Intra- and Inter-Rater Reliability of the Test: Anteromedial Position of the Talus in the Talocrural Joint

Master Thesis zur Erlangung des Grades
Master of Science in Osteopathie

an der **Donau Universität Krems**,

niedergelegt an der

Wiener Schule für Osteopathie

von **Ralf DORNIEDEN**

Wien, September 2008

Statistical Evaluation and Translation: Dr. Gebhard Woisetschläger

“By falsification of our expectations, we come in true contact with ‘reality’. The disproof of our misapprehensions is the ‘positive’ experience, we acquire from reality.”

(Sir Karl Raimund Popper, 1998, 359)

Preface

Einen großen Dank an Dr. Gebhard Woisetschläger für die statistische Betreuung und die Übersetzung. Danke für die Gespräche und die konstruktive Hilfe.
Auch einen Dank an Kathi Musil von der WSO für die Betreuung meiner Masterarbeit während des Studiums.
Einen speziellen Dank an Daniel Kauffmann vom Kantonsspital St.Gallen. Haufenweise sind seine organisierten und bestellten Artikel und Bücher meiner Literaturliste bei mir eingetroffen.
Ron Clijssen (Akademie Thim van der Laan), Hans-Josef Haas (spt-education), Frank Rouls (SKOM) und Roman Pfister (HC Bruggen) einen grossen Dank für die Unterstützung an den jeweiligen Testorten.
Meinen Kommilitonen, besonders Carsten Pflüger möchte ich für die langen Gespräche und kreativen Stimulationen während der gemeinsamen Zeit in Wien danken.
Meiner Familie möchte ich Danke sagen, für die lange Zeit, die sie ohne mich verbracht, mich aber trotzdem immer unterstützt hat. Danke meiner Frau Meike und meinen Söhnen Jonas, Simon und Linus.

Heiden, im September 2008

Eidesstattliche Erklärung

Hiermit versichere ich, Ralf Dornieden, geboren am 19.10.1973 in Hildesheim, ehrenwörtlich,

(1) dass ich meine Masterarbeit mit dem Titel:

„Intra- and Inter-rater Reliability of the Test: Anteromedial Position of the Talus in the Talocrural Joint “

selbständig und ohne fremde Hilfe angefertigt habe und keine anderen als in der Abhandlung aufgeführten Quellen benutzt habe.

(2) dass ich die Übernahme wörtlicher Zitate aus der Literatur sowie die Verwendung der Gedanken anderer Autoren an den entsprechenden Stellen innerhalb der Arbeit gekennzeichnet habe.

(3) dass die Arbeit noch keiner anderen Prüfungsbehörde vorgelegen hat.

Ich bin mir bewusst, dass eine falsche Erklärung rechtliche Folgen haben wird.

Ralf Dornieden Heiden, den 23.09.2008

Abstract

Context: Establishing reliable palpatory tests continues to be a critical, yet elusive step in osteopathic medical research and evidence-based clinical practice.

Background: Ankle joint injuries in the region of the apparatus of the lateral ligamental capsule belong to the most frequent sporting injuries. Their proportion is 15-25% (KEEMANN, 1990; RENSTRÖM, 1993; UMANS, 2002). Altogether they belong to the most frequent injuries of the musculoskeletal system (VAN DIJK, 2002).
During a follow-up, six years after the initial trauma, VERHAGEN (1995) found feelings of instability in 39% of all patients, a recidivating predisposition to turgor in 24% and pain syndromes in further 18%.
KONRADSEN ET AL. (2002) come upon 32% chronic complaints of pain, swelling or recurrent sprains in a seven-year follow-up after inversion trauma of the ankle. 72 % of the subjects with residual disability reported that their ankle functionally impaired them.
In the face of the innumerable foot injuries per year and the successive complaints, osteopathic literature about the diagnosis of the foot is almost not existent.

Questions: Is the talus-test according to Ebenegger and Tixa (2004) universally reliable for osteopathic diagnosis of anteromedial dysfunctions of the talus in the talocrural joint or is it only relevant for a subjective diagnosis for each individual osteopath.

Design and Methods: Methodological double-blinded research. Repeated measures design for inter- and intra-rater reliability testing.
20-40 symptomatic as well as asymptomatic subjects (in total 121) have been tested with the test according to Ebenegger and Tixa at four different test locations by three or four osteopaths, each (in total 12). Osteopaths and subjects have been blinded.
Anamnestic data of the test persons (actual complaints of the feet, operations/accidents, known inversion traumata of the ankle) as well as data of the osteopaths (school-leaving year, school, renownedness of the test) have been surveyed by means of questionnaires in advance of the tests.
The intra- and interrater reliability was investigated in two test runs and evaluated by Cohen's kappa.

Results: In the actual study, as maximum kappa indices, kappa= 0.12, “poor” was reached in the case of inter-examiner reliability, and kappa= 0.19, “poor” in the case of intra-examiner reliability.
Furthermore, no causal relationship between the investigated anamnestic data of the test persons and the test findings can be revealed, results of the first test run are more reliable than the outcomes of the second one and practical experience of the examiners has no influence on the results.

Conclusions: The results of this study indicate, that the described talus test is not universally reliable. However, this test has its justification in osteopathy, and is administered despite the low reliability in manual diagnostics.

Key Words: reliability, ankle sprain, osteopathy, talus test

Table of Contents

1. Introduction and Goal of this Study

Ankle joint injuries in the region of the apparatus of the lateral ligamental capsule belong to the most frequent sporting injuries. Their proportion is 15-25% (KEEMANN, 1990; RENSTRÖM, 1993; UMANS, 2002). Altogether they belong to the most frequent injuries of the musculoskeletal system (VAN DIJK, 2002).

Main causes for these injuries to the upper ankle joint as most frequent traumatic lesion of the lower limb are specific causalities, whereby a special risk can be observed in sports with a high jump- and rotational load, as there are basketball, soccer, and volleyball (MENKE, 2000; EBRAHEIM ET AL., 2006), handball, ballet, football, running, snowboarding (CLANTON, 1999; CASILLAS, 2002; CLANTON/WOOD, 2002), and high jump (SCHMITT ET AL., 2003).
Incidence of ankle joint injuries varies substantially within the different sport disciplines.

According to BOZIC (1991), sport is the main reason for collateral ligament ruptures with a proportion of 42%, followed by occupational injuries with 24%, and casual accidents of different kinds.
According to literature, the age peak is described relatively consistently with 15-20 years of age (85% are between 10 and 30 years old), whereby in this age group men prevail.

There are almost no differences in the lateral distribution of the injuries. Recurrence rates are high (YEUNG ET AL., 1994) and many patients experience long term residual symptoms that limit lifestyle (BRAUN, 1999) and affect athletic performance (ANANDACOOMARASAMY AND BARNSLEY, 2006, YEUNG ET AL., 1994).

According to KATCHERIAN (1994), injuries to the talocrural articulation happen with an incidence of 1/10.000 persons per day, according to COLVILLE (1995) with an incidence of seven of 1000 persons per year.
According to ZEEGERS (1995) resulting costs are estimated with Euro 40 million per one million citizens and year.
Thousands of sportsmen are provided for in outpatient clinics every year.

In osteopathic praxis, depending on orientation, the treating osteopath has to cope with acute lesions of the foot, or during anamnesis medical conditions after foot injuries can be found.

Even years after an inversion trauma of the ankle, there is still discomfort in the foot or the region of the lower limb. During a follow-up, six years after the initial trauma, VERHAGEN (1995) found feelings of instability in 39% of all patients, a recidivating predisposition to turgor in 24% and pain syndromes in further 18%.

Other authors report residual complaints in 10-30% of all patients after acute trauma (LIU, 1995; LOFVENBERG, 1994).

DENEGAR ET AL. (2002) examined the range of motion, ligament laxity and posterior gliding of the talus of sportsmen with and without experienced inversion trauma of the ankle. The residual ligamentous laxity was commonly found allowing lateral ankle sprain. Dorsiflexion range of motion was restored in the population studied despite evidence of restricted posterior glide of the talocrural joint. Although restoration of physiological range of motion was achieved, residual joint dysfunction persisted.

Physical therapists, athletic trainers and others involved in the athletes care often recommend weight-bearing and non weight-bearing stretching exercises for the gastrocnemius-soleus complex to restore dorsiflexion after lateral ankle sprain. DENEGAR's results suggest that these therapeutic exercises and the passage of time restore dorsiflexion range of motion but not normal talocrural joint arthrokinematics. The results suggest that a restriction of talocrural arthrokinematics may be common following lateral ankle sprain, the restriction may persist despite restoration of dorsiflexion range of motion and treatment of such restrictions may need to be considered in the rehabilitation following lateral ankle sprain.

KONRADSEN ET AL. (2002) come upon 32% chronic complaints of pain, swelling or recurrent sprains in a seven-year follow-up after inversion trauma of the ankle in 648 participants of this long-term study. 72 % of the subjects with residual disability reported that they were functionally impaired by their ankle - in most cases a question of not

performing sports at a desired level. 4% experienced pain at rest and were severely disabled.

In the face of the innumerable foot injuries per year and the successive complaints, osteopathic literature about the diagnosis of the foot is almost not existent.

The actual study is dealing with the inter- and intrarater reliability of a specific talus test (cf. chapter 4) with an introduction and discussion of the relevant literature (cf. chapter 5) as well as the presentation and discussion of the results of the reliability tests (cf. chapters 7 and 8).
After the introduction to the backgrounds of the study, presented above (chapter 1), I will summarise important information about the foot (cf. chapter 2). Scientific definitions of terms and procedures are presented in chapter 3.
In chapter 6, the methods used in this study will be described. A comprehensive bibliography can be found in chapter 10 and finally, the documents and records, typical for this kind of studies are added in the appendix.

In total, 12 osteopaths have tested altogether 121 test persons twice each in four different locations (Landquart, CH; St.Gallen, CH; Ulm, D; und Magglingen, CH). Intra- and inter-rater reliability was evaluated by means of Cohen`s kappa.
The test - anteromedial position of the talus in the talocrural joint - is a test for the determination of a so-called osteopathic dysfunction[1], as can emerge from an inversion trauma of the ankle.
This test, for example is taught at the Swiss School of Osteopathy in Lausanne and in similar fashion at the school of classical osteopathic medicine (SKOM).

[1] “The osteopathic dysfunction of all bones is named after the direction of the higher mobility”. (LIEM, 2003: Introduction).
“An osteopathic dysfunction manifests in an alteration of the physiological relationship within a tissue or between different organ structures. […] Each dysfunction is highly individual, e.g. depending on the kind and intensity of its cause, of the state of the tissue, the activity, the age and the emotional status of man. Osteopathic dysfunctions have to be palpable, otherwise they can not be treated” (LIEM, 1998: 10-11).

I want to find out with this methodological study, whether different osteopaths come to the same results, when they carry out this test independently. Can this test be applied reliably for osteopathic diagnosis of a restricted talocrural mobility of the talus?

It is the goal of the study to find out, whether the talus-test is universally reliable for osteopathic diagnosis or only relevant for a subjective diagnosis for each individual osteopath. This shall make a reliable decision about the selection of therapeutic possibilities feasible, serving the safety of patients and the success of therapy.
Proceeding on the assumption, that the healing of the primary lesion is of main importance for an enduring solution of a health problem, methods to locate this primary lesion are of great importance.

According to the osteopathic concept of COYSTEN (2008), the talus is the only bone in the human body without insertion of muscles and thus it is not able to correct dysfunctions autonomously. Therefore, according to COYSTEN, dysfunctions of the talus are always primary dysfunctions[2], causing ascendent dysfunctional chains. From this point of view, it is of crucial importance, to diagnose these lesions in order to treat them afterwards.
Criticism of manual palpation techniques often is used to question their clinical relevance. A proof of good agreement, also under restricted but clearly defined conditions with regard to consistency, reproducibility, and reliability, would support the osteopaths in their selection of therapeutic means. By that, an optimised therapy would serve the safety of the patients and be a basic for a successful treatment.
“An important step in establishing the efficacy of any diagnostic procedure is the investigation of its reliability” (HAAS, 1991, 199-208).
BOLINE (1988) writes that palpatory examinations must be accurate if they are to be useful in clinical practise and clinical research. One indicator of accuracy is the interexaminer reliability of measurement, which is evaluated by determining the degree of association between the findings of two or more examiners. He adds that validity and reliability of palpatory evaluations are important issues for clinicians and investigators,

[2] Definitions of a primary dysfunction: „A primary dysfunction most often is of traumatic nature, remains monosegmental and is caused by exogenous influences”. (LIEM, 1998, 11)
„...the primary dysfunction, potentially disequilibrating the organism and health in general most. It can be the oldest dysfunction or the dysfunction having caused most of the secondary dysfunctions.” (CROIBIER , 2006, 26)

since palpatory examination are very commonly used by chiropractic and osteopathic physicians to detect manipulable lesions and to evaluate patient's progress.

For the osteopathic profession it is important to face up to the reliability of osteopathic techniques and procedures. In this respect, osteopathic literature is still very young. Up until now, most studies about the reliability of specific techniques, and procedures have been published by chiropractors (cf. chapter 5) and predominately, they are dealing only with the spine.

Actually, manual techniques or approaches constitute the chief elements in osteopathy, in diagnosis as well as treatment. Achieving definite palpatory skills for assessment takes a great part in the education of the osteopath. This is why the check of the grade of correctness and accuracy of the student's ascertained findings is of essential importance for the student. The knowledge about the reliability of the repertoire of the used manual techniques can make explanations to patients easier for the osteopath and also can make the quantitative assessment of the validity of manual diagnostic tools the dialogue with colleagues who don't know much about the principles of osteopathy easier. As manual ascertained data have a big influence on the therapeutic intervention in osteopathic treatment, the examination of the validity of the used techniques by quantitative and objective means of measurement seems necessary and useful with respect to clinical efficiency and safety.

Nevertheless, anticipating this fact, reliability studies for manual techniques lead to problems in osteopathy but also in other medical fields. Miscellaneous studies describe a poor agreement on the findings of passive motion palpation. But passive motion testing takes a prominent part in osteopathic assessment procedures.

In the course of this, the methodological strategy is crucial for evaluability and a sustainable approval of the study (cf. chapter 6).

2. The Foot in Osteopathy

Accounting for the holistic aspect of osteopathy, in this chapter, the foot will be studied integrated in the “master plan man”.

In my experience, the region of the foot is one of many important sections of the human body, which is often wrongly neglected.

During biped walk, the feet are the only contact points of the body to the ground. A unique co-ordinated biomechanical interaction of soft tissue and bones resorbs impacts, enables different types of locomotion and movements, adapts to the subsurface without any problems and proximally has to bear and balance the body parts and organ systems.

This is a permanent dynamic process. During walking, the feet have to resorb the 4.5-fold, during jogging, even the nine-fold of the bodyweight.

According to KLEIN/SOMMERFELD (2004) the foot consists of a variety of complex articulations, whose interaction is hardly explicable to its whole extent.

Perhaps, this is a reason, why many osteopaths tackle only superficially with the foot and its function and why it is only a subject of minor interest in osteopathic education.

Alterations of the biomechanics in the region of the foot bring about, that other systems, as e.g. the knee, hip, pelvis, and lumbar spine are forced to compensate[3].

From a kinematical view (MATHIS, 2006), the whole lower limb, including the pelvis and - via the iliolumbal ligament – also the lumbar spine can be counted to the kinematical movement chain of the foot. Thus, static alterations can proceed up to the head.

[3] “Compensation designates the adaption of the body to abnormal situations or to disease-causing influences. In compensation, there subsists no loss of mobility, but a tendency to dysfunction.” (LIEM, 1998: 12)

However, progressing compensation mechanisms can be observed not only in the parietal field. Alterations may occur in all anatomically associated structures and systems.

Two osteopaths wrote in the early nineteen fifties about the comprehension and diagnostics of the foot. For both authors, anatomical, physiological and functional fundamentals are vital prerequisites for the treatment of the foot:

BEAL (1951) writes: “The basic principle of the osteopathic treatment of mechanical disorders of the foot is that of the restoration of the normal function. The restoration of normal function implies freedom of movement in the articulation of the foot. […] A knowledge of the anatomy and physiology of the foot forms the basis for specific management of foot problems.” (BEAL, 1951, 98-111)

NORTHUP (1953) is proceeds from the same prerequisites for assessing a foot: “Knowledge of the normal foot under normal conditions is essential to the proper evaluation of abnormal functional conditions” (NORTHUP, 1953, 150-156). In order to understand the foot and its possibilities, it is essential to face up to the functional anatomy, the biomechanics of the foot, in order to recognise dysfunctions and compensation mechanisms.

According to KLEIN/SOMMERFELD (2004), the three-dimensional movement in the upper ankle joint can be looked at from two reference systems: On one hand, the movement of the talus in relation to the lower leg and on the other hand, the movement of the lower leg in relation to the talus. Since the foot remains on the ground for a longer time during walking and the lower leg moves in relation to the fixed foot, it is important to consider both reference systems in the kinematical analysis.
By this approach, the influence on proximal articulations can be elucidated and be explained more precisely with regard to function.

When inspecting the movement of the talus in relation to the tibia and fibula, supinations and adductions can be observed as associated movements. Adduction and supination can be observed during the plantar flexion and pronation during dorsal extension.

According to KLEIN/SOMMERFELD (2004), the head of the talus moves further medial during plantar flexion, representing the adduction component. Additionally it subsides a little bit to lateral, which is complying with the supination component.

Thus, altogether, the talus moves to anteromedial during plantar flexion. The other way round, during dorsal extension the talus moves to posterolateral.

Concerning the actual test, this means, that (a) the dorsal extension will be restricted and (b) the talus will be restricted in its posterolateral movement in relation to the malleolar fork, if the talus is forced into an anteromedial position in relation to the malleolar fork by external or internal forces (as e.g. during an inversion trauma) and if it is retained that way.

DENEGAR (2002) describes that dorsal extension, but not normal talocrural joint play, recovers frequently after an inversion trauma of the ankle.

If a plantar flexion in the upper ankle joint can be noticed during the inspection of the movements of the tibia and fibula in relation to the talus, the lower leg will move posteriorly. Additionally, it will rotate outwards and tilt medially. During dorsal extension, the lower leg moves anteriorly with an interior rotation and a lateral tilt.

The previous considerations can be assigned to everyday movements, too. During the movement from stance into squat, both knees and hips are bent simultaneously.

The upper ankle joint has to allow a dorsal flexion. The tibia glides anteriorly, associated with an internal rotation. Additionally, flexion in the knee is associated with an internal rotation. Ideally, the internal rotation is regularly distributed on both articulations.

Accordingly, compensation has to take place within the chain of the participating joints after changes of these relations. The same accounts for the hip joint, the pelvis, and the lumbar spine. With this thought model, compensations after an inversion trauma of the ankle can be explained parietally.

Therefore, complaints and symptoms can be theoretically explained with these considerations in anamnesis and the subsequent inspection, and eventually they can be causally traced back to a dysfunction of the foot, e.g. after an inversion trauma of the ankle.

3. Requirements for Tests from a Scientific Point of View

It is indispensable to examine standards of approved studies and approaches before performing a scientific study. In this chapter, I will explain the most important terms and laws and their application in the present study.

Principally, the significance of a testing method depends on two factors, reliability, and validity.
For a scientific approach, reproducibility, consistency, and reliability are of great importance for all ways for gaining information in the diagnostic field. Naturally, this also applies to manual diagnostic techniques. A decisive factor for the course of therapy for a patient should fulfil all these criteria.

Reliability
Reliability characterises the formal exactness of scientific investigations. Reliable scientific results are free of random errors, i.e., the same results are obtained from repeated experiments under the same general conditions.
Thus, reliability is a measure for the replicability of the results under the same conditions. Reliability (accuracy of measurement) describes one of the three most important quality criteria of empirical investigations aside of validity and objectivity (agreement of examiners=inter-rater reliability).

Validity implies the resilience of the operationalisation (“To what extent is the test measuring, what it is supposed to measure?“) and on the other hand, the resilience of the predications or conclusions basing on the measurements (“To what extent does it apply that X influences Y?”)

The concept of reliability refers to the steadiness of the measuring process and device. It implies the characteristics of reproducibility and consistency under the same or similar conditions.

Inter-rater reliability refers to the agreement between testers, investigated at the same moment and in reference to the same test objects. Common values are the coefficient of agreement according to Holst or Cohen's kappa. In the present study, inter-rater reliability is determined in the first and second test run, each, and evaluated by means of Cohen's Kappa.

Intra-rater reliability refers to the agreement on results gained during repeated measurements, performed by a single examiner.
In the current study, a second trial was performed with all test persons in randomly changed order and then the conformity of the results of the two test runs was evaluated.

Reproducibility
Reproducibility demarks the repeatability of empiric scientific research methods. It is a basic requirement for scientific experiments, tests, and analyses.

Under equal test conditions the same results should be obtained (within the measuring error, which has to be taken in account). In order to verify the reproducibility, an adequate recording of the experimental setup and of the performance of the test are standard.

Totally equal conditions are only difficult to ensure in the dynamic systems of the patients as well as of the therapists. In these cases, when the results cannot be repeated as often or exact as wanted due to influences, which are not checked or difficult to check, statistical methods have to be used (SOMMERFELD 2000.4).

Influences on the patients by postulated therapy effects (RUSSEL, 1983 and GOLDSTEIN, 1978) by palpation or influences on the diagnostician by other information flows de facto cannot be precluded. Therefore, apriori, lower reproducibility is adjudged to manual methods than to instrumental ones. Nevertheless, some studies in the

literature show, that seemingly “subjective” methods are more reliable than “objective” ones, among them also instrumental (KEATING ET AL., 1990; BOLINE ET AL., 1993; BOLINE ET AL., 1988 and MOOTZ ET AL., 1989).

In order to ensure, that no or as least as possible information can be exchanged between the examiners, studies have to be blinded.

In the present study, raters tell the results their recording assistants. After each test run, test logs are collected and kept save by the director of the study.

A visual cover, ensuring that only the legs were visible, and the randomly changed order of test persons in the second test run prevented recognition of the test persons. The osteopaths had no access to the records and did not communicate with others.

Therapy effects by touching the patient during diagnosis can only be precluded by simultaneous diagnosis by all therapists, which, though, hinders the blinding. Thus, the prerequisite of the same conditions might not be fulfilled for the patients as well as the therapists. HAAS (1991) doubts the possibility of blinding of one and the same examiner against the same patient.

Consistency

Consistency is a measure for the exemption from contradictions, which can be tested, if different contradictory statements occur, which are equally registered by different individual observers (SOMMERFELD, 2000).

In the actual study consistence is reached, when each osteopath finds the same dysfunction at one and the same test person, and another dysfunction at another test person without contradictions of the individual therapists.

Sensitivity

Sensitivity characterises the capacity (sensibility) of the test, to find an attribute, when it is present for real, i.e. to deliver correct positive results. If a test is highly sensitive, normally probability increases, that individual persons with certain characteristics are

identified correctly (correct positive), but also that persons without these characteristics are identified as positive, as well (false positive) (SOMMERFELD, 2006).

Specifity

Specifity describes the capacity of a test, to indicate a negative result, if the symptom is not present, that means to produce correct negative results. For tests with a high specifity, normally, probability increases that persons without certain attributes are identified correctly (correct negative), but also that subjects with this attributes are identified as negative (false negative) (SOMMERFELD, 2006).

4. Introduction into the Test: Anteromedial Position of the Talus in the Talocrural Joint

The test assessed in the present study is described by Tixa and Ebenegger (2004). Both authors are lecturers at the Ecole d`osteopathique (osteopathic school) in Lausanne and freelancers with own praxis.

In osteopathic literature, no description of a functional, commonly known gold standard for the determination of an anteromedial dysfunction of the talus in the talocrural joint can be found. Therefore, it was crucial for selecting this specific test according to Tixa and Ebenegger, to find it in writing in an osteopathic book The second reason was the easy teaching and administration.
In the following, the test will be described for the right foot:

The left hand encompasses the calcaneus, the second finger lies on the posteromedial tuberculum of the talus. The right ring- or the second finger is placed on the neck of the talus. The foot is in contact with the sternum of the osteopath (cf. Fig. 1 and 2).

Fig. 1: Grip technique

Fig. 2: Grip technique from lateral

The therapist moves the right foot with his sternum in dorsal extension and assesses the quality of the movement of the talus by the aid of his both second fingers. Testing is done via a rotation of the torso around the right hip. By this movement, only little tension is generated in the arm and the fingers can sense well (cf. Fig. 3).

Fig. 3: Grip technique and performance of the test

The medial tuberculum of the talus is palpable best during posteriorisation of the talus. The test assesses the quantity and quality of the movement of the talus. It is performed comparing side by side.

Starting from the biomechanical features, described in chapter 2, the talus should glide posteriorly and laterally during dorsal extension. If there is an osteopathic dysfunction in the sense of an anteromedial talocrural dysfunction, the talus cannot glide to posterior and remains anterior or this movement is distinctly restricted.

This restriction of movement is distinctly noticeable during a passive dorsal extension initiated by the osteopath.

In order to perform the test under ideal conditions, it should be paid attention to the height of the massage table and taken care, that the patient's feet are resting on the foot of the massage table.

5. Manual Diagnosis

In this chapter, research about the reliability of manual diagnosis (inter- and intraexaminer reliability) and its consequences for the present study will be presented. In this respect, no literature was found for the topic foot.

In order to document the strategy for literature research, the procedure is described first.

The following medical databases were browsed for relevant literature:

- One of the most important English-speaking databases for medical biochemical literature is MEDLINE, an acronym for "medical literature online". It comprises about 10 million articles from 3.900 periodicals of approximately 70 countries since 1966.

- The National Library of Medicine (NLM in Bethesda, USA) in co-operation with the National Centre for Biotechnology Information (NCBI) provides the cost-free Internet version PUBMED, with worldwide more than 4.600 biomedical journals since 1950. More than 60% of the papers covered in this database are summarised in abstracts, which facilitate the aim of electronic literature research, to find a manageable number of relevant articles.

- The Cochrane Collaboration is an international not-for-profit and independent organization, dedicated to making up-to-date, accurate information about the effects of healthcare readily available worldwide. It produces and disseminates systematic reviews of healthcare interventions and promotes the search for evidence in the form of clinical trials and other studies of interventions. The Cochrane Collaboration was founded in 1993 and named after the British epidemiologist, Archie Cochrane.The major product of the Collaboration is the Cochrane Database of Systematic Reviews which is published quarterly as part of The Cochrane Library. Those who prepare the reviews are mostly healthcare professionals who volunteer to work in one of the many Cochrane Review Groups, with editorial teams overseeing the preparation and maintenance of the reviews, as well as application of the rigorous quality standards for which Cochrane Reviews have become known.

Relevant papers were searched with the aid of electronic database queries and by manual monitoring of selected references. The research implicated English- German and French speaking literature.

In order to obtain sound results, literature research was done with English search terms and under keyword search without temporal limitations. Key words were “foot”, “ankle”, “ankle sprain”, “reliability”, “reproducibility”, “inter-/ intraexaminer”, “diagnostic tests”, “manual diagnosis,“ and “osteopathy”.
Literature research was complemented by a manual inspection of relevant periodicals, web pages of relevant professional associations, and abstracts.

Furthermore, references and names of authors were found in relevant literature, which could be used as concrete research criteria in the databases.

Finally, the selected articles and books were ordered via inter-library loan in the Hospital St. Gallen /CH.

Especially during the search for talus tests, I talked with many osteopaths about their experiences in talus tests and the treatment of the foot, additionally.

With the coming of the evidenced medicine in osteopathy, osteopathic techniques and modes of action are increasingly scrutinised and examined. The discussion about the reliability of manual diagnosis in osteopathy has begun just in the last years.
The classical osteopathic treatment is administered to the patients by manual techniques.

SOMMERFELD (2005) writes that passive motion testing, which means assessments of joint-mobility, which is induced and executed by the therapist, takes a great part in functional osteopathic diagnostics. In many cases it represents the deciding factor, which structure in relation to another should finally be treated.

Concerning reliability in manual diagnosis, many studies have been published. Most of them have been performed by chiropractors and are dealing with tests or diagnostic procedures for the spine and the pelvis. In this respect, no studies could be found concerning the foot.

It has to be distinguished between reviews, as for example by ALLEY (1983), RUSSEL (1983), SEFFINGER (2004), HAAS (1991), HESTBAEK AND LEBOEUF-YDE (2000) and STOCHKENDAHL (2006) and “normal” studies. The papers cited here, were covered by reviews, too, but are mentioned, in order to provide an exemplary insight in the studies and study designs as well as to point out consequences for the current study.
In the reviews it was made clear, that pain provocation studies feature a higher reliability (“acceptable”) than palpatory tests of landmark, positional asymmetry, movement

asymmetry and muscle tension (“poor”) (SEFFINGER, 2004; HESTBAEK and LEBOEUF-YDE, 2000).
Altogether, manual diagnostic techniques feature a high variability of reliability. Many studies could not be enclosed in the reviews due to deficient methodology, cutting down the number of studies evaluated in the reviews (ALLEY, 1983; SEFFINGER, 2004; HAAS, 1991; HESTBAEK AND LEBOEUF-YDE, 2000 and STOCHKENDAHL, 2006).
It became obvious, that studies always should be blinded for the examining osteopaths (a) against the results of the other examiners, (b) the medical findings and (c) the test persons, as well as (d) for the test persons against the medical findings (HAAS.1991; HAWK, 1999; PATIJN, 2002; STOCHKENDAHL, 2006).

The test group should comprise asymptomatic as well as symptomatic test persons (KEATING, 1989). It is suggestive to perform a pilot session as a preliminary test, in order to recognise possible errors in methodology and in the course of the study (HAWK, 1999; PATIJIN, 2002). The statistical evaluation (of anonymised data) should be done by means of Cohen’s kappa by an independent person (HAAS, 1991; PATIJN, 2002; STOCHKENDAHL, 2006).
Preceding technique trainings are discussed controversially (McCONNELL, 1980; KELSO, 1981; BEAL ET AL., 1982; JONSTON ET AL., 1982; SCHÖPS ET AL., 2000; PATIJN, 2002; SEFFINGER, 2004; DEGENHARDT, 2005). Furthermore, it becomes obvious, that the testers’ experience does not influence the results (MIOR, 1990; HESTBAEK and LEBOEUF-YDE, 2000). In the study by LUNDIN (1999) it was demonstrated, that testing under stress or in noisy environment has a negative influence on the reliability. Another important item is the possibility that administration of diagnostic techniques might change test results, and a therapeutical effect of diagnosis cannot be precluded (FJELLNER, 1999).
In the following, results of the reviews and studies as well as connections to the present one will be presented. At the end of this chapter, the most important aspects will be summarised again.

In literature, it is distinguished between four different fundamental types of palpatory tests (DINNAR ET AL., 1980, 1982, KUCHERA and KAPPLER, 2002): Differentiation of

tissue textures, evaluation of static landmark positional asymmetry, evaluation of motion asymmetry and assessment of tenderness.

SEFFINGER ET AL. (2004) performed a systematic review to determine the quality of the research and assess of the interexaminer and intraexaminer reliability of spinal palpatory tests. The authors used 13 electronic databases and manually searched the literature from January 1, 1966 to October 1, 2001. Forty-nine (6%) of 797 primary research articles met the inclusion criteria. Two blinded, independent reviewers scored each article. Consensus or a content expert reconciled discrepancies. The quality scores ranged from 25 to 79/100. Subject description, study design, and presentation of results were the weakest areas. The 12 highest quality articles found pain provocation, motion, and landmark location tests to have acceptable reliability (κ=0.40 or greater), but they were not always reproducible by other examiners under similar conditions. In those that used kappa statistics, a higher percentage of the pain provocation studies (64%) demonstrated acceptable reliability, followed by motion studies (58%), landmark (33%) and soft tissue studies (0%). Regional range of motion is more reliable than segmental range of motion and intraexaminer reliability is better than interexaminer reliability. Overall, examiners discipline, experience level, consensus on procedure used, training just before the study, or use of symptomatic subjects do not improve reliability. The conclusion was that the quality of the research on interreliability and intrareliability of spinal palpatory diagnostic procedures needs to be improved. Pain provocation tests are most reliable. Soft tissue paraspinal palpatory diagnostic tests are not reliable (SEFFINGER ET AL., 413-425).

Consistently with literature, no effect of the practical experience of the osteopaths on the agreement could be observed in the actual examination. Nevertheless, a distinct difference in the acting with the tissue could be observed in the sense of a more secure, self-evident acting of the osteopaths with higher experience.

DEGENHARDT ET AL. (2005) investigate the interobserver reliability of common osteopathic palpatory tests used to evaluate the lumbar spine in a double-blinded study. 119 test persons are examined by three experienced osteopathic medical examiners in two subgroups. The osteopaths perform palpatory tests of tenderness and tissue texture

changes, as well as – in three planes - vertebral positional asymmetry and motion asymmetry. In the first subgroup (n=42) Kappa-indices ranging from –0.02 to 0.34 (within the poor-to-fair reliability range) are reached. After a following consensus training reliability increases in the second subgroup (n=77), rising to kappa (κ)= 0.45 for tissue texture changes (“moderate”), to κ=0.68 for tenderness (“substantial”). The Reliabilities for positional asymmetry in the transverse plane and for rotational motion asymmetries improve only slightly (κ=0.34, κ = 0.20, respectively).
The authors conclude from the results of the consensus training, that osteopathic medical educators need to modify to better calibrate and standardize palpatory diagnostic skills (DEGENHARDT ET AL., 465-473).

DEGENHARDT ET AL. (2005) propose that acceptable interobserver reliability was not found in earlier studies of palpatory diagnostic tests for several reasons. First, human beings are not static entities. Homeostatic mechanism, such as heart rate, respiratory rate, blood pressure, and neuromotor reflexes are responsible for constant inherent neurophysiologic variability. The authors add that the neuromusculoskeletal system changes on some level each second according to the impulses or stresses that individuals experience. This inherent neurophysiologic variability occurs in both the examiner and subject. The dynamic nature of the human body could thus challenge clinicians’ ability to reliably perform palpation because, by definition, reliability determines the reproducibility of findings when a test is repeated to evaluate an unchanged attribute. Second, the authors mean, some current educational systems may not provide students or clinicians with the necessary skills for reliably performing palpation. Third, because of the relative isolation of many private practises, experts in manual medicine may not perform palpatory procedures in similar ways, limiting the likelihood of reliability.
They hypothesize that examiners would have improved interobserver reliability after participating in a rigorous consensus-training program. Because palpation that induced repetitive motion would be more likely than positional palpation to change the characteristics being examined, they also predicted that positional asymmetry tests would require less training than motion asymmetry testing to obtain at least moderate interobserver reliability.

At this point, I want to say, that consensus training does not comply with everyday praxis. In praxis, a test is performed according to the educational level and eventually varied depending on practical experience.
A range of techniques and methods is adjusted to the patients individually, implicating, that possibly some of them have not been used for a longer time.

A consensus-training program of a technique or a scheme in order to improve the interrater reliability does not comply with everyday praxis and can be implemented only for scientific reasons. Therefore, the instruction for the actual test was limited to written information and a demonstration, only.

FJELLNER ET AL. (1999) assess the interexaminer reliability in physical examinations of the cervical spine at forty-eight subjects, age range 18 to 63 years. Two physiotherapists independently evaluated a number of clinical tests of passive general and intersegmental movement. Acceptable kappa (κ) values were obtained in several of the clinical tests of passive general motion range but in few of the clinical tests of passive intersegmental movement. More clinical tests had acceptable reliability and less bias in symptomatic subjects than asymptomatic subjects. Many of the clinical tests of passive general motion range were shown to be reliable. The increased number of acceptable kappa (κ) values obtained in the symptomatic subjects indicates that further studies of the reliability of the clinical tests of passive intersegmental movement should be performed on patients (FJELLNER ET AL., 511-516).

FJELLNER ET AL. discuss, that it cannot be precluded, that poor reliability might be caused by a change of the clinical evidence by the examinations. Another possible reason could arise from the study design with standardised, repetitive examinations during stress.

The purpose of the review by HAAS (1991) is to evaluate the conclusions of reliability studies conducted on diagnostic procedures in terms of the appropriateness of the statistical analyses and experimental designs. He points out the difficulty in controlling for conscious and unconscious memory cues such as perfume, clothing, or unique

physical features. Because of these cues it may be virtually impossible to ensure sufficient blinding of raters to accurately assess intraexaminer reliability.

In the present study, the subjects had to undress only so far, that the osteopaths could see only two thirds of the thighs, the lower legs, and the feet below the visual cover. Exclusion criteria (cf. chapter 6) were defined in a way, that recognition should not be possible.

HAAS claims standardised courses for the analysis, the design, and the interpretation of the results with regard to reliability. He recommends the following items:

1. Kappa is the statistic of choice for nominal data, 2. Reliability should be reported segment by segment. Data should not be collapsed to give regional reliability without compelling clinical justification, 3. In reviewing the literature, the appropriateness of analysis, design and conclusions should be noted, 4. Comparisons of reliability studies should be confined to investigations with like concordance indices, 5. A discussion of clinical and (or research) utility should be included in the evaluation of the strength of concordance, 6. An attempt should be made to use a representative sample of subjects, 7. Examiners and subjects should be blinded and it is the best to use subjects with no previous experience with the diagnostic procedure under investigation.

HAWK (1999) performs a preliminary study, assessing the intra- and interexaminer reliability for the indication of chiropractic manual therapy for the segments of the lumbar spine (Th12/L1 to L5/S1). Four chiropractors trained in flexion - distraction technique with different experience palpate 18 test persons by means of static and motion palpation and visual observation. In order to reproduce the “real-life” clinical setting, each examiner is allowed to use his individual combination of techniques.
The Kappa-indices are calculated for all comparisons. The intra-examiner reliability is higher than inter-examiner reliability, which generally is in the “poor” to “slight” categories (HAWK ET AL., 382-389)

In HAWK’s study, the clinicians are blinded to the other clinicians’ findings. The order of re-examination is varied, and clinicians do not record their own findings to blind them as

much as possible to their first assessment of each subject. The investigator analysing the data is blinded to the identity of all subjects and clinicians. To standardize the method of assessment, all clinicians are provided for with a written and verbal explanation of the features specified in their definitions.
A pilot session with three subjects and three clinicians preceeds the study to test the study protocol and provide information on whether the examination procedures appeared to cause changes in the subject's pattern of subluxation.
The question for the chiropractors for facilitating the decision-making is: "Would you adjust this vertebra?"

In the present study, the procedure was similar: First, the performance of a pilot session in order to test the records and to uncover deficiencies and only then the main examinations. All, examiners, test persons and statistician, were blinded to all results.

HESTBAEK AND LEBOEUF-YDE (2000) review studies about the reliability and validity of chiropractic tests used to determine the need for spinal manipulative therapy of the lumbopelvic spine, which were performed between 1976 and 1995, taking into account the quality of studies.
In summary, only studies focusing on palpation for pain (palpation for tenderness) have consistently acceptable reliability values. These results are patient-induced, whereas clinician-induced interpretations are worse.
Palpatory tests of landmark positional asymmetry, movement asymmetries, and muscle tension result in poor reliability.
Concerning the palpation for muscle tensions only BOLINE (1988) and KEATING (1990) meet the quality criteria of the authors (HESTBAEK AND LEBOEUF-YDE, 258-275).

An interesting detail in this review is the comparison of results obtained by students, experienced clinicians, and experts. Some researchers have explained poor results by citing the examiner's lack of experience (BOLINE, 1988), but the opposite has been demonstrated. The review shows that when comparing students with clinicians, the students consistently did best. When comparing experts with ordinary clinicians, the experts did not produce better results than ordinary clinicians and on some cases

produced worse results. This may occur because the clinician develops idiosyncratic standards for the procedure.
Keating (1989) criticized the over-reliance on asymptomatic students as palpatory subjects.

For this reason, the questionnaires had to be returned via Internet in advance of the present study, in order to get an idea about the distribution of symptomatic and asymptomatic patients.

MIOR ET AL. (1990) conclude, that experience does not play a significant role in the diagnostic test analysed, but rather that clinicians may establish their own criteria by which to determine the standards of a given test.
Their argumentation for the differences in reliability are, that we always will have to accept some differences, even within a wide level of tolerance, due to human nature.

LUNDIN ET AL. (1999) investigate the reliability of distal pulse palpation (A. dorsalis pedis and A. tibialis posterior). Nine examiners palpate the pulse of 25 patients with suspected lower limb arterial disease. The palpation findings are compared to the ankle/brachial index (ABI). The palpation technique is not standardised. The investigations are performed under two different conditions, an undisturbed examination situation with sufficient time, and a hectic outpatient clinic.
The agreement of the palpation results with the ABI is estimated by percent agreement and kappa statistics. Different ankle/brachial indices are used as reference points for the separation of palpable from non-palpable arteries.
The proportion of underdiagnosis is 33% with an ABI of 0.96 as criterion of disease. Overdiagnosis occurs at 19%. With an ABI of 0.71 the results are 19% for underdiagnosis and 34% for overdiagnosis, respectively. With a mixed concept (ABI>0.96, underdiagnosis; ABI<0.71, overdiagnosis), the overall proportion of misdiagnosis is between 9.7% and 32.3%.
An undisturbed examination situation with sufficient time to examine each patient proves to be of primary importance. Under quiet conditions, the examiners reach a kappa of k= 0.68, whereas the agreement is lower with k =0.38 in the busy outpatient

clinic. The results emphasize the importance of optimal examination circumstances and concentration on the procedure (LUNDIN ET AL., 252-255).

In the present study, purposely time for the test has not been limited, in order to cause no additional stress.

LICHT ET AL. (2007) performed an interrater reliability test with myofascial trigger points (MTrP). A doctor and a physiotherapist examined a total of 304 muscles in 38 test subjects with extensive experience in the field of MTrP following a standardized procedure. The evaluation was done using kappa statistics. The reliability of the clinical examination of the MTrP ranges from “good” to “excellent”, with kappa values up to 0.82.

SCHÖPS ET AL. (2000) investigate the reliability of manual examination at the cervical spine. Twenty patients suffering from neck diseases and twenty asymptomatic test persons are randomised and assessed by five examiners blind to patients’ histories. The cervical zygapophysial joints and the superficial neck muscles are tested for pressure sensitivity and a segmental function test of the segment C0/C1 to C7/Th1 is performed regarding to hypomobility and pain.
Significant findings are described only for the palpation of tension in joint-facets and superficial neck muscles as well as for induced kinesialgia. No significant relationship can be found between the patient’s health status and the findings from muscle palpation and functional examination of the motion segments.
The reliability between the examiners turns out to be “fair” to “moderate”. (0.2< Kappa< 0.6) (SCHÖPS ET AL., 2-7).
In the discussion, SCHÖPS ET AL. point out, that several authors, as KELSO (1981), McCONNELL ET AL. (1980), BEAL ET AL. (1982) und JONSTON ET AL. (1982) have shown, that a consensus training in advance of the examinations with precise guidelines of the techniques being applied, the definition of benchmark assessment and consistent documentation improves interrater agreement.

In the actual study, the examiners were asked to specify the nature of the connective tissue in the sense of “soft” or “tense”, for additional information apart from the specification of the presence of a dysfunction. Since no training was performed with

regard to these judgements, it became obvious, that many of the osteopaths could not do this job adequately or even at all.

STOCHKENDAHL ET AL. (2006) performed an systematic review and meta-analysis on relevant literature published from 1965 to 2005, identified using the electronic databases MEDLINE, MANTIS, and CINAHL and checking of reference lists.
Two reviewers extracted descriptive data from included articles independently. A six-point scale was constructed to assess the methodological quality of original studies. A meta-analysis was conducted among the high-quality studies to separately examine the consistency of data on motion palpation, static palpation, osseous pain, soft tissue pain, soft tissue changes, and global assessment. A standardized method was used to determine the level of evidence.

Poor reproducibility of spinal palpation is continuously being reported and authors of recent reviews have criticized the quality of studies. This article critically analyses the literature pertaining to the inter- and intraobserver reproducibility of spinal palpation to investigate the consistency of study results and assess the level of evidence for reproducibility.
The quality score of 48 included studies ranged from 0% to 100%. There was strong evidence that the interobserver reproducibility of osseous and soft tissue pain is clinically acceptable (kappa > or = 0.4) and that intraobserver reproducibility of soft tissue pain and global assessment are clinically acceptable. Other spinal procedures are either not reproducible or the evidence is conflicting or preliminary (STOCHKENDAHL ET AL., 475-485).

According to STOCHKENDAHL ET AL., there is no standardised and validated method for judging the methodological quality of studies for reproducibility. In order to be able to appraise such studies, a 6-point scale was designed by the authors, basing on approved requirements for clinical studies for reproducibility and on the common recommendations for systematic assessment of the accuracy of tests.

A high quality of a study was defined, if the score for methodological quality (6-point system described below) was at least 50% (3 points), low quality with a score <50%.

The quality score reflects the relevance and the adequacy of three dimensions, which might affect the interpretation of the results: the sample characteristics and the design of the study as well as the statistical evaluation.

The operational definitions of the quality criteria are:

- Order of observers conducting the test(s) randomised (1).
- Case mix: Both symptomatic and asymptomatic subjects. For studies with an intentional case mix, the case mix must represent a natural clinical population, i.e. a population consisting of a variety of patients with regards to gender, age, and problems in different part of the spine. (If the purpose is to look at subgroups, such as symptomatic patients only, credit will be given) (1).
- Observers blind to other observers findings (1).
- Observers blind to confounding information (such as patient history) (1).
- Subjects blind to observers findings (1 point for true/complete blinding. 0.5 point for no blinding but subjects unable to affect results) (1).
- Kappa or ICC (interclass correlation coefficient) used for analysis (1).

In the actual study, the points 2 through 6 are met (cf. chapter 6). Due to the design of the study, a randomisation (point 1) could not be implemented.

In the second version of the protocol about reproducibility and validity of the FIMM (International Federation for Manual/Musculoskeletal Medicine), PATIJN (2002) points out, that a sufficient training for the standardisation of the test method and its interpretation is missing in many studies.

A problem with more than one examiner is the possible therapeutic effect of the test method, which might influence the test results of the following examiners. Additionally, it is called for, that the inspection is performed as a well-defined blind test for the therapists versus the test persons, as well as for the examiners among each other.
The whole examination has to be performed in a standardised way and it has to be paid attention to a consistent judgement of the test results. PATIJN points out, that training is of crucial importance for the standardisation of reproducibility studies.

According to PATIJN (2002), fourty test persons are sufficient for a reproducibility study. For evaluation, the calculation of kappa values is the method of the choice. During the training phase, a preliminary study should be performed with ten participants, in order to screen the methodology and the test for deficiencies.

In the current study, a preliminary study with 20 test persons was performed.
Altogether, no studies with reliability tests of manual techniques for the lower limb could be found. Most of the studies refer to spinal segments.

Further authors, listed below have faced up to the reliability of manual inspections and techniques. In order not to break the mould of this study, they will only be invoked by name: ALLEY (1983), BOLINE (1993, 1998), CARMICHAEL (1987), KORAN (1975), LAWSON and CALDERON (1997), MC CONNELL (1980), MOOTZ ET AL. (1989), MORAN and GIBBONS (2001), PANZER (1992), RUSSEL (1983). The insights and results are in accord with the studies presented above or are covered by the reviews.

5.1. Summary

Recapitulating, the following points can be derived from the present scientific literature:

- The literature summarised above shows that manual diagnosis methods cover a high variability of reliability.
- In the literature it is distinguished between four different fundamental types of palpatory tests: Differentiation of tissue textures, evaluation of static landmark positional asymmetry and assessment of tenderness (DINNAR ET AL., 1980; KUCHERA and KAPPLER, 2002; SEFFINGER ET AL., 2002).
- Most of the studies dealing with reliability have been performed by chiropractors.
- It makes sense to perform a preliminary study, for checking the methodology for errors (HAWK, 1999; PATIJN, 2002).
- In reviews, predominately statistical and methodological deficiencies are criticised (ALLEY, 1983; RUSSEL.1983; HAAS.1991; HESTBAEK and LEBOEUF-YDE, 2000; SEFFINGER, 2004; STOCHKENDAHL ET AL., 2006).
- A preceding consensus training of the technique is discussed controversially (KELSO, 1981; McCONNELL Et AL., 1980, BEAL ET AL., 1982; JONSTON ET

AL. (1982); DEGENHARDT, 2005; SCHÖPS ET AL., 2000; PATIJN, 2002; SEFFINGER, 2004).

- The sample must enclose symptomatic as well as asymptomatic test persons (KEATING, 1989).
- Osteopaths and test persons must be blinded (HAAS, 1991; HAWK, 1999; PATIJN, 2002; STOCHKENDAHL, 2006): The osteopaths against the results of the other examiners, the medical findings of the subjects, and the test persons themselves, as well as the test persons against the medical findings.
- The anonymised results should be evaluated by means of Cohen`s Kappa and by an independent person (HAAS, 1991; PATIJN, 2002; STOCHKENDAHL, 2006).
- The practical experience of the examiners does not affect test results (MIOR, 1990; HESTBAEK and LEBOEUF-YDE, 2000).
- Testing under stress or in noisy environments has a negative impact on the reliability (LUNDIN, 1999).
- Low agreements can arise from changes of the medical findings by diagnostic techniques (FJELLNER, 1999; PATIJN, 2002).
- Pain provocation studies demonstrated acceptable reliability. Palpatory tests of landmark, positional asymmetry, movement asymmetry and muscle tension result in poor reliability. Regional range of motion is more reliable than segmental range of motion. Intraexaminer reliability is better than interexaminer reliability (SEFFINGER, 2004; HESTBAEK and LEBOEUF-YDE, 2000).

6. Methodology

The word “methodology” has Greek roots and means „teaching about the procedure” (WIKIPEDIA[4]). In this sense, the development and performance of the study will be presented in this chapter.

Twelve osteopaths performed tests for anteromedial dysfunctions of the talus in the talocrural joint according to Ebenegger/Tixa (2004) (cf. chapter 4) on 121 test persons.

The tests took place in four different locations in Switzerland (CH) and Germany (D) on different samples of test persons, in order to ensure a high variability of test persons with and without anteromedial dysfunctions. Additionally, by this splitting, osteopaths with different practical experience could be accessed (cf. Table 1).

Test	Osteopaths	Test persons	Date
Landquart (CH) (preliminary test)	2	20	2008-02-19
St. Gallen (CH)	3	43	2008-03-10
Ulm (D)	4	30	2008-03-16
Magglingen (CH)	3	28	2008-04-14

Table 1: Places, dates and numbers of osteopaths as well as test persons taking part in the study.

The first test (Landquart) was performed as a preliminary test, in order to check the procedure with regard to difficulties, which were considered in the main examinations. The other tests were performed with always the same procedure.

6.1. Preparation of the Examination

In December 2007, possible locations for the tests were selected and contact was established to the contact persons. Finally, four places were found, where the contact persons were ready to overtake the on-site organisation and to provide a location, where the tests could be performed.

The locations are the “School for Physiotherapy Thim van der Laan” in Landquart/Switzerland, the “Handball Club Bruggen-St.Gallen”/Switzerland, the “SKOM”

[4] Wikipedia [www.de.wikipedia.org/wiki/Methodologie] (last access: 22.08.2008): Methodologie (German): methodology.

in Dornstadt-Ulm/Germany and the “Physical Education College” in Magglingen/ Switzerland.
In the end of January 2008, the questionnaires and descriptions of the test were sent to the osteopaths, who had signaled readiness to take part in the study and to the contact persons, who forwarded them to the test persons.
All questionnaires were sent and - partially - returned by e-mail. Two weeks before the tests, approximately 60% of the questionnaires for the test persons had been returned. The residual ones were completed on the spot. The osteopaths received an additional written description of the test.

The preliminary study took place on February 19, 2008 in Landquart, Switzerland. Two osteopaths tested 20 students in the planned manner, in order to find out methodological deficiencies, to analyse the data and to prepare the main examinations. The whole testing procedure took approximately 1.5 hours.
The main examinations took place on March 10, 2008 in the gym “Boppartshof” in St.Gallen/ Switzerland, where three osteopaths tested 43 handball players (duration: approx. 1.5 hours), on March 16, 2008 in the school for osteopathy of the SKOM in Dornstadt-Ulm/Germany, where four osteopaths tested 30 test persons (duration: approximately one hour) and on April, 14, 2008 in the physical education college in Magglingen/Switzerland, where three osteopaths tested 28 students in sports physiotherapy (duration: approximately one hour).

6.2. Procedure of the Investigation

Osteopaths as well as test persons were invited in writing and obtained a written introduction to the procedure of the test. Additionally, the test persons were asked to fill a questionnaire concerning the history of dysfunctions of their feet (cf. Annex 1a). The osteopaths were asked about their osteopathic education, their practical experience and the knowledge and usage of the test by another questionnaire (cf. Annex 1b).

The examination rooms were completely equipped, when the osteopaths arrived 30 minutes before the examinations. The osteopaths became acquainted to the rooms and they received another introduction into the course of the study.

An assistant for the recordings was assigned to each single osteopath and missing data of the questionnaire were completed.

Subsequently, the test was demonstrated. During the examinations, the osteopaths and their assistants were fully isolated from the test persons. Moreover, the osteopaths did not communicate among themselves. The records, kept invisible for the examiners and test persons, were collected and kept safe by the director of the study after each test run.
The blinding of the examiners and test persons with regard to the findings of the other examiners (double blinding) was met during the whole examination.

After arrival of the test persons, they were thoroughly instructed in writing and verbally about the procedure.

Then random numbers were assigned to the test persons, which were unknown to the osteopaths.

Incomplete questionnaires of the subjects were completed. The test persons were asked to be punctually ready in gym shorts or underwear in the vestibule.
For a clear assignment of the results, the test persons carried a sheet with their ID in the left hand.

On the massage table, the test persons should hold the ID in the left hand beside the pelvis, so that the assistant could identify it.

In the first test run the massage tables were filled in numerical order (e.g. with four massage tables: 1-4, 5-8, 9-12, …).

The test persons laid down on the massage table and slipped below the sheets until their heels were placed on the foot of the massage table. The osteopaths could only see the lower limbs. The trunks and heads were not visible.

Each osteopath tested the talus of each foot under consideration of the other foot.

The assistant was told the result (within the defined possibilities), who noted it in the records (cf. Annex 1c/1d). None of the other participants could access these data. The possible answers were: Anteromedial dysfunction yes/no and as additional information: tense or soft connective tissue.

Then, the osteopath switched to the next test person. When all of the test persons had been tested, the examiners drew back and the director of the study invited the next test persons to take place on the particular massage tables. This procedure was repeated, until the first test run was completed.
The same procedure was repeated with a randomly changed order of the test persons.
In the second test run, the succession of test persons was drawn by lots and noted on a starting list. The test persons laid down in this order on the massage tables following the procedure above. After this test run, the examination was completed.

After a check, whether the submitted record sheets and questionnaires were completed, the osteopaths as well as test persons separately were asked for their feedback and impressions.

Afterwards they were collectively thanked and departed.
Data were entered in a readily prepared software mask by the director of the study and sent to the statistician by e-mail, who received the results only with the IDs of the osteopaths (O1, O2, …) and test persons (1, 2, ...). The evaluation of the inter-examiner reliability of the individual pairs of examiners on the individual test places was done by means of Cohen's kappa.

6.3. The Test Persons

6.3.1. Selection of the Test Persons

The test persons were provided by the particular institutions, after explicit agreement of the test persons.
Exclusion criteria were features, which would enable a recognition of the test persons or feet, as e.g. painted toenails, scars, disturbance of growth, alterations in the region of the toe nails, skin pigment disorders, dominant birth marks, piercing and tattoos.

Additionally, all test persons had to be able to take up supine position and had to wear underwear or gym shorts. They received the questionnaire online and had to return it the same way in order to guarantee beforehand, that symptomatic as well as asymptomatic subjects take place.

The test persons were recruited from different strata of the population (cf. Table 2).

Landquart	Physiotherapist students of the school for physiotherapy
St. Gallen	Handball players of the HC Bruggen (St. Gallen)
Ulm	Osteopathy students of the SKOM Ulm
Magglingen	Sport physiotherapist students of the spt-education

Table 2: Characteristics of the samples of test persons.

The test persons were informed about the course of the investigation and it was stressed that they would not be treated during the tests. They took part on their free will and did not receive any financial compensation for their participation.
The participants were blinded against their medical findings and neither had access to the results, nor were they informed about them.
Test persons for the preliminary test were 20 physiotherapist students (12 female, eight male) between 21 and 32 years. Only five of them have never knowingly had a lateral ankle sprain before. Three subjects stated, that they had problems with the foot at present.
In St. Gallen/Bruggen, 43 handball players (16 female, 27 male) between 14 and 37 years were tested. Only two of them declared, that they have never experienced a lateral ankle sprain before.
In Ulm/Dornstadt, 30 osteopathy students (19 female, 11 male) between 27 and 61 years were tested. Eight of them have never knowingly twisted an ankle, before.
In Magglingen, 28 students in sports physiotherapy (19 female, 9 male) between 24 and 40 years were tested. Eleven of them declared, that they have never had an inversion trauma of the ankle before.

6.3.2. Characteristics of the Test Persons

The characteristics of the test persons, collected by means of the questionnaire (cf. Annex 1b) are summarised in Fig. 4 – Fig. 7 for the individual samples (raw data cf.

Annex 2). In total, 18% of the test persons state to have actual dysfunctions of their feet, 79% have sprained an ankle in the past, 38% have had injuries to their legs and 28% have undergone a leg or foot operation.

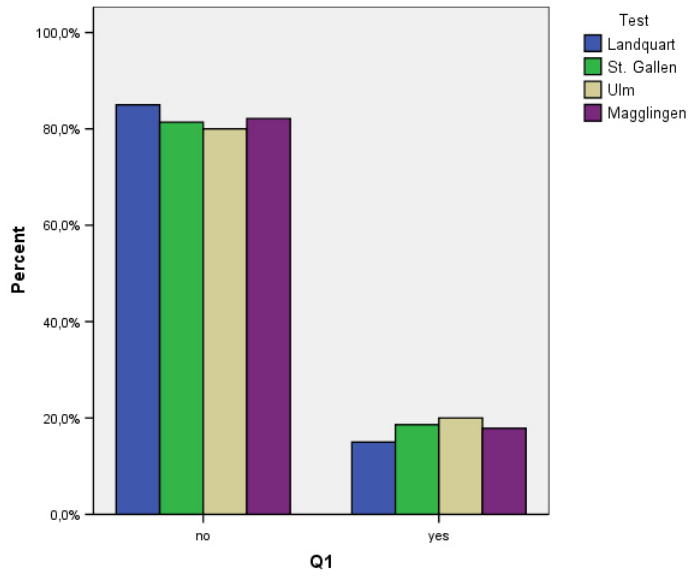

Fig. 4: Distribution of actual dysfunctions of the feet in the four samples.

According to Fig. 4, test persons with actual dysfunction of the feet are most frequent in the sample of Ulm (20%) and least frequent in Landquart (15%). Distribution of these persons does not differ significantly in the four places, where the tests were performed (Kruskal-Wallis test: χ^2= 0.21, p= 0.98).

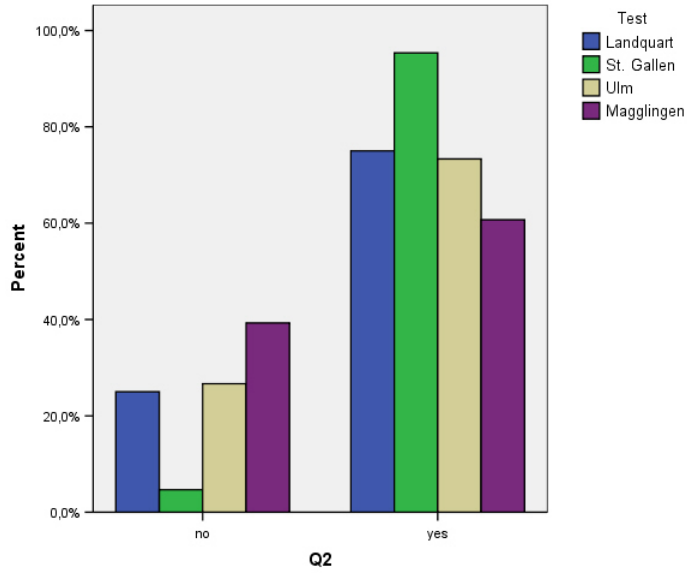

Fig. 5: Distribution of test persons with past ankle sprains in the four samples.

According to Fig. 5, test persons with past ankle sprains are most frequent in the sample of St. Gallen (95%) and least frequent in Magglingen (61%). Results of a Kruskal-Wallis test indicate the presence of significant differences in the distribution of test persons with sprains within the four groups (χ^2= 13.0, p= 0.005). Differences are significant between the samples of Magglingen and St. Gallen (Mann-Whitney test: p<0.001), St. Gallen and Ulm (p=0.008), as well as St. Gallen and Landquart (p= 0.02).

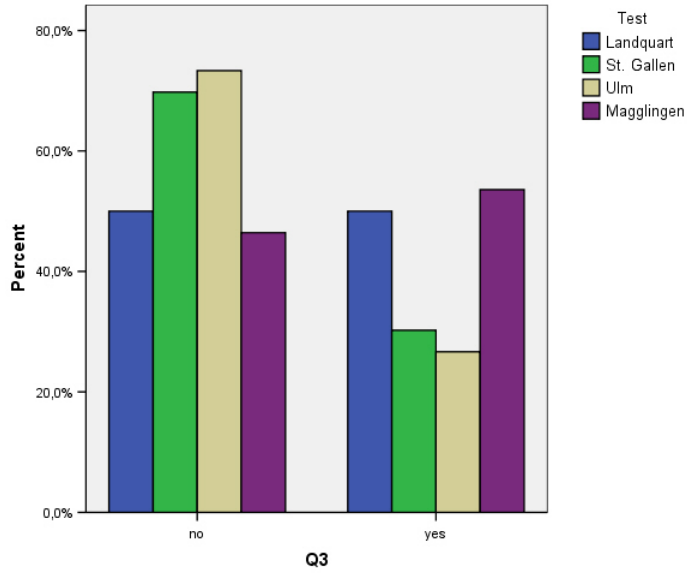

Fig. 6: Distribution of test persons with injuries to the legs in the four samples.

According to Fig. 6, test persons with injuries to their legs are most frequent in the sample of Magglingen (54%) and least frequent in Ulm (27%). Results of a Kruskal-Wallis test indicate the presence of distinct differences in the distribution of these persons within the four groups (χ^2= 6.78, p= 0.08). Differences are significant between the samples of Magglingen and Ulm (Mann-Whitney test: p=0.04) and Magglingen and St. Gallen (p=0.05). The other samples do not differ significantly.

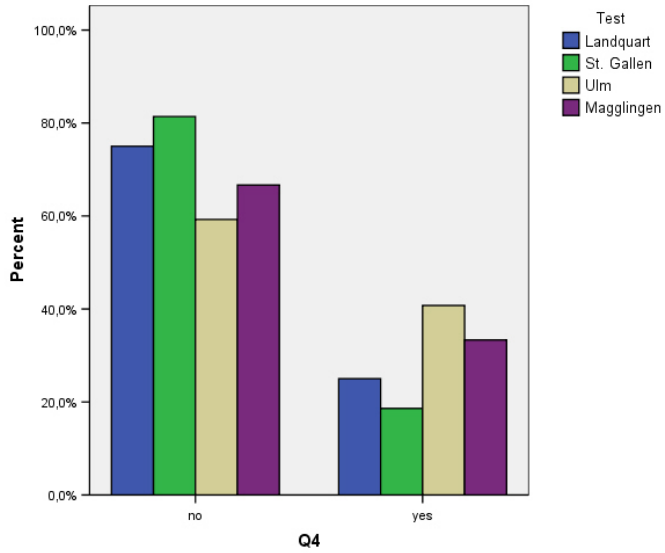

Fig. 7: Distribution of test persons with past leg or foot operations in the four samples.

According to Fig. 7, test persons with past leg or foot operations are most frequent in the sample of St. Gallen (41%) and least frequent in Ulm (19%). Nevertheless, there are no significant differences in the distribution of test persons leg or foot operations within the four groups (Kruskal-Wallis test: $\chi^2 = 4.47$, $p = 0.22$).

6.4. The Osteopaths

The osteopaths have been invited to the particular tests from the surrounding and at the SKOM from the current teachings in osteopathy, respectively. It has been taken care, that osteopaths with low practical experience (1-5 years) with high practical experience (>10 years) as well as examiners, who know and use the test for certain are considered. Ahead of the tests, each examiner has had to fill an online-questionnaire concerning examinations, the attended school for osteopathy, and knowledge and practical usage of the particular test. Additionally, they have bee instructed in the test in written form.
Prior to testing, the examiners received another oral instruction, an explanation of the procedure of the documentation and a demonstration of the test.
During the tests, the osteopaths are blinded with regard to the medical findings of the other examiners and they have never access to the records. The director of the study keeps the records safe. Additionally, the examiners cannot take insight into the

questionnaires of the test persons and thus do not know, whether they are symptomatic or asymptomatic.
The osteopaths have finalised their studies in different institutions. Two of them have had their training at the IWGS Paris, one at the IWGS Antwerp, four at the SKOM, two at the College Sutherland and two at the Swiss School for Osteopathy in Lausanne. Another examiner is OMT-Senior Instructor and has passed several manual therapeutic educations.
In Landquart, two male osteopaths have been invited as examiners, in St. Gallen one female and two male, in Ulm four male and in Magglingen two female and one male. No examiner took part twice.

There is a big difference in the practical experience of the osteopaths in St. Gallen and Ulm. The osteopaths in Ulm are lecturers at the SKOM with up to 25 years of practical experience in osteopathy. The two female osteopaths in Magglingen have been trained at the school in Lausanne, where Mr. Ebenegger and Mr. Tixa among others are teaching the presented talus-test (cf. chapter 4).
All testers have been informed in writing beforehand and again 30 minutes before the start of the first test run instructed in written form, verbally and practically. They took part voluntarily and without financial remuneration.
The characteristics of the therapists, collected by means of the questionnaire (cf. Annex 1a) in advance of the study are summarised in Table 3.

Osteopath	Age	Sex	Handedness	Q1: Last examination (year)	Q2: Practical experience (years)	Q3: Osteopathic training centre	Q4: Knowledge of the test	Q5: Regular use of the test
Landquart								
O1	36	m	right	2008	5	SKOM	No	No
O2	34	m	right	2006	2	SKOM	Yes	Yes
St. Gallen								
O3	32	w	right	2005	3	SKOM	No	No
O4	41	m	right	2004	5	C.Sutherl.	Yes	No
O5	39	m	left	2005	5	C.Sutherl.	No	No
Ulm-Dornstadt								
O6	47	m	right	1994	14	IWGS Antwerpen	No	No
O7	51	m	right	1988	20	IWGS Paris	No	No
O8	57	m	right	1985	23	IWGS Paris	No	No
O9	48	m	right	2005	4	SKOM	No	No
Magglingen								
O10	55	m	left-right	1983	27	Different schools	Yes	Yes
O11	31	w	right	2003	5	Lausanne	Yes	No
O12	29	w	right	2007	2	Lausanne	Yes	Yes

Table 3: Characteristics of the osteopaths.

On the basis of the answers of Table 3, the therapists can be aggregated in four different groups (cf. Table 4).

The experience of the osteopaths was assessed by the question whether the talus test was known, unknown (Q4), used in praxis or not (Q5). Additionally, the final year of the osteopathic education (Q1) and the years of osteopathic praxis (Q2) were relevant for this classification. The threshold for the final year is the mean value 1997.9 and for the years of osteopathic praxis the mean value of 10.8 years.

	Landquart		St. Gallen			Ulm				Magglingen		
Q1,Q2	1	2	3	4	5	6	7	8	9	10	11	12
Q4		+		+						+	+	+
Q5		+								+		+
Class.			**1**	**4**	**1**	**2**	**2**	**2**	**1**	**3**	**4**	**3**

Table 4: Classification of the therapists by their practical experience (blue cells…high experience, yellow ones…lower experience).

These four groups are formed by therapists with:

1 low practical experience and no usage of test

2 high practical experience and no usage of test

3 usage of test (different practical experience)

4 low practical experience but knowledge of test (no usage)

6.5. The Rooms

The rooms used during the study were provided by the particular institutions free of charge. In Ulm/Germany (cf. Fig. 11) and Landquart/Switzerland (cf. Fig. 8) they were normal classrooms, in St. Gallen/Switzerland (cf. Fig. 9) a gym and in the sport school Magglingen/Switzerland (cf. Fig. 12) a large corridor.

Fig. 8: Preliminary study in Landquart.

Fig. 9 and Fig. 10: Main study in Bruggen.

Fig. 11: Main study in Ulm

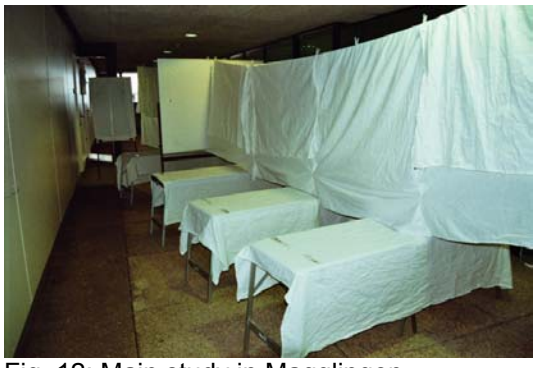

Fig. 12: Main study in Magglingen.

Per osteopath, one massage table was set up, plus an additional one in order to grant a fluent course of the experiments (cf. preliminary study). Sheets were attached to the cladding of the ceiling, forming a visual cover between osteopath and test person. The massage tables were positioned centrically below the sheets, so that during testing only

the lower limb was visible. “Left”- and “right”- marks were affixed on the sheets covering the massage tables, in order to prevent a possible uncertainty in the side of the body.

6.6. Preliminary Test

In order to uncover methodological deficiencies, a preliminary test was conducted with two osteopaths and 20 students in physiotherapy in the “Academy Thim van der Laan” in Landquart/CH on February 19, 2008.
After this preliminary test, the procedure was adapted with regard to the feedback of the therapists and test persons:

- The distance between the massage tables was increased; otherwise therapists could possibly hear others.
- Since body sides might have been confused during the preliminary study, they were marked on the therapy beds in the further test settings.
- Since testing was exhausting (therapists felt tired after testing 40 times), the therapists could have a break between the two test runs.
- Due to wrong or missing declarations in the questionnaires (e.g. the side of dysfunctions of the foot), the instructions for filling the questionnaires had to be refined.
- Test persons with small scars on the feet were not excluded any longer, because several test persons had such scars and thus no risk for recognition was assumed.
- With only one massage table per osteopath, there were delays, because the examiners had to wait until the test person became vacant. In order to avoid this idle time of the examiner, at least one additional massage table was added during the main examinations.

6.7. Evaluation of the Results

6.7.1. Description of the Statistic used

Cohen's Kappa (κ) was used for the evaluation of the inter-rater and intra-rater reliability. For the explanation of Cohen's Kappa (κ),a schematic 2x2 contingency table of the test results according to ROSNER (1986) and FEINSTEIN (s.a.) is shown in Table 5. The numbers of agreements (P, S) between the two examiners in the diagonal are marked dark blue, the number of differing results light blue (Q, R). The orange cells represent calculated values, where C_i are the column sums and Row_i the row sums.

Results		Examiner 1				Row totals (R)
		yes		**no**		
Examiner 2	**yes**	Number of agreements on presence	Q	Number of disagreements	S	Row_1
	no	Number of disagreements	R	Number of agreements on absence	T	Row_2
Column totals (C):			C_1		C_2	N

Table 5: 2x2 Contingency table (P, Q, R, S…numbers of findings).

The formulas for the calculation of kappa are:

$N = Q + R + S + T$ (total number of comparisons)

The probability of the *actual agreement* p_0 is calculated by:

$$P_0 = \frac{Q+T}{N}$$

The probability of agreement *expected on chance* p_e is calculated by the formula:

$$P_e = \left[\frac{Q+R}{N}\right]\left[\frac{Q+S}{N}\right] + \left[\frac{R+T}{N}\right]\left[\frac{S+T}{N}\right]$$

Kappa (κ) usually is expressed in the following standardized way:

$$\kappa = \frac{P_0 - P_e}{1 - P_e}$$

That means, Kappa is the ratio of the actual agreement beyond chance and the potential agreement beyond chance.

In WOODWORD, 1999 a clear idea of the kappa-index, as well as its context with the agreements is given (cf. Fig. 13):

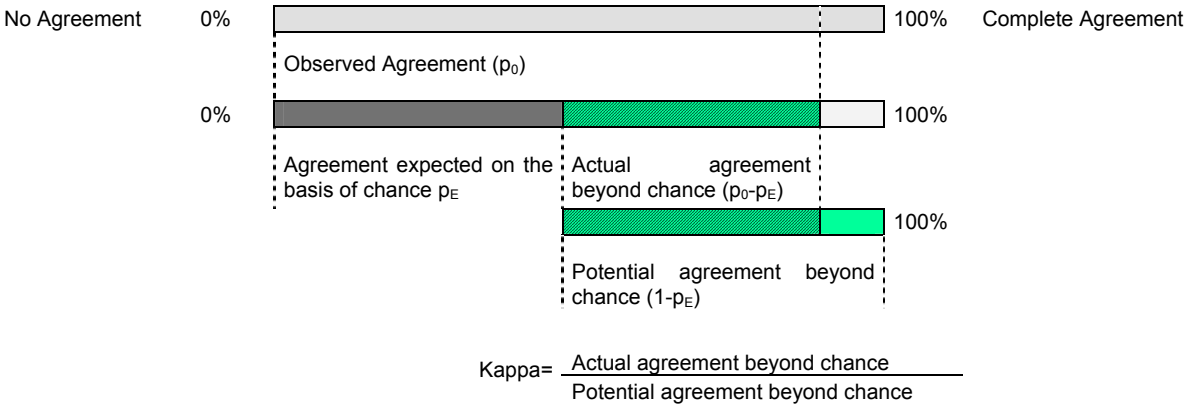

Fig. 13: Explanation of the meaning of the κ-index.

6.7.2. Statistical Evaluation

Data gained during the investigation were collected in Microsoft® Excel® 2000 spreadsheets and imported into a data bank for further classifications (Microsoft® Access® 2000). Cohen's Kappa (κ) was calculated by SPSS 14.0.0 for the agreements of all pairs of osteopaths (inter-examiner reliability) and for the agreements of all individual therapists in the two examinations (intra-examiner reliability). Calculations were performed under consideration of all test persons. Subsequently, in both cases, mean values, standard deviations[5], and the estimated 95% - confidence intervals[6] were calculated for different groups of therapists (cf. Table 6). For the calculation of these statistics, κ-indices with negative signs (κ <0) were taken as zero.

[5] The standard deviation is a measure of the dispersion in a distribution. It is equal to the square root of the arithmetic mean of the squares of the deviations from the arithmetic mean.

[6] 95%-confidence intervals for the mean are interval estimates for the mean, giving an indication of how much uncertainty there is in the estimate of the true mean. The narrower the interval, the more precise is the estimate. 95%-confidence intervals get narrower with increasing sample size and with decreasing sample standard deviation.

1	(Pairs of)* therapists examining the same sample of test persons (Landquart, St. Gallen, Ulm, Magglingen), individually.
2	All (pairs of)* osteopaths (independent from the different samples).
3	(Pairs of)* therapists with comparable practical experience

Table 6: Data bases for the computation of the κ-indices for reliability (*…For inter-examiner reliability, agreement of pairs of osteopaths and for intra-examiner reliability, agreement of single therapists in two tests is considered).

The aggregation of the therapists with similar practical experience (3) is described in Chapter 6.4.

The therapists forming the four groups are:

- Osteopaths with low practical experience and no usage of test: O3, O5, O9
- Osteopaths with low practical experience but knowledge of the test (no usage): O4, O11
- Osteopaths with high practical experience and no usage of test: O6, O7, O8
- Osteopaths who use the test (with different practical experience): O10, O12

The possible influence of the curricula of different schools has not been evaluated, because the test is unknown to most of the osteopaths.
Additionally, data about the tightness of the tissue were not considered in the evaluation, since they turned out to be too subjective and thus error prone.

The results of κ were interpreted with the degrees of agreement commonly used in the later literature (LANDIS AND KOCH, 1977, 159-174):

$\kappa < 0.20$	poor
$0.20 < \kappa < 0.40$	fair
$0.40 < \kappa < 0.60$	moderate
$0.60 < \kappa < 0.80$	substantial
$0.80 < \kappa < 1.00$	almost perfect

According to FJELLNER ET AL., 1999, 511-516 κ-indices higher than at least 0.4 are considered as indicator for an acceptable interobserver-reliability.

7. Results

In this chapter, results for inter- and intra-examiner reliability at first will be presented for the different samples of test persons and therapists, then summarised for all examinations together and finally organised by experience of the osteopaths.

7.1. Preliminary Test (Landquart)

During the preliminary test in Landquart, two therapists performed the tests for anteromedial dysfunctions on 20 test persons, recruited from physiotherapist students. Both therapists have short practical experience (final year after 2006) and one of them uses the test regularly, the other has not known this test before. In order to test the procedure, I took part as examiner, personally.

7.1.1. Inter-Examiner Reliability

The κ-indices for the inter-examiner reliability of this examination are summarised in Table 7.

Foot	left		right	
Examination	Ex1	Ex2	Ex1	Ex2
TH1 vs. TH2	0.03	0.17	0.35	*
Mean kappa	0.18			
SD	0.16			
95%-CI	-0.22 – 0.58			

Table 7: Individual κ-indices for the inter-examiner reliability for each foot and examination as well as mean values, standard deviations and 95%-confidence intervals of the means calculated under consideration all pairs of osteopaths. *… κ could not be calculated, because results of one examiner are constant.

The mean degree of agreement (both feet, both examinations) between the two examiners is “poor” (maximum “fair”).

Since several problems became evident in the discussion after the test, the procedure was adapted for the following investigations (cf. chapter 6.6).

7.1.2. Intra-Examiner Reliability

The κ-indices for the intra-examiner reliability of this examination are summarised in Table 8.

Foot	left	right
TH1	<0	0.21
TH2	0.38	*
Mean kappa	0.20	
SD	0.19	
95%-CI	-0.28 – 0.67	

Table 8: Individual κ-indices for the intra-examiner reliability for each foot as well as mean values, standard deviations and 95%-confidence intervals of the means calculated under consideration of all therapists. *… κ could not be calculated, because results of one examiner are constant.

As usual, mean intra-examiner reliability is higher than inter-examiner reliability. Nevertheless, reliability is still “poor” (maximum “fair”).

7.2. St. Gallen

During the test in St. Gallen, three therapists performed the tests for anteromedial dysfunctions on 43 test persons, recruited from handball players of the HC Bruggen.
Since handball players are at high risk of foot injuries, there should be a high number of positive testing. All of the therapists have finished their osteopathic studies after 2004 and do not use the test in daily praxis.

7.2.1. Inter-Examiner Reliability

The κ-indices for the inter-examiner reliability of this examination are summarised in Table 9.

Foot	left		right	
Examination	Ex1	Ex2	Ex1	Ex2
TH3 vs. TH4	0.04	0.08	0.28	<0
TH3 vs. TH5	<0	<0	<0	<0
TH4 vs. TH5	0.26	0.16	0.45	0.14
Mean kappa	0.12			
SD	0.15			
95%-CI	0.02 – 0.21			

Table 9: Individual κ-indices for the inter-examiner reliability for each foot and examination as well as mean values, standard deviations and 95%-confidence intervals of the means calculated under consideration of all pairs of therapists.

In spite of the changes in the procedure, mean inter-examiner reliability (both feet, both examinations) is still “poor” (maximum: “moderate”). During the first examination, mean κ= 0.19, during the second one κ= 0.06. Agreement is higher on the right foot, than on the left one (κ= 0.15 and κ=0.09, respectively).

7.2.2. Intra-Examiner Reliability

The κ-indices for the intra-examiner reliability of this examination are summarised in Table 10.

Foot	left	right
TH3	0.19	<0
TH4	<0	<0
TH5	<0	0.45
Mean kappa	0.11	
SD	0.18	
95%-CI	-0.09 – 0.30	

Table 10: Individual κ-indices for the intra-examiner reliability for each foot as well as mean values, standard deviations and 95%-confidence intervals of the means calculated under consideration of all therapists.

In this case, intra-examiner reliability is even worse than inter-examiner agreement and distinctly worse than intra-examiner reliability in Landquart.
Agreement is higher on the right foot, than on the left one (κ= 0.15 and κ= 0.06, respectively)

7.3. Ulm

During the test in Ulm, four therapists performed the tests for anteromedial dysfunctions on 30 test persons, recruited from osteopathy students of the SKOM Ulm. Three of the therapists have finished their osteopathic studies before 1994. The other therapist (TH9) has finished in 2005. None of the osteopaths uses the test in daily praxis.

7.3.1. Inter-Examiner Reliability

The κ-indices for the inter-examiner reliability of this examination are summarised in Table 11.

Foot	left		right	
Examination	Ex1	Ex2	Ex1	Ex2
TH6 vs. TH7	<0	0.17	<0	0.38
TH6 vs. TH8	0.20	<0	0.30	<0
TH6 vs. TH9	<0	<0	0.51	0.22
TH7 vs. TH8	<0	0.13	<0	<0
TH7 vs. TH9	0.36	<0	0.00	0.31
TH8 vs. TH9	<0	0.11	0.36	0.29
Mean kappa	0.14			
SD	0.16			
95%-CI	0.07 – 0.22			

Table 11: Individual κ-indices for the inter-examiner reliability for each foot and examination as well as mean values, standard deviations and 95%-confidence intervals of the means calculated under consideration of all pairs of therapists.

The results are almost the same as in St. Gallen (on average “poor”, maximum “fair”).
During the first examination, mean κ= 0.14, during the second one κ= 0.13. Agreement is higher on the right foot, than on the left one (κ= 0.20 and κ=0.08, respectively)

7.3.2. Intra-Examiner Reliability

The κ-indices for the intra-examiner reliability of this examination are summarised in Table 12.

Foot	left	right
TH6	<0	0.38
TH7	0.50	<0
TH8	0.22	<0
TH9	<0	0.38
Mean kappa	0.19	
SD	0.21	
95%-CI	0.01 – 0.36	

Table 12: Individual κ-indices for the intra-examiner reliability for each foot as well as mean values, standard deviations and 95%-confidence intervals of the means calculated under consideration of all therapists.

In spite of the changes in the procedure after the preliminary test, intra-examiner reliability is slightly lower than in Landquart. On average, κ-indices are “fair” (maximum “moderate”). In contrary to the data of St. Gallen, intra-examiner reliability is higher than inter-examiner reliability. Agreement is slightly higher on the right foot, than on the left one (κ= 0.19 and κ= 0.18, respectively).

7.4. Magglingen

During the test in Magglingen, three therapists performed the tests for anteromedial dysfunctions on 28 test persons, recruited from spt-education.
All therapists know the test and two use it regularly in daily praxis. One of them has finished his studies in 1983, the other two osteopaths after 2003.

7.4.1. Inter-Examiner Reliability

The κ-indices for the inter-examiner reliability of this examination are summarised in Table 13.

Foot	left		right	
Examination	Ex1	Ex2	Ex1	Ex2
TH10 vs. TH11	<0	<0	0.28	0.13
TH10 vs. TH12	0.52	<0	0.21	<0
TH11 vs. TH12	0.09	0.47	0.07	<0
Mean kappa	0.15			
SD	0.19			
95%-CI	0.03 – 0.27			

Table 13: Individual κ-indices for the inter-examiner reliability for each foot and examination as well as mean values, standard deviations and 95%-confidence intervals of the means calculated under consideration of all pairs of therapists.

Also in Magglingen, the results are almost the same as in the other examinations with the other samples (on average “poor”, maximum “fair”).

During the first examination, mean κ= 0.20, during the second one κ= 0.10. In this case, agreement is lower on the right foot, than on the left one (κ= 0.12 and κ=0.18, respectively)

7.4.2. Intra-Examiner Reliability

The κ-indices for the intra-examiner reliability of this examination are summarised in Table 14.

Foot	left	right
TH10	0.13	<0
TH11	0.63	0.60
TH12	0.24	0.11
Mean kappa	0.29	
SD	0.27	
95%-CI	0.00 – 0.57	

Table 14: Individual κ-indices for the intra-examiner reliability for each foot as well as mean values, standard deviations and 95%-confidence intervals of the means calculated under consideration of all therapists.

Intra-examiner reliability is distinctly higher than in the other investigations and than inter-examiner reliability (on average: “fair”, maximum “substantial”). Agreement is lower on the right foot, than on the left one (κ= 0.24 and κ= 0.33, respectively).

7.5. Reliability under Consideration of all Examinations

7.5.1. Inter-Examiner Reliability

In Table 15, the descriptive statistics of the κ-indices for the inter-examiner reliability, calculated with the results of all tests of the therapists TH3-TH12 (both feet, both examinations) can be observed.

Mean kappa	0.14
SD	0.16
95%-CI	0.09 – 0.19

Table 15: Mean values, standard deviations and 95%-confidence intervals of the means of the κ-indices for inter-examiner reliability calculated under consideration of all pairs of therapists (TH3-TH12).

Since the results gained with the different samples of test persons and therapists do not vary distinctly, the grand mean of kappa is similar to the results of the individual experiments. On average, the κ-index is “poor”.

7.5.2. Intra-Examiner Reliability

In Table 16, the descriptive statistics of the κ-indices for the intra-examiner reliability, calculated with the results of all tests of the therapists TH3-TH12 (both feet) can be observed.

Mean kappa	0.19
SD	0.21
95%-CI	0.10 – 0.28

Table 16: Mean values, standard deviations and 95%-confidence intervals of the means of the κ-indices for intra-examiner reliability calculated under consideration of all therapists (TH3-TH12).

Again, reliability is only “poor”.

7.6. Influence of the Experience of the Osteopaths

Experiments were performed with different samples of test persons and only a low number of therapists. Therefore, the results classified by different experience of the

therapists, presented in the next chapters may only be interpreted as estimates and must not be generalised.

7.6.1. Inter-Examiner Reliability

By classification of the therapists by different experience, three rather homogeneous groups can be formed. However, it was not possible to form adequate pairs of examiners with the therapists TH4, TH9 and TH11.

The κ-indices of the relevant comparisons as well as group means, standard deviations and 95%-confidence intervals are presented in Table 17.

Experience	Examiners	L_ex1	L_ex2	R_ex1	R_ex2	mean	SD	95%-CI
Low experience/ No usage or knowledge	TH3 vs. TH5	<0	<0	<0	<0	0.00	0.00	-
High experience/ No usage or knowledge	TH6 vs. TH7	<0	0.17	<0	0.38	0.10	0.14	0.01 – 0.18
	TH6 vs. TH8	0.20	<0	0.30	<0			
	TH7 vs. TH8	<0	0.13	<0	<0			
Mixed experience /usage	TH10 vs. TH12	0.52	<0	0.21	<0	0.18	0.25	-0.21 – 0.57

Table 17: Individual κ-indices for the inter-examiner reliability for each foot and examination. Mean values, standard deviations and 95%-confidence intervals of the means are calculated for all test persons under consideration of the results of all pairs of therapists with similar practical experience (TH2-TH12).

The therapists with low practical experience who additionally do not know the test do not agree in their findings beyond chance. The highest agreement can be observed among the therapists, who use the test regularly in praxis. Nevertheless, mean reliability can be interpreted as “poor”, only and $\kappa<0$ can be observed in the results of half of the test runs. The therapists with longer practical experience, but who do not use the test are halfway between. It has to be stressed, that in two cases kappa was calculated with the results of two therapists, only. Generally, in case of results of $\kappa<0$, these can be observed on both feet, either during the first or second examination.

7.6.2. Intra-Examiner Reliability

For assessment of the intra-examiner reliability it was possible to form four groups of osteopaths with different practical experience. The κ-indices of the comparisons

between the findings in the first and second examination as well as group means, standard deviations and 95%-confidence intervals are presented in Table 18.

Experience	Osteopath	Left	Right	Mean	SD	l	u
Low experience/no knowledge or usage	TH3	0.19	<0	0.17	0.20	-0.05	0.38
	TH5	<0	0.45				
	TH9	<0	0.38				
Low experience/ knowledge of test	TH4	<0	<0	0.31	0.36	-0.26	0.87
	TH 11	0.63	0.60				
High experience/no knowledge or usage of test	TH6	<0	0.38	0.18	0.22	-0.05	0.41
	TH7	0.50	<0				
	TH8	0.22	<0				
Mixed experience/usage of test	TH10	0.13	<0	0.12	0.10	-0.04	0.28
	TH12	0.24	0.11				

Table 18: Individual κ-indices for the intra-examiner reliability for each foot. Mean values, standard deviations and 95%-confidence intervals of the means are calculated for all test persons under consideration of the results of all therapists with similar practical experience (TH2-TH12).

There are only two therapists (TH11 and TH12), who agree relatively well in their findings during the two examinations on both feet. In all other cases, there is at least one foot, where no agreement beyond the probability of chance can be observed.
This fact on its own indicates, that intra-examiner reliability of this test is low and depends on the individual therapists. These results also show that no conclusions may be drawn from the mean κ-indices, which group of therapists achieves the highest reliability: For example, in the group of osteopaths with the highest mean κ-index, therapist TH11 has excellent agreements, whereas TH4 fails on both feet.

7.7. Summary of the Results

7.7.1. Inter-Examiner Reliability

Inter-rater reliability is poor, independent from the sample of test persons (cf. Table 19), as can be read from the mean values and 95%-confidence intervals of the different test locations.

	(Landquart)	St. Gallen	Ulm	Magglingen	Total
All test persons					
Mean	0.18	0.12	0.14	0.15	0.14
SD	0.16	0.15	0.16	0.19	0.16
95%CI	-0.22 – 0.58	0.02 – 0.21	0.07 – 0.22	0.03 – 0.27	0.09 – 0.19

Table 19: Summary of the results (inter-examiner reliability).

The mean κ-index for all pairs of therapists (TH3-Th12) is 0.14 (95%-confidence interval: 0.09 - 0.19).
In contrary, high standard deviations of the individual tests indicate, that agreement between different pairs of osteopaths is individually different.

Differences between the two examinations

In Table 20 mean κ-indices are displayed for the individual examinations. During the second examination, lower agreement can be observed than during the first one.

Test	κ 1st examination	κ 2nd examination	ratio 2nd/1st examination
St. Gallen	0.19	0.06	0.32
Ulm	0.14	0.13	0.93
Magglingen	0.20	0.10	0.50

Table 20: Mean κ-indices of the two examinations (both feet).

Differences between the two feet

In Table 21, mean κ-indices are displayed for the two feet, individually. In Magglingen better agreement is higher on the left foot, in the other places on the right foot.

Test	Examination	κ Left foot	κ Right foot	Ratio left/right
St. Gallen	1	0.09	0.15	0.6
	2	0.06	0.15	0.4
Ulm	1	0.08	0.20	0.4
	2	0.18	0.19	0.9
Magglingen	1	0.18	0.12	1.5
	2	0.33	0.24	1.4

Table 21: Mean κ-indices for the two feet, individually (both examinations).

Experience of the Osteopaths

There are hints, that therapists who use the test regularly in praxis agree in their findings more often. Nevertheless, mean reliability can be only interpreted as “poor” (mean κ = 0.18). Therapists with low practical experience, who additionally do not know the test, do not agree in their findings beyond chance (κ=0). The therapists with longer practical experience, but who do not use the test are halfway between (κ=0.10). κ<0 can be observed in the results of at least half of the test runs in each group of therapists with different practical experience. These results can be observed on each of the feet, either during the first or second examination. Improving experience with the test and tiring of other therapists, respectively, might cause these patterns.

7.7.2. Intra-Examiner Reliability

On average, intra-rater reliability is slightly higher than inter-rater reliability, but still poor. The mean κ-index for all therapists (TH3-TH12) is 0.19 (95%-confidence interval: 0.10 - 0.28).

	(Landquart)	St. Gallen	Ulm	Magglingen	Total
All test persons					
Mean	0.20	0.11	0.19	0.29	0.19
SD	0.19	0.18	0.21	0.27	0.21
95%CI	-0.28 – 0.67	-0.09 – 0.30	0.01 – 0.36	0.00 – 0.57	0.10 – 0.28

Table 22: Summary of the results (intra-examiner reliability).

In particular, results of Magglingen and St. Gallen are conspicuous. In St. Gallen, intra-examiner reliability is slightly lower than inter-examiner reliability and in Magglingen a mean κ exceeding the upper bond of the 95%-confidence interval of the total sample can be observed.

Only two of the ten therapists (TH11 and TH12), agree relatively well in their findings during the two examinations on each of the two feet. In the other cases, there is at least one foot, where no agreement beyond the probability of chance can be observed.

Differences between the two feet

In Table 23, mean κ-indices are displayed for the two feet, individually. Similar to the results for the inter-examiner reliability, in Magglingen, agreement is higher on the left foot, in the other places on the right foot.

Test	κ Left foot	κ Right foot	Ratio left/right
St. Gallen	0.06	0.15	0.4
Ulm	0.18	0.19	0.9
Magglingen	0.33	0.24	1.4

Table 23: Mean κ-indices for the two feet, individually.

Experience of the Osteopaths

Obviously, the experience of the therapists does not play a major part for agreement on the findings of the first and second examination.

For example, in the group of osteopaths with the highest mean κ-index (osteopaths with low practical experience, who know but do not use the test, mean $\kappa = 0.31$), therapist TH11 has excellent agreements (left: $\kappa = 0.63$, right: $\kappa = 0.60$), whereas TH4 fails on both feet. Paradoxly, the lowest agreement can be observed in the group of osteopaths, who use this test regularly ($\kappa = 0.12$). Since groups of therapists with different experience comprise only few therapists, results are only estimates and must no be generalised.

7.8. Results of the Exchange of Experiences after the Tests

At the end of each examination, the osteopaths, test persons and secretaries discussed their experiences. In the following, I want to summarise their feedback.

Already during the preliminary test in Landquart/CH, both examiners noted, that performance and result of the tests depend on individual conditions as e.g. a) foot size, b) tension of the dorsal leg muscles, c) tissue tonus of the patient as well as d) hand size, e) hand force, f) body mechanics and g) experience in sensing and interpretation of the osteopaths.

Feedback of the test persons during the preliminary study touched changing movement amplitudes and different exertion of force of the examiners. By contact of the foot with the sternum (cf. chapter 4), influences by the points a-c (test persons) and d-f (examiners) could be minimised.

Feedback of the test persons after the main examinations aimed rather on the variance of the time taken for the tests in the two test runs and a more secure and routinised impression during the second test run. Some subjects described, that examiners applied different force during testing, so that the end feel was different. In one case, that much force was applied, that the movement was sensible up to the head.
The examiners in St. Gallen/Bruggen did not raise objections or suggestions to the course of the study. After enquire, they did not have the impression that their concentration has lessened. They noticed a more customary proceeding at the end of the study, only.
The examiners in Ulm/Dornstadt gave a very differentiated and experienced feedback. They recommended, that a subdivision in block sensation and restriction of movement would have been easier to assess, because some medical findings had been difficult to classify.

Another interesting fact is, that they realised a change in the tests “from testing to routine”. All of them had the impression that they became more secure. Going hand in hand, also the time needed for the tests became shorter. They suggested, that a specified time frame would be helpful, in order to be able to better empathise with the tissue. This would be “more osteopathic” than testing only. None of them was able to make use of the additional information about tenseness or softness of the connective tissue.

The alumni of the school for osteopathy in Lausanne, who tested in Magglingen confirmed, that the test was taught as used in this study, and that they use it in the same way.

Without exception, all examiners stated that this test in combination with anamnesis data is very suitable for the determination of anteromedial dysfunctions, and that they felt very certain.

8. Discussion

8.1. Discussion of the Results

Inter- as well as intra-examiner reliability of the test is low. The according mean κ-indices are κ= 0.14 (95%-confidence interval: 0.09 - 0.19) and κ= 0.19 (95%-confidence interval: 0.10 - 0.28).
Inter-examiner reliability does not differ distinctly among the different samples of test persons. Therefore, different sample characteristics (number of former ankle sprains or injuries to the legs or feet) obviously have no influence on the results.

During all tests, higher inter-examiner κ-indices can be observed during the first **test run**. Maybe this is caused by a lack of concentration or by tiring of some therapists during the second examination.

Inter- and intra-examiner agreement is higher for the results of the left **foot** in Magglingen, whereas in the other places it is higher for the right foot.

The influence of the **practical experience** of the therapists can only be estimated by the data gained by the experiments, since groups with different experience comprise only a low number of therapists.
Pairs of therapists who use the test regularly in praxis agree most often in their findings. Nevertheless, mean reliability can be interpreted as “poor”, only (mean κ= 0.18; both feet, both examinations). Therapists with low practical experience, who additionally have not known know the test, do not agree in their findings beyond chance (κ=0). The therapists with longer practical experience, but who do not use the test are halfway between (κ=0.10).

According to the results, there is no major influence of the experience of the therapists on intra-examiner agreement on the findings of the first and second examination. However, it has to be taken into account, that only few therapists could be allotted to the different experience groups and results might not represent the population of therapists.

A possible influence of the **handedness of the therapists** on the test results was discussed after the preliminary study. However, since only one of the osteopaths taking part in the study was lefthander, it could not be assessed.

Summing up the results, a universal reliability of the test could not be proved by this investigation. Under test conditions, even experienced therapists do not achieve an agreement, which can objectively be considered as acceptable for diagnosis.

8.2. Discussion of the Method

Questionnaire data

Questionnaire data concerning the characteristics of the test persons have been filled incompletely. Especially, the time of the occurrence of the dysfunctions, the body side of the affected leg of foot and of operations are frequently missing.
A question about the handedness was missing in the questionnaires for the therapists. Therefore these data were requested at the end of the study.

Blinding

Osteopaths and test persons were totally separated by a visual cover, so that no contact could take place. The results were told the personal assistant only and the other examiners had no access to them. No examiner had access to the questionnaire data, either. During the tests, between-examiner contact was not allowed.

The results were made anonymous and sent to the statistical consultant by e-mail in a pre-arranged data mask.

Exclusion of ambiguity errors

Since body sides might have been confused during the preliminary study, they were marked on the therapy beds during the further examinations.

Exterior influences on the results

Since testing was exhausting (therapists felt tired after testing 40 times in the preliminary study), there was a break after the first test runs of the further examinations. Nevertheless, agreement is higher during the first examination than in the second one.

In the discussion after the preliminary study, the therapists stated, that testing large feet and muscular patients needs more effort; the first due to leverage and the second due to high tension in the dorsal chain, which restrains the dorsal extension.

Tests on different samples of test persons

Due to different characteristics of the test persons in the different places, it is difficult to compare the results of therapists with similar practical experience. Even though, the results of inter-examiner reliability are homogeneous within the different tests, a final conclusion about influences of experience cannot be drawn, because samples might have been differently difficult to test.

Standardisation of the judgement of the findings

During the examination in Ulm, it became notable, that a more precise instruction concerning the judgement of the findings and classification would have been suggestive.

In advance of further examinations, the osteopaths should be trained in the medical assessment and in the documentation of the results in a standardised way.

Number of therapists

Without consideration of the preliminary test, total number of therapists performing the test is ten. Even though practical experience of the therapists covers a wide range, a generalisation of the results may only be done with caution. This accounts even more for the results for groups with different experience, which comprise even fewer therapists.

Recording of the therapeutic skill

The osteopaths specified the osteopathic skill very differently in the questionnaires. One osteopath who finished in 2008 declared to work osteopathically already for five years. Therefore, the final year of osteopathic training was used as the basis for the practical osteopathic experience in order to gain comparable results.

Accustoming problems of the therapists

Especially therapists, who were not used to this test but also some of those, who are regularly using the test, brought up, that they needed 5-10 tests to “get a feeling” for the test and the feet. Since these experiments should simulate everyday praxis of the therapists, no training phase in advance of the tests was designed. In the end, the test results indicate, that regular use of the tests does not necessarily improve the reliability of the test.

Realisation of the test

The test was performed as described in chapter 4 in extension position of the knee. Admittedly, this posture enhances the tension via the dorsal musculature and may lead to misinterpretations.

Additional information of the tension in the connective tissue

Apart from the test, the examiners were asked to record the general tension in the connective tissue in the sense of “soft” and “tense”. In the end, these data were impractical, because the idea of the term “tension in connective tissue” was not clear for all osteopaths. This fact became obvious not before the end of the study. The tension in the connective tissue was estimated very inhomogeneously.

Diagnosis

In the study of HAWK (1999) the examiners were asked, whether they would treat a structure or not („Would you adjust this vertebra - yes or no?”). I think, that this approach eventually is more interesting and expressive for diagnosis than the question “Is there a dysfunction - yes or no?”. Dysfunctions have a broad range and for the examiners it might be easier to decide whether treatment is necessary or not. The osteopaths testing in Ulm gave a similar feedback.

9. Synopsis

It was the goal of the study to find out, whether the talus test is universally reliable for osteopathic diagnosis or only relevant for a subjective diagnosis for each individual osteopath.

Hence, twelve osteopaths were asked to test 121 symptomatic and asymptomatic subjects by means of the test according to Ebenegger/Tixa (2004) in four different institutions. Osteopaths and test persons were blinded and the results were evaluated by means of Cohen’s kappa by a blinded statistical consultant.

Anamnestic data of the test persons (actual complaints of the feet, operations/accidents, known inversion traumata of the ankle) as well as data of the osteopaths (school-leaving year, school, renownedness of the test) were surveyed by means of questionnaires. The intra- and interrater reliability was investigated in two test runs.

As a result of the evaluation of the data, it became certain, that a universal reliability of the described talus test could not be proved by this actual investigation.

In the actual study, as maximum kappa indices, κ = 0.14, “poor“ is reached for inter-examiner reliability, and κ = 0.19, “poor“ for intra-examiner reliability.
Further results of this study are,

- that no causal relationship between the investigated anamnestic data of the test persons and the test findings can be revealed,
- that the results of the first test run are more reliable than the outcomes of the second one and
- that the practical experience of the examiners has no influence on the results.

However, it has to be noted, that the test has its justification in osteopathy, and that it is administered despite the bad reliability in manual diagnostics.
Inter- and intraexaminer reliability can be regarded as an essential problem for manual methods of assessment and treatment. Studies on interexaminer reliability for manual assessment techniques within the musculoskeletal system show moderate to poor agreement, especially for passive mobility testing, which represents an essential part of the osteopathic assessment.

How could this study be further improved and what should be modified in further studies?
The osteopaths would have to be trained in concerns of recording and judgement of the findings, in order to minimise divergences in these respects.
The problem would have to be specified with regard to the definiteness of the classification (“Is there a dysfunction - yes or no?”), or alternatively, the sense of the question would have to be changed in “Would you administer treatment yes or no?”, what might better comply with osteopathic philosophy.

However, finally and personally, I do not believe, that reliability is low due to the design of the study or the chosen procedure, but that osteopathy and manual treatment and diagnostics, respectively, represent a system, which is only individually accessible for the therapists, and which is affecting the patients individually, too. Actually, this study confirms, what other scientists have verified before, i.e. that these individual approaches cannot be compared in this kind of problem.

10. Bibliography

Alley, J.R. (1983): The clinical value of motion palpation as a diagnostic tool: a review. J. Can. Chiro Assoc 27: 97-100.

Anandacoomarasamy, A. and Barnsley, L. (2006): Long term outcome of ankle inversion injuries. British Journal of Sports Medicine, 39: 3, e14.

Beal, M.C. (1951): The manipulative treatment of mechanical disorders of the foot. AAO Yearbook 1991, 98-111.

Beal, M.C. and Patriquin, D.A. (1995): Interexaminer agreement on palpatory diagnosis and patients self-assessment of disability: A pilot study. Journal of the American Osteopathic Association 95 (2): 97-100, 103-6.

Beyer, L. (2007): Reproduzierbarkeit manueller Untersuchungen, Manuelle Medizin, 45: 426.

Boline P.D.; Keating J.C.; Brist, J. and Denver, G. (1988): Interexaminer reliability of palpatory evaluations of the lumbar spine. Am. J. Chiro Med 1 (1): 5-11.

Boline, P.D.; Haas, M.; Meyer, J.J.; Kassak, K.; Nelson, C. and Keating, J.C. Jr. (1993): Interexaminer reliability of eight evaluative dimensions of lumbar segmental abnormality, Part II. J. Manipulative Physiol. Ther. 16 (6): 363-74.

Bozic, R. and Weiser, J. (1991): Epidemiologische Daten zur Aussenbandruptur des oberen Sprunggelenkes. Traumatologie 21:118-120.

Braun, B.L. (1999): Effects of ankle sprain in a general clinic population 6 to 18 month after medical evaluation. Archives of Family Medicine 8: 143-148.

Burgess, P.R. and Perl, E.R.(1973): Cutaneous Mechanoreceptors and Nociceptors. In: Iggo, A. (Ed): Handbook of Sensory Physiology, Vol II. New York: Springer: 30.

Carmichael, J.P. (1987): Inter- and intraexaminer reliability of palpation for sacroiliac joint dysfunction. J. Man. Physiol Ther 10: 164-171.

Casillas, M.M. (2002): Ligament injuries of the foot and ankle in the athlete. In: DeLee, J.C. and Drez, D. Jr. (Ed.): DeLee & Drez`s orthopaedic sports medicine; principles and practice. 2. Ed., Philadelphia: Saunders.

Clanton, T.O. (1999): Athletic injuries to the soft tissues of the foot and ankle. In: Coughlin, M.J., Mann, R.A. (Ed.): Surgery of the foot and ankle, 7.Ed. St.Louis: Mosby.

Clanton, T.O. and Wood, R.M. (2002): Etiology of injury to the foot and ankle. In: DeLee, J.C. and Drez, D. Jr. (Ed.): DeLee & Drez`s orthopaedic sports medicine; principles and practice. 2. Ed., Philadelphia: Saunders: 2224-2247.

Cloet, E., Ranson, G. and Schallier, F. (1999): Praxis der Osteopathie, Stuttgart: Hippokrates

Cloet, E. and Muts, R. (2005): Wie wichtig ist die visuelle Wahrnehmung in der Osteopathie. Contra - Standpunkt. DO (2) 28 – 29: 63.

Colville, M.R. (1995): Reconstruction of the lateral ankle ligaments. Instr. Course Lect 44, 341-348.

Coysten, T. (2008): unpublished written note; WSO-Masterkurs, Maidstone.

Croibier, A. (2006): Diagnostik in der Osteopathie, München: Urban & Fischer.

Degenhardt, B.F. et al. (2005): Interobserver reliability of osteopathic palpatory diagnostic tests of the lumbar spine: improvements from consensus training. Journal of the American Osteopathic Association 105 (10): 465-473.

Denegar, C.R., Hertel, F. and Fonseca, J. (2002): The effect of lateral ankle sprain on dorsiflexion range of motion, posterior talar glide, and joint-laxity. J Orthop Sports Phys Ther. Apr 32 (4): 166-73.

Dinnar, U. et al. (1980): Classsification of diagnostic tests used with osteopathic manipulation. J. Am. Osteopath. Assoc. 81: 314-321.

Dinnar, U. et al. (1982): Description of fifty diagnostic tests used with osteopathic manipulation. J. Am. Osteopath. Assoc. 79: 451-455.

Dornieden, R. (2005): Das Supinationstrauma aus sportphysiotherapeutischer und manualtherapeutischer Sicht. Bachelor thesis in sports sciences, University of Konstanz.

Ebraheim, N.A. et al. (2006): Anatomical evaluation and clinical importance of the tibiofibular syndesmosis ligaments. Surgical and Radiology Anatomy 28: 142-149.

Echaute, C., Vaes, P., Duquet, W. and Gheluwe, van B. (2007): Test-retest reliability of sudden ankle inversion measurements in subjects with healthy ankle joints, J Athl. Train. Jan-Mar, 42(1):60-5.

Feinstein, A.R.: Clinical epidemiology: the architecture of clinical research. Philadelphia: WB Saunders: 184-6.

Fischer, R.A. and Yates, F. (1982): Statistical Tables for Biological, Agricultural and Medical Research 6th ed. Harlow: Longman, 146.

Fjellner, A.; Bexander, C.; Faleji, R. and Strender, L.E. (1999): Interexaminer reliability in physical examination of the cervical spine. J. Manipulative Physiol Ther. 22: 511-516.

Fleiss, J.L. (1981): Statistical Methods for Rates and Proportions, 2nd ed. New York: Wiley, 321.

Haas, M. (1991): Statistical methodology for reliability studies. J. Man Physiol Ther 14:119-132.

Haas, M. (1991): The reliability of reliability. J. Man Physiol Ther 14: 199-208.

Hawk, C.; Phongphua, C.; Bleecker, J.; Swank, L.; Lopez, D. and Rubley, T. (1999): Preliminary study of the reliability of assessment procedures for indications for chiropractic adjustments of the lumbar spine. Journal of Manipulative and Physiological Therapeutics 22 (6): 382-389.

Hestbaek, L. and Leboeuf-Yde, C. (2000): Are chiropractic tests for the lumbopelvic spine reliable and valid? A systematic critical literature review. J. Man. Physiol Ther. 23: 258-275.

Hubka, M.J. and Phelan, S.P. (1994): Interexaminer reliability of palpation for cervical spine tenderness. Journal of Manipulative and Physiological Therapeutics 17 (9): 591-595.

Jensen Stochkendahl, M. et al. (2007): Manuelle Untersuchung der Wirbelsäule – ein systematischer, kritischer Review zur Reproduzierbarkeit; Manuelle Medizin 45: 301-308.

Katcherian, D.A. (1994): Soft-tissue injuries of the ankle. In: Dutter, L.D., Mizel, M.S. and Pfeffer, G.B. (Ed.): Orthopaedic Knowledge Update: Foot and Ankle. American Academy of Orthopaedic Surgeons, Rosemont, Illinois, 241253.

Keating, J.C. Jr.; Bergmann, T. F.; Jacobs, G. E.; Finer, B. A. and Larson, K. (1990): Interexaminer reliability of eight evaluative dimensions of lumbar segmental abnormality. Journal of Manipulative and Physiological Therapeutics 13 (8): 463-470.

Keemann, J.N. (1990): Commentaar Enkelspecial, Reuma en Trauma, 1:34-5.

Keenan, A.M. and Bach, T.M. (2006): Clinicians assesment of the hindfoot: A study of reliability; Foot Ankle Int. 27(6): 451-60.

Kerkhoffs, G.M.M.J., Rowe, B.H., Assendelft, W.J.J., Kelly, K., Struijs P.A.A. and van Dijk, C.N. (2004): Immobilisation and functional treatment for acute lateral ankle ligament injuries in adults (Cochrane Review). In: The Cochrane Library, Oxford: Update Software, Issue.

Klein, P. and Sommerfeld, P. (2004): Biomechanik der menschlichen Gelenke. Grundlagen, Becken und untere Extremität; München: Urban und Fischer.

Konradsen, L., Bech, L., Ehrenbjerg, M. and Nickelsen, T. (2002): Seven years follow-up after ankle inversion trauma, Scand J. Med. Sci Sports 12: 129-135.

Koran, L.M. (1975): The reliability of clinical methods, data and judgements II. N Engl. J. Med 293: 695-701.

Kuchera, W.A. and Kappler, R.E. (2002): Musculoskeletal examination of somatic dysfunction. In: Ward, R.C. (Ed.): Foundations for Osteopathic Medicine, 2nd Ed. Philadelphia: Lippincott Williams and Wilkins: 633-659.

Landis J.R. and Koch, G.G. (1977): The measurement of observer agreement for categorical data. Biometrics 33: 159-174.

Lawson, A. and Calderon, L. (1997): Interexaminer agreement for applied kinesiology from an interexaminer reliability study design. Perceptual and Motor Skills 84 (2): 539-546.

Licht, G., Müller-Ehrenberg, H., Mathis, J., Berg, G. and Greitemann, G. (2007): Untersuchung myofaszialer Triggerpunkte ist zuverlässig. Manuelle Medizin 45: 402-408.

Liem, T. and Dobler,T.K. (2002): Leitfaden Osteopathie, München:Urban & Fischer

Liem, T. (2003): Praxis der kraniosakralen Osteopathie, Stuttgart: Hippokrates.

Liem, T. (1998): Kraniosakrale Osteopathie, Stuttgart: Hippokrates.

Ligner, B. and Van Assche, R. (1994): Bildatlas der Osteopathie, Band 1. Kötzting/Bayer. Wald: Verlag für Osteopathie: 23-24.

Lundin, M. et al. (1999): Distal pulse palpation: Is it reliable? World J. Surg. 23 (3):252-255.

Liu, S.H. and Jacobson, K.E. (1995): A new operation for chronic lateral ankle instability. J Bone Joint Surg Br 77, 55-59.

Lofvenberg, R., Karrholm, J. and Lund, B. (1994): The outcome of nonoperated patients with chronic lateral instability of the ankle: A 20-year follow-up study. Foot ankle Int 15: 165-169.

Matthijs, O., van Paridon-Edauw, D. and Winkel, D. (2006): Manuelle Therapie der peripheren Gelenke 3, München: Elsevier GmbH.

McConnell, D.G. et al. (1980): Low agreement of findings in neuromusculoskeletal examinations by a group of osteopathic physicians using their own procedures. Journal of the American Osteopathic Association 79 (7): 441-450.

Menke, W. (2000): Spezielle Sportorthopädie und Sporttraumatologie. Wiebelsheim: Limpert.

Mior, S.A.; McGregor, M. and Schut, B. (1990): The role of experience in clinical accuracy. J. Man. Physiol Ther. 13: 68-71.

Mootz, R.D. et al (1989): Intra- and Interobserver Reliability of Passive Motion Palpation of the Lumbar Spine. Journal of Manipulative and Physiological Therapeutics 12: 440-445.

Moran, R. W. and Gibbons, P. (2001): Intraexaminer and interexaminer reliability for palpation of the cranial rhythmic impulse at the head and sacrum. Journal of Manipulative and Physiological Therapeutics 24 (3): 183-190.

Muzzi, D. (2005): Palpation und Forschung. Osteopathische Medizin 2005 (1): 13-16.

Netter, F. (2006): Atlas der Anatomie des Menschen. Stuttgart: Thieme Verlag.

Northup, T.L. (1953): Foot Technique. AAO Yearbook 1953, 150-156.

Panzer, D. M. (1992): The reliability of lumbar motion palpation. J. Man. Physiol Ther 15: 518-524.

Patijn, J. (2002): Studien zur Reproduzierbarkeit und Validität diagnostischer Verfahren in der Manuellen Medizin, Manuelle Medizin 40: 339-351.

Popper, K.R. (1998): Objektive Erkenntnis. Ein evolutionärer Entwurf. Hamburg: Hoffmann und Campe.

Renström, P. and Theis, M. (1993): Die Biomechanik der Verletzungen der Sprunggelenksbänder, Sportverletzung Sportschaden 7: 29-35.

Rosner, B. (1986): Fundamentals of biostatics, 2nd Ed. Boston: PWS Publishing.

Russel, R. (1983): Diagnostic palpation of the spine: a review of procedures and assessment of their reliability. J. Man. Physiol. Ther. 6: 195-198.

Sacket, D.L.; Brian Haymes, R.; Guyatt, G.H. and Tugwell, P. (1991): Clinical Epidemiology, 2nd Ed., Philadelphia: Lippincott Williams & Williams.

Schallier, F. (2006): Osteopathische Diagnostik des oberen und unteren Sprunggelenks. DO 10: 21.

Schmitt, H. et al. (2003): Degenerative changes in the ankle in former elite high jumpers. Clin J of Sports Med 13: 6-10.

Schöps, P.; Pfingsten, M. and Siebert, U. (2000): Reliabilität manualmedizinischer Untersuchungstechniken an der Halswirbelsäule. Studie zur Qualitätssicherung in der manuellen Diagnostik. Zeitschrift für Orthopädie und ihre Grenzen 138 (1): 2-7.

Seffinger, M.A. et al. (2004): Reliability of Spinal Palpation for Diagnosis of Back and Neck Pain – A Systematic Review of the Literature. Spine, Vol. 29, 19: E413-425.

Smith, D. and Craige, E. (1980): Enhancement of Tactile Perception in Palpation. Circulation 62 (5): 1114-1118.

Sommerfeld, P. (2006): Touching Reliability, Wien: Masterthesis, WSO.

Sommerfeld, P. (2006): Methodologie, 3. Ed., Script. Wien: WSO.

Sommerfeld, P., Kaider, A. and Klein, P. (2004): Inter- und Intratester-Reliabilität der Palpation des "Primären Respiratorischen Mechanismus" innerhalb des "Cranialen Konzeptes". Osteopathische Medizin 2004(4): 4-10.

Stochkendahl, M.J. et al. (2006): Manual examination of the spine: A systematic critical literature review of reproducibility. J Manipulative Physiol Ther 29: 475-485.

Tixa, S. and Ebenegger, B. (2004): Angewandte Osteopathie, Bd. 1, Stuttgart: Hippokrates: 200-201.

Van Dijk, C.N. (2002): Management of the sprained ankle. British journal of sports medicine 36: 83-84.

Verhagen R.A., de Keizer, G. and van Dijk, C.N. (1995): Long-term follow up of Inversion trauma of the ankle. Arch Orthop Trauma Surg 114, 92-96.

Weileder, A. (2007): Die „physiologische Inkongruenz“ des oberen Sprunggelenkes – Ein optimiertes biologisches Bauprinzip auch in nicht-sphärischen Gelenken? München: Dissertation, Ludwig-Maximilian-Universität.

Woodword, M. (1999): Epidemiology. Study design and data analysis. Boca Rata-London-New York: Chapman and Hall.

Yeung, M.S., Chan, K.M., So, C.H. and Yuan, W.Y. (1994): An epidemiological study on ankle sprains. British Journal of Sports Medicine, 28: 112-116.

Zeegers, A. (1995): Het supinatieletsel injury van de enkel [NL]. Utrecht: Univ. of Utrecht.

11. Table of Illustrations

Fig. 1 – Fig. 3: Dornieden, R. (2008)

Fig. 8 – Fig. 12: Dornieden, R. (2008)

Annex 1a

Questionnaire (Osteopaths)

Fragebogen Osteopath **Studienort:**

Name:	Nr: O	Alter:	Geschlecht: m w

Wann haben Sie Ihren Abschluss gemacht? (Jahr der letzten Prüfung)	
Wie lange arbeiten Sie schon osteopathisch?	

An welcher Schule?

Kannten Sie den Test bereits vorher?	Ja	Nein

Wenn ja, wenden Sie ihn regelmässig am Patienten an?	Ja	Nein

(regelmässig = mehrmals in der Woche)

Vielen Dank, dass Sie an meiner Studie mitarbeiten. Ihre Angaben werden vertraulich behandelt.

Annex 1b

Questionnaire (Test Persons)

Fragebogen Proband **Studienort:**

Vielen Dank, dass Sie an der Studie teilnehmen. Bitte füllen Sie den Fragebogen sorgfältig aus.

Name:	**Nr.:**	**Geschlecht: m w** **Alter:**

Haben Sie aktuell Fussbeschwerden?	**Ja**	**Nein**
Wenn Ja: welcher Art? Bitte beschreiben Sie.		

Sind Sie schon mal umgeknickt?	**Ja**	**Nein**
Wenn ja? Wann/ welche Seite?		

Hatten Sie andere Unfälle im Bereich der Beine?	**Ja**	**Nein**
Wenn ja: Wann/ welcher Art?		

Wurden Sie im Bereich der Beine operiert?	Ja	Nein
Wenn Ja: Wann/ Wo?		

Bitte beachten Sie,dass für die Untersuchung wichtig ist, dass Sie:

keine bemalten Fussnägel
keine Narben im Beinbereich
keine Wachstumsstörungen/Veränderungen im Nagelbereich
keine Pigmentstörungen/ dominanten Muttermale
keine Piercings, Tatoos im Beinbereich haben und
die Rückenlage zum Testing einnehmen können.

Hiermit bestätige ich, dass ich keine der o.g. Merkmale aufweise: Ja Nein

Bitte ziehen Sie für die Untersuchung eine kurze Sporthose an. Vielen Dank.

Annex 1c

Record Sheet (1st Test Run)

Studienort

Masterarbeit: Ralf Dornieden
Thema: Inter- und intrarater Reliabilität Talustest anteromediale Dysfunktion

Stichprobengrösse: Durchgang: 1
Name Osteopath (Tester):

Anteromediale Dysfunktion

Bitte ankreuzen

Pat. Nr	linker Fuss (ja/nein)	rechter Fuss (ja/nein)	festes BG	weiches BG	
1					1
2					2
3					3
4					4
5					5
6					6
7					7
8					8
9					9
10					10
11					11
12					12
13					13
14					14
15					15
16					16
17					17
18					18
19					19
20					20
21					21
22					22
23					23
24					24
25					25
26					26
27					27
28					28
29					29
30					30

Annex 1d

Record Sheet (2nd Test Run)

Studienort

Masterarbeit: Ralf Dornieden
Thema: Inter- und intrarater Reliabilität Talustest anteromediale Dysfunktion

Stichprobengrösse: Durchgang: 2
Name Osteopath (Tester):

Anteromediale Dysfunktion

Bitte ankreuzen

Pat. Nr	linker Fuss (ja/nein)	rechter Fuss (ja/nein)

festes BG	weiches BG	

Annex 2

Questionnaire Data and Test Results (Raw Data)

Landquart				Left foot				Right foot			
Test person	**Age**	**Sex**	**ID**	**th1_ex1**	**th1_ex2**	**th2_ex1**	**th2_ex2**	**th1_ex1**	**th1_ex2**	**th2_ex1**	**th2_ex2**
2	24	w	2	no	no	no	no	yes	yes	no	yes
3	21	m	3	yes	no	yes	no	no	yes	no	yes
4	21	m	4	yes	no	yes	no	no	no	no	yes
5	22	w	5	no	no	no	no	no	no	no	no
6	22	w	6	yes	no	yes	no	no	no	no	no
7	24	w	7	no	no	no	no	no	no	no	no
8	22	w	8	yes	yes	no	no	no	no	no	yes
9	23	w	9	no	no	yes	yes	no	no	no	no
10	21	m	10	no	yes	no	no	no	no	no	yes
11	21	w	11	no	no	no	no	no	no	no	no
12	26	m	12	no	no	yes	no	no	no	no	no
13	21	w	13	no	yes	no	no	no	no	no	no
14	22	m	14	no	no	no	no	no	no	no	no
15	26	w	15	yes	no	no	no	no	no	no	no
16	23	w	16	no	no	no	no	no	no	no	yes
17	24	w	17	no	no	no	no	no	yes	no	no
18	23	m	18	no	no	no	no	no	no	no	no
19	32	m	19	no	no	no	yes	no	yes	no	no
20	24	m	20	yes	no	no	no	no	no	no	no
21	22	w	21	no	no	no	no	no	no	no	no

	Q1: actual dysfunction (foot)				Q2: ankle sprain			
Test person	**yes/no**	**dysfunction**	**foot**		**yes/no**	**year**	**foot**	**else**
2	no				yes		r	
3	no				yes	2000	r	
4	yes	pain right foot	r		yes		r	long ago
5	no				yes			occasionally
6	no				yes		l/r	
7	no				yes	2004	r	
8	no				no			
9	yes	pain at load	l		yes	2007	l	occasionally
10	no				yes	2007	l/r	occasionally
11	no				yes	2002	l	
12	no				no			
13	no				yes	2003	r	
14	yes	Morton's neuralgia	l/r		yes		l/r	occasionally
15	no				yes		?	
16	no				no			
17	no				yes	2008	r	often
18	no				no			
19	no				yes	1985		
20	no				yes	1994	l/r	
21	no				no			

	Q3: accidents (legs)				Q4: surgery (legs)		
Test person	**yes/no**	**Foot**			**yes/no**	**Seite**	
2	no				no		
3	yes	r	greenstick fracture 2000		no		
4	no				no		
5	yes	r	ant. cruc. lig rupture		yes	r	ant. cruc. lig r: 2004, meniscus r: 2006
6	no				no		
7	yes	?	ant. cruc. lig/meniscus		yes	?	meniscus: 2004, ant. cruc. lig+ meniscus: 2006, meniscus: 2007
8	yes	l/r	Patella lux.		yes	r	Patella lux. l: 2000, r: 1991
9	no				no		
10	yes	l	ant. cruc. lig-rupture: 2003		no		
11	no				no		
12	yes	?	Fracture metatarsus: 1995		no		
13	no				no		
14	yes	l/r	Talus fracture		no		
15	no				no		
16	no				no		
17	yes	l/r	Patella lux. l 2001, Meniscus r: 2006		yes	l/r	meniscus refixation l: 2001, r: 2006
18	no				no		
19	yes	?	toboggan acc.: 1985		no		
20	yes	l	Patella inflamation		yes	l	2x Resection M.Osgood Schlatter l
21	no				no		

Ulm-Dornstadt				Left foot							
Test person	**Age**	**Sex**	**ID**	**th6_ex1**	**th7_ex1**	**th8_ex1**	**th9_ex1**	**th6_ex2**	**th7_ex2**	**th8_ex2**	**th9_ex2**
1	28	w	1	no	no	no	no	yes	no	no	no
2	32	w	2	no	no	no	no	no	no	no	no
3	34	w	3	no	no	no	no	no	no	no	no
4	32	m	4	no	no	yes	no	no	no	no	no
5	45	m	5	no	yes	yes	no	no	yes	yes	no
6	38	w	6	no	no	no	no	no	no	no	no
7	37	w	7	yes	no	yes	no	no	no	yes	no
8	45	w	8	no	yes	no	no	no	no	no	no
9	39	w	9	no	yes	no	no	no	no	no	no
10	36	w	10	no	yes	no	yes	yes	yes	no	no
11	32	m	11	no	no	yes	no	no	no	yes	yes
12	35	m	12	no	no	no	no	no	no	yes	no
13	33	m	13	no	no	no	no	no	no	no	no
14	50	m	14	no	yes	no	yes	no	yes	yes	no
15	30	w	15	no	yes	yes	no	no	yes	no	no
16	42	w	16	no	no	no	no	no	no	yes	no
17	42	w	17	no	no	no	no	no	no	no	no
18	30	w	18	no	no	no	no	no	no	no	no
19	27	w	19	no	no	no	no	no	no	no	no
20	50	w	20	no	no	no	no	no	no	yes	no
21	38	w	21	no	no	no	no	no	no	no	no
22	30	w	22	no	no	yes	no	no	no	yes	no
23	38	m	23	no	yes	no	yes	no	yes	yes	no
24	38	m	24	no	yes	no	no	no	no	no	no
25	27	w	25	no	no	yes	no	no	no	no	no
26	43	w	26	no	no	no	no	no	no	no	no
27	38	m	27	no	no	no	no	no	yes	no	no
28	40	w	28	no	no	no	no	no	no	no	no
29	32	m	29	no	yes	no	no	no	no	yes	no
30	61	m	30	no	yes	no	no	no	no	yes	no

	Right foot							
Test person	**th6_ex1**	**th7_ex1**	**th8_ex1**	**th9_ex1**	**th6_ex2**	**th7_ex2**	**th8_ex2**	**th9_ex2**
1	no	no	no	no	no	no	no	no
2	no	no	no	no	no	no	yes	no
3	no	no	no	no	no	no	no	no
4	no	no	no	no	no	yes	no	no
5	no	yes	no	no	no	no	no	no
6	no	no	no	no	yes	no	no	no
7	no	yes	no	no	no	no	no	no
8	no	no	no	no	no	no	yes	yes
9	no	no	no	no	no	no	no	no
10	yes	no	no	no	yes	yes	no	no
11	no	yes	no	yes	no	no	no	no
12	no	no	yes	no	no	no	no	no
13	yes	no	yes	yes	no	yes	no	yes
14	no	no	no	no	no	no	no	no
15	yes	no	no	yes	yes	no	no	yes
16	no	no	no	no	no	no	no	no
17	no	no	no	no	no	no	no	no
18	no	yes	no	no	no	no	no	no
19	no	no	no	yes	no	no	yes	yes
20	no	no	no	no	no	no	no	no
21	no	no	no	no	no	no	no	no
22	no	no	no	no	no	no	no	no
23	yes	no	no	no	no	no	no	no
24	no	no	no	no	yes	yes	no	yes
25	no	no	no	no	no	no	no	no
26	yes	yes	no	no	no	yes	no	yes
27	no	no	no	no	no	no	no	no
28	no	no	no	no	no	no	yes	no
29	yes	no	yes	yes	no	yes	no	no
30	yes	no	no	yes	yes	yes	no	no

	Q1: actual dysfunction (foot)				Q2: ankle sprain				
Test person	**yes/no**	**dysfunction**	**foot**		**yes/no**	**year**	**foot**	**frequency**	**else**
1	no				yes	2006	r		l/r
2	no				yes	1996	l/r	occasionally	
3	no				yes	1998	r		
4	no				no				
5	no				no				
6	no				no				
7	no				yes	1992	l		
8	no				yes	1987	r		l/r
9	no				yes	1986	r		
10	no				yes		l/r	occasionally	
11	no				yes	1999	l		
12	no				yes	1993	l		
13	no				yes	1995	r		l/r
14	yes	big toe	r		no				
15	no				yes		l/r		
16	no				no				
17	yes	pain	r		yes	1978	?		
18	no				yes		l/r		
19	no				no				
20	no				yes		r		
21	yes	instability	r		yes	2003	r		
22	no				yes	2004			
23	no				yes		l/r		
24	yes	congestion	l		yes		l/r		l>r
25	yes	block feeling	r		no				
26	no				yes	1997	l		
27	no				yes	2002	l		
28	yes	blockage	r		no				
29	no				yes	2000	r		l/r
30	no				yes		l/r	often	

	Q3: accidents (legs)				Q4: surgery (legs)		
Test person	**yes/no**	**Foot**			**yes/no**	**Foot**	
1	yes	r	ant. cruc. lig 1998		yes	r	knee: 1998
2	no				no		
3	yes	r	intraarticular ligament knee, 1986		yes	r	knee: 1998
4	no				no		
5	yes	r	fracture of patella: 1976, intaarticular ligament knee: 2001				
6	no				no		
7	no				yes	r	knee: 1996
8	no				no		
9	yes	r	intraarticular ligament knee: 1987		yes	r	knee: 1987
10	yes	l	knee: 1987		yes	l	knee: 1987, OSG: 1990
11	no				no		
12	yes	l/r	intraarticular ligament left knee: 1995, Fibula pain l: 1979, Fibula pain, r: 1985				
13	yes	r	ant. cruc. lig: 1999		yes	l/r	OSG l: 1993, r: 1995, ant. cruc. lig r: 1999
14	no				yes	l/r	Miniscus r: 2003, l: 1998
15	no				yes	l/r	knee r: 1993, l: 1995
16	no				no		
17	no				no		
18	no				no		
19	no				no		
20	no				yes	r	veins: 2005
21	no				no		
22	no				no		
23	no				yes	l	meniscus: 1996
24	no				no		
25	no				no		
26	no				no		
27	yes	l/r	Achilles tendon, r: 1988, l: 2006, ant. cruc. lig l: 2008				
28	no				no		
29	no				no		
30	no				yes	l/r	meniscus r: 2004, l: 2005

St Gallen				Left foot					
Test person	**Age**	**Sex**	**ID**	**th3_ex1**	**th3_ex2**	**th4_ex1**	**th4_ex2**	**th5_ex1**	**th5_ex2**
1	37	m	1	no	no	no	no	no	no
2	21	m	2	no	no	no	no	no	no
3	22	m	3	no	no	no	no	no	no
4	16	m	4	no	no	no	yes	no	no
5	21	m	5	no	no	yes	no	no	no
6	24	m	6	no	yes	no	no	no	yes
7	26	m	7	no	no	no	no	no	no
8	25	m	8	no	no	no	no	no	no
9	21	m	9	no	yes	yes	no	no	no
10	21	m	10	yes	no	no	no	no	yes
11	30	m	11	no	no	no	yes	no	no
12	24	m	12	no	no	no	no	no	no
13	24	m	13	yes	no	no	no	no	no
14	35	m	14	no	no	no	no	no	no
15	17	m	15	no	no	no	yes	no	no
16	29	m	16	no	no	no	no	no	no
17	17	m	17	no	no	yes	yes	no	no
18	17	m	18	no	yes	no	no	no	no
19	33	m	19	no	no	no	no	no	no
20	13	w	20	no	no	yes	no	no	no
21	26	m	21	yes	no	no	no	yes	no
22	25	m	22	no	yes	yes	yes	no	no
23	24	m	23	no	no	no	no	no	no
24	19	w	24	no	no	no	no	no	no
25	18	w	25	no	yes	no	no	no	no
26	20	w	26	no	no	no	no	no	no
27	19	w	27	no	no	no	no	no	no
28	21	w	28	yes	no	yes	yes	no	no
29	17	w	29	yes	no	yes	no	no	no
30	17	m	30	no	no	yes	no	no	no
31	15	w	31	no	no	no	no	no	no
32	16	w	32	no	no	no	no	no	no
33	15	w	33	no	no	yes	no	no	no
34	19	w	34	no	no	yes	yes	no	no
35	15	w	35	no	yes	no	yes	no	no
36	14	w	36	no	no	no	no	no	no
37	30	m	37	no	no	no	no	no	no
38	36	m	38	no	no	yes	no	no	no
39	19	w	39	no	no	no	no	no	no
40	31	w	40	no	no	yes	no	no	no
41	30	m	41	yes	no	no	yes	no	no
42	34	m	42	no	no	no	no	no	no
43	25	w	43	no	yes	no	no	no	no

	Right foot					
Test person	**th3_ex1**	**th3_ex2**	**th4_ex1**	**th4_ex2**	**th5_ex1**	**th5_ex2**
1	no	no	no	no	no	no
2	no	no	no	no	no	no
3	no	no	no	no	no	no
4	no	no	no	no	no	no
5	no	yes	no	no	no	no
6	no	no	no	yes	no	no
7	no	no	no	no	no	no
8	no	yes	yes	no	no	no
9	yes	no	no	no	yes	no
10	no	no	no	no	no	no
11	yes	no	no	no	yes	no
12	no	yes	no	no	no	no
13	no	no	no	no	no	no
14	no	no	no	no	no	no
15	yes	no	yes	no	no	no
16	no	no	no	no	no	no
17	no	no	no	no	no	no
18	no	no	no	no	no	no
19	no	no	no	no	no	no
20	no	no	no	no	no	no
21	no	yes	no	no	no	yes
22	no	no	no	no	yes	yes
23	no	no	no	yes	no	no
24	no	no	yes	no	no	no
25	no	no	no	no	yes	yes
26	no	yes	no	no	no	no
27	no	no	no	no	yes	no
28	no	no	no	no	no	no
29	no	no	no	no	no	no
30	no	no	no	no	no	no
31	no	no	no	no	no	no
32	no	no	no	no	no	no
33	no	no	no	no	no	no
34	no	no	no	no	no	no
35	no	no	no	no	no	no
36	no	no	no	no	no	no
37	no	no	no	no	no	no
38	no	yes	no	no	no	no
39	no	no	no	no	no	no
40	no	no	no	no	no	no
41	no	no	no	no	no	no
42	no	no	no	no	no	no
43	no	no	no	no	no	no

	Q1: actual dysfunction (foot)				Q2: ankle sprain				
Test person	**yes/no**	**dysfunction**	**foot**		**yes/no**	**year**	**foot**	**frequency**	**else**
1	no				yes	2000	l		
2	no				yes	2007	l		
3	no				yes	2008	l/r		
4	no				yes	2005	l		
5	no				yes	2007	l/r		r: 2007, l: 2006
6	no				yes		l/r	occasionally	
7	yes	ligament laxation, toe pain	l/r		yes		l/r	occasionally	
8	no				yes	2004	r		
9	no				yes		l/r	very oftenen	
10	no				yes	2006	r		
11	no				yes		l/r	often	
12	no				yes	2003	l/r	occasionally	
13	no				yes	2007	r	occasionally	
14	no				yes	2008	r		l: 1998
15	no				yes	2007	r		
16	no				yes		l/r	often	r>l
17	no				yes				long ago
18	no				yes		l/r	occasionally	
19	yes	joint inflammation	l/r		yes	2008	r		
20	no				yes	2008	r		
21	no				yes	2007	r	very oftenen	
22	no				yes		l/r	very oftenen	
23	yes	pain fifth toe	r		yes	2001	l		
24	yes	ankle sprain			yes		l/r	often	
25	no				yes				long ago
26	yes	Achilles tendon	l/r		yes	2006	l		l/r
27	no				yes	2003	l		
28	no				yes	2007	l/r	often	
29	no				yes	2008	r		
30	no				yes	2006	r		
31	no				yes	2007	r		
32	no				yes	2007	r		
33	no				no				
34	no				yes	2006	l/r	occasionally	
35	no				yes	1999	l/r		
36	no				yes	2007	l		
37	yes	swollen ankle	l		yes	2008	l	occasionally	
38	no				yes	2007	r		
39	no				yes	2007	r		
40	no				no				
41	yes	foot lat. lig.	r		yes	2003	l		
42	yes	Achilles tendon	l		yes	2002	r		
43	no				yes	1990	l		l/r

	Q3: accidents (legs)				Q4: surgery (legs)		
Test person	**yes/no**	**Foot**			**yes/no**	**Foot**	
1	yes	l/r	ant. cruc. lig, meniscus		yes	l/r	arthroscopy: 2000
2	yes	r	ant. cruc. lig tear 2007		yes	r	ant. cruc. lig: 2007
3	no				no		
4	no				no		
5	no				no		
6	no				no		
7	no				no		
8	no				no		
9	no				no		
10	yes	r	toe fracture: 2006		yes	r	toes: 2006
11	yes	r	Achilles tendon		yes	r	Achilles tendon: 2004
12	yes	r	ant. cruc. lig 2003		yes	r	ant. cruc. lig: 2003
13	yes	l/r	ligament tear: 2007		no		
14	no				no		
15	no				no		
16	no				no		
17	no				no		
18	no				no		
19	yes	?	Patella luxation		no		
20	no				no		
21	no				no		
22	no				no		
23	no				no		
24	yes	l	leg fracture: 1980		no		
25	yes	r	double tibia fracture: 1996		no		
26	no				no		
27	yes	r	lat. lig. Knee: 2006		yes	r	lat. lig.: 2006
28	no				no		
29	no				no		
30	no				no		
31	no				no		
32	no				no		
33	no				no		
34	no				yes	l	foot: 2000
35	no				no		
36	no				no		
37	no				no		
38	no				no		
39	yes	r	swollen knee		no		
40	no				no		
41	no				no		
42	yes	r	ligament tear 2x		yes	r	Achilles tendon: 2006
43	yes	l	ligament tear: 1990		no		

Magglingen				Left foot						
Test person	**Age**	**Sex**	**ID**	**th10_ex1**	**th10_ex2**	**th11_ex1**	**th11_ex2**	**th12_ex1**	**th12_ex2**	**th10_ex1**
1	29	w	1	no	no	no	no	no	no	no
2	25	w	2	no	yes	no	no	no	no	no
3	29	w	3	yes	no	no	no	yes	no	no
4	25	m	4	no	no	no	yes	no	no	no
5	28	w	5	no	no	no	no	no	no	no
6	24	w	6	no	no	no	no	no	no	no
7	36	m	7	no	no	no	no	no	no	no
8	33	m	8	no	no	yes	no	yes	no	no
9	34	w	9	yes	yes	no	no	yes	no	no
10	30	w	10	no	no	no	no	no	no	no
11	26	w	11	no	yes	no	no	no	no	yes
12	27	w	12	no	no	no	no	no	no	no
13	28	m	13	no	no	no	no	no	no	no
14	40	m	14	no	no	no	no	yes	no	no
15	30	m	15	no	no	no	no	yes	no	yes
16	33	m	16	no	no	no	no	no	no	no
17	25	w	17	no	no	no	no	no	no	no
18	30	w	18	no	no	yes	no	no	no	no
19	32	m	19	no	no	yes	no	no	no	no
20	33	w	20	no	no	no	no	no	no	no
21	28	w	21	no	no	no	no	no	no	no
22	33	w	22	no	no	no	no	no	no	no
23	31	w	23	no	no	no	yes	no	no	no
24	28	w	24	no	no	no	no	no	yes	no
25	25	w	25	no	no	no	no	no	no	no
26	28	m	26	no	no	no	yes	no	no	no
28	25	w	28	yes	yes	no	no	yes	no	no
29	35	w	29	yes	no	no	no	no	no	no

Right foot					
Test person	**th10_ex2**	**th11_ex1**	**th11_ex2**	**th12_ex1**	**th12_ex2**
1	no	no	no	no	yes
2	no	no	yes	no	no
3	no	no	yes	no	no
4	no	no	yes	yes	yes
5	no	yes	no	no	no
6	no	no	yes	yes	yes
7	no	yes	no	no	no
8	no	yes	no	no	no
9	no	no	no	no	yes
10	no	no	no	no	yes
11	no	yes	no	yes	no
12	no	no	no	no	no
13	no	no	no	no	no
14	no	no	no	no	no
15	no	yes	no	no	yes
16	no	no	yes	no	no
17	no	no	yes	no	no
18	no	yes	no	yes	no
19	no	yes	no	no	no
20	no	yes	yes	no	yes
21	no	no	no	no	yes
22	yes	yes	no	no	no
23	no	no	yes	no	no
24	no	no	no	no	no
25	no	no	yes	yes	no
26	no	no	yes	no	no
28	no	no	no	no	no
29	no	no	no	no	no

	Q1: actual dysfunction (foot)				Q2: ankle sprain				
Test person	**yes/no**	**dysfunction**	**foot**		**yes/no**	**year**	**foot**	**frequency**	**else**
1	no				no				
2	no				yes	2007	r		r/l
3	no				yes		r/l	often	
4	no				yes	2003	l	occasionally	
5	yes	pain	r		yes		r/l		rezidivating
6	yes	at inversion, Achilles tendon	r		yes	2007	r	occasionally	r/l
7	no				no				
8	no				yes		r/l		
9	no				no				
10	yes	?	r		yes	1988	l		
11	no				yes	2007	r		
12	no				no				
13	no				yes	2007	r/l		
14	no				yes	2006	r/l		
15	no				no				
16	no				no				
17	yes	plantar tendon			no				
18	no				no				
19	no				yes	1992	r		
20	no				no				
21	no				no				
22	no				yes		r	occasionally	
23	yes	arch of foot	r		yes	2004	l		
24	no				yes		r/l		
25	no				yes		l		
26	no				yes	2006	l		
28	no				yes	1998	r		
29	no				no				

	Q3: accidents (legs)				Q4: surgery (legs)		
Test person	**yes/no**	**Foot**			**yes/no**	**Foot**	
1	no				no		
2	yes	l	meniscus: 1995		yes	r/l	knee l: 1995, knee r: 2005
3	yes	r	intraarticular ligament knee: 2004		no		
4	yes	l	fibula pain: 1988		no		
5	no				no		
6	no				no		
7	yes	l	meniscus		yes	l	knee: 2002
8	yes	l	injury of bursa: 2007		yes	r	foot: 1986
9	no				no		
10	no				no		
11	yes	l	ant. cruc. lig: 2005		yes	l	ant. cruc. lig: 2005
12	yes	?	knee distortion: 2006		no		
13	yes	r	ant. cruc. lig, meniscus: 1999		yes	r	t. cruc. lig, meniscus: 19
14	yes	r	ant. cruc. lig: 2007		yes	r	big toe: 1998
15	no				no		
16	no				no		
17	no				no		
18	yes	l	knee pain: 2003		no		
19	no				no		
20	yes	?	tibia- and fibula pain: 1991		yes	?	osteosynthesis
21	yes	r/l	ant. cruc. lig l: 1994, pain in first toe r: 1996				
22	yes	r	ant. cruc. lig: 1998		yes	r	ant. cruc. lig: 1998
23	no				no		
24	yes	?	naviculare pain: 2005		yes	?	suprapaletar cut
25	no				no		
26	no				no		
28	no				no		
29	yes	r/l	nt. cruc. lig l: 2000, pain in Achilles tendon: 2002		?		