the subjects were first examined from was median 40 mm (range 0-90) and from the other side median 40 mm (range 0-80). The pain level on M-VAS after the test sessions the first day was median 39 mm (range 0-90), after the second day 46 mm (range 0-72) and after the last day 37 mm (range 0-69).

The ANOVA showed no significant interaction between groups and PTs regarding the magnitude of F_{total} or time interval of force exposure when testing from the right-hand side or from the left; nor was there any interaction between groups and force plates. However, for both sides the magnitude of F_{total} was significantly smaller and the time interval of force exposure significantly shorter in the sacroiliitis group than in the healthy-subjects group (Figure 13). The mean magnitude, SD and 95% CI of F_{total} and time interval of force exposure summarised for both force plates are shown for each repetition of the distraction test in Table IX.

The sensitivity was 0.55 bilaterally calculated for all three PTs and varied between them, range 0.55-0.82 when tested from the right-hand side and 0.55-0.73 when tested from the left hand side. The specificity was 1.0. Two subjects with sacroiliitis experienced no pain during the test period.

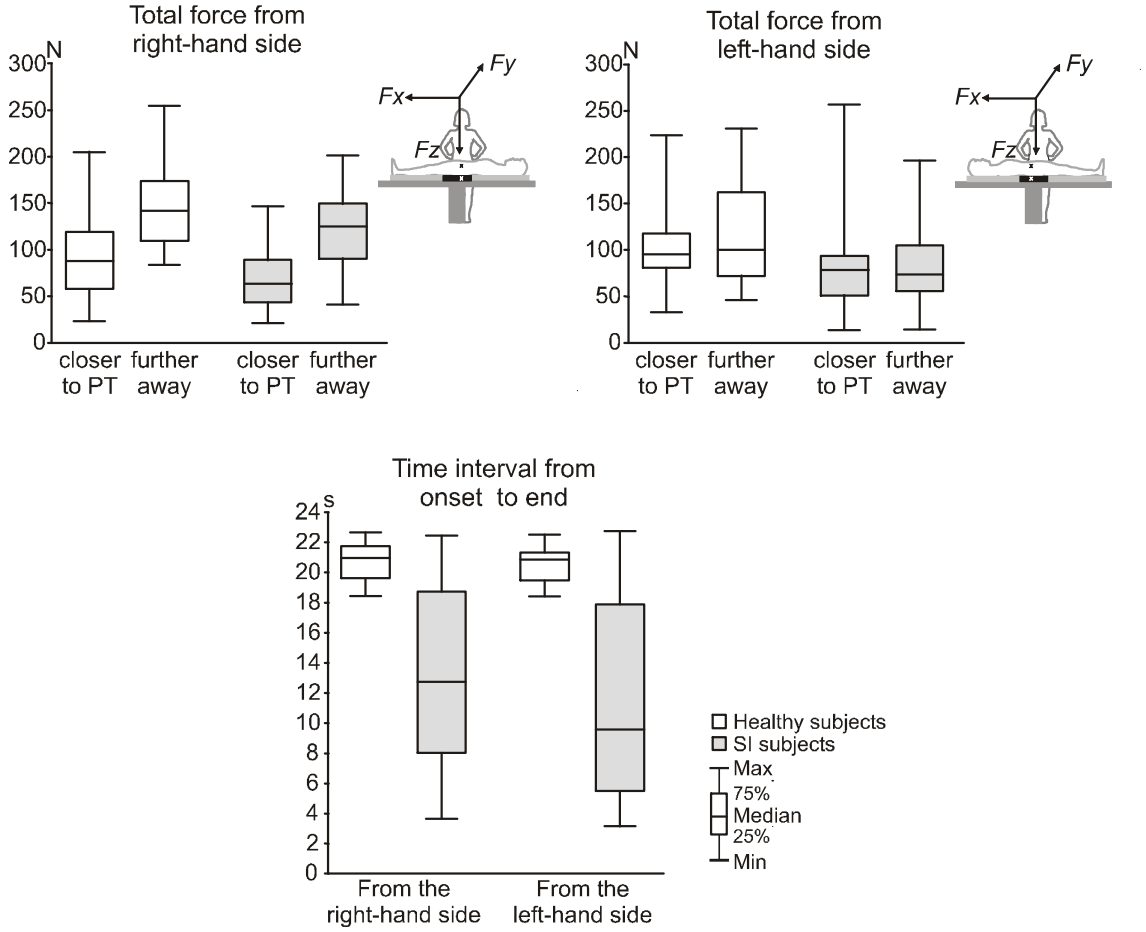


Figure 13 Magnitude of total force (N) on the force plate closer to the physiotherapist (PT) and the force plate further away, and time interval (s) from onset of force until end of application during distraction test for healthy subjects and subjects with sacroiliitis (SI subjects), respectively.

Table IX Means, standard deviation (SD) and 95% confidence intervals (CI) of the magnitude of total force vector (F_{total}) for the right (R) and left (L) force plates and of time interval of force exposure for each repetition in the distraction test of healthy subjects (n=11), for sacroiliitis (SI) subjects (n=11) and for the tests of SI subjects with pain signals. Subjects were tested three times by three physiotherapists from both sides.

	Distraction test from					
	right-hand side			***left-hand side***		
	Mean	SD	95% CI	Mean	SD	95% CI
Repetition 1						
F_{total} R and L (N)						
All tests healthy n=33/33	217.5	76.2	190.5 -244.5	247.5	64.9	224.5 -270.6
All tests SI subjects n=33/32	173.3	63.2	150.9 -195.7	184.1	51.0	165.1 -203.1
Tests with pain signal n=24/20	168.2	62.9	141.6 -194.7	186.7	52.9	161.9 -211.4
Time interval (s)						
All tests healthy n=33/33	21.3	1.5	20.8 -21.9	20.3	2.1	19.6 -21.1
All tests SI subjects n=33/32	14.1	5.4	12.2 -16.0	13.0	5.4	11.0 -15.0
Tests with pain signal n=24/20	11.5	3.7	9.9 -13.0	9.6	2.2	8.6 -10.6
Repetition 2						
F_{total} R and L (N)						
All tests healthy n=33/33	232.4	59.8	210.8 -254.0	211.9	59.8	190.7 -233.1
All tests SI subjects n=33/32	236.2	74.8	209.7 -262.7	215.0	78.4	187.2 -242.8
Tests with pain signal n=10/9	179.0	46.0	146.1 -211.8	162.8	62.1	115.1 -210.5
Time interval (s)						
All tests healthy n=33/33	21.0	1.5	20.4 -21.5	20.6	1.2	20.2 -21.0
All tests SI subjects n=33/32	16.4	7.2	13.8 -18.9	16.5	7.2	13.9 -19.0
Tests with pain signal n=10/9	6.0	2.5	4.2 - 7.8	5.6	3.0	3.3 - 8.0
Repetition 3						
F_{total} R and L (N)						
All tests healthy n=33/33	232.6	67.7	208.5 -256.6	201.1	64.1	177.9 -224.2
All tests SI subjects n=33/32	159.4	73.7	133.3 -185.6	127.2	51.2	109.1 -145.4
Tests with pain signal n=25/29	149.9	81.1	116.5 -183.4	120.2	49.7	101.2 -139.1
Time interval (s)						
All tests healthy n=33/33	20.6	1.6	20.1 -21.2	20.5	1.6	19.9 -21.1
All tests SI subjects n=33/32	10.3	7.5	7.6 -13.0	8.0	5.9	5.9 -10.1
Tests with pain signal n=25/29	6.8	4.6	4.9 - 8.7	6.5	4.1	4.9 - 8.1

4.4 Study IV

Only results from the test performed with the PT standing on the right-hand side of the subjects are reported since testing from the left-hand side resulted in lower sensitivity (Study III).

Means, SD, 95% CI and S_W of the magnitudes of *Fy* and *Fz*, respectively, and time intervals of force exposure for all subjects with sacroiliitis, are shown for the two analysis periods in Table X. The variation between subjects with sacroiliitis (SD) was higher than the variation between the three repetitions (S_W) for *Fz* and *Fy* for both analysis periods for each PT.

Table X Means, SD, 95% confidence interval (CI) and standard error of measurement (S_W) of the magnitude (N) of lateral (*Fy*) and vertical (*Fz*) force components and time intervals (s) recorded from the force plates in distraction test for three repetitions by three physiotherapists (PTs) on subjects with sacroiliitis (n=11).

	Force plate							
	closer to PT				***further away from PT***			
	Mean	SD	95% CI	S_W	Mean	SD	95%CI	S_W
Onset to plateau								
***Fy* (N)**								
PT1	13.2	7.4	8.2-18.1	4.6	17.8	5.3	14.2-21.3	2.8
PT2	11.1	4.3	8.2-14.0	3.5	7.9	4.7	4.8-11.1	2.0
PT3	18.6	9.5	12.2-25.0	7.2	14.1	4.3	11.2-16.9	3.5
***Fz* (N)**								
PT1	52.1	17.5	40.4-63.9	16.4	86.5	14.2	77.0-96.1	13.3
PT2	41.2	18.2	29.0-53.4	12.1	52.0	17.1	40.5-63.5	9.6
PT3	65.1	32.9	43.0-87.3	20.7	90.4	17.9	78.4-102.5	15.0
s								
PT1	4.6	1.6	3.5-5.7	1.0	4.3	1.5	3.2-5.3	1.4
PT2	2.8	0.9	2.2-3.3	0.8	3.0	0.7	2.6-3.5	1.2
PT3	5.2	2.1	3.8-6.6	1.4	4.1	1.7	3.0-5.2	1.2
Plateau								
***Fy* (N)**								
PT1	31.2	12.6	22.7-39.6	5.7	40.7	8.9	34.8-46.7	2.5
PT2	16.6	8.1	11.2-22.0	4.2	14.9	7.7	9.8-20.1	2.4
PT3	32.2	13.3	23.2-41.1	8.3	25.2	10.5	18.1-32.2	4.3
***Fz* (N)**								
PT1	106.2	37.6	80.9-131.4	18.8	179.3	36.6	154.7-203.9	19.2
PT2	57.5	27.3	39.1-75.8	12.6	109.1	32.3	87.4-130.9	14.2
PT3	106.4	51.1	72.1-140.7	30.9	164.3	37.0	139.5-189.1	25.4
s								
PT1	5.9	5.1	2.5-9.3	3.0	6.3	5.2	2.8-9.8	3.4
PT2	12.0	6.9	7.3-16.6	3.7	11.5	6.6	7.1-15.9	3.8
PT3	10.5	6.2	6.4-14.7	3.6	11.5	6.6	7.0-15.9	3.3

For the period from onset of force exposure until the plateau was reached (Figure 9), statistically significant differences were found for the magnitudes of *Fz* impulse and *Fy* impulse between PTs and between repetitions. During the plateau (Figure 9) statistically significant differences were found for the *Fz* impulse between PTs and between the force plates, and for the *Fy* impulse between PTs. During the plateau statistically significant decreases were found in the amplitude of *Fz* for two of the PTs. The PT with the highest value for sensitivity showed least decrease in amplitude of *Fz*.

The median level of pain after the test sessions was 39 mm (range 0-55) when tested by PT1, 20 mm (range 0-72) when tested by PT2 and 31 mm (range 0-87) when tested by PT3. Two subjects experienced no pain during the test period.

Time intervals from onset to end of force exposure and magnitudes of *Fz* during the plateau for three repetitions for each PT and for all subjects are shown in Figure 14.

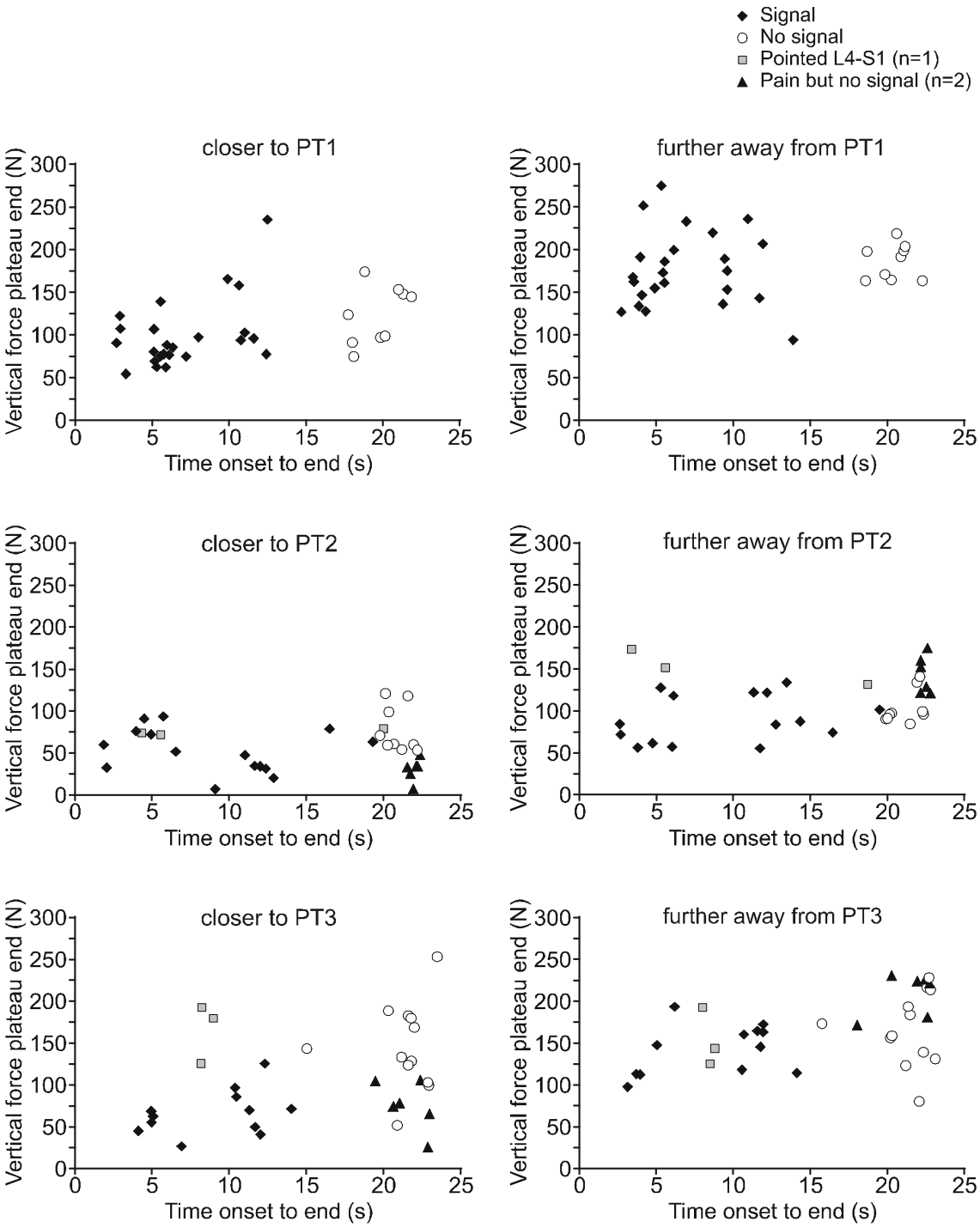


Figure 14 Time intervals (s) from onset to end of force exposure and magnitudes of vertical force component from start of plateau to end of force exposure in distraction testing three times on eleven subjects (n=33) by each physiotherapist (PT). Force plate closer to PT and force plate further away. Each subject (n=11) is represented by three plots. Two subjects did not signal the pain they felt (pain but no signal).

5 DISCUSSION

The main focus in the present work was to investigate force and time interval of force exposure and their influence on the reliability, validity, sensitivity and specificity of SI pain-provocation tests. The studies were the first to systematically evaluate applied force and time interval of force exposure in SI pain-provocation tests.

The examination table with digital scales used in Study I was originally constructed to ensure that the same force was applied in SI pain-provocation tests before and after a physiotherapy intervention (Levin, 1989). The applied force was read from a display placed beside the patient. Testing was performed from one side of the patient and the force was maintained for 20 s unless discontinued earlier because of pain. The study raised questions that had to be answered – as might have been the case in many other physiotherapy situations: 'Could the same force have been applied without the scales; was it enough to perform the test only from one side of the patient; and was it relevant to maintain applied force for 20 s?'

5.1 Results

Although the PTs had the same instructions and tested the same person, great variations were found in Study I within and between PTs in the magnitude of the vertical force recorded by the scales. It seems likely that pain is a crucial factor normally determining the amount of force tolerated. Consequently the examinee's lack of pain could be supposed to cause uncertainty, some PTs being afraid of hurting the healthy person and this being one explanation for the varying forces within and between PTs. Most of the PTs, however, regarded their performance as being the same as when testing a patient. It appeared reasonable to assume that lack of pain influenced more experienced PTs to a greater degree. This is because they regularly used the SI pain-provocation tests clinically and consequently had more difficulties in repeating the same level of force with a healthy person. Regarding the variation in applied forces the results corresponded in general with those of studies evaluating forces during the posteroanterior (PA) vertebral pressure test in the lumbar spine. Simmonds et al. (1995) concluded that the wide range of forces applied by PTs under controlled 'joint' conditions suggests that the individual PT is the primary source of variability in force application, and that PTs' perceptions of applied forces are inaccurate. Harms et al. (1997) found that the variation in forces used by experienced therapists was substantial, and most therapists were inconsistent in replicating the amplitude of force. Latimer et al. (1998) studied therapists' ability to judge PA vertebral stiffness and presumed that one likely reason for poor reliability was that the therapists used varying forces.

When two AMTI force plates were used to record the forces applied separately for each SI joint (Study II) the magnitude of force, recorded by the force plate closer to the PT, was significantly less although applied by the dominant hand in the distraction test. This can hardly be explained by the examined person's lack of pain. It is more likely to be explained ergonomically by the performance of the test with cross-armed pressure, the PT producing more force by pushing her arm away with a rotation of her trunk when applying force on the ASIS further away, but only applying force through her arm muscles on the closest ASIS. The distraction test has sometimes been performed without cross-armed pressure and seeking to direct the force only vertically (Laslett et al., 2003). In a future study it might be useful to see whether there was less variation between forces applied by the right hand and the left hand if

the test was performed in this way. However, the findings in Study II of less recorded force from the force plate closer to the PT support the recommendation to perform the tests standing on the pain-free side to ensure sufficient provocation of the joint. Accordingly, when both SI joints are suspected to be involved the tests should routinely be performed from both sides of the patient.

The variation between PTs in the distribution of force components (Study II) was probably partly a consequence of the differing operational descriptions of SI pain-provocation tests, not only in the literature but also when learning the tests. This is because the PTs were instructed to perform the tests in their habitual manner with a free choice of working postures and positions in relation to the examined person certainly causing variation in the distribution of force components, but also in the magnitude of applied force. The PTs' working postures and positions are most likely not similar each time they perform the same test. Force plates underneath the PT's feet would probably have given valuable information about this. Another reason for the variation in force distribution might be a consequence of hand placement or difficulties in locating anatomical landmarks. It seems not unlikely that the PTs' hand placement differed in pressure on apex sacralis, but it seems improbably that they had difficulties in locating apex sacralis. The ASIS could supposedly be more difficult to palpate. However, Riddle et al. (2002) strongly suggested that not even overweight is a source of error when palpating bony landmarks on the pelvis.

During the distraction test on subjects with sacroiliitis (Study III) almost 20 s could elapse before pain appeared, although this time interval was analysed from the onset of applied force, not from the very beginning when the PT put her hands on the ASIS. The different time intervals of force exposure were probably due to varying inflammatory activity, but also to the fact that joint surfaces and soft tissues were more involved or less, needing a shorter or longer time to react. Probably, individual variation in anatomical configuration of the SI joints also influenced the pain response. Viner et al. (1995) measured the directions of forces applied during assessment of PA vertebral stiffness in pain-free subjects. They concluded that the inter-examiner variations in force directions may result in different pain responses from a vertebral level when examined by different PTs. This possible explanation might also apply to the SI joints. Personal factors probably also influenced how much force could be applied until change in pain and was probably an explanation of why the two 'heroic' subjects did not signal their pain. Lee (1999) recommends that SI pain-provocation tests be maintained for 20 s, which many PTs appear to regard as too long. The present results support Lee's recommendation, but also the suggestion by Laslett (1998) that insufficient force, a consequence not only of the magnitude of force but also of too short a time of force exposure, has been one reason for many false negatives in earlier studies.

The M-VAS ratings of pain intensity were not always higher after the test sessions than before (Study III). Testing was immediately discontinued when the subject indicated a change in pain, and this might have been one reason for not rating pain higher after the test sessions. However, decreased pain levels after the testing cannot easily be explained. One possible explanation might be that the effect of the test on soft tissues sometimes gave *pain relief.* Another explanation might be rating difficulties. In a study comparing pain ratings on a verbal pain scale (VPS) with ratings on a VAS after postoperative anaesthesia, 3% of the patients rated 'worse pain' on the VPS but less pain on the VAS, while 5.4% rated 'better' or 'much better' on VPS but increased pain on the VAS (DeLoach et al., 1998). The primary reason for choosing the M-VAS in Study III was to minimize recalling bias, as the subjects rated pain

three times within some minutes and on three consecutive days. Presumably it is easier to recall the same adjective or adverb on a VPS than to recall the last rating on a VAS.

Comparing M-VAS with a VPS in patients with rheumatological conditions, Clark et al. (2003) found both rating scales to be reliable and valid, but over half their patients preferred the VPS and fewer than 30% preferred the M-VAS. The most frequent reason for preferring the VPS was that it was easier to understand and more convenient to use words. On the other hand, patients who preferred the M-VAS found it more objective and precise in explaining pain, and more comprehensible. The authors advocated taking into account the patient's own choice and preference of scale when choosing an instrument for pain rating, feasible only in the clinical situation. Anyhow, a 'yes' seems safer than VAS ratings for establishing whether the SI joint test does provoke pain, but excludes information about the intensity, important when evaluating pain over time. Pressing a button indicating provoked pain during SI pain-provocation tests seemed both convenient and comprehensible and as certain as a 'yes'. If the magnitude of applied force and the time interval of force exposure are recorded simultaneously during testing, pressing a button could also be supposed to give values comparable over time.

The magnitude of recorded force and the time interval of force exposure in the distraction test did discriminate subjects with sacroiliitis from healthy subjects (Study III). This indicated face validity and acceptable known-groups construct validity, even though the self-reported scores for disease activity and consequences were reported as low-to-moderate by the subjects with sacroiliitis, and even though two subjects did not give a pain signal despite feeling provoked pain. However, the sensitivity varied from fair-to-good depending on which PT examined the subjects. The highest previously reported value for sensitivity for the distraction test was 0.21, i.e. poor, discriminating patients with ankylosing spondylitis from patients with mechanical/degenerative low-back pain (Blower et al., 1984; van der Wurff et al., 2000b). The sensitivity was higher in Study III, but evaluating sensitivity and specificity among subjects who clearly have the actual impairment and subjects who clearly do not is, though efficient, not a method of choice even where the PTs are unaware of which group the examinees belong to. The test may be able to distinguish very well between these extremes which, however, seldom occur in clinical situations (Fletcher et al., 1996). Accordingly the specificity was the highest possible in Study III as half of the tested subjects were healthy.

As well as the methodological problems mentioned above concerning the values for sensitivity and specificity, there may be other reasons for the varying values between the PTs. The two subjects who unfortunately, from the viewpoint of Study III, were pain-free when they attended the first test session were not excluded from the study as pain sometimes varies from day to day in patients with spondyloarthropathy. They were however pain-free throughout the test sessions, giving a lower value for sensitivity, but to the same extent for all PTs. Another factor, causing variation in sensitivity between PTs might have been the subjects' ability to relax during testing, seldom possible for the PT to observe. Soft tissue is unavoidably compressed between the PT's hands and the ASIS, and some subjects complained of discomfort from this area. They may have reacted with reflex muscle tension, giving less provocation of the SI joints. Consistent neuromuscular reflex responses were found by Colloca et al. (2001) while applying forces of approximately 150 N to spinal landmarks in patients with low-back pain. The amplitude and frequency of the reflex responses corresponded to self-reported measures of pain. The reflex response is probably less in the distraction test, as the applied force is slow and not a thrust as in spinal manipulative therapy, but may have varied as a

consequence of differences in the PTs' examination techniques later found in Study IV. The reflex response might also have been influenced by the size of the PT's hands as a pressure (force per unit area) of a constant magnitude of force applied by a small, thin hand is higher than that applied by a large, muscular hand (Özkaya et al., 1998).

It was not surprising to find significantly less force recorded by the force plate closer to the PT when testing subjects with sacroiliitis in Study IV, as was the case when testing the healthy person in Study II: the examinee's pain probably cannot influence the magnitude of forces applied by the PT's right and left hands, respectively. The variation in the magnitude of recorded force between subjects can partly be explained by different pain intensities. This, however, ought to have less influenced the onset-of-force/plateau period which was seldom interrupted by pain signals.

The variation in recorded force between the three repetitions with each examinee is more difficult to explain. It seemed probable that the examinee's pain would result in a similar amount of applied force, so it was unexpected to find that the variation in applied force in subjects with sacroiliitis essentially corresponded with that in the healthy person in Study II. Even though test performance may differ somewhat between PTs, it seems logical that pain should have been provoked either by a briefer, higher magnitude or by a longer, lower magnitude when examining the same subject (Study IV). No consistent pattern of this kind was found. In addition it was not possible to determine the lowest or highest magnitude of force provoking pain, because the force magnitudes and time intervals recorded from the force plate closer to the PT and the force plate further away differed. However, the median level of pain intensity was lowest when the subjects were examined by the PT who applied the least magnitude of force. This indicated that the structures might not always be adequately affected if the applied forces are not large enough. Furthermore, less ability on the part of the PT to maintain the amplitude of force throughout the test period seemed to result in lower sensitivity, as did higher variation in magnitude between repetitions during the plateau. The preconception that pain might be the crucial factor normally determining the amount of applied force, as was stated when variation was found testing a healthy person, does not seem to be the truth and, anyway, the variation was probably not due to the experimental conditions. Rather, the results indicated that variation in examination technique within and between PTs may explain varying outcomes of SI pain-provocation tests and, accordingly, their varying sensitivity.

The present results render it unlikely that the same force can be applied in SI pain-provocation tests before and after an intervention unless the forces are recorded. It is evidently not enough to perform the test unilaterally or to record forces that do not distinguish between the two SI joints. However it seems relevant to maintain the applied force for 20 s unless discontinued earlier because of pain. So far it can be concluded that recording applied force and time interval of force exposure, simultaneously monitored by the investigator, can supplement standardized manual assessment of SI joint pain.

5.2 Methodological considerations

SI joint tests are intended to provoke pain. It may therefore seem irrelevant to evaluate the reliability of applied force on a person without SI joint complaints (Studies I-II). The reliability investigations were limited to one healthy person mainly in order to achieve the same conditions for all PTs. Further, repeated tests by several PTs on a single patient with pain were considered to be unethical, and the pain intensity would probably have changed from one

week to another. When evaluating the PA vertebral pressure test Simmonds et al. (1995) had PTs perform the test on a spinal model in plastic, covered with high-density rubber under which one spinal vertebra was mounted. This was ethical but an extreme way to minimize the sources of variation, and not very similar to the normal testing situation during mobilisation in vivo. The sensory feeling of testing a healthy person, rather than an inanimate model, must better approximate the clinical situation even though our PTs knew that the examinee had no SI joint complaint.

The Study-I and Study-II results showed that the reliability of force application was unsatisfactory in the tests. It was therefore interesting to investigate whether variations within and between PTs had any importance for the test outcome in subjects with pain (Studies III-IV). Pain caused by sacroiliitis was chosen because disorders belonging to spondyloarthropathy must fulfil special classification criteria (Dougados et al., 1991), while SI joint dysfunction is a more questionable diagnosis and accordingly more difficult to establish (Levangie, 1999; Riddle et al., 2002). To ensure that the pain was as stable as possible in the sacroiliitis subjects during the investigation period, each was examined on consecutive days, and with one SI joint test only. Somewhat surprisingly the pain was not worse from the second side or on the second or third day (Studies III-IV).

It has been assumed that an examination of SI joint pain should rely on more than one test (Kokmeyer et al., 2002; Laslett et al., 2003; Young et al., 2003). The main reason for the choice of compression test, distraction test and pressure on apex sacralis (Study I) was because they permit the subject to remain stable and immobile during the test procedure without movement in the hip- and knee joints. This promotes standardized testing on the sensitive scales and force plates. Also these tests were among those reported in earlier studies to be at least acceptably reliable (Potter et al., 1985; Laslett et al., 1994; Albert et al., 2000; van der Wurff et al., 2000a). The thigh thrust, originally developed for pain with onset during pregnancy or delivery, has excellent reliability (Albert et al., 2000) and high sensitivity (Östgaard et al., 1994; Broadhurst et al., 1998; Albert et al., 2000; Vleeming et al., 2002), but its applicability in possible hip and knee arthritis may be questioned.

In Study II, even though the subject was healthy, the tests were reduced from three to two for ethical reasons. The compression test was eliminated as it showed the greatest variation in Study I and further, side-lying is not as stable a position as those used in the other two tests. However, the healthy person in Studies I and II underwent >500 repetitions with no sign of developing SI joint complaints or low-back pain. It is probable that SI pain-provocation tests do not cause pain in intact SI joints. In Studies III and IV only one test was evaluated since repeated tests by several PTs on a patient with pain were considered to be unethical. Also, testing with more than one test could affect the pain intensity from one test to another. In addition, when using a cluster of tests, a prerequisite should be to determine the overall validity of each, so as to provide the examiner with proper information without loss of time (Sackett et al., 1991). The distraction test was chosen as it was more reliable than pressure on apex sacralis (Table III), but also as it supposedly involves the lateral force component more highly.

To resemble a clinical situation, no training sessions were held. Freburger et al. (2001) criticised McCombe et al. (1989) for not standardising their test procedures. However, despite standardizing the test procedure according to the examinees, it seems hard to standardize the PT's working posture and position in relation to the examinee and – consequently – the performance of SI pain-provocation tests including applied force. To achieve roughly the

same amount of force, Kokmeyer et al. (2002) held training sessions where the examiners got feedback on applied forces between 150 and 300 N depending on which of six SI pain-provocation tests were performed. In the clinical situation there is normally no time for training sessions, and the present results indicate that such session would be of scant value.

The examination table with digital scales (Study I) was developed to allow recordings of applied force without disrupting the PT's mode of performance of the tests or affecting the examinee. This was the reason not using force transducers under the hands such as the thin, flexible pressure mat previously used for measuring perpendicular exerted force during spinal manipulative therapy (Herzog et al., 1993). Although only the vertical force component was quantified the AMTI force plates were most appropriate for recording the applied forces, simultaneously for each SI joint and still not disrupting the test (Studies II-IV). As force application during manual SI pain-provocation tests inevitably comprises forces in different directions, it seemed important to evaluate all force components intended for each SI joint when comparing the performance of testing within and between PTs.

However the examination tables had limitations. The working posture was not always optimal as the table with scales could not be set low enough for some PTs; and the examinee could only be placed in the middle of the table because the scales' position was fixed. The table with AMTI force plates had to be very stable because of the sensitive plates and its top was therefore thicker than normal. Some PTs claimed that they could not bend their knees to adopt their habitual working posture. Hence table design might have influenced test performance and, consequently, outcomes.

The reason for the contradiction in results between Study I and Study IV regarding the ability to maintain the amplitude of applied force during the exposure was most probably that the scales were less sensitive than the force plates. This may also partly explain the differences in the magnitudes of force recorded by the scales and the force plates respectively. Another explanation may be that the range of the analysis periods differed in the studies.

In the distraction test a lateral force was applied simultaneously on the right and left ASIS in opposite directions. Different lateral force magnitudes were recorded from the force plates closer to and further from the PT, respectively (Studies II and IV). This indicated unevenly applied magnitudes counter-balanced by frictional forces. Except for the effects of the properties and conditions of the surfaces in contact, the magnitude of friction is directly proportional to the force acting perpendicularly on the surfaces (Özkaya et al., 1998). A vertical force was inevitably applied perpendicularly on the force plates by part of the PT's body weight as she leaned over the examinee when performing the test. Although zero-levelled in the data analysis, the mass of the examinee´s lower torso also acted perpendicularly on the force plates. The lower torso is considered to represent 12.5% of a woman's weight and 11.2% of a man's (Enoka, 2002).

It would of course have been desirable to explore which anatomical structures were provoked and with how much force: but this probably requires an invasive method, beyond the present aim. Using digital scales or force plates in evaluating SI pain-provocation tests can never verify which structures are involved, how much a force affects them or what happens to them; it can only ascertain differences in recorded forces within and between the examiners. For example, the recorded vertical force component was substantial in the distraction test, although most PTs felt they had applied most force laterally. The lateral and/or caudal/cranial force components could have been rather large in the tests without being reflected by the force plates. Thus it would have been more correct in Study II to state 'the force recorded from the

force plates' instead of 'the applied force intended for the SI joints'. The lateral and caudal/cranial force components might possibly be quantified using additional force transducers in the PTs' hands. While interposing a pressure mat between the hands and the examinee could be supposed to record the applied force more directly, this technique only permits recording the force perpendicular to the mat. Such mats are made of compliant material which tend to suffer from use, affecting the reliability of the device (Harms et al., 1995).

Several other techniques have been used, except for force transducers under the hands (Herzog et al., 1993), when measuring forces applied on the column in spinal manipulative therapy. Lee et al. (1990) used a Kistler force plate with a superposed steel platform on which an examination table was placed. The tester stood on the platform. In other studies the tester stood directly on the force plate (Keating et al., 1993; Petty et al., 1996; Petty, 2000). Load cells measuring vertical and horizontal forces were incorporated at the short ends of the frame of the examination table used by Harms et al. (1997). One AMTI force plate was mounted about 250 mm (as estimated from a figure) under the surface supporting the upper part of a specially constructed split-surface examination table used by Triano et al. (1997) and Rogers et al. (2003). The use in the present work of two force plates directly underneath the examinee's pelvis probably provided more detailed information of forces applied.

Examples of factors – mainly context-related – other than those investigated, but certainly influencing the outcome of assessment methods and interventions of pain are shown in Figure 15. For example, all communication between the examinee and the PT is of essential importance, but communication is never 'normal' in the experimental situation. In Studies III-IV the examinees had to conceal their identity from the PT, by lying on the examination table before the PT entered. Further there was no verbal communication, except for greetings. This and the fact that different PTs tested each day may have influenced the examinees' expectations and trust, some reacting to the tests differently from how they would in the normal clinical situation. Such reactions could be due to earlier pain experience.

Physiotherapy has been described as including elements of both art and science (Peat, 1981). Scientific components related to SI pain-provocation tests, have been explored in detail in the present work. Artistic components are certainly also involved. Two artists may create works of art with excellent results but with completely different technique and interpretation from the same scenario. Similarly two PTs, depending on their experience and 'tacit knowledge', may reach the same treatment outcome with different physiotherapeutic methods or with the same method performed somewhat differently. However, when evaluating treatment outcome, the present work indicated that standardized methods are needed.

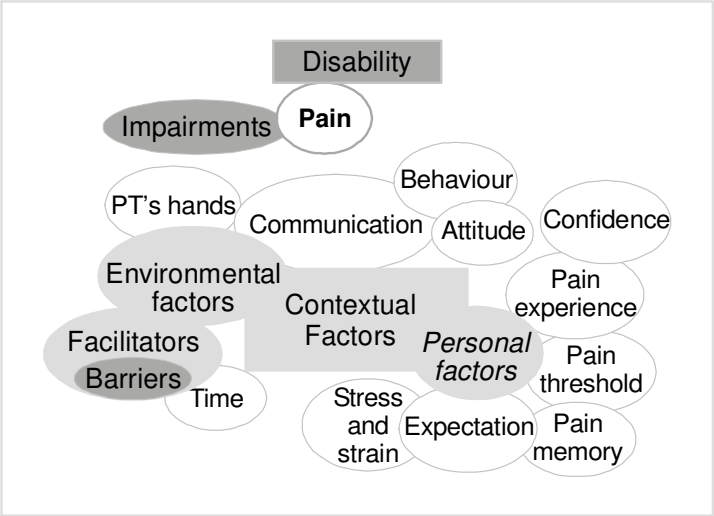


Figure 15 Examples of factors certainly influencing outcomes of assessment methods in evaluating pain.

5.3 Clinical implications

The present results suggest that the primary source of variability in outcomes of SI pain-provocation tests is individual variation in examination technique. It is recommended that the tests be performed standing on the opposite side of unilateral SI joint pain and routinely be performed from both sides of the patient when both SI joints are suspected to be involved. Further, the tests should be maintained for 20 s unless discontinued earlier because of pain or change in pain, simultaneously attempting to keep the applied force steady throughout the test period. As pain was not provoked each time it is recommended that the tests should be repeated at least twice.

The differences shown between forces recorded from the force plate closer to the PT and that further away, and the inability to keep steadiness during the time interval of force exposure, indicated a need for recording the forces intended for each SI joint, simultaneously monitored by the investigator. A more reliable and valid pain evaluation would probably be accomplished through the simple and cheap technique with the scales and the applied force read from a display (Levin, 1989). This technique has proved useful in a clinical study (Nilsson-Wikmar et al., 1999). It could be a first step towards standardizing SI pain-provocation tests until more sensitive equipment feasible for the clinic is developed. A portable pressure mat, which could be placed on the examination table, and equipped with displays for applied forces and a button for pain signals, is one such equipment. Without this technique the clinical validity of SI pain-provocation tests when used for evaluative purpose is probably weak. Thus the recommendation in accordance with the results of the present work is that a negative SI pain-provocation tests should be interpreted with caution in clinical situations.

5.4 Further research

Based on the present findings, force and time recording would be useful in further research regarding methodological issues, clinical usefulness and educational needs related to the SI pain-provocation tests.

Further methodological issues of interest would be whether less variation occurs in recorded force application between the right and left force plates in the distraction test if testing is performed without cross-armed pressure and with the force applied only vertically. Another methodological issue relates to whether the positions of PTs in relation to the examined person influence the directions and the amount of applied force and accordingly the sensitivity of the SI pain-provocation test. The occurrence of reflex muscle tension and whether it varies when examined by different PTs might also be of interest.

As to clinical usefulness there is a need for exploring whether the magnitude of recorded force and time interval of force exposure increase with decreased pain intensity in patients with sacroiliitis and could thus be used in pain evaluation. Other clinical issues relate to whether the compression tests and/or pressure on apex sacralis confer higher sensitivity than the distraction test and whether a combination of two or three tests afford a higher sensitivity than a single test when applied forces are monitored.

Concerning educational usefulness there is a need in further studies to explore how the present findings can be implemented in basic and higher education of PTs. For example will recording of applied force monitored by the investigator be used as a teaching aid, to increase the ability to maintain a given force amplitude when learning the methodology of SI joint tests?

6 CONCLUSIONS

- The PTs maintained a relatively constant force during the time interval of force exposure when recorded with digital scales. The intra-examiner reliability of applied force was acceptable despite the individual differences irrespective of experience. Inter-examiner reliability was unsatisfactory.
  (Studies I)

- Less force was recorded from the force plate closer to the PT than that from the force plate further away. The distribution of force components varied within and between PTs. The PTs´ opinions regarding the directions of forces applied did not always correspond with the recorded force distribution.
  (Study II)

- Less than 2 s to almost 20 s could elapse before pain was provoked in patients with sacroiliitis. The applied force and the time interval of force exposure in the distraction test did discriminate subjects with sacroiliitis from healthy subjects but the sensitivity varied from fair-to-good according to which PT examined the subjects.
  (Study III)

- Variations in recorded forces and time intervals of force exposure were found between repetitions in testing the same subject with sacroiliitis and between PTs when examining the same subjects. The amplitude of force decreased during testing. Subjects rated pain differently depending on the magnitude of force applied.
  (Study IV)

- The accumulated results of the present work indicate that variation in examination technique within and between PTs, irrespective of experience, explain varying outcomes of SI pain-provocation tests and, accordingly, varying sensitivity of the tests. Negative SI pain-provocation testing should be interpreted with caution in clinical situations, unless simultaneously monitoring of applied force is available.
  (Studies I-IV)

7 ACKNOWLEDGEMENTS

I would like to express my sincere thanks to all those who have supported me during the years spent on this work and made this thesis possible. In particular I thank:

All you physiotherapists and all of you testees who willingly gave your time to participate in the studies. Thank you!

Christina H. Stenström, my supervisor with great warmth and intellectual capacity, for generously sharing your broad scientific knowledge and for thoroughly guiding me through the research process, with enthusiasm and never-ending patience. I very much appreciate your confidence in me; but at the same time you gave me necessary supervision and were always there when needed.

Karin Harms-Ringdahl, my co-supervisor, thank you for your always positive attitude and for inspiring discussions, good advice and support.

Thomas Lundeberg, my co-author, for your friendly attitude, believing in my capacity and encouraging me to continue with research after my study in the research preparatory course in physiotherapy.

Lena Nilsson-Wikmar, my co-author, for inspiring research discussions. Thank you for proving that SI joints tolerate an apparently endless number of provocation tests.

Helga Hirschfeld for generously providing me with research resources at the Motor Control and Physiotherapy Research Laboratory and for being so warm-hearted and artistic in preparing surprises.

Ingmarie Apel for all your help with the experiments at the laboratory, teaching me how to analyse the data files and to work with different data programs and for always listening and helping when needed no matter how busy you were.

Erika Jonsson, whose office I share, Tina Jesinkey and Marketta Henriksson for friendly and inspiring sharing of immense knowledge about everything in the laboratory, for always helping at once and for friendship and constructive criticism.

Elisabeth Olsson, for generously providing departmental time for me to complete parts of my work.

Lennart Löfqvist, for helping me to get the scales to work with the ink recorder, and for always being on hand when the ink dried up.

Elisabeth Berg, for pedagogically explained and valuable statistical advices.

Tim Crosfield, for your wise language revision in English and encouraging sense of British humour and for always making me explain what I really meant. Thank you so much for always being on hand!

Inger Björsell, Karin Eriksson, Elisabeth Lundgren-Wirséen, Charlotte Lüning Bergsten, Madeleine Walander, my dear colleagues for covering up for me working extra hard with the students and so making it possible for me to complete this thesis.

All my dear colleagues and the staff at the Division of Physiotherapy for your cheerful support.

Ingiäld Hafström, Kyllikki Widberg and Helene Langeland for invaluable help with patient recruitment.

Helene Alexanderson, Mia André, Gabriele Biguet, Carina Boström, Margareta Börjesson, Stina Lundgren, Anne-Marie Norén and Nina Sturk in the research group I belong to, for valuable and inspiring discussions and for constructive comments on the manuscript of this thesis.

Ulla Evers Larsson for some valuable comments after devoting part of her holiday to reading the manuscript.

Britt Elfving for being of the opinion that statistics is the most enjoyable topic to discuss.

My beloved mother and my late father for giving so much love and support during all my life and for always showing such a great interest in whatever I have done.

My sisters Monica, Nicke, Snippe and Dodo and my brothers Bengt and Lars and all their family members for being such a wonderful big family always backing me up.

My wonderful family: Claes, my husband, for endless love, patience and support; my sons Tom, Jens, Jon and Dick and my daughters-in-law Frida, Charlotte and Silla for love, and never ending support. My grandchildren Isak, Irma and Pixie for giving so much joy and for interrupting my work, so creating new energy.

Soon I will spend lots of time together with all my family and friends.

Financial support for the work of this thesis was gratefully received from the Swedish Rheumatism Association, from the Neurotec Department, Division of Physiotherapy and from the Research Committee for the Health and Care Sciences at Karolinska Institutet.

8 REFERENCES

Adams M.A., Bogduk N., Burton K., Dolan P., 2003. The lumbar vertebral column and sacrum. In: The Biomechanics of back pain, first ed. Churchill Livingstone Longman Group, Edinburgh, pp 27-29.

Albert, H., Godskesen, M., Westergaard, J., 2000. Evaluation of clinical tests used in classification procedures in pregnancy-related pelvic joint pain. Eur Spine J. 9, 161-166.

Altman D.G., 1991. Some common problems in medical research. In: Practical statistics for medical research, Chapman & Hall, London, pp 409-419.

American Pain Society, 2002. Pain in osteoarthritis, rheumatoid arthritis, and juvenil chronic arthritis. American Pain Society, Glenview.

Blower, P.W., Griffin, A.J., 1984. Clinical sacroiliac tests in ankylosing spondylitis and other causes of low back pain--2 studies. Ann Rheum Dis. 43, 192-195.

Borg G., 1982. A category scale with ratio properties for intermodal and interindividual comparisons. In: Geissler, H., Petzold, P., (Eds), Psychophysical judgement and the process of perception. VEB Deutsche Verlag der Wissenschaften, Berlin, pp 25-34.

Bradley, K.C., 1974. The anatomy of backache. Aust N Z J Surg. 44, 227-232.

Braun, J., Sieper, J., Bollow, M., 2000. Imaging of sacroiliitis. Clin Rheumatol. 19, 51-57.

Broadhurst, N.A., Bond, M.J., 1998. Pain provocation tests for the assessment of sacroiliac joint dysfunction. J Spinal Disord. 11, 341-345.

Calin, A., Garrett, S., Whitelock, H., Kennedy, L.G., O'Hea, J., Mallorie, P., Jenkinson, T., 1994. A new approach to defining functional ability in ankylosing spondylitis: the development of the Bath Ankylosing Spondylitis Functional Index. J Rheumatol. 21, 2281-2285.

Carlsson, A.M., 1983. Assessment of chronic pain. I. Aspects of the reliability and validity of the visual analogue scale. Pain. 16, 87-101.

Chinn, S., 1991. Statistics in respiratory medicine. 2. Repeatability and method comparison. Thorax. 46, 454-456.

Clark, P., Lavielle, P., Martinez, H., 2003. Learning from pain scales: patient perspective. J Rheumatol. 30, 1584-1588.

Colloca, C.J., Keller, T.S., 2001. Electromyographic reflex responses to mechanical force, manually assisted spinal manipulative therapy. Spine. 26, 1117-1124.

Daum, W.J., 1995. The sacroiliac joint: an underappreciated pain generator. Am J Orthop. 24, 475-478.

DeLoach, L.J., Higgins, M.S., Caplan, A.B., Stiff, J.L., 1998. The visual analog scale in the immediate postoperative period: intrasubject variability and correlation with a numeric scale. Anesth Analg. 86, 102-106.

Dougados, M., van der Linden, S., Juhlin, R., Huitfeldt, B., Amor, B., Calin, A., Cats, A., Dijkmans, B., Olivieri, I., Pasero, G., 1991. The European Spondylarthropathy Study Group preliminary criteria for the classification of spondylarthropathy. Arthritis Rheum. 34, 1218-1227.

Dreyfuss, P., Michaelsen, M., Pauza, K., McLarty, J., Bogduk, N., 1996. The value of medical history and physical examination in diagnosing sacroiliac joint pain. Spine. 21, 2594-2602.

Enoka, R. M., 2002. Neuromechanics of human movement, third ed. Human Kinetics, Champaign, pp 22-27, p 73.

Ferrante, F.M., King, L.F., Roche, E.A., Kim, P.S., Aranda, M., Delaney, L.R., Mardini, I.A., Mannes, A.J., 2001. Radiofrequency sacroiliac joint denervation for sacroiliac syndrome. Reg Anesth Pain Med. 26, 137-142.

Finch, E., Brooks, D., Stratford, P. W., Mayo, N., 2002. Physical rehabilitation outcome measures, second ed. BC Decker Inc., Hamilton, Ontario.

Fleiss, J. L., 1986. The Design and Analysis in Clinical Experiments. Wiley & Sons, New York.

Fletcher R.H., Fletcher S.W., Wagner E.H., 1996. Diagnosis. In: Clinical Epidemiology: The Essentials, third ed. Williams & Wilkins, Baltimore, pp 43-74.

Fortin, J.D., Kissling, R.O., O'Connor, B.L., Vilensky, J.A., 1999. Sacroiliac joint innervation and pain. Am J Orthop. 28, 687-690.

Freburger, J.K., Riddle, D.L., 2001. Using published evidence to guide the examination of the sacroiliac joint region. Phys Ther. 81, 1135-1143.

Garret, S., Jenkinson, T., Kennedy, L.G., Whitelock, H., Gaisford, P., Calin, A., 1994. A new approach to defining disease status in ankylosing spondylitis: the Bath Ankylosing Spondylitis Disease Activity Index. J Rheumatol. 2286-2291.

Gerlach, U.J., Lierse, W., 1992. Functional construction of the sacroiliac ligamentous apparatus. Acta Anat (Basel). 144, 97-102.

Gray, H., 2000. Anatomy of the human body, 20 ed. Bartleby. com, New York.

Grieve, E.F., 1983. Mechanical dysfunction of the sacro-iliac joint. Int Rehabil Med. 5, 46-52.

Harms, M.C., Bader, D.L., 1997. Variability of forces applied by experienced therapists during spinal mobilization. Clin Biomech (Bristol, Avon). 12, 393-399.

Harms, M.C., Cusick, G., Bader, D.L., 1995. Measurement of spinal mobilisation forces. Physiotherapy. 81, 599-604.

Harrison, D.E., Harrison, D.D., Troyanovich, S.J., 1997. The sacroiliac joint: a review of anatomy and biomechanics with clinical implications. J Manipulative Physiol Ther. 20, 607-617.

Herzog, W., Conway, P.J., Kawchuk, G.N., Zhang, Y., Hasler, E.M., 1993. Forces exerted during spinal manipulative therapy. Spine. 18, 1206-1212.

Huskisson, E.C., 1974. Measurement of pain. Lancet. 1127-1131.

Jacob, H.A., Kissling, R.O., 1995. The mobility of the sacroiliac joints in healthy volunteers between 20 and 50 years of age. Clin Biomech (Bristol, Avon). 10, 352-361.

Johnston, M.V., Keith, R.A., Hinderer, S.R., 1992. Measurement standards for interdisciplinary medical rehabilitation. Arch Phys Med Rehabil. 73, S3-23.

Kandel, E. R., 2000. Principles of neural science, fourth ed. McGraw-Hill cop, New York.

Keating, J., Matyas, T., Bach, T.M., 1993. The effect of training on physical therapists´ ability to apply specified forces of palpation. Phys Ther. 73, 38-46.

Kerns, R.D., Turk, D.C., Rudy, T.E., 1985. The West Haven-Yale Multidimensional Pain Inventory (WHYMPI). Pain. 23, 345-356.

Kirk R.E., 1995. Experimental design: Procedures for behavioural sciences, third ed. Brooks-Cole publishing company, Pacific Grove.

Kokmeyer, D.J., van der Wurff, P., Aufdemkampe, G., Fickenscher, T.C., 2002. The reliability of multitest regimens with sacroiliac pain provocation tests. J Manipulative Physiol Ther. 25, 42-48.

Lahey, M.A., Downey, R.G., Saal, F.E., 1983. Intraclass correlation: There´s more there than meets the eye. Psychol Bull. 93, 586-595.

Laslett, M., 1998. The value of the physical examination in diagnosis of painful sacroiliac joint pathologies. Spine. 23, 962-964.

Laslett, M., Williams, M., 1994. The reliability of selected pain provocation tests for sacroiliac joint pathology. Spine. 19, 1243-1249.

Laslett, M., Young, S.B., Aprill, C.N., McDonald, B., 2003. Diagnosing painful sacroiliac joints: A validity study of a McKenzie evaluation and sacroiliac provocation tests. Aust J Physiother. 49, 89-97.

Latimer, J., Lee, M., Adams, R.D., 1998. The effects of high and low loading forces on measured values of lumbar stiffness. J Manipulative Physiol Ther. 21, 157-163.

Lee D., 1999. The pelvic girdle, second ed. Churchill Livingstone Longman Group, Edinburgh.

Lee, M., Moseley, A., Refshauge, K., 1990. Effect of feedback on learning a vertebral joint mobilization skill. Phys Ther. 70, 97-102.

Levangie, P.K., 1999. The association between static pelvic asymmetry and low back pain. Spine. 24, 1234-1242.

Levin, U., 1989. Transcutaneous electrical nerve stimulation (TENS) on pain in the sacroiliac joint. *In Swedish*: Transkutan elektrisk nervstimulering (TENS) vid smärta i sacroiliacaleden. Project in research preparatory course in physiotherapy. Karolinska Institutet, Stockholm, Sweden.

Levin, U., 2003. Sacroiliac joint pain-provocation tests and force registration. Licentiate thesis. Neurotec Department, Division of Physiotherapy, Karolinska Institutet, Stockholm, Sweden.

Magee D., 2002. Pelvis. In: Orthopedic physical assessment, fourth ed. Saunders, Philadelphia, PA, pp 567-605.

Maigne, J.Y., Aivaliklis, A., Pfefer, F., 1996. Results of sacroiliac joint double block and value of sacroiliac pain provocation tests in 54 patients with low back pain. Spine. 21, 1889-1892.

Margolis, R.B., Chibnall, J.T., Tait, R.C., 1988. Test-retest reliability of the pain drawing instrument. Pain. 33, 49-51.

McCombe, P.F., Fairbank, J.C., Cockersole, B.C., Pynsent, P.B., 1989. 1989 Volvo Award in clinical sciences. Reproducibility of physical signs in low-back pain. Spine. 14, 908-918.

McLauchlan, G.J., Gardner, D.L., 2002. Sacral and iliac articular cartilage thickness and cellularity: relationship to subchondral bone end-plate thickness and cancellous bone density. Rheumatology (Oxford). 41, 375-380.

Melzack, R., 1975. The McGill Pain Questionnaire: major properties and scoring methods. Pain. 1, 277-299.

Merskey, H.; Bogduk, N., 1994. Classification of chronic pain. Descriptions of chronic pain syndromes and definitions of pain terms, second ed. IASP Press, Seattle.

Muche, B., Bollow, M., Francois, R.J., Sieper, J., Hamm, B., Braun, J., 2003. Anatomic structures involved in early- and late-stage sacroiliitis in spondylarthritis: a detailed analysis by contrast-enhanced magnetic resonance imaging. Arthritis Rheum. 48, 1374-1384.

Nilsson-Wikmar, L., Harms-Ringdahl, K., Pilo, C., Pahlback, M., 1999. Back pain in women post-partum is not a unitary concept. Physiother Res Int. 4, 201-213.

Norman, G.F., 1968. Sacroiliac disease and its relationship to lower abdominal pain. Am J Surg. 116, 54-56.

Peat, M., 1981. Physiotherapy: art or science? Physiother Can. 33, 170-176.

Petty, N.J., 2000. The effect of posteroanterior mobilisation on sagittal mobility of the lumbar spine. Man Ther. 1, 25-29.

Petty, N.J., Messenger, N., 1996. Can the force platform be used to measure the forces applied during a PA mobilisation of the lumbar spine? J Manual Manipulative Ther. 4, 70-76.

Pool-Goudzwaard, A., Hoek van Dijke, G., Mulder, P., Spoor, C., Snijders, C., Stoeckart, R., 2003. The iliolumbar ligament: its influence on stability of the sacroiliac joint. Clin Biomech (Bristol, Avon). 18, 99-105.

Potter, N.A., Rothstein, J.M., 1985. Intertester reliability for selected clinical tests of the sacroiliac joint. Phys Ther. 65, 1671-1675.

Price, D.D., Bush, F.M., Long, S., Harkins, S.W., 1994. A comparison of pain measurement characteristics of mechanical visual analogue and simple numerical rating scales. Pain. 56, 217-226.

Puhakka, K.B., Melsen, F., Jurik, A.G., Boel, L.W., Vesterby, A., Egund, N., 2004. MR imaging of the normal sacroiliac joint with correlation to histology. Skeletal Radiol. 33, 15-28.

Riddle, D.L., Freburger, J.K., 2002. Evaluation of the presence of sacroiliac joint region dysfunction using a combination of tests: a multicenter intertester reliability study. Phys Ther. 82, 772-781.

Rogers, C.M., Triano, J.J., 2003. Biomechanical measure validation for spinal manipulation in clinical settings. J Manipulative Physiol Ther. 26, 539-548.

Russell, A.S., Maksymowych, W., LeClercq, S., 1981. Clinical examination of the sacroiliac joints: A prospective study. Arthritis and Rheum. 24, 1575-1577.

Sackett D.L., Haynes R.B., Guyatt G.H., Tugwell P., 1991. The Interpretation of Diagnostic Data. In: Clinical Epidemiology: A Basic Science for Clinical Medicine, second ed. Little Brown and Company, Berlin, pp 69-152.

Sashin, D., 1930. A critical analysis of the anatomy and patologic changes of the sacroiliac joint. J Bone and Joint Surg. 12, 891-910.

Schwarzer, A.C., Aprill, C.N., Bogduk, N., 1995. The sacroiliac joint in chronic low back pain. Spine. 20, 31-37.

Shrout, P.E., Fleiss, J.L., 1979. Intraclass correlation: Uses in assessing rater reliability. Psychol Bull. 86, 420-428.

Simmonds, M.J., Kumar, S., Lechelt, E., 1995. Use of a spinal model to quantify the forces and motion that occur during therapists' tests of spinal motion. Phys Ther. 75, 212-222.

Slipman, C.W., Sterenfeld, E.B., Chou, L.H., Herzog, R., Vresilovic, E., 1998. The predictive value of provocative sacroiliac joint stress maneuvers in the diagnosis of sacroiliac joint syndrome. Arch Phys Med Rehabil. 79, 288-292.

Solonen, K.A., 1957. The sacroiliac joint in the light of anatomical, roentgenological and clinical studies. Acta Ortop Scand. supplement XXVII,

Strender, L.E., Sjöblom, A., Sundell, K., Ludwig, R., Taube, A., 1997. Interexaminer reliability in physical examination of patients with low back pain. Spine. 22, 814-820.

Sturesson, B., Uden, A., Vleeming, A., 2000. A radiostereometric analysis of the movements of the sacroiliac joints in the reciprocal straddle position. Spine. 25, 214-217.

Toussaint, R., Gawlik, C.S., Rehder, U., Ruther, W., 1999. Sacroiliac joint diagnostics in the Hamburg Construction Workers Study. J Manipulative Physiol Ther. 22, 139-143.

Triano, J., Schultz, A.B., 1997. Loads transmitted during lumbosacral spinal manipulative therapy. Spine. 22, 1955-1964.

van der Wurff, P., Hagmeijer, R.H., Meyne, W., 2000a. Clinical tests of the sacroiliac joint. A systemic methodological review. Part 1: Reliability. Man Ther. 5, 30-36.

van der Wurff, P., Meyne, W., Hagmeijer, R.H., 2000b. Clinical tests of the sacroiliac joint. A systematic methodological review. Part 2: Validity. Man Ther. 5, 89-96.

Vilensky, J.A., O'Connor, B.L., Fortin, J.D., Merkel, G.J., Jimenez, A.M., Scofield, B.A., Kleiner, J.B., 2002. Histologic analysis of neural elements in the human sacroiliac joint. Spine. 27, 1202-1207.

Viner, A., Lee, M., 1995. Direction of manual force applied during assessment of stiffness in the lumbosacral spine. J Manipulative Physiol Ther. 18, 441-447.

Vleeming, A., de Vries, H.J., Mens, J.M., van Wingerden, J.P., 2002. Possible role of the long dorsal sacroiliac ligament in women with peripartum pelvic pain. Acta Obstet Gynecol Scand. 81, 430-436.

Walker, J.M., 1992. The sacroiliac joint: a critical review. Phys Ther. 72, 903-916.

WHO, 2001. International Classification of Functioning, Disability and Health: ICF, WHO Library Cataloguing-in-Publication Data, Geneva.

Yin, W., Willard, F., Carreiro, J., Dreyfuss, P., 2003. Sensory stimulation-guided sacroiliac joint radiofrequency neurotomy: technique based on neuroanatomy of the dorsal sacral plexus. Spine. 28, 2419-2425.

Young, S., Aprill, C., Laslett, M., 2003. Correlation of clinical examination characteristics with three sources of chronic low back pain. Spine J. 3, 460-465.

Östgaard, H.C., Zetherström, G., Roos-Hansson, E., 1994. The posterior pelvic pain provocation test in pregnant women. Eur Spine J. 3, 258-260.

Özkaya, N.; Nordin, M., 1998. Force vector In: Fundamentals of biomechanics, second ed. Springer, New York.

Errata Study I

Levin U, Nilsson-Wikmar L, Stenström CH, Lundeberg T. Reproducibility of manual pressure force on provocation of the sacroiliac joint. Physiotherapy Research International 1998; 3: 1-14.

Page 2, second paragraph, line 4 below the heading 'Introduction', page 12, third paragraph, line 4 below the heading 'Discussion' and page 13 below the heading 'References' *For* Huskinson *read* Huskisson

Page 8, line 6 below the heading 'Data analysis' *For* mean of each second and the mean of each occasion *read* mean of each occasion and the mean of three occasions

Page 8, line 12 below the heading 'Data analysis' *For* SEM *read* standard error of the mean (SEM)

Page 8, line 3 below the heading 'Results' *For* $p = 0.6671$ *read* $p = 0.6716$

Page 13, line 1 below the heading 'Discussion' *For* in six *read* in three

Photoghraphs – Figure 1, Figure 2, Figure 3, Figure 4, Figure 5: J Garsten HS-Bild

Clarification Study I

Page 8, line 9 below the heading 'Data analysis' *For* the standard deviation (SD) *read* standard error of measurement (S_W)

Page 8, second paragraph, line 3 below the heading 'Results' *For* SD *read* S_W

Page 11, Table 2 *For* SD *read* S_W

Errata Study II

Levin U, Nilsson-Wikmar L, Harms-Ringdahl K, Stenström CH. Variability of forces applied by experienced physiotherapists during provocation of the sacroiliac joint. Clinical Biomechanics 2001; 16: 300-306.

Page 302, line 1 below the heading 'Data analysis' *Delete* converted into text files and

Errata Study III

Levin U, Stenström CH. Force and time recording for validating the sacroiliac distraction test. Clinical Biomechanics 2003; 18: 821-826.

Page 821, second paragraph, line 3 below the heading 'Introduction' *For* (Levin et al., 1998) *read* (Levin et al., 2001).

Page 823, first paragraph, line 1 below the heading 'Data analysis' *Delete* converted into text files and

Page 823, second paragraph, line 6 below the heading 'Data analysis' *For* all subjects and *read* for healthy subjects, for all subjects with sacroiliitis and

Page 824, last paragraph, line 3 below the heading 'Results' *For* of L4-S1 was excluded from the calculation *read* of L4-S1 was graded as testing negative in the calculation

Page 826, below the heading 'References' *Delete* Levin, U., Nilsson-Wikmar, L., Stenström, C.H., Lundeberg, T., 1998. Reproducibility of manual pressure force on provocation of the sacroiliac joint. Physiother. Res. Int. 3, 1-14.

Clarification Study III

Page 824, Table 1, table text line 3 *for* either side *read* both sides