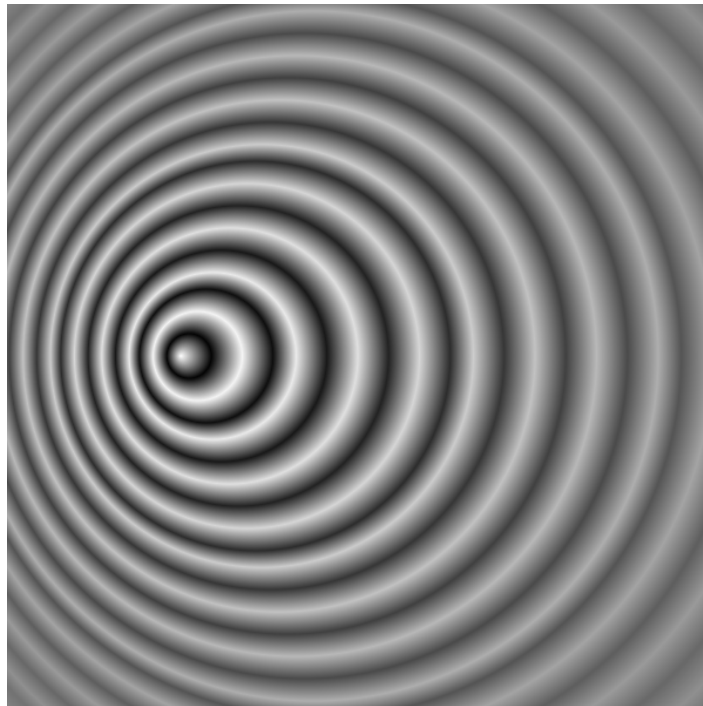


Accuracy of ultrasound Doppler measurements in assessment of disease activity in patients with rheumatoid arthritis

**PhD thesis by Karen Ellegaard, PT, MSc
The Parker Institute
Frederiksberg Hospital
The Capital Region of Copenhagen
Denmark
October 2009**



**This Thesis is submitted for the degree of Doctor of Philosophy
at the International Doctoral School
In Biomedical Science and Engineering
Aalborg University, Denmark**

CONTENTS

PREFACE.....	4
ABBREVIATIONS	5
ACKNOWLEDGMENTS	6
INTRODUCTION	8
Hypotheses.....	8
Aim	9
RHEUMATOID ARTHRITIS.....	9
Epidemiology.....	9
Aetiology and pathogenesis.....	9
Diagnosis	10
Disease course	10
Assessment of disease activity, progression and treatment response.....	11
<i>Joint count</i>	11
<i>Laboratory tests (CRP, ESR)</i>	11
<i>Visual Analog Scale (VAS)</i>	11
<i>Morning stiffness</i>	12
<i>Health assessment questionnaire and SF-36</i>	12
<i>Disease indices</i>	12
Treatment of RA	12
ULTRASOUND	13
US transducer.....	13
Grey-scale (B-mode)	13
Doppler	14
<i>Doppler signal</i>	14
<i>Insonation angle/Doppler angle</i>	15
<i>Spectral Doppler</i>	15
<i>Colour Doppler and Power Doppler (energy Doppler)</i>	15
<i>Choice of Doppler</i>	15
Doppler parameters/setting.....	16
<i>Doppler frequency</i>	16
<i>Colour box</i>	16
<i>Pulse repetition frequency</i>	16
<i>Gain</i>	16
<i>Focus</i>	17
<i>Colour priority (threshold)</i>	17
<i>Wall filters</i>	17
<i>Persistence-frame averaging</i>	17
Artefacts.....	17
<i>Random noise</i>	17
<i>Motion</i>	17
<i>Mirror</i>	17
<i>Blooming</i>	18
<i>Reverberation</i>	18
<i>Aliasing</i>	18
<i>Pressure and positions</i>	18
ASSESSMENT OF PERFUSION BY DOPPLER US	18
MEASUREMENT VALIDITY	20
Face and content validity	21

Criterion validity.....	22
Construct validity.....	22
OMERACT.....	23
RELIABILITY OF DOPPLER MEASUREMENTS IN RA.....	24
Study I - Test-retest reliability.....	25
Reliability of analysis of static images.....	26
Doppler measurements during the heart cycle.....	27
Scoring technique.....	27
Summary.....	27
VALIDITY OF DOPPLER MEASUREMENTS.....	28
Study II - Concurrent validity.....	29
Study III - Accuracy.....	29
Assessment of concurrent validity.....	30
Assessment of accuracy.....	32
Summary.....	33
LEARNING COURSE IN ULTRASOUND EXAMINATION IN RA.....	33
Study IV - Learning curve in musculoskeletal US.....	34
The framework for a US training program.....	35
Utility of various training approaches in US.....	36
Summary.....	37
CONCLUSION.....	38
FUTURE CONSIDERRATIONS.....	39
ENGLISH SUMMARY.....	40
DANISH SUMMARY.....	42
REFERENCE LIST.....	44

APPENDIX: Study I-IV 52

PREFACE

The present thesis is based on the four papers listed below (I-IV). The original papers are attached as appendices 1-4. All four studies were carried out at The Parker Institute, Frederiksberg Hospital in the period from 2006-2009.

Study I

Ellegaard K, Torp-Pedersen S, Lund H, Henriksen M, Terslev L, Jensen PS, Danneskiold-Samsøe B, Bliddal H

Quantification of colour Doppler activity in the wrist in patients with rheumatoid arthritis - the reliability of different methods for image selection and evaluation.

Ultraschall Med. 2008 Aug;29(4):393-8.

Study II

Ellegaard K, Torp-Pedersen S, Terslev L, Danneskiold-Samsøe B, Henriksen M, Bliddal H.

Ultrasound colour Doppler measurements in a single joint as measure of disease activity in patients with rheumatoid arthritis - assessment of concurrent validity.

Rheumatology (Oxford). 2009 Mar;48(3):254-7

Study III

Ellegaard K, Torp-Pedersen S, Henriksen M, Lund H, Danneskiold-Samsøe B, Bliddal H

Influence of Recent Exercise and Skin Temperature on US Doppler Measurements in Patients with Rheumatoid Arthritis - an Intervention Study.

Rheumatology (Oxford). 2009 Sep 18. [Epub ahead of print] PMID:197673388

Study IV

Ellegaard K; Torp-Pedersen S; Stoltenberg M; Hansen A; Lorenzen T; Jensen D.V; Lindegaard H; Juul L; Røgind H; Bülow P; Cheysides S; Kowalski M; Danneskiold-Samsøe B; H.Bliddal

Standardized ultrasound examination of the wrist and metacarpophalangeal joints in patients with rheumatoid arthritis - A multicentre learning experience: Education in Ultrasound of Rheumatoid Arthritis- EURA.

Submitted

ABBREVIATIONS

ACR	American College of Rheumatology
CCP	Cyclic citrullinated peptides
CRP	C- reactive protein
ESR	Erythrocyte sedimentation rate
EULAR	European League against Rheumatism
HAQ	Health assessment questionnaire
IL-1	Interleukin-1
IL-6	Interleukin-6
MCP	Metacarpophalangeal
MRI	Magnetic resonance imaging
PIP	Proximal interphalangeal
PRF	Pulse repetition frequency
RA	Rheumatoid arthritis
ROI	Region of interest
RF	Rheumatoid factor
TNF α	Tumour necrosis factor alpha
OMERACT	Outcome Measures in Rheumatology (formerly: Outcome Measures in rheumatoid arthritis clinical trials)
US	Ultrasound
VAS	Visual analogue scale

ACKNOWLEDGMENTS

The summer day in 2004 when I arrived at my first appointment at the Parker Institute, I did not have the slightest idea that this meeting would be the beginning of a complete change in my career. I am very happy it did so, but it has only been possible with the help, support and encouragement I have had from a lot of people not least at the Parker Institute. Without the help from all these people, my Ph.D. study would never have been possible and this thesis would not have been a reality.

The two people whom I first of all want to thank are Bente Danneskiold-Samsøe and Hans Lund. It was their positive attitude and their confidence in me that summer day in 2004, which made it possible for me to take the first steps into the field of medical science.

I also want to thank Søren Torp-Pedersen for gently guiding me into the area of musculoskeletal ultrasound imaging and for providing the ideas behind the studies in the thesis. Without Søren's support, patience, open-mindedness and capacity to make me feel that every single of my endless stream of questions was relevant, I would have stumbled a lot of times in the process.

I am also very thankful for the support and positive attitude from Henning Bliddal, who has been of great importance in the conception and conduction of the studies forming this thesis. His impressive knowledge of the rheumatological field has been a great help and inspiration for me.

I want to thank all my colleagues at the Parker Institute for creating a very dynamic, positive and inspiring working environment. Several colleagues have contributed directly to my work, particularly the group of people who have been part of the ultrasound group in the five years I have been at the Parker Institute, especially I want to thank Mette Gad for her support in the ultrasound room and for her proof-reading of my manuscripts and other texts and Peter, Kira, and Katrine for their help with the image analysis, and in addition Peter for his willingness to solve my various computer problems. I want to thank Marius Henriksen, Robin Christensen and Christian Cato Holm for their help with systematizing and analysing my data. I also want to thank Salomea Hirschorn, Jette Nielsen and the rheumatologists at clinic H for the collaboration and finally Line Rustad and Claus Bomhoff for both professional assistance and personal support.

The studies on which this thesis is based would have been impossible to perform without the financial support of *The OAK foundation, The Danish Physiotherapy Association, Aase and Ejnar Danielsens Foundation, Beckett-Foundation, The Danish Rheumatism Association, Direktør Ib Henriksens Foundation, C.C. Klestrup Foundation, Civilingeniør Frode V. Nyegaards Foundation, Toyota Foundation, Støtteforeningen for gigtbehandling og forskning ved Frederiksberg Hospital, The A.P. Møller Foundation for the Advancement of Medical Science and H.H Holm's foundation*. I am grateful to all of them.

I thank all the patients who participated in the various projects, for their patience, willingness and their impressive ability to focus on the benefit for all patients – even when the study resulted in inconvenience for themselves. Without their help the whole project would of course have been impossible.

I would like to thank my friend Annette for initially encouraging me to re-start my studies and for personal support during the whole process.

Last but not least I am very thankful to my sweet and patient husband John and our clever and beautiful daughters Katrine and Ida for always believing in my skills even when I doubted myself.

Thank you!

Karen Ellegaard

Summer 2009

INTRODUCTION

Rheumatoid arthritis (RA) is a chronic autoimmune disease with a pathogenesis intimately associated with the synovial membrane. RA is characterized by pain on movement and most patients develop weakness and wasting of muscles. Also, RA often leads to irreversible cartilage and bone destruction, resulting in varying degrees of impairment of physical ability (1;2). No pathognomonic feature exists for the diagnosis of RA, which is therefore based on a set of classification criteria (3). No single measurement of disease activity exists thus the evaluation of disease activity is made by assessment of a combination of various measurements.

The use of ultrasound (US) imaging in the rheumatologic area has increased considerably within the last years and is becoming an integral part of the management of patients with RA in many rheumatologic settings (4). In an US examination, morphological changes are assessed by grey-scale US, and blood flow is investigated by US Doppler (4).

Despite increasing use of US Doppler in the assessment of synovial perfusion in RA, evidence of the validity of the measurement is sparse. Thus, the validity of Doppler US has to be further explored in order to optimise its use as a diagnostic tool for detection of inflammation and assessment of the degree of inflammation in RA (5).

Sufficient treatment in the initial stages of RA is crucial for the course of the disease and may prevent joint destruction (6-9). The early treatment of RA in combination with the introduction of new effective – and expensive - medications has increased the requirement for valid measurements for both diagnosis and monitoring of treatment in patients with RA. The use of Doppler US for these purposes has been demonstrated in several studies (6;10-15). Furthermore, recent studies have shown that patients in clinical remission may still display disease activity when assessed with Doppler US (16-18). In addition, one study showed that patients displaying subclinical Doppler activity were more likely to develop bone erosions (18) suggesting a superiority of US Doppler measurements to clinical assessment in the assessment of ongoing inflammation in patients with RA.

Thus, Doppler US is a promising measurement in the rheumatological area. However, before the full benefit of Doppler US measurements may be achieved, the validity has to be fully clarified.

Hypotheses

The hypotheses of the thesis are

- US Doppler is a valid measurement of disease activity in RA
- US Doppler is applicable in assessment of disease activity in RA

Aim

The aim of this Ph.D. project was to investigate the accuracy of US Doppler as a measurement of inflammation/disease activity in patients with RA.

The accuracy was pursued by investigation of the:

- reliability of the US Doppler
- validity of the US Doppler
- learning experience of US Doppler examination

RHEUMATOID ARTHRITIS

Epidemiology

Rheumatoid arthritis is an autoimmune disorder primarily attacking the synovial tissue. The prevalence of RA in the western world is approximately 1% and the female-to-male ratio is 2.5:1. The skewed distribution between men and women decreases with increasing age. The disease is less common in Asian and African parts of the world (1;2). The peak age for disease onset is between 40 and 60 years and the prevalence increases steadily with increasing age (1;2).

In many patients with RA the ability to work is affected, causing early retirement, indicating the seriousness of the disease for both the individual and the society as a whole (19).

Aetiology and pathogenesis

Despite intensive investigation the fundamental aetiology of the development of autoimmunity in RA is still unknown. The pathogenesis is intimately associated with the synovial membrane; the synovial membrane is the connective tissue lining the synovial joints, tendon sheaths and bursae. However, some patients with especially aggressive RA develop extra-articular manifestations of the disease (1;2). After onset most patients have a chronic fluctuating course of the disease, but the severity of the disease differs considerably from a single or few attacks causing no permanent injury to an aggressive development with more or less permanent, severe attacks, which leads to progressive joint destruction and disability (1;2).

In RA, the inflammatory attack in the synovial membrane results in hypertrophy, oedema and hyperaemia (20). The hyperaemia is caused by vessel dilatation, decreased flow velocity and vessel proliferation by neoangiogenesis (21-23). Furthermore, an accumulation of inflammatory cells, especially macrophages, is present in the synovial membrane. The macrophages are the primary source of production of the proinflammatory cytokines (Interleukin) IL-1, IL-6 and tumour necrosis factor alpha (TNF- α). These cytokines play an important role in both initiation and maintenance of the synovial inflammation (24). The most frequently affected joints in RA are the wrist, metacarpophalangeal (MCP) and proximal interphalangeal (PIP) joints (25).

Diagnosis

There is no exclusive, pathognomonic feature, which can be used for the diagnosis of RA, while certain classification criteria have been agreed upon. In 1956 the first set of diagnostic criteria for RA were developed by the American College of Rheumatology (ACR), and in 1987 these were revised to the criteria used today (3).

The classification contains seven criteria:

1. Morning stiffness of at least one hour
2. Arthritis of at least three joint areas out of 14 possible (right or left PIP, MCP , wrist, elbow, knee, ankle and MTP joints)
3. Arthritis of PIP, MCP or wrist joints
4. Symmetric joint swelling
5. Rheumatic nodules
6. Positive rheumatoid factor (RF)
7. Radiographic changes (erosions) in hand and/or wrist joints.

In order to diagnose RA, at least four of the seven criteria must be present. Criteria one to four must be present for at least six weeks (3).

Within the last years, by consequence of the increased evidence for the benefit of early intensive treatment, it has been argued that the ACR criteria are not sensitive enough for detection of early RA and that other criteria should be applied in order to decrease the diagnostic delay and enable an immediate initiation of appropriate treatment (9; 26).

Recently, studies have emphasised that anti-CCP (cyclic citrullinated peptides) antibody tests are more sensitive than RF in the early diagnosis of RA (9; 26) and various image modalities have proven sensitive in detection of signs of inflammation and early destructions of the joints in RA (6;18).

Disease course

Both the onset and clinical presentation of the disease are heterogeneous, and it is impossible to predict if a patient will follow a moderate or a more severe disease course. However, some markers are associated with poor disease outcome (26;27). These are:

1. A large number of involved joints
2. Poor functional status
3. Presence and high titre of RF
4. Presence and high titre anti-CCP antibody
5. Low haemoglobin level
6. High platelet (thrombocyte) count
7. Elevated Erythrocyte sedimentation rate (ESR) and/or C- reactive protein (CRP)
8. Presence of bone erosions

Assessment of disease activity, progression and treatment response

In parallel to the lack of pathognomonic tests for RA, no single feature for assessment of disease activity in RA can be given. Instead, several measurements are combined for the assessment of disease activity, including both single measurements and disease indices (28-30).

The most commonly used measurements are:

1. Joint count, i.e. number of tender and swollen joints
2. Laboratory tests (CRP, ESR)
3. Qualitative assessment of pain and disease activity on a Visual Analog Scale (VAS)
4. Morning stiffness
5. Health assessment questionnaire (HAQ)
6. SF-36
7. Disease indices

Joint count

In patients with RA a joint is classified as active if tenderness on pressure or passive movement and/or soft tissue swelling along the joint margin are present (31). In the estimation of disease activity various joint scores containing different number of joints are used (32;33). Assessment of 28 joints in evaluation of disease activity is as valid as more extensive joint counts (34). However, interobserver reliability of joint counts has a moderate Intraclass Correlation Coefficient (ICC) around 0.5 (35), and the rather low interobserver reliability is only slightly improved after training (36).

Laboratory tests (CRP, ESR)

In RA the most commonly used laboratory tests are tests of acute phase proteins (CRP and ESR), RF and anti-CCP. RF and anti-CCP are mainly used as diagnostic tools (26).

Acute phase proteins are primarily produced in the liver and their production is initiated by pro-inflammatory cytokines such as IL-6, IL-1 and TNF- α , which are all produced in the inflamed synovium. Both CRP and ESR are used for monitoring disease activity (37). CRP is one of the most responsive acute phase proteins, responding to inflammatory stimuli within hours. It peaks within 1 to 2 days and returns to normal in a few days after end of inflammation (37;38). In healthy controls serum CRP is below 10 mg/l. ESR is an indirect measure of inflammation and is affected by other factors than inflammation such as haemoglobin levels, renal function, age and sex. ESR rises slowly, within days, and may not return to normal until weeks after the end of the inflammation (38). The normal value of ESR is <20mm/h. Due to these differences CRP is preferred in the assessment of disease activity in RA (37-39). Furthermore, CRP was shown to have a higher correlation than ESR with other measures of disease activity (40)

Visual Analogue Scale (VAS)

Pain is of central importance for assessment of RA and in the absence of other markers of pain; the subjective sensation is assessed by a VAS mostly indicating global pain. The VAS is also used for

assessment of self reported general health status, often filled in by both the patient and the rheumatologist (29;30).

Morning stiffness

Morning stiffness is a characteristic feature of RA, presumably caused by accumulation of inflammatory products in the soft tissues of the joint (41) and by consequence may be used as indicator of disease activity. The patient is asked for the duration of stiffness in the morning. The morning stiffness is recorded in minutes.

Health assessment questionnaire and SF-36

Self reported health assessment has been shown to have a very significant prognostic value in RA (42). Various instruments are recommended all of which are commonly filled in by patients followed in the outpatients' clinic. The patient's functional ability is assessed by a standardized health assessment questionnaire (HAQ), in which the patient has to estimate the degree of ability to perform common daily activities (43;44). Finally, the patients health related quality of life is assessed by the questionnaire SF-36 (45).

Disease indices

Various disease indices have been developed in order to assess disease activity and/or improvement and the fluctuation of the disease in patients with RA (28-30). The most commonly used measure of disease activity score based on 28-joint assessment is the DAS28, which is used both as a measure of disease activity and as a measure of improvement (29;33). Another commonly used disease index used only to assess improvement in RA is the American College of Rheumatology-Preliminary definition of improvement in rheumatoid arthritis (30).

The DAS28 is calculated by an equation containing four elements; tender and swollen joint count assessed separately, the patient's assessment of global disease activity and ESR or CRP (29;33;46). Levels of disease activity defined by the DAS28 score are: Not active (<2.6), low activity (2.6 - 3.2), and high activity (>5.1)(47). Significant improvement requires a change in score of at least 0.6 (46). In the ACR criteria improvement in joint count, patient and physician assessment of pain and global disease activity and ESR or CRP are assessed. The improvement is categorized as being either 20%, 50% or 70% improvement in symptoms from the baseline assessment for instance in response to a new treatment (30). A good correlation in improvement assessment by DAS28 and the ACR criteria has been found (47). With a DAS28 score <2.6 the patient is classified to be in remission (46). The ACR has developed a set of remission criteria as well. According to this set of remission criteria, 4 out of 6 common symptoms in RA must have been absent for at least 2 months (48). The ACR criteria for improvement correlate with the definition of remission using DAS28 (48).

Treatment of RA

Medical therapy of RA has changed dramatically with the introduction of biological drugs. The first and currently most popular medications are directed against TNF α , a cytokine of central importance for the

inflammatory process in RA (49). Even with these highly advanced medications, disease activity cannot be controlled in about 30% of the patients with RA. These non-responders may be detected clinically with the use of e.g. follow-up on DAS28 (50), or by more advanced methods e.g. imaging (6;11;15;18). Apparently, a patient may be regarded to be in clinical remission in spite of still having subclinical ongoing disease activity (16-18). For this reason, imaging might be suggested as an additional tool for monitoring of therapy. If persisting synovitis is demonstrated by imaging techniques despite clinical remission, adjustments in therapy may be called for in order to reach both clinical and paraclinical remission.

ULTRASOUND

Improved quality of ultrasound equipment within the last decade has enabled use of higher US frequency, which has increased the quality of US examinations of superficial structures. This improvement of US has made it suitable for assessment of patients with rheumatic diseases (51-54).

Two types of US are used; grey-scale US with which it is possible to detect morphological changes, e.g. synovial thickening and bone erosions (55;56) and Doppler US by which the blood flow in the synovium and surrounding tissue is demonstrated (4;10;57;58).

US transducer

The US transducer is built from piezoelectric material, which has the ability to transform electrical energy into mechanical vibrations and vice versa. This allows the transducer to serve as both an emitter and receiver (59-61). The linear array transducers contain numerous individual piezoelectric elements (between 150 and 500 elements), which increases the quality of the image. In examination of patients with rheumatic disease, high frequency linear array transducers are used, with frequencies from 7.5 to 20 MHz (53;59). The US frequency and the penetration are inversely proportional, thus the choice of transducer/examination frequency should be the highest possible frequency allowing a satisfactory depiction of the structures to be examined.

Grey-scale (B-mode)

As the main investigation topic for the four studies in this thesis is Doppler US, the principle for grey-scale will only be described briefly.

The B-mode image demonstrates the morphology of the tissues, thus morphological changes in RA such as synovial hypertrophy and erosions can be seen in the B-mode examination.

On the US monitor, the grey-scale image shows the tissues in grey tones, the various shades of grey depicting the reflective ability of the various tissues. This ability depends on the acoustics impedance; Z ; of the tissues ($Z = \text{the density} * \text{the speed of the sound}$). Every time a US beam crosses a border between two tissues with different acoustic impedance, called an acoustic interface, an echo is generated (59).

Good reflectors, as bone or tendons, are shown as white (when insonated orthogonally) and at the other

end of the scale; a homogenous structure, such as synovial fluid, without reflectors is shown as black on the monitor.

Doppler

The Doppler is added to the grey-scale image (58;59). Doppler is superimposed on the grey scale image. The Doppler registers movement in the scanned tissues and when transducer and patient are immobile, the movement of erythrocytes is registered (58-61). Doppler is used to assess the tissue perfusion, which may be increased under inflammatory conditions, and is thus used as an indirect measure of the degree of inflammation in RA (10).

Application of Doppler mode makes it possible to distinguish between thickened synovium with inflammation and thickened synovium due to previous inflammatory attacks, which will not display Doppler activity (6;62).

Doppler signal

The Doppler signal is generated on the basis of the Doppler effect. The Doppler effect is a change in frequency (wavelength) resulting from a motion of source, receiver or reflector. In US imaging, the transducer is the stationary source and receiver, and the moving source is the blood, mainly the erythrocytes (57-60). When a pulse is reflected from the moving erythrocytes, the received frequency differs from that transmitted from the transducer. This change of the frequency is known as the Doppler shift and was first described for light by the Austrian physicist Chr. A. Doppler in 1843. He described the Doppler equation:

$$f_D = f_t - f_r = f_t v \cos \Theta / c$$

f_D is the Doppler shift, f_t is the transmitted frequency, f_r is the received frequency, v the velocity of the moving source (erythrocytes), Θ is the Doppler angle (insonation angle=the angle between the sound beam and the moving source) and c is the speed of the sound. The Doppler shift is directly proportional to the transmitted frequency, the velocity of the moving source and cosine to the insonation angle.

In the US Doppler examination, two successive Doppler shifts are generated. The first one is generated when the US beam from the stationary transducer is received by the moving erythrocyte. In the second shift, the moving erythrocyte acts as the moving transmitter, and the transducer is now the stationary receiver (58). This explains the multiplication with the factor 2 in the Doppler equation:

$$f_D = f_t - f_r = 2f_t v \cos \Theta / c$$

The frequencies of Doppler shifts are within the audible range with frequencies up to 15 KHz (58;60).

Insonation angle/Doppler angle

The insonation angle is the angle between the course of the moving objects, the erythrocytes in the vessel, and the US beam. As seen from the Doppler equation, this angle is of importance as it affects the Doppler shift.

In all newer US equipment the Doppler angle is assumed to be zero, which is seldom correct, as the blood most frequently will move in some angle (58). If the velocity of the flow has to be measured, it is important to adjust the insonation angle in order to get a reliable value (61;63). In small vessels, as in the synovial tissues, angle adjustment is not possible because the course of the vessel is not detectable, however, it is of no importance as it is the amount of flow and not the velocity which is of interest in rheumatological US (58).

Spectral Doppler

The Doppler shift within a region of interest on a single line (Doppler gate) is continuously recorded displaying a Doppler spectrum. The Doppler spectrum is displayed as a spectral curve. The spectral curve shows the flow profile in the measured area (Doppler gate). The spectral curve plots the Doppler shift (y-axis) against time (x-axis) (57;60;64). In older equipment the y-axis will be kHz whereas in newer equipment it will be m/s because the machine calculates a velocity assuming 0 degrees insonation. The spectral curve demonstrates abnormal flow; in the rheumatological area inflammatory flow is characterized by sustained blood flow in the diastole (64).

Colour Doppler and Power Doppler (energy Doppler)

In Colour Doppler and Power Doppler, real time presentation of flow is depicted. The flow analysis is performed in the colour box, which is a framed area superimposed on the grey-scale image. The colour box is divided into numerous cells/colour pixels. In each of the cells, an analysis of the Doppler spectrum is performed. In Colour Doppler, the mean frequency shift in each cell is computed and displayed as a colour on the monitor (65). The colours that arise from the Doppler shift indicate direction of the measured flow. In general red colours indicate flow directed towards the transducer and blue colours flow away from the transducer. Different hues of red or blue indicates differences in velocity. Lighter hues are used to display higher frequency shifts (58). In Power Doppler the power of the Doppler shifts is displayed (65). The power is the sum of the amplitudes of the different Doppler shifts in each cell. Light colours indicate high power (57;66). In Power Doppler no direction or velocity of flow is measured. As Power Doppler does not measure velocity, but adds the power of the obtained Doppler shifts, it is theoretically more sensitive to low flow than is Colour Doppler (57;58;66).

Choice of Doppler

The theoretical advantage of Power Doppler in the detection of low velocity flow indicates that Power Doppler is more suitable in rheumatology than Colour Doppler. However, superior sensitivity of the Power Doppler is only true in older US equipment; in newer high-end machines the trend is that Colour Doppler is now more sensitive to low flow than Power Doppler. A satisfactory explanation for this

change in sensitivity has not yet been put forward (58). Thus, the choice of Doppler mode depends on the type of equipment available.

Doppler parameters/setting

The most important issue in the adjustment of Doppler parameters in rheumatology is that the Doppler should be as sensitive as possible for low velocity flow. This will make it sensitive for both low and high velocity flow – all flow. The Doppler parameters of particular importance in rheumatological US are listed below.

Doppler frequency

A low Doppler frequency allows high penetration making assessment of flow in deeper structures possible. Low Doppler frequency will make the Doppler image more grainy; however, the resolution of the Doppler image is not of great importance. The setting of the Doppler frequency is a balance between the enhanced penetration with lower frequencies and on the other hand the enhancement due to the increased Doppler shift when applying higher frequencies. Which frequency produces the most sensitive Doppler images varies with different US equipments (58).

Colour box

The colour box is the area in which the Doppler analysis is made, so the larger the box is, the bigger is the area in which the Doppler analysis is made. The size of the box affects the frame rate; the larger the size of the box the more reduction of the frame rate. Therefore, in general it could be recommended only to include the area under assessment in the box. However, to be aware of possible sources of reverberation artefacts, it is important always to include the upper limit of the image in the box (58).

Pulse repetition frequency

Pulse repetition frequency (PRF) is the Doppler sampling frequency of the transducer. The Doppler shift is determined indirectly by comparing the distance of the wave tops between the transmitted and received sound in a number of pulses.

The Doppler sensitivity of both Colour and Power Doppler is affected by PRF adjustment (58). It could be assumed that the higher PRF, the more sensitive the machine would be to detect any flow in the tissues. However, in most Doppler systems there is a linked control between PRF and sensitivity to low flow as it is assumed that when the PRF is set high, the investigator is not interested in low velocity flow, and filters are applied removing noise as well as erasing signals from low velocity flow. In rheumatology the flow investigated is primary low velocity flow, and sensitivity to this flow is therefore important and by consequence the PRF applied should be as low as possible.

Gain

The gain setting determines the system's sensitivity to noise. If the gain setting is too low, low amplitude flow information may be overlooked, but on the other hand a too high Doppler gain setting will result in random noise (63;66;67). The optimal Doppler gain is made by turning up the gain until random noise is encountered and then lowering the gain to the point where the noise disappears (58;66;67).

Focus

The Doppler focus setting is very important as a Doppler focus placed incorrectly may lead to a misleading change in flow information. The Doppler focus should be focused in the very middle of the region of interest (58).

Colour priority (threshold)

As Doppler information is added to the grey-scale image, the machine must choose if grey-scale or Doppler information should be expressed in the various areas of the screen. What modality will be expressed depends on the setting of the colour priority/threshold. In rheumatology, the threshold function should be set in favour of the Doppler according to the requirements for sensitivity to low flow (58).

Wall filters

All Doppler instruments have filters which eliminate the lowest Doppler shift in order to avoid random noise. However, in rheumatology where low velocity flow is examined, the wall filters should be adjusted as low as possible in order not to overlook any flow (58). As mentioned, in most US equipment PRF and wall filter settings are linked, thus low wall filter is linked to low PRF setting (58).

Persistence-frame averaging

In frame averaging, colour information in a number of frames is averaged making the flow profile smoother. However, in this way the dynamic nature of the flow is compromised. Thus, when the nature of the flow has to be assessed the persistence setting should be low (58).

Artefacts

In imaging, artefact means components of the image generated in the image process which is not properly indicative for the structures under investigation (68). Many artefacts results from errors in the scanning technique and machine settings and can be avoided once the operator is aware of the artefacts.

Random noise

Random noise is produced in all electrical circuits. The artefacts produced by random noise are seen as colour foci appearing randomly in the image, opposite to true flow which is seen constantly in the same area (58).

Motion

Movement of the patient, the investigator, the scanned tissue or vessel wall during Doppler imaging causes low Doppler shifts appearing as random flashes at the screen (58;63;68). When the Doppler is adjusted to be sensitive to low flow, motion artefacts cannot be avoided completely, but minimized if both the patient and investigator are relaxed during the Doppler examination (58)

Mirror

Mirror artefacts appear in the presence of highly reflective surfaces. In rheumatology these reflectors are nearly always bone surfaces (58). The bone surface will act as a mirror showing a reflection of the true vessel below the highly reflective bone surface (58;63).

Blooming

Blooming artefact is the phenomenon that colour is seen outside the vessel walls making the vessels appear larger than they really are. The phenomenon is dependent on the gain setting (58;68). As the gain in rheumatology is adjusted to be sensitive to low velocity flow, blooming must be accepted as a systematic error (58).

Reverberation

Reverberations occur when sound is repeatedly reflected and bounces back and forth between parallel interfaces. In US examination the reverberation appear as reflections between the transducer and an interface or between various interfaces. The reverberation can be both simple and complex. Simple reverberation in Doppler imaging is seen if the vessel is repeated lower in the image. Complex reverberations produce a showering of colours behind the vessel. The reverberation echoes decrease with distance from the reflective interface (58).

Aliasing

Aliasing is related to the PRF setting which is the Doppler sampling frequency of the transducer. Aliasing artefact is an incorrect display of the direction and velocity of the blood flow, due to undersampling of the re-reflected Doppler signal. If the PRF is too low compared to the velocity of the erythrocytes, aliasing will occur. The maximal Doppler shift frequency that can be sampled without aliasing is half the PRF ($PRF/2 = \text{Nyquist limit}$) (58;60). As direction and velocity of the flow in the vessels in the synovial tissues is of no interest in a rheumatologic setting, aliasing is not an issue in rheumatological US. Furthermore, as mentioned earlier, the wall filter and the PRF settings are linked controls, thus high PRF results in proportionally high wall filter which is a drawback because high wall filter decreases the sensitivity to low flow (58)

Pressure and positions

It is important to avoid mechanical pressure when a Doppler examination is performed, as pressure may diminish the flow by compression of the vessels. Pressure is avoided by application of a generous amount of gel and comfortable positions of both the patient and the investigator (58).

ASSESSMENT OF PERFUSION BY DOPPLER US

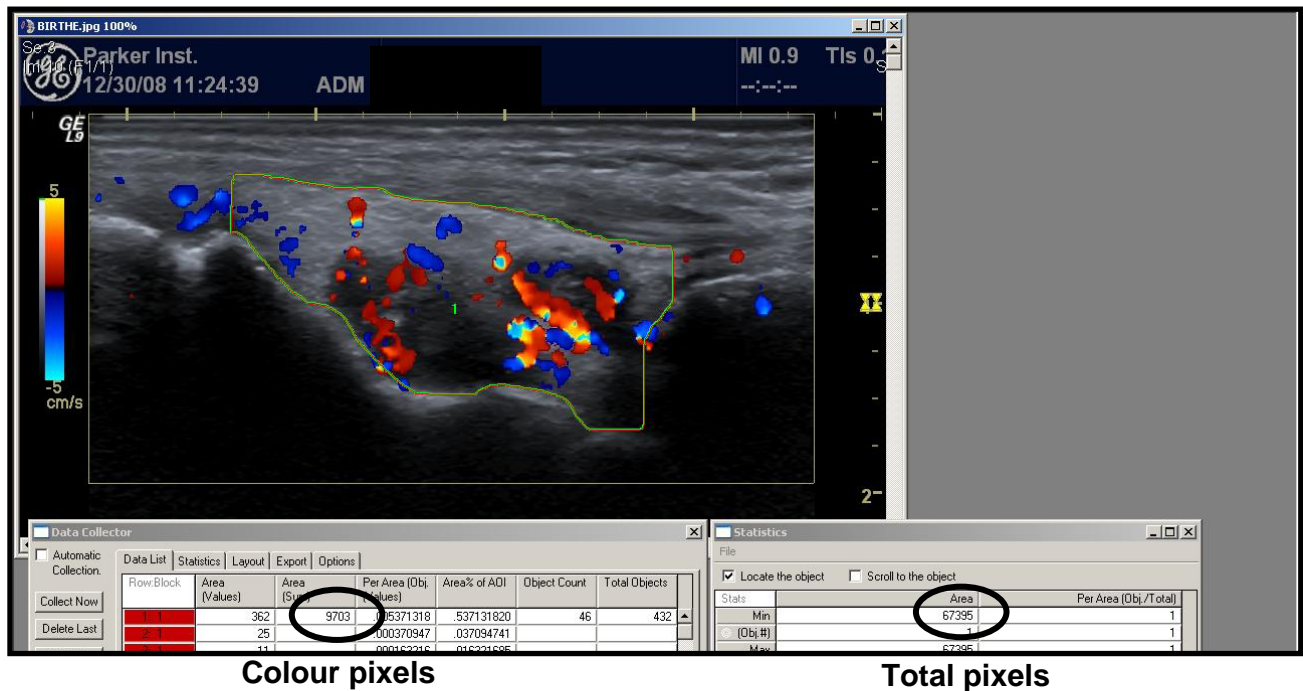
The principle of assessment of disease activity in RA by Doppler US is that Doppler displays blood flow in the tissues and since increased blood flow is part of the inflammatory response the amount of Doppler activity can be an indirect measure of inflammation (65;69).

Various methods and scoring systems have been introduced to assess Doppler activity in RA. Thus both quantitative and semi-quantitative scoring systems have been applied (6;11;12;14;70-73). The underlying principle in all scoring systems is to evaluate the amount of colour pixels in relation to the total amount of pixels in a specific region of interest (ROI). In the semi-quantitative systems this evaluation relies on a

subjective assessment of the percentage of colour pixels in ROI. The percentage of colour is scored on an ordinate scale typically containing 4 levels (0-3). In the quantitative systems the amount of Doppler activity is estimated on a computer where an exact pixel count is made. On the basis of this count the exact percentage of colour pixels is calculated. In a quantitative scale the Doppler activity is measured on a ratio-interval scale.

There are some obvious problems in the use of a semi-quantitative scoring system. First the subjective evaluation of the amount of colour pixels; it is difficult to estimate the percentage of colour just by looking. This difficulty is underlined by the fact that when we assess the amount of colour quantitatively, we find that only a modest number of joints display more than 50% of colour pixels, which is often the criterion required for a score of 3. This finding is not in accordance with the relatively high number of joints scored with a 3 in many studies using semi-quantitative scores (12;13;73). Another problem concerning assessment with a score on an ordinate scale is that no uniform definition between the different grades exists, and that the distance between the various grades is heterogeneous. Thus an increase in score from 2 to 3 is not comparable to one from 1 to 2. This adds up to the fact that scores from different studies are difficult if not impossible to compare. The moderate correlation in Doppler assessment among experts in US underlines this difficulty (74). The reliability of quantitative Doppler assessment between US experts has not been tested by now. One last difficulty with the semi-quantitative scores is that only an improvement or worsening of a certain degree can be registered because smaller changes may result in unchanged grade. In contrast, all changes of Doppler activity will be registered on a quantitative scoring scale.

In all studies in this thesis a quantitative scoring system is applied. The percentage of colour pixels is expressed by the Colour Fraction (CF)(Colour Fraction =colour pixels/total number of pixels*100) (see figure 1).



Colour pixels

Total pixels

Colour Fraction $9703/67395 \times 100 = 14.39\%$

Figure 1: An US image of the central position of the wrist displaying Doppler activity. The ROI is framed and the sum of both total number of pixels and colour pixels are shown and the calculation of the Colour Fraction (CF) is shown.

MEASUREMENT VALIDITY

The validity of a measurement concerns the extent to which it measures what it is supposed to measure (75). The validity of a measurement is not all or nothing, but may be graded, for instance as the degree to which it correlates with other measurements assessing the same phenomenon, or its ability to predict future events (75-77).

As there are no universal measurements for the diagnosis of RA and no single test to assess the level of disease activity or disease progression, the assessment of RA is based on a more or less subjective grading of various measurements, thus validation of measurements in RA is highly relevant (76;78;79). Validity assessment is divided into three categories; Face/content validity; criterion validity and construct validity. It is important to emphasise that the different categories of validity are mutually related and are partly overlapping and that the definitions of the different types of validity are not totally stringent. Thus, in various definitions of validity, the three groups do not contain exactly the same assessment parameters, which may partly be due to the blurred character of the validity terms. However, despite the inconsistency in the various validity criteria, there seems to be total agreement as to which aspects are of importance when applying and assessing a new measurement in the rheumatological area, (75-77;79). The aspects required to evaluate a given measurement are shown in figure 2.

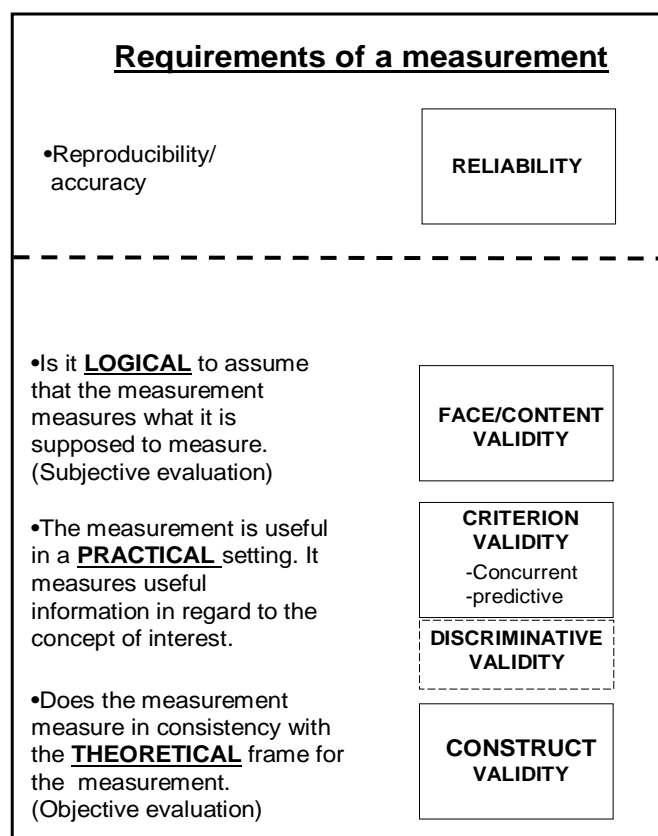


Figure 2: The various aspects which have to be investigated in the assessment of a new measurement. The left column describes the various requirements in words. The right column shows the different categories of validity.

There is total agreement that good validity of a measurement always implies that the measurement is reliable. The reliability of a measurement refers to the ability to obtain the same results when the same condition is assessed in a comparable context, i.e. to what extent the measure is free of random error (75;79). There are various ways to evaluate reliability of a measurement. The *intraobserver reliability* assesses an investigator's ability to obtain the same result if the same phenomenon is assessed several times, and *interobserver reliability* measures two investigators' ability to obtain the same result if they assess the same phenomenon. *Test-retest reliability* refers to whether the same results can be obtained when the same phenomenon is repeatedly investigated in exactly the same setting. In contrast to intra/interobserver reliability, test-retest reliability assesses a more complex situation where e.g. both image acquisition and image analysis are necessary for the diagnostic result in assessment of a US examination.

Face and content validity

Face validity refers to whether it is plausible to assume that a measurement measures what it is supposed to measure. Face validity is essentially subjectively based, mainly on an educational and experience level. Content validity is based on evidence as well, but this validity term assesses if the measurement exposes all important aspects of the condition assessed (75;76;80). One of the main symptoms in RA is inflammation of the synovial tissue, and a cardinal aspect of inflammation is increased blood flow

(10;20;81). Furthermore, angiogenesis is part of the pathogenesis of RA (22;23). Thus it is reasonable to assume that US Doppler reflects (65) disease activity in patients with RA. These conditions indicate that the US Doppler measurements have face validity in assessment of RA. In addition, B-mode US depicts synovial hypertrophy and to some extent erosions, meaning that content validity of the combination of these US measurements in RA is present as well.

Criterion validity

Criterion validity is the practical and objective term of validity (75;76;80). First, part of criterion validity refers to the correlation of a measurement with an estimated test or a “gold standard” for assessment of a certain condition. This is called concurrent validity. When a single criterion or "gold standard" for a condition, as for instance in RA, has not been established, then a series of validated measurements must be used as the criterion in order to interpret the validity/usefulness of new measurements. To test the concurrent validity of a newly introduced method, you often need to know the degree of concurrent validity with several established measurements. An integral part of the assessment of criterion validity is the testing of accuracy (75;79). If a measurement is inaccurate, no validity can be obtained. The accuracy can also be defined as the measurement precision, or to what extent it is affected by both systematic and random errors (79). The accuracy is closely related to reliability.

The second part of criterion validity is predictive validity. Predictive validity assesses to what extent a measurement is able to predict future events, for instance disease progression (75). In some validity definitions predictive validity is termed discriminative validity (76;79). In this thesis the term predictive validity is used. A subterm of predictive validity is prescriptive validity, which assesses the ability of measurement to predict which patients will benefit from a given treatment or intervention (75).

Construct validity

Construct validity is closely related to content validity as it estimates the ability of a measurement to measure what it theoretically is supposed to measure (75;76;79;80). Thus, in construct validity it is actually tested if the hypothesis underlying the content validity can be verified scientifically. This can be done by looking into whether the measurement changes as it may be expected, when assessing for instance response to a treatment with a known treatment effect. In comparison with concurrent validity, construct validity is not merely correlation assessment, but tests if the theory underlying the measurement can be verified (75;76). In the rheumatological area, the definition of construct validity and concurrent validity is blurred, because in some definitions correlation with established measurements it is defined as construct validity (76;79), and in other definitions construct validity must be more strictly theoretically founded (75;77). In this thesis, correlation with other measurements of disease activity is termed concurrent validity.

OMERACT (Outcome Measures in Rheumatology)

The organisation/forum for Outcome Measures in Rheumatology (OMERACT) was established in 1992. OMERACT is an organization including various interest groups who represent different areas of rheumatology. The participants in the interest groups are appointed experts in various areas of rheumatology, and the groups work out frameworks for the use of measurements in their specific fields of interest. The various groups are organised either in relation to a certain disease or to various measuring tools, i.e. US or Magnetic resonance imaging (MRI) (78). All decisions and recommendations made within the OMERACT setting are created in consensus among the experts in the interest group (78). The OMERACT definitions and recommendations are not evidence-based, but rather reflect how the majority of performers in the field understand the interaction between the disease and the tests.

As a result of discussions in the OMERACT forum, a set of evaluation tools for the assessment and applicability of measurements were worked out. These evaluation tools are drawn up in the “OMERACT filter” (78). The OMERACT filter divides the applicability of a measurement into three areas:

Truth: does the measurement really measure what it is assumed to do. I.e. is the measurement valid.

Discrimination: can the measurement discriminate between various groups, and can it be used to measure changes over time in a situation of interest. This item concerns reliability, predictive validity and sensitivity to change.

Feasibility: how are the practical applications of the measurement; costs, accessibility and difficulty in performance and evaluation.

In 2003, a Special Interest Group in musculoskeletal US was established in the context of OMERACT. The first focus of this group has been to evaluate and optimize the use of US assessment in inflammatory arthritis including RA (82). Since the establishment of the group, various initiatives concerning the use of US in RA have been made. An attempt to make definitions of US-based pathology in inflammatory conditions has been made (5), and investigation of reliability in examination techniques and assessment of pathology among US experts has been performed (83;84).

The four studies in this thesis investigate various parts of the applicability of Doppler US in the assessment of patients with RA. The four studies are classified according to the OMERACT filter in order to make an overview over which components of the applicability were investigated. The classification is shown in figure 3.

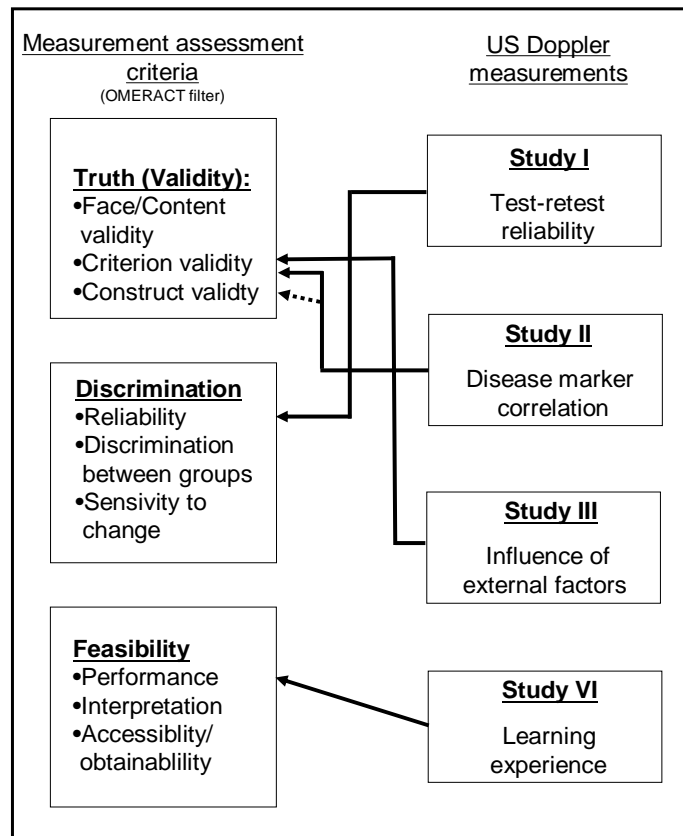


Figure 3: The four studies depicted according to the three OMERACT filter categories for assessment of a measurement.

RELIABILITY OF DOPPLER MEASUREMENTS IN RA

Study I tested the test-retest reliability of the Doppler measurement. Evaluation of the reliability of a measurement is associated to the second part of the OMERACT filter (78) (figure 3). The second part of the OMERACT filter assesses the ability of a measurement to discriminate between situations of interest. In relation to reliability, this means whether a measurement is sensitive. However, there is an obvious problem by placing the reliability in this part of the filter, as no measurement can be validated if the data cannot be collected in a precise manner. Thus a precondition for all measurements is reliability.

The reliability of US Doppler measurements in patients with RA has only been investigated to a modest extent. The assessment of reliability in US includes both the image acquisition and image analysis phase (85). However, most of the reliability studies conducted have only studied reliability of analysis of static US images (13;70;72-74;86). Beside the study in this thesis, the test-retest reliability is investigated in only three other studies (12;71;87).

The Doppler reliability testing can be divided into assessment of either Colour or Power Doppler, and into reliability of quantitative or semi-quantitative scoring systems and examination of various joint types. Taking into consideration the sparse number of studies investigating the reliability, all types of reliability studies will be evaluated in the following.

Study I - Test-retest reliability

(See appendix 1 for details)

The reliability study in this thesis assessed the reliability of both image acquisition and the subsequent image assessment technique of colour Doppler activity in the wrist of patients with RA. We did so by evaluating two image acquisition techniques combined with two different image analysis methods (91). In both image acquisition techniques the selection was guided by specific landmarks in the grey-scale image. In the first technique the landmarks were identified, whereupon the Doppler was switched on and a cine-loop was stored. In the second the landmarks were identified and the Doppler switched, but in this technique the transducer was moved while keeping the landmarks in the image. When the area with most pronounced Doppler activity was identified, a cine-loop was stored. Both acquisition techniques were repeated after 30 minutes. In both techniques the two images with minimum and maximum Doppler activity were stored for analysis, in order to assess Doppler in both the systole and diastole. In the two analysis techniques the Colour Fraction (CF) was calculated in a ROI defined by either the synovial tissue, or by specific anatomic landmarks surrounding the synovium.

The best test-retest reliability was obtained when both landmarks in the grey-scale image and the amount of Doppler activity were used in the image acquisition, and the Colour Fraction (CF) was calculated in a ROI framed by anatomic landmarks. With this combination the intraobserver ICC was 0.95, which indicates an excellent reliability (89). Subsequently we assessed the interobserver reliability in image evaluation for this combination of image acquisition and analysis technique in a limited number of patients, and in accordance with the intraobserver reliability this was excellent with an ICC of 0.97 (these data are not presented in study I). The superiority of the chosen combination is not surprising: first, the Doppler activity is not uniform and the ultrasound image obtained is only approximately 1 mm thick, and by consequence the activity may change with only a little movement of the transducer. By selecting the site with most pronounced activity it was possible to reselect the same site at repetition. Second, the borders of the synovial tissue are often difficult to define, which is likely to lead to variation in the measured size of the synovial ROI between exams. This is reflected by the rather poor ICC value of 0.48, when the assessment was based on only the grey-scale image in the acquisition and synovium tissue as ROI

At the moment we seem to be the only centre using specific landmarks to guide image acquisition. In the US examination guidelines from the European league against rheumatism (EULAR) US group, which is referred to as the examination technique used in most studies, positions are described but without defining landmarks (51).

Three studies have investigated the test-retest reliability of Doppler measurements by Power Doppler in patients with RA (12;71;87), in all three studies a semi-quantitative scoring was applied. In two of the studies, the EULAR examination guidelines were used (12;87). In the third study, no description of image

acquisition technique was described (71). In two studies, a trained and an untrained investigator examined the same patient (71;87). In these studies the results of the reliability testing were an ICC of 0.72 (good reliability (89)) and a kappa of 0.87 (lower limit of excellent reliability (90)), respectively. In the third study, all examinations were carried out by two experienced examiners, and in this study the Kappa values of the Doppler assessment were 0.90 and 0.95 (12). The better reliability in our study (91) and in the study of Filippucci et al. (12) may be explained by the long scanning experience of the investigators. However, the other image acquisition and analysis procedures assessed in our study showed poorer reliability with ICC values between 0.48 and 0.82(91). This, together with the poorer reliability in two of the other studies (71;87), indicates that the more standardized an examination procedure, the better the reliability. This is not in agreement with the results of Filippucci et al. (12), however a more elaborate analysis should include a head-to-head comparison between the two different scoring principles, which remains to be performed. Until otherwise stated, we would expect our quantitative analysis to be of more scientific value in distinguishing small differences, as it is more sensitive than a four point score.

In our study (91), the wrist joint was examined, which is a much more complicated joint than the simple MCP and PIP joints, which were mainly examined in the other test-retest studies (12;71;87). That it is easier to assess simple joints is supported by the results from reliability studies of static images. In the studies investigating images of simple joints (MTP, PIP and MTP joints), higher kappa values were obtained (mean 0.85) in comparison to the studies investigating the wrist joint (mean 0.69). The fact that we achieved excellent reliability when we used the strictly standardized examination procedure, despite using the wrist as target joint, may further indicate the improved reliability of a standardized method.

Reliability of analysis of static images

Five of the studies with analysis on static images used Power Doppler images and a semi-quantitative scoring system (13;72-74;86). Only one assessed a quantitative scoring system on Colour Doppler (70). Surprisingly, the reliability on static images is not superior to the ones in the test-retest studies. In the Power Doppler studies, the kappa and ICC values span from 0.53 to 0.9 (mean 0.63) for interobserver reliability and from 0.72 to 0.98 (mean 0.85) for intraobserver. In the Colour Doppler study where the amount of Colour Doppler was estimated by the Colour Fraction (CF), the inter- and intraobserver reliability had Intraclass Correlation Coefficient (ICC) values of 0.81 and from 0.82 to 0.97, respectively. With only one study of a quantitative scoring system (70), no conclusions can be made in relation to which type of scoring system has the best reliability when evaluating stored images.

In the studies by Koski et al., the reliability was tested between 17 physician sonographers from various European countries (74). The physicians investigated 99 video clips with both healthy and pathological conditions. The clips were examined twice with an interval of three to four months. In this study a semiquantitative scoring system was applied. The intraobserver reliability was moderate (kappa 0.72) and the interobserver reliability fair (kappa 0.53).

The results from this study strongly emphasize that if it is to be possible to perform valid multicentre studies, much work needs to be done concerning development of standardized examination procedures. This need is further emphasized by the fact that comparable kappa values were obtained, when US experts assessed the reliability of grey-scale examinations (83;84).

Doppler measurements during the heart cycle

In study I (91) we assessed the Doppler activity in both the systole and the end diastole, and calculated the average in order to depict the flow during the heart cycle. By measuring both the systolic and diastolic Doppler activity, we wanted to investigate variation throughout the heart cycle. We did so because we had the subjective impression that patients display various flow patterns, and that changes over time may be due to change in various parts of the heart cycle. We found a difference in the Colour Fraction (CF) in the systole and diastole of around 50% (91). There is very little information of changes in flow pattern during treatment of RA, while preliminary data might suggest a recurrence of systolic flow, while not of the diastolic component (11). In a study by Teh et al. they measured the Doppler activity throughout three heart cycles. They showed that patients display various flow patterns (14). These results underline that it is important to measure in a well defined part of the heart cycle, especially in prospective studies with repeated Doppler examinations. However, a more elaborate description of evolution in the flow pattern changes depending on treatment of RA still needs to be investigated in longitudinal studies.

Scoring technique

Only one study has investigated the reliability of various Doppler evaluation techniques applied on a group of patients with RA (87). In this study, both a quantitative and a semi-quantitative scoring system were tested in both 2D and 3D US images. Similar levels of correlation were seen in all evaluation techniques with kappa and ICC values between 0.8 and 0.88. The quantitative pixel count score applied in the study of Albrecht et al. (87) is not completely comparable to the technique we investigated. Firstly, the scanning positions are not well defined, and secondly, the ROI was defined with a fixed frame setting in contrast to our ROI which differs from image to image as we defined ROI in each image as either the borders of the synovial tissue or by some specific anatomic landmarks. This may explain that we had a better ICC, 0.95 compared to 0.8. Furthermore, we obtained a much higher correlation between the pixel count (Colour Fraction) and CRP (92). We obtained an r-value of 0.5 compared to $r=0.24$ in the study by Albrecht et al. (87).

Summary

In general, all studies investigating the reliability of Doppler measurements found a good reliability, regardless of the use of Doppler technique and scoring system. Most studies obtained Doppler measurements in the MCP joints. In all studies conducted at the same clinic, Kappa and ICC values were either in the good or excellent category with values above 0.70 (12;13;70-73;86-87) However, the study which investigated the reliability of Doppler images among 17 experts in different centres using

musculoskeletal US (74) showed only moderate interobserver reliability. This strongly suggests that more elaborately developed standards are required before results may be compared from centre to centre.

A better reliability was obtained in test-retest studies where the US examinations were performed by experienced sonographers (12;91). Interestingly, reliability in test-retest studies had as good reliability as in the assessment of still images, although one might have expected a better reliability in evaluation of still images, since the variation in image acquisition is excluded.

The majority of reliability studies are performed with Power Doppler applying a semi-quantitative scoring system, thus more studies assessing the reliability of Colour Doppler and a quantitative scoring are needed, before any conclusions regarding reliability of various Doppler modality and scoring system can be made.

VALIDITY OF DOPPLER MEASUREMENTS

Study II and III investigate the criterion validity, i.e. concurrent validity and accuracy, of Doppler measurements in RA (figure 3)

Evaluation of the validity of a measurement is linked to the first part of the OMERACT filter (78), which is about the assessment of the *truth* of a measurement. Assessment of the truth concerns the investigation of the various types of validity i.e. does the theoretical frame for the measurement seem reasonable (content validity) and does the measurement e.g. follow degrees of inflammation - can the theory be verified scientifically (construct validity). In addition, does the measurement correlate with other measurements assessing the same phenomenon (criterion validity). The latter two validities may be subject to scientific investigations, while the first may be argued based on common knowledge of disease mechanisms and technical insight in the measurement.

The most intensively investigated validity in RA is the correlation to other measurements of disease activity; concurrent validity. The concurrent validity has been investigated in comparison to so-called “gold standards”. A “gold standard” is a measurement which is assumed to measure the true condition in the phenomena investigated. Furthermore, the concurrent validity has been investigated in relation to other commonly used measurements of disease activity in RA. In some studies “gold standards” in the assessment of Doppler US as a measurement of synovial blood flow has been either contrast-enhanced MRI or histopathological analyses of synovial tissue from RA patients (93-97). However, the use of both measurements as “gold standard” is associated with some difficulty. First it is unknown if the degree of increased blood flow is associated with severity of the disease. This problem also concerns the US Doppler measurement even though studies suggest that Doppler activity is associated with the development of bone erosion (6;18). Second, no consensus exists as to which method should be applied in both MRI and histopathological assessment of perfusion (93-97) and finally both measurements are sparsely investigated in a limited number of centres. Since there are no specific measures of disease

activity in RA another possibility is that e.g. DAS28, which is a disease activity score including both clinical and laboratory abnormalities (see page 12), could serve as a gold standard.

It should be noted that some investigators have categorized correlation with measurements of disease activity as construct validity (76;79).

The accuracy is very sparsely investigated. Besides the study in this thesis only two other studies with a limited number of patients have investigated this aspect of validity (98-99).

Study II - Concurrent validity

(See appendix 2 for details)

In study II the concurrent validity was investigated by correlating commonly used measurements of disease activity in RA with the Doppler activity (measured as Colour Fraction) in the wrist joint with most pronounced Doppler activity. The measurements used in correlation with the Doppler activity were: tender and swollen joint counts, VAS, ESR, CRP and DAS28.

All data including US was obtained at baseline, i.e. when the patient initiated treatment with an anti-TNF- α drug, and the RA may be assumed to have been in an active state at this time. The Colour Fraction (CF) correlated significantly with the following disease measurements: swollen joint count, ESR, CRP and DAS28 (figure 4).

In addition to the theoretical content validity, this study demonstrated moderate significant concurrent validity. That it is only moderate is not discouraging since we do not regard the other measurements of disease activity to be precise. Albeit moderate, a significant correlation between the Colour Fraction (CF) and the accepted standards may be regarded as an indicator of reasonable concurrent validity and suggest that the Colour Fraction (CF) should be tested in a more longitudinal setting against robust final outcomes in RA as e.g. erosions. Although this remains to be studied, preliminary data indicate that persistent Doppler activity may – in spite of absent clinical signs of disease activity – be a predictor of erosions in RA (6;18).

Study III - Accuracy

(See appendix 3 for details)

Before accepting Colour Fraction (CF) as a measurement of disease activity, a possible influence of non-disease factors on the result have to be clarified. As Colour Fraction (CF) measures perfusion as indicator of the disease, some obvious confounders might be changes in perfusion caused by temperature or activity. These external factors were tested systematically in study III. We evaluated the change in Colour Fraction (CF) in the wrist joint following five interventions; two heating, two cooling and one exercise intervention. In the first heating intervention the hand was placed between two warm packings until a skin temperature of 37°C was reached. In the second, the hand was placed in a paraffine bath (50°C) for 15 minutes. In the first cooling intervention the hand was cooled down to 22°C between two bags of crushed ice and in the second it was placed in a bucket of crushed ice for 10 minutes. In the exercise intervention,

five times of maximal grip strengths were performed. The interventions were carried out in random order on three different days.

Before and immediately after each intervention a US examination was performed. The change in Colour Fraction (CF) between the two examinations was determined to see if the various interventions affected the amount of Doppler activity. The only intervention which affected the Doppler activity in a systematic way was the cooling of the skin. In the intervention a modest decrease of the Doppler activity was seen. However, the decrease was too modest to be of clinical relevance, i.e. there is no need for special considerations regarding skin temperature of the examination area.

Assessment of concurrent validity

The “gold standards” in the assessment of Doppler US as a measurement of synovial blood flow has been either contrast-enhanced MRI or histopathological analyses of synovial tissue from RA patients with an active disease. The concurrent validity with contrast-enhanced MRI have been investigated both between Colour Fraction (CF) and area of contrast-enhanced synovium and between presence of Power Doppler signal and speed of enhancement in the synovium (93;94). Only a moderate agreement was seen when correlation between presence of contrast-enhancement and presence of Doppler was estimated, with a kappa value of 0.45. In joints with Power Doppler signal, the speed of enhancement was increased compared to joints with no Power Doppler signal, thus the specificity and sensitivity of fast enhancement ($\geq 1\%/sec$) were 88.8 and 97.9% (94). When comparing the area of contrast-enhanced synovium with the Colour Fraction (CF), a slightly better correlation was seen with a significant correlation coefficient of 0.59 (93).

Three studies have investigated the correlation between Doppler activity and histopathological findings (95-97). In all studies there was a positive relation between Doppler signals in the synovium and histopathological findings. The relation was only modest and not significant in one study (97). In another study a positive relation between a semi-quantitative Doppler score and a broad definition of pannus was seen. In this study no correlation was calculated (95). In a third study significant correlation between Power Doppler and histopathological parameters of inflammation was shown with correlation coefficients between 0.81 and 0.89 (96).

In conclusion, when assessing the concurrent validity of Doppler UL in comparison to “gold standards” an ambiguous picture appears. This may partly be explained by the above described difficulties with the applied “gold standards”. Thus, to get additional evaluation of the correlative (concurrent) validity of US measurements, correlation with other commonly used measurements of disease activity in RA is necessary. Actually it could be argued that these measurements are as valid as “gold standard” as MRI and histopathological evaluation.

Consistency in the changes in Doppler US and other measurements of disease activity is demonstrated in many studies (10-15;100-102). However, correlation is only determined in a smaller number of studies (10:62;70;92;103-105).

As mentioned in study II we investigated the concurrent validity of Doppler measurements (Colour Fraction) in a single wrist joint of patients with RA (92). In the other studies investigating the concurrent validity, the Doppler activity is correlated to various number of different disease activity measurements. Thus the Doppler measurements are correlated to tender and swollen joint counts (10;92;103;105), CRP(62;92;103-105), ESR (10;70;92;103-105), VAS (70;92;103), DAS28 (62;92;103)and HAQ(62;70;104)(se figure 4).

The correlation between a quantitative Doppler score (Colour Fraction) and other measures of disease activity is only sparsely investigated (10;70;92). All other correlation studies used semi-quantitative Doppler scoring systems. The number of joints assessed to calculate the degree of Doppler activity varies considerably between all studies. Thus, no direct comparison between results in the various studies can be made, only a general view may be applied.

In general, the correlation coefficients in all studies are moderate or even low. The highest r-values are around 0.5. Study II in this thesis is the only study, which assessed the correlation with a single joint (wrist joint) and interestingly, we found correlations comparable to more comprehensive joint counts, thus no relation between the number of joints assessed with US and the correlation with other measurements of disease activity is proven. The correlation with biochemical markers (CRP and ESR) is similar or even higher when a single target joint is evaluated compared to a comprehensive joint count (10;92). In our study (92) we found higher correlation with CRP ($r=0.5$) than compared to the only other correlation study also investigating RA patients starting up a new treatment (62). In this study (62) they made a 28 US joint count of the joints used for calculation of DAS28. Their correlation to DAS28 was higher than in our study. We also found acceptable correlations with ESR ($r=0.5$), which was equal or above the correlation in other studies. We showed significant but low correlations with DAS28 and swollen joint count. Our results of the correlation with US measurements in a single joint may increase the feasibility (OMERACT filter) as the time needed to perform a US examination will be less making the examination technique more suitable for daily clinical praxis.

The moderate or low correlation to other disease markers may be explained by the relatively inconsistent disease parameters, which are present in various degrees in patients with RA and in addition not always are in accordance with the severity of the disease.

The limited utility of the separate measurements of disease activity is emphasized by the fact that disease indices are widely used in assessment of both disease activity and improvement (29;30;48).

When we made the correlation analysis in study II, we calculated the correlation between all the measured disease markers, and the only r-value comparable to these between Colour Fraction (CF) and ESR and CRP was the correlation between tender and swollen joint count ($r=0.43$). All other correlations were statistically insignificant and below 0.25.

The insensitivity of the disease measurements is emphasized by results from newer studies showing Doppler activity and sustained progression in joint destruction in patients assessed as being in clinical

remission by DAS28 and the ACR criteria (16-18). This indicates that it could be relevant to include Doppler US information in the assessment of disease activity.

	ESR	CRP	VAS	Swollen Joint count	Tender Joint count	DAS28	HAQ
Ellegaard(92)	r=0.5 p<0.001*	r=0.5 p<0.001*	r= -0.11 p=0.26	r=0.35 p<0.001*	r=0.07 p=0.64	r=0.29 p<0.001*	NM
Naredo* (105)	r=0.5 p<0.001	r=0.6 p<0.001	NM	r=0.49 p<0.001	NM	NM	NM
Hameed ⁺ (103)	r=0.37 p=0.007	r=0.23 p=0.06	r=NC	r=NC	r=0.26 p=NS	r=0.32 p=0.03	NM
Kiris (104)	r=NC	r=NC	NM	NM	NM	NM	r=0.46 p<0.05
Naredo# (62)	NM	r=0.32-0.57 p=0.01-0.001	NM	NM	NM	r=0.43-0.56 p=0.01-0.001	r=0.35-0.41 p=0.05
Terslev (10)	r=0.34 p=0.02	NM	NM	NM	r=0.6 p=0.02	NM	NM
Qvistgaard (70)	r=0.53 p=0.03	NM	NS	NM	NM	NM	NS

*Six joint assessment ; ⁺PDJC = Power Doppler Joint count; #USJIPD=US joint index for Power Doppler signal; NM=not measured; NS=not significant NC=no correlation

Figure 4: The assessed disease parameters and associated correlations with Doppler measurements (r-values) in the seven correlation studies

Assessment of accuracy

In study III (108) the influence of extraneous factors on the Doppler measurement was investigated. Because the US Doppler measures blood flow, it is possible that other factors not related to inflammation, which affect the blood flow, may influence the Doppler measurements. Everyday conditions affecting the blood flow are body temperature and physical activity. We assessed Doppler activity after both decrease and increase of the skin temperature and after exercise of the muscles of the hand and forearm. Some groups have indicated that temperature might influence the Doppler by addressing the temperature of the room in which the US examinations were performed (6;97). However, beside our study only two studies

with a small number of patients with RA have investigated the effect of external factors on the Doppler measurements, namely skin cooling (98;99). Both studies found decreased Doppler signal after cooling of the skin. This is in accordance with our results as we found that skin cooling was the only intervention affecting the amount of Doppler activity in the synovium. Before we conducted study III we made a test study on 14 patients with RA, and here we found an effect of skin cooling as well (107). Thus all studies performed suggest that Doppler is affected by intensive skin cooling. However, the decrease in Doppler signal study after cooling in our study was only modest and much less than the variation in Doppler measurements between various examinations, thus the detected decrease do not seem to be of any clinical relevance.

Summary

In general, validity of the Doppler measurements is difficult to test due to a lack of precise measurements of disease activity in RA. Thus published results, including our own, show only moderate correlation to MRI, histopathology and commonly used blood tests or scores for disease activity in RA.

The methods used to assess the Doppler activity differ both in relation to the scoring system used and number of joints assessed. Despite that, it does not seem to influence the magnitude of the correlation when various numbers of joints are assessed. In study II (92) where we investigated the concurrent validity with Doppler measurements in only one joint, the wrist joint, the r-values were similar or above the scores of more comprehensive joint counts. No definitive conclusion about the concurrent validity of Doppler US can be made. First, due to the variation in the studies as previously mentioned and second because measurements of disease activity in RA have an imprecise character, thus not always in accordance with the patient's symptoms. This is a serious difficulty in the implementation of all new measurements in RA and a pitfall in all validity assessment in the area of RA. In the assessment of the accuracy (study III) (108) we only found a modest influence of no clinical relevance when the skin of the wrist was cooled down. None of the other interventions affected the Doppler measures in a systematic way. Our results are in accordance with two other small studies investigating the effect of skin cooling (98;99). This may indicate that the effect of external factors on Doppler measurements is nearly nonexistent. However, our results on the influence of various external factors on Doppler measurements should be confirmed by other groups to substantiate these conclusions.

LEARNING COURSE IN ULTRASOUND EXAMINATION IN RA

Study IV investigates the learning progress of a standardized US examination procedure of the hand in patients with RA. Investigation of the practical use and learning process in obtaining skills in a standardized US examination is associated with the third part of the OMERACT filter (78)(figure 3). The third part of the filter is about the implementation of a measurement for instance if the costs make it

realistic to acquire the device, if skills in collecting and assessing the measurement can be obtained and if the examination procedure is harmless to the patient

Ultrasound is supposed to be the most operator dependent image modality (109) and therefore extensive training and supervision are required. Despite this there are still no formalised training programs in rheumatology in Europe or the rest of the world.

The need for a standardized training program is underlined by the fact that when US experts from different European countries have tested their interobserver reliability only moderate kappa values are obtained when testing interpretation of grey-scale images (83;84). This relative lack of agreement is even more pronounced when Doppler examinations are tested, with kappa values around 0.55 (74). What is also of interest in the Doppler study is that the results of the intra-reader agreement is only good (kappa=0.8), although you could have expected it to be excellent in a group of experts in musculoskeletal US. Nearly all of the experts who participated in the three reliability studies state that they used the guidelines for US examination made by the EULAR US group (51). The moderate reliability may indicate that these guidelines use a difficult approach to the interpretation of images. One problem with the guidelines might be that only general landmarks are described. Consequently not only standardized training programs are needed but also strictly standardized international examination procedures are required. It could be presumed that a standardization of examination procedures will cause both a better intra- and interobserver reliability. The results in study I (91) indicate that standardized examination and analysis procedures improve the intraobserver reliability with an increase of the ICC from 0.48 to 0.95.

Study IV - Learning curve in musculoskeletal US

(See appendix 4 for details)

In study IV (110) we investigated the learning progress for 10 rheumatologists in performing a standardized US examination of the hand in patients with RA. Nearly 100 patients with RA were included and a total of 378 US examinations were evaluated. The examination contained both a grey- scale and a Doppler evaluation (110). The examination procedure taught in study IV included the method with excellent reliability in study I. The ten rheumatologists participated in two short courses in basic skills in US performance and afterwards each of them had two days of individual training in the standardized examination procedure. For one year the rheumatologists enrolled patients with RA in the study. The patients were enrolled as they initiated treatment with an anti-TNF α drug. The patients were followed for 1 year and in the study period five US examinations were performed. All examinations were sent to a central unit for evaluation; the evaluation was made according to a scoring sheet assessing the quality of the US examination. In each evaluation all scores were summarized and a total score for the examination calculated. Furthermore, the time spent on the examination was registered. A written evaluation of each US examination was sent to the participant. On the basis of the scores obtained in each evaluation a learning curve was constructed for all participants.

The results in study IV (110) showed that good skills with high and satisfactory scores in a standardized US examination can be obtained within a few days of training and interestingly, there was no correlation between the time applied to get satisfactory skills and previous scanning experience.

A remarkable result in the study was that no or only modest progress in the scores was seen throughout the study period. This may partly be caused by the high scores archived already in the beginning of the study period. But actually some participants had decreasing scores throughout the study period, maybe caused by lost concentration and development of some kind of carelessness.

For all the participants in the study, the time spent on an examination decreased with 50% throughout the study period (se figure 5).

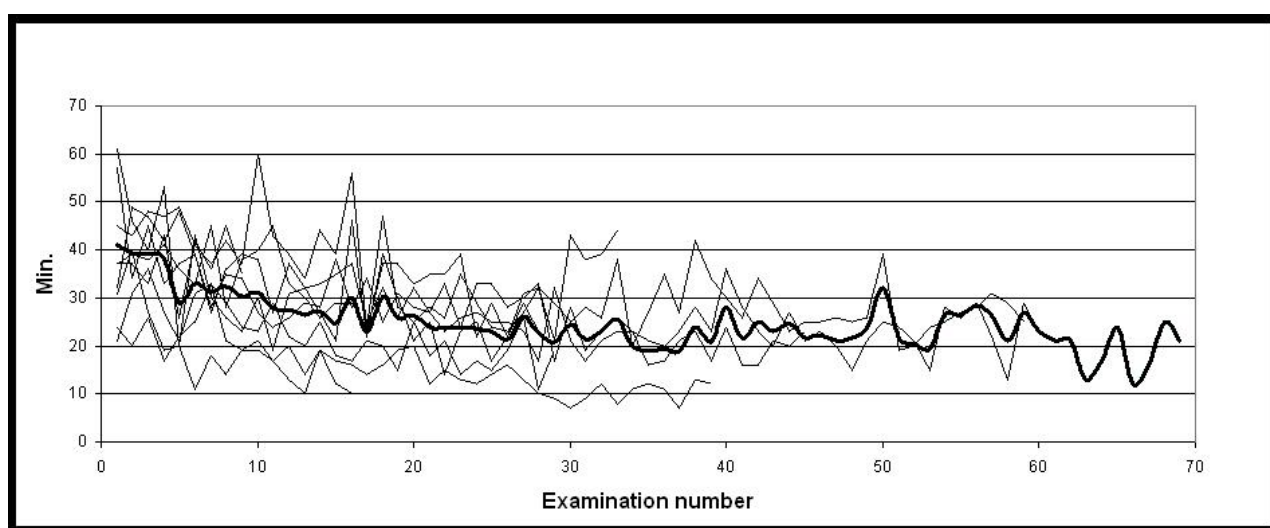


Figure 5: The development in examination time for ten rheumatologists in the performance of an US examination of the wrist and MCP joints throughout a two year period. The fat line is mean examination time for the group.

The framework for a US training program

A group of English rheumatologists have carried out an extensive work to investigate the area of musculoskeletal US in Europe. First they investigated to what extent persons performing musculoskeletal US throughout Europe participated in formalised education programs (111). They found that formal training regimes and competency assessment procedures were very limited. In continuation of this work, they carried out an inquiry among European experts in musculoskeletal US to uncover what an expert panel would suggest an education program in musculoskeletal US should include and which skills they found are required to perform satisfactory musculoskeletal US (112;113). Furthermore, they carried out a survey among 48 rheumatologists to reveal what was of importance for them if they were to undergo an education program in US. The overall conclusion of the survey was that a fair trade-off between clinical usefulness and the time spent to achieve competency was essential (114). With this in mind, the results of study IV are interesting as the examination skills in performing an US examination in an anatomic area

highly relevant for rheumatologists were achieved rapidly, and the examination time decreased markedly in the study period (110).

Both the experts panel and the 48 participants in the survey pointed out that it is of specific interest in the rheumatological area to use US to monitor disease activity and progression, and that the joints of the hand are an important region to investigate (112-114). This together with the fact that the wrist and finger joints are the joints most frequently involved in RA (25) make it relevant to investigate the learning progress for US examination of these joints. However, the joints of the hand do not appear to be the easiest joints to examine (74;115). With all this in mind, we found it of specific relevance to study the learning progress of performing US of the joints of the hand in patients with RA when an assessment of learning progress of a standardized examination procedure was carried out.

Based on consensus procedure, the EULAR US group has established an US education program. This program concerns both the content and organising of the EULAR US courses and the skills required to participate in the courses (109;116). It is stated that the optimal training program in musculoskeletal US contains both theory, practical instruction by a tutor and evaluation of performed US examinations. We tried to incorporate all these aspects in study IV (110), as it was a prerequisite that the participants had participated in two courses in basic US and that all had an individual training before starting to examine patients. Finally, all their examinations were evaluated by a tutor.

Utility of various training approaches in US

The research made in training programs in US is modest and most of the work derives from one group (115;117-119). Two of the studies are not actually investigating learning progress but describe the organization, training and implementation of US in daily practice (117;119). The other studies from this group describe different training concepts; a web-based learning program (118) and a self-teaching program (115). In the web-based learning program only 23% of the participants (60 rheumatologists) passed the competency test performed after a 6 month training program (118). This percentage is much lower than in one of the implementation studies (119) in which they also conducted a competency assessment. In study IV (110) the participants started out with a high percentage of correct scores after participating in the face-to-face part of the training program, but subsequently the scores increased only modestly despite intensive corrections, and this might suggest that personal supervision is more effective than web-based education and evaluation (110;118). This point of view is further supported by the results in the self-learning program (115). In this study, a novice was trained in US examination guided by digital images of standard scanning positions. However, in each scanning session all the images stored by the novice were evaluated together with a tutor and this study is therefore not totally web-based. In this study satisfactory skills in US were obtained in only 24 hours of training within one month.

A study has investigated the learning-curve for the detection of synovitis for three rheumatologists with varying experience in US (120). What is of interest in this study is that the two persons with limited experience in performing US had increasing scores throughout the study period, while the person with

prior US experience showed decreasing scores. This is in accordance with the experience in our study where some of the participants obtained excellent scores in the beginning of the study but had modestly decreasing scores in the end of the study period primarily due to incorrectness in the image acquisition.

Summary

Despite agreement on the necessity for both training and standardized examination procedures in rheumatological US and the fact that experts in US have only modest reliability in their US performance (74;83;84) only few studies have been performed to elucidate an optimal training procedure. Our study is to our knowledge the first to investigate the learning progress in a large group of rheumatologists (110). We found that skills in performing US of the hand in patients with RA can be achieved rather easily. However, some persons do not maintain their good performance over time. A similar decrease in performance skills is also seen in another study (120). This suggests that constant training and evaluation is essential. However, more studies are needed to define the optimal training and evaluation program in rheumatological US.

CONCLUSION

Ultrasound Doppler has now been accepted widely as a practical tool in the clinic. Until recently, studies have been lacking for a validation of the use of Doppler US in more elaborate evaluations of the RA activity.

The present thesis is based on four studies, which in various ways elucidate the validity of Doppler US. The OMERACT group has developed the OMERACT filter which is to be applied in the assessment of the applicability of a new measurement. The filter assesses the *truth*, *discriminative ability* and *feasibility* of a measurement. The studies in this thesis have investigated Doppler US measurement according to the various parts of the OMERACT filter. The *truth* was investigated by studying the correlation of US Doppler in the wrist joint with other measurements of disease activity (concurrent validity) and by investigating the influence of extraneous factors on the amount of US Doppler in the synovium (accuracy). The correlation to other measurements of disease activity was only moderate, but still similar to or higher than correlations obtained in other studies. The moderate correlation may be explained by the lack of a single valid disease measurement and the heterogeneity of patients with RA, making assessment of disease activity complicated. We found that the only extraneous factor affecting the amount of Doppler activity was cooling of the skin and that this effect was too small to have any clinical relevance. Part of the *discrimination* was explored by assessing the test-retest reliability in image acquisition and analysis techniques. The reliability was excellent when a strictly standardized procedure in both image acquisition and analysis was applied. The *feasibility* was investigated by examining the learning process for ten rheumatologists trained in the standardized examination procedure assessed in the test-retest study (91). Satisfactory skills in this examination procedure were obtained after a focused, while rather short period of training. However, hereafter only minimal or no improvement was seen in the following year indicating that a follow-up program maintaining the good spirit is mandatory; consistent reporting of results with correction may not be enough. Furthermore, the *feasibility* was verified by the fact that assessment of a single inflamed wrist joint displayed as good validity as more comprehensive joint counts, when correlated with various measurements of disease activity (concurrent validity)(92). In conclusion, the results in the four studies of this thesis contribute to the evidence of Doppler US measurement as a useful tool in evaluation of disease activity in patients with RA. Thus Doppler US measurements may serve as a valuable contribution to the evaluation of patients with RA with the ultimate goal of providing optimal treatment of this important disease.

FUTURE CONSIDERATIONS

There are already several applications of Doppler US in rheumatology, and since both the validity of the technique and the quality and versatility of ultrasound machines increase, it is likely that more applications will appear.

The diagnostic value of Doppler US is still unexplored and as the need for early and precise diagnosis in RA is growing it is of great interest to look further into the diagnostic abilities of Doppler US.

Recently, a number of studies have shown that patients assessed clinically to be in remission had inflammation on US and MRI (16-18;121), which indicates superiority of imaging in comparison to clinical assessment of RA. Of particular interest is that in one of the studies, the group of patients with inflammation on imaging was followed for 1 year, and it was found that these patients had a more pronounced development of erosion than other patients (18). The best predictor for erosive progression in both MRI and US was Doppler US. The remission criteria used in the studies were DAS28 and/or ACR remission criteria, showing that they may not be sufficient in the assessment of remission in patients with RA. With this in mind, an incorporation of Doppler measurements in a future disease index seems to be relevant.

The technical improvement of US equipment is developing fast, improving the image quality and thereby the accuracy of the US examinations. The development of new US techniques such as 3D and image fusion may increase the applications of US in the future.

ENGLISH SUMMARY

Rheumatoid arthritis is a chronic autoimmune disease, which primarily attacks the synovial membrane. More often than not, rheumatoid arthritis is a chronic disease and can lead to extensive destruction of joints and a resulting deterioration of motor function.

The aetiology of rheumatoid arthritis is yet unknown and no pathognomonic test exists.

Recent research and treatment results show that fast and intensive treatment is important to avoid permanent joint destruction. On this background it is important to have reliable diagnostic tools and valid measurements for evaluation of disease activity. Doppler US measurements have gained ground in the evaluation of patients with RA. The measurement is demonstrated to be useful for assessment of disease activity, and has proven to be more sensitive than clinical evaluation in the assessment of disease activity. Despite the growing use of Doppler US in patients with RA, the validity of the method is still undetermined. The aim of this PhD thesis has been by way of 4 studies to answer questions related to validity and usability of Doppler US measurements in patients with RA.

The 4 studies concern

- I. Reproducibility of Doppler measurements
- II. The correlation of Doppler measurements with other measures for disease activity in patients with RA
- III. The influence of external factors on Doppler measurements
- IV. Learning curve when being taught a standardised ultrasound examination procedure

In Study I, 14 patients with rheumatoid arthritis in the wrist were US scanned twice with an interval of 30 minutes. All images were analysed with two methods. The best test-retest reliability was obtained when the images for analysis were selected based on both anatomical structures and the amount of Doppler activity in the US image, and when the subsequent analysis was guided by specific anatomical structures. The amount of Doppler activity for each investigation method was assessed by determining the Colour Fraction (CF), which is the relation between colour pixels and the total amount of pixels.

In Study II the correlation between Colour Fraction (CF) and other measurements used for evaluation of disease activity was investigated in 109 patients with RA. We found a moderate correlation between acute phase protein measurements and a modest correlation to the score of swollen joints and DAS28.

Correlations between the other measurements of disease activity were lower. The best correlation was seen between Colour Fraction (CF) and the acute phase protein measurements.

In Study III, a possible effect on Doppler measurements from skin temperature and exercise was investigated. Twenty-nine patients with RA who had Doppler activity in the wrist joint underwent an US examination before and after cooling and heating of the hand and before and after performing hand exercises. Colour Fraction (CF)s before and after each intervention were compared. Only cooling had a

modest effect on Colour Fraction (CF), causing it to decrease slightly. However, the decrease was so slight that it is regarded to be of no clinical relevance.

In Study IV the learning curves for 10 rheumatologists were examined. The rheumatologists were taught the standardised scanning of hand and finger joints, which was tested in Study I. The study showed that good skills in a standardised ultrasound examination procedure can be obtained quite easily and quickly. At the same time it turned out that previous scanning experience was of no importance for how quickly the skills were acquired. However, the development of ultrasound scanning skills quickly plateaued in spite of intensive evaluation.

The results from the four studies show that the validity of Doppler ultrasound examinations is very good if the collection and analysis of images is done in a systematic way and that skills for performing a standardised ultrasound examination can be obtained from a short training period, which renders a standardisation of examination procedures realistic. We also showed that Doppler measurements were influenced very slightly by decreased skin temperatures. No other external circumstances assessed influenced the measurements. Concurrently we demonstrated a moderate correlation between Doppler ultrasound and biochemical markers for disease activity and modest correlation with joint count and disease index (DAS28) and Doppler measurements. The fact that we demonstrated only a moderate correlation with other disease parameters does not necessarily indicate a low validity of Doppler US measurements, rather that there is a general problem with exact markers of disease activity in patients with RA.

The results in this thesis demonstrate good and valid options for applying Doppler US measurements in patients with RA.

DANISH SUMMARY

Leddegigt er en kronisk autoimmun sygdom, der primært angriber ledslimhinden. Leddegigt har som oftest et kronisk forløb og kan medføre udtalt leddestruktion og heraf følgende nedsat fysisk funktion. Årsagen til leddegigts opståen er fortsat ukendt og intet enkelt diagnostisk kriterium for sygdommen eksisterer.

Nyere forskning og behandlingsresultater viser at hurtig intensiv behandling er vigtig for at undgå varige leddskader. På baggrund af dette er pålidelige diagnoseredskaber og valide målemetoder til vurdering af sygdomsaktivitet vigtige. Doppler-ultralysmålinger har vundet udbredelse i vurderingen af patienter med leddegigt og har vist sig anvendelige til vurdering af sygdomsaktivitet og har vist større følsomhed end klinisk vurdering af denne.

Trods den øgede anvendelse af Doppler-ultralyd hos patienter med leddegigt er pålideligheden af undersøgelsesmetoden stadig uafklaret. Formålet med denne ph.d.-afhandling har været gennem 4 delstudier at besvare spørgsmål i relation til pålidelighed og anvendelighed af ultralydsmålinger hos patienter med leddegigt. De 4 studier omhandler:

- I. Reproducerbarheden af Dopplermålinger
- II. Dopplermålingernes korrelation med andre mål for sygdomsaktivitet hos patienter med leddegigt
- III. Ydre faktorerers påvirkning på Dopplermålingerne
- IV. Indlæringsforløb ved oplæring i en standardiseret ultralydsundersøgelse

I studie I blev fjorten patienter med leddegigt i håndleddet undersøgt to gange med en halv times mellemrum. Alle billeder blev analyseret med to metoder. Den bedste reproducerbarhed blev opnået når billederne til analyse blev udvalgt både på baggrund af anatomiske strukturer og mængden af Doppler-aktivitet på ultralydsbilledet, og den efterfølgende billedanalyse blev lavet vejledt af specifikke anatomiske strukturer. Graden af Doppler-aktivitet i de enkelte undersøgelsesmetoder blev bestemt ved måling af Colour Fraction (CF) som er forholdet mellem farvede pixels og det samlede antal pixels i det analyserede område.

I studie II blev korrelationen mellem Colour Fraction (CF) og almindelig anvendte målemetoder til vurdering af sygdomsaktivitet hos 109 leddegigtpatienter undersøgt. Der fandtes moderat korrelation til akut fase protein-målinger og beskeden korrelation til score for antallet af hævede led og til DAS28. Korrelationen mellem de øvrige sygdomsparametre indbyrdes var ikke mere udtalt. Den bedste korrelation var mellem Colour Fraction(CF) og akut fase protein-målingerne.

I studie III blev en eventuel påvirkning af hudtemperatur og fysisk aktivitet på Doppler-målingerne undersøgt. Nioogtyve leddegigtpatienter med Doppler-aktivitet i håndleddet blev skannet før og efter nedkøling og opvarmning af hånden og før og efter udførelse af håndøvelser. Colour Fraction (CF)) før og efter de enkelte interventioner blev sammenlignet. Kun nedkøling havde en beskeden effekt på Colour

Fraction (CF)) idet den viste et beskedent fald. Det var dog et så beskedent fald, at det vurderes til at være uden klinisk betydning.

I studie IV blev indlæringsforløbet for 10 reumatologer undersøgt. De blev oplært den standardiserede ultralydsundersøgelse af hånd og fingerled som blev testet i studie I. Studiet viste at gode færdigheder i en standardiseret ultralydsskanning kan opnås relativt nemt og hurtigt. Samtidig viste det sig at forkundskaber ikke havde betydning for, hvor hurtigt færdighederne blev opnået. Udviklingen i ultralydsfærdigheder fladede dog hurtigt ud trods intensiv evaluering.

Resultaterne fra de 4 studier viser at reproducerbarheden af Doppler-ultralydsundersøgelser er god hvis billedindsamling og analyse foretages på en systematisk måde og at færdigheder i en standardiseret undersøgelsesmetode kan opnås gennem et kort træningsforløb, hvilket gør en sådan standardisering af undersøgelsesprocedurer realistisk. Samtidig viste det sig at Dopplermålingerne kun i meget ringe grad blev påvirket af nedkøling af huden. Ingen andre ydre forhold påvirkede Dopplermålingerne. Der fandtes en acceptabel korrelation mellem Doppler-ultralyd og biokemiske markører for sygdomsaktivitet samt en beskedent korrelation mellem Doppler-målinger og ledscore og sygdomsindex (DAS28). Den kun moderate korrelation med andre sygdomsparametre peger ikke nødvendigvis på beskedent validitet af Doppler-ultralydsmålinger, men på et generelt problem med præcise diagnose- og sygdomsmarkører hos patienter med leddegigt.

Resultaterne i denne afhandling peger på gode og pålidelige anvendelsesmuligheder for Dopplerultralydsmålinger hos patienter med leddegigt.

REFERENCE LIST

- (1) ACR ad hoc committee. Guidelines for the management of rheumatoid arthritis. American College of Rheumatology Ad Hoc Committee on Clinical Guidelines. *Arthritis Rheum* 1996 May;39:713-22.
- (2) Abdel-Nasser AM, Rasker JJ, Valkenburg HA. Epidemiological and clinical aspects relating to the variability of rheumatoid arthritis. *Semin Arthritis Rheum* 1997 Oct;27:123-40.
- (3) Arnett FC, Edworthy SM, Bloch DA, McShane DJ, Fries JF, Cooper NS, et al. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. *Arthritis Rheum* 1988 Mar;31:315-24.
- (4) Bliddal H, Boesen M, Christensen R, Kubassova O, Torp-Pedersen S. Imaging as a follow-up tool in clinical trials and clinical practice. *Best Pract Res Clin Rheumatol* 2008 Dec;22:1109-26.
- (5) Wakefield RJ, D'Agostino MA, Iagnocco A, Filippucci E, Backhaus M, Scheel AK, et al. The OMERACT Ultrasound Group: status of current activities and research directions. *J Rheumatol* 2007 Apr;34:848-51.
- (6) Taylor PC, Steuer A, Gruber J, Cosgrove DO, Blomley MJ, Marsters PA, et al. Comparison of ultrasonographic assessment of synovitis and joint vascularity with radiographic evaluation in a randomized, placebo-controlled study of infliximab therapy in early rheumatoid arthritis. *Arthritis Rheum* 2004 Apr;50:1107-16.
- (7) Mottonen T, Hannonen P, Korpela M, Nissila M, Kautiainen H, Ilonen J, et al. Delay to institution of therapy and induction of remission using single-drug or combination-disease-modifying antirheumatic drug therapy in early rheumatoid arthritis. *Arthritis Rheum* 2002 Apr;46:894-8.
- (8) Emery P, Breedveld FC, Hall S, Durez P, Chang DJ, Robertson D, et al. Comparison of methotrexate monotherapy with a combination of methotrexate and etanercept in active, early, moderate to severe rheumatoid arthritis (COMET): a randomised, double-blind, parallel treatment trial. *Lancet* 2008 Aug 2;372:375-82.
- (9) Combe B. Early rheumatoid arthritis: strategies for prevention and management. *Best Pract Res Clin Rheumatol* 2007 Feb;21(1):27-42.
- (10) Terslev L, Torp-Pedersen S, Qvistgaard E, Danneskiold-Samsøe B, Bliddal H. Estimation of inflammation by Doppler ultrasound: quantitative changes after intra-articular treatment in rheumatoid arthritis. *Ann Rheum Dis* 2003 Nov;62:1049-53.
- (11) Terslev L, Torp-Pedersen S, Qvistgaard E, Kristoffersen H, Rogind H, Danneskiold-Samsøe B, et al. Effects of treatment with etanercept (Enbrel, TNRF:Fc) on rheumatoid arthritis evaluated by Doppler ultrasonography. *Ann Rheum Dis* 2003 Feb;62:178-81.
- (12) Filippucci E, Farina A, Carotti M, Salaffi F, Grassi W. Grey scale and power Doppler sonographic changes induced by intra-articular steroid injection treatment. *Ann Rheum Dis* 2004 Jun;63:740-3.
- (13) Filippucci E, Iagnocco A, Salaffi F, Cerioni A, Valesini G, Grassi W. Power Doppler sonography monitoring of synovial perfusion at wrist joint in rheumatoid patients treated with adalimumab. *Ann Rheum Dis* 2006 Nov;65:1433-7 Feb.

- (14) Teh J, Stevens K, Williamson L, Leung J, McNally EG. Power Doppler ultrasound of rheumatoid synovitis: quantification of therapeutic response. *Br J Radiol* 2003 Dec;76:875-9.
- (15) Naredo E, Moller I, Cruz A, Carmona L, Garrido J. Power doppler ultrasonographic monitoring of response to anti-tumor necrosis factor therapy in patients with rheumatoid arthritis. *Arthritis Rheum* 2008 Jul 30;58:2248-56.
- (16) Brown AK, Quinn MA, Karim Z, Conaghan PG, Peterfy CG, Hensor E, et al. Presence of significant synovitis in rheumatoid arthritis patients with disease-modifying antirheumatic drug-induced clinical remission: Evidence from an imaging study may explain structural progression. *Arthritis Rheum* 2006 Nov 28;54:3761-73.
- (17) Ozgocmen S, Ozdemir H, Kiris A, Bozgeyik Z, Ardicoglu O. Clinical evaluation and power Doppler sonography in rheumatoid arthritis: evidence for ongoing synovial inflammation in clinical remission. *South Med J* 2008 Mar;101:240-5.
- (18) Brown AK, Conaghan PG, Karim Z, Quinn MA, Ikeda K, Peterfy CG, et al. An explanation for the apparent dissociation between clinical remission and continued structural deterioration in rheumatoid arthritis. *Arthritis Rheum* 2008 Sep 29;58:2958-67.
- (19) Jantti J, Aho K, Kaarela K, Kautiainen H. Work disability in an inception cohort of patients with seropositive rheumatoid arthritis: a 20 year study. *Rheumatology (Oxford)* 1999 Nov;38:1138-41.
- (20) Tak PP, Bresnihan B. The pathogenesis and prevention of joint damage in rheumatoid arthritis: advances from synovial biopsy and tissue analysis. *Arthritis Rheum* 2000 Dec;43:2619-33.
- (21) FitzGerald O, Bresnihan B. Synovial membrane cellularity and vascularity. *Ann Rheum Dis* 1995 Jun;54:511-5.
- (22) Taylor PC, Sivakumar B. Hypoxia and angiogenesis in rheumatoid arthritis. *Curr Opin Rheumatol* 2005 May;17:293-8.
- (23) Pap T, Distler O. Linking angiogenesis to bone destruction in arthritis. *Arthritis Rheum* 2005 May;52:1346-8.
- (24) Harris ED, Jr. Rheumatoid arthritis. Pathophysiology and implications for therapy. *N Engl J Med* 1990 May 3;322:1277-89.
- (25) Fleming A, Crown JM, Corbett M. Incidence of joint involvement in early rheumatoid arthritis. *Rheumatol Rehabil* 1976 May;15:92-6.
- (26) van Venrooij WJ, van Beers JJ, Pruijn GJ. Anti-CCP Antibody, a Marker for the Early Detection of Rheumatoid Arthritis. *Ann N Y Acad Sci* 2008 Nov;1143:268-85.
- (27) van Riel PL, Schumacher HR, Jr. How does one assess early rheumatoid arthritis in daily clinical practice? *Best Pract Res Clin Rheumatol* 2001 Mar;15:67-76.
- (28) van der Heijde DM, 't Hof MA, van Riel PL, Theunisse LA, Lubberts EW, van Leeuwen MA, et al. Judging disease activity in clinical practice in rheumatoid arthritis: first step in the development of a disease activity score. *Ann Rheum Dis* 1990 Nov;49:916-20.
- (29) Prevoo ML, 't Hof MA, Kuper HH, van Leeuwen MA, van de Putte LB, van Riel PL. Modified disease activity scores that include twenty-eight-joint counts. Development and validation in a

- prospective longitudinal study of patients with rheumatoid arthritis. *Arthritis Rheum* 1995 Jan;38:44-8.
- (30) Felson DT, Anderson JJ, Boers M, Bombardier C, Furst D, Goldsmith C, et al. American College of Rheumatology. Preliminary definition of improvement in rheumatoid arthritis. *Arthritis Rheum* 1995 Jun;38:727-35.
- (31) Scott DL, Houssien DA. Joint assessment in rheumatoid arthritis. *Br J Rheumatol* 1996 Sep;35 Suppl 2:14-8.
- (32) Prevoo ML, van Riel PL, 't Hof MA, van Rijswijk MH, van Leeuwen MA, Kuper HH, et al. Validity and reliability of joint indices. A longitudinal study in patients with recent onset rheumatoid arthritis. *Br J Rheumatol* 1993 Jul;32:589-94.
- (33) Smolen JS, Breedveld FC, Eberl G, Jones I, Leeming M, Wylie GL, et al. Validity and reliability of the twenty-eight-joint count for the assessment of rheumatoid arthritis activity. *Arthritis Rheum* 1995 Jan;38:38-43.
- (34) Fuchs HA, Brooks RH, Callahan LF, Pincus T. A simplified twenty-eight-joint quantitative articular index in rheumatoid arthritis. *Arthritis Rheum* 1989 May;32:531-7.
- (35) Hernandez-Cruz B, Cardiel MH. Intra-observer reliability of commonly used outcome measures in rheumatoid arthritis. *Clin Exp Rheumatol* 1998 Jul;16:459-62.
- (36) Scott DL, Choy EH, Greeves A, Isenberg D, Kassiror D, Rankin E, et al. Standardising joint assessment in rheumatoid arthritis. *Clin Rheumatol* 1996 Nov;15:579-82.
- (37) Otterness IG. The value of C-reactive protein measurement in rheumatoid arthritis. *Semin Arthritis Rheum* 1994 Oct;24:91-104.
- (38) Wolfe F. Comparative usefulness of C-reactive protein and erythrocyte sedimentation rate in patients with rheumatoid arthritis. *J Rheumatol* 1997 Aug;24:1477-85.
- (39) Kushner I. C-reactive protein in rheumatology. *Arthritis Rheum* 1991 Aug;34(8):1065-8.
- (40) Mallya RK, de Beer FC, Berry H, Hamilton ED, Mace BE, Pepys MB. Correlation of clinical parameters of disease activity in rheumatoid arthritis with serum concentration of C-reactive protein and erythrocyte sedimentation rate. *J Rheumatol* 1982 Mar;9:224-8.
- (41) Cutolo M, Maestroni GJ, Otsa K, Aakre O, Villaggio B, Capellino S, et al. Circadian melatonin and cortisol levels in rheumatoid arthritis patients in winter time: a north and south Europe comparison. *Ann Rheum Dis* 2005 Feb;64:212-6.
- (42) Wolfe F, Pincus T, O'Dell J. Evaluation and documentation of rheumatoid arthritis disease status in the clinic: which variables best predict change in therapy. *J Rheumatol* 2001 Jul;28:1712-7.
- (43) Bruce B, Fries JF. The Stanford Health Assessment Questionnaire: a review of its history, issues, progress, and documentation. *J Rheumatol* 2003 Jan;30:167-78.
- (44) Thorsen H, Hansen TM, McKenna SP, Sorensen SF, Whalley D. Adaptation into Danish of the Stanford Health Assessment Questionnaire (HAQ) and the Rheumatoid Arthritis Quality of Life Scale (RAQoL). *Scand J Rheumatol* 2001;30:103-9.
- (45) Bjorner JB, Thunedborg K, Kristensen TS, Modvig J, Bech P. The Danish SF-36 Health Survey: translation and preliminary validity studies. *J Clin Epidemiol* 1998 Nov;51:991-9.

- (46) Wells GA, Becker JC, Teng J, Dougados M, Schiff MH, Smolen JS, et al. Validation of the Disease Activity Score 28 (DAS28) and EULAR response criteria based on CRP against disease progression in patients with rheumatoid arthritis, and comparison with the DAS28 based on ESR. *Ann Rheum Dis* 2009 Jun;68:954-60 19.
- (47) van Gestel AM, Anderson JJ, van Riel PL, Boers M, Haagsma CJ, Rich B, et al. ACR and EULAR improvement criteria have comparable validity in rheumatoid arthritis trials. *American College of Rheumatology European League of Associations for Rheumatology. J Rheumatol* 1999 Mar;26:705-11.
- (48) Balsa A, Carmona L, Gonzalez-Alvaro I, Belmonte MA, Tena X, Sanmarti R. Value of Disease Activity Score 28 (DAS28) and DAS28-3 compared to American College of Rheumatology-defined remission in rheumatoid arthritis. *J Rheumatol* 2004 Jan;31:40-6.
- (49) Feldmann M, Brennan FM, Foxwell BM, Maini RN. The role of TNF alpha and IL-1 in rheumatoid arthritis. *Curr Dir Autoimmun* 2001;3:188-99.
- (50) Ostergaard M, Unkerskov J, Linde L, Krogh NS, Ravn T, Ringsdal VS, et al. Low remission rates but long drug survival in rheumatoid arthritis patients treated with infliximab or etanercept: results from the nationwide Danish DANBIO database. *Scand J Rheumatol* 2007 Mar;36:151-4.
- (51) Backhaus M, Burmester GR, Gerber T, Grassi W, Machold KP, Swen WA, et al. Guidelines for musculoskeletal ultrasound in rheumatology. *Ann Rheum Dis* 2001 Jul;60:641-9.
- (52) Ostergaard M, Szkudlarek M. Imaging in rheumatoid arthritis--why MRI and ultrasonography can no longer be ignored. *Scand J Rheumatol* 2003;32:63-73.
- (53) Kane D, Balint PV, Sturrock R, Grassi W. Musculoskeletal ultrasound--a state of the art review in rheumatology. Part 1: Current controversies and issues in the development of musculoskeletal ultrasound in rheumatology. *Rheumatology (Oxford)* 2004 Jul;43:823-8.
- (54) Kane D, Grassi W, Sturrock R, Balint PV. Musculoskeletal ultrasound--a state of the art review in rheumatology. Part 2: Clinical indications for musculoskeletal ultrasound in rheumatology. *Rheumatology (Oxford)* 2004 Jul;43:829-38.
- (55) Scheel AK, Hermann KG, Ohrndorf S, Werner C, Schirmer C, Detert J, et al. Prospective 7 year follow up imaging study comparing radiography, ultrasonography, and magnetic resonance imaging in rheumatoid arthritis finger joints. *Ann Rheum Dis* 2006 May;65:595-600.
- (56) Wakefield RJ, Gibbon WW, Conaghan PG, O'Connor P, McGonagle D, Pease C, et al. The value of sonography in the detection of bone erosions in patients with rheumatoid arthritis: a comparison with conventional radiography. *Arthritis Rheum* 2000 Dec;43:2762-70.
- (57) Wakefield RJ, Brown AK, O'connor PJ, Emery P. Power Doppler sonography: improving disease activity assessment in inflammatory musculoskeletal disease. *Arthritis Rheum* 2003 Feb;48:285-8.
- (58) Torp-Pedersen ST, Terslev L. Settings and artefacts relevant in colour/power Doppler ultrasound in rheumatology. *Ann Rheum Dis* 2008 Feb;67:143-9.
- (59) Schmidt WA, Backhaus M. What the practising rheumatologist needs to know about the technical fundamentals of ultrasonography. *Best Pract Res Clin Rheumatol* 2008 Dec;22:981-99.
- (60) Rubin JM. Spectral Doppler US. *Radiographics* 1994 Jan;14:139-50.

- (61) Burns PN. Principles of Doppler and color flow. *Radiol Med (Torino)* 1993 May;85(5 Suppl 1):3-16.
- (62) Naredo E, Collado P, Cruz A, Palop MJ, Cabero F, Richi P, et al. Longitudinal power Doppler ultrasonographic assessment of joint inflammatory activity in early rheumatoid arthritis: predictive value in disease activity and radiologic progression. *Arthritis Rheum* 2007 Feb 15;57:116-24.
- (63) Pozniak MA, Zagzebski JA, Scanlan KA. Spectral and color Doppler artifacts. *Radiographics* 1992 Jan;12:35-44.
- (64) Terslev L, Torp-Pedersen S, Qvistgaard E, Bliddal H. Spectral Doppler and resistive index. A promising tool in ultrasonographic evaluation of inflammation in rheumatoid arthritis. *Acta Radiol* 2003 Nov;44:645-52.
- (65) Albrecht K, Muller-Ladner U, Strunk J. Quantification of the synovial perfusion in rheumatoid arthritis using Doppler ultrasonography. *Clin Exp Rheumatol* 2007 Jul;25:630-8.
- (66) Bude RO, Rubin JM. Power Doppler sonography. *Radiology* 1996 Jul;200:21-3.
- (67) Martinoli C. Gain Setting in Power Doppler. *Radiology* 1997;202:284-5.
- (68) Nilsson A. Artefacts in sonography and Doppler. *Eur Radiol* 2001;11:1308-15.
- (69) Schmidt WA. Doppler sonography in rheumatology. *Best Pract Res Clin Rheumatol* 2004 Dec;18:827-46.
- (70) Qvistgaard E, Rogind H, Torp-Pedersen S, Terslev L, Danneskiold-Samsøe B, Bliddal H. Quantitative ultrasonography in rheumatoid arthritis: evaluation of inflammation by Doppler technique. *Ann Rheum Dis* 2001 Jul;60:690-3.
- (71) Szkudlarek M, Court-Payen, Jacobsen S, Klarlund M, Thomsen HS, Ostergaard M. Interobserver agreement in ultrasonography of the finger and toe joints in rheumatoid arthritis. *Arthritis Rheum* 2003 Apr;48:955-62.
- (72) Naredo E, Bonilla G, Gamero F, Uson J, Carmona L, Laffon A. Assessment of inflammatory activity in rheumatoid arthritis: a comparative study of clinical evaluation with grey scale and power Doppler ultrasonography. *Ann Rheum Dis* 2005 Mar;64:375-81.
- (73) Strunk J, Heinemann E, Neeck G, Schmidt KL, Lange U. A new approach to studying angiogenesis in rheumatoid arthritis by means of power Doppler ultrasonography and measurement of serum vascular endothelial growth factor. *Rheumatology (Oxford)* 2004 Dec;43:1480-3.
- (74) Koski JM, Saarakkala S, Helle M, Hakulinen U, Heikkinen JO, Hermunen H, et al. Assessing the intra- and inter-reader reliability of dynamic ultrasound images in power Doppler ultrasonography. *Ann Rheum Dis* 2006 Dec;65:1658-60
- (75) Johnston MV, Keith RA, Hinderer SR. Measurement standards for interdisciplinary medical rehabilitation. *Arch Phys Med Rehabil* 1992 Dec;73(12-S):S3-23.
- (76) Ostergaard M, Wiell C. Ultrasonography in rheumatoid arthritis: a very promising method still needing more validation. *Curr Opin Rheumatol* 2004 May;16:223-30.

- (77) Sim J, Arnell P. Measurement validity in physical therapy research. *Phys Ther* 1993 Feb;73:102-10.
- (78) Boers M, Brooks P, Strand CV, Tugwell P. The OMERACT filter for Outcome Measures in Rheumatology. *J Rheumatol* 1998 Feb;25:198-9.
- (79) Tugwell P, Bombardier C. A methodologic framework for developing and selecting endpoints in clinical trials. *J Rheumatol* 1982 Sep;9:758-62.
- (80) Bland JM, Altman DG. Statistics Notes: Validating scales and indexes. *BMJ* 2002 Mar 9;324:606-7.
- (81) Rooney M, Condell D, Quinlan W, Daly L, Whelan A, Feighery C, et al. Analysis of the histologic variation of synovitis in rheumatoid arthritis. *Arthritis Rheum* 1988 Aug;31:956-63.
- (82) Wakefield RJ, Balint PV, Szkudlarek M, Filippucci E, Backhaus M, D'Agostino MA, et al. Musculoskeletal ultrasound including definitions for ultrasonographic pathology. *J Rheumatol* 2005 Dec;32:2485-7.
- (83) Scheel AK, Schmidt WA, Hermann KG, Bruyn GA, D'Agostino MA, Grassi W, et al. Interobserver reliability of rheumatologists performing musculoskeletal ultrasonography: results from a EULAR "Train the trainers" course. *Ann Rheum Dis* 2005 Jul;64(7):1043-9.
- (84) Naredo E, Moller I, Moragues C, De Agustin JJ, Scheel AK, Grassi W, et al. Interobserver reliability in musculoskeletal ultrasonography: results from a "Teach the Teachers" rheumatologist course. *Ann Rheum Dis* 2006 Jan;65:14-9.
- (85) Joshua F, Edmonds J, Lassere M. Power Doppler ultrasound in musculoskeletal disease: a systematic review. *Semin Arthritis Rheum* 2006 Oct;36:99-108.
- (86) Rosenberg C, Etchepare F, Fautrel B, Bourgeois P. Diagnosis of synovitis by ultrasonography in RA: A one-year experience is enough for reliability on static images. *Joint Bone Spine* 2009 May;76:265-7.
- (87) Albrecht K, Grob K, Lange U, Muller-Ladner U, Strunk J. Reliability of different Doppler ultrasound quantification methods and devices in the assessment of therapeutic response in arthritis. *Rheumatology (Oxford)* 2008 Oct;47:1521-6 Aug 2.
- (89) Fleiss JL. The design and analysis of clinical experiments. New York: Wiley; 1986.
- (90) Landis J R, Koch G G. The measurements of observer agreement for categorical data. *Biometrics* 1977;33:159-74.
- (91) Ellegaard K, Torp-Pedersen S, Lund H, Henriksen M, Terslev L, Jensen PS, et al. Quantification of colour Doppler activity in the wrist in patients with rheumatoid arthritis - the reliability of different methods for image selection and evaluation. *Ultraschall Med* 2008 Aug;29:393-8.
- (92) Ellegaard K, Torp-Pedersen S, Terslev L, Danneskiold-Samsøe B, Henriksen M, Bliddal H. Ultrasound colour Doppler measurements in a single joint as measure of disease activity in patients with rheumatoid arthritis - assessment of concurrent validity. *Rheumatology (Oxford)* 2009 Mar;48:254-7

- (93) Terslev L, Torp-Pedersen S, Savnik A, von der RP, Qvistgaard E, Danneskiold-Samsøe B, et al. Doppler ultrasound and magnetic resonance imaging of synovial inflammation of the hand in rheumatoid arthritis: a comparative study. *Arthritis Rheum* 2003 Sep;4:2434-41.
- (94) Szkudlarek M, Court-Payen, Strandberg C, Klarlund M, Klausen T, Ostergaard M. Power Doppler ultrasonography for assessment of synovitis in the metacarpophalangeal joints of patients with rheumatoid arthritis: a comparison with dynamic magnetic resonance imaging. *Arthritis Rheum* 2001 Sep;4:2018-23.
- (95) Schmidt WA, Volker L, Zacher J, Schlafke M, Ruhnke M, Gromnica-Ihle E. Colour Doppler ultrasonography to detect pannus in knee joint synovitis. *Clin Exp Rheumatol* 2000 Jul;18:439-44.
- (96) Walther M, Harms H, Krenn V, Radke S, Faehndrich TP, Gohlke F. Correlation of power Doppler sonography with vascularity of the synovial tissue of the knee joint in patients with osteoarthritis and rheumatoid arthritis. *Arthritis Rheum* 2001 Feb;44:331-8.
- (97) Koski JM, Saarakkala S, Helle M, Hakulinen U, Heikkinen JO, Hermunen H. Power Doppler ultrasonography and synovitis. Correlating ultrasound imaging with histopathological findings and evaluating the performance of ultrasound equipments. *Ann Rheum Dis* 2006 Dec;65:1590-5.
- (98) Strunk J, Strube K, Klingenberg P, Muller-Ladner U, Lange U. Two- and three-dimensional Doppler sonographic evaluation of the effect of local cryotherapy on synovial perfusion in wrist arthritis. *Rheumatology (Oxford)* 2006 May;45:637-40.
- (99) Kelly S, Taylor P, Pitzalis C. Three Dimensional Power Doppler Signal is More Sensitive to Changes Than Two Dimensional Signal Under a Physiological Stimulus. *Arthritis Rheum* 58[9], S467. 9-9-2008.
Ref Type: Abstract
- (100) Hau M, Kneitz C, Tony HP, Keberle M, Jahns R, Jenett M. High resolution ultrasound detects a decrease in pannus vascularisation of small finger joints in patients with rheumatoid arthritis receiving treatment with soluble tumour necrosis factor alpha receptor (etanercept). *Ann Rheum Dis* 2002 Jan;61:55-8.
- (101) Iagnocco A, Filippucci E, Perella C, Ceccarelli F, Cassara E, Alessandri C, et al. Clinical and Ultrasonographic Monitoring of Response to Adalimumab Treatment in Rheumatoid Arthritis. *J Rheumatol* 2007 Nov 15.
- (102) Iagnocco A, Perella C, Naredo E, Meenagh G, Ceccarelli F, Tripodo E, et al. Etanercept in the treatment of rheumatoid arthritis: clinical follow-up over one year by ultrasonography. *Clin Rheumatol* 2008 Apr;27:491-6 Oct 19.
- (103) Hameed B, Pilcher J, Heron C, Kiely P D W. The relation between composite ultrasound measures and the DAS28 score, its components and acute phase markers in adult RA. *Rheumatology (Oxford)* 2008 Apr;47:476-80.
- (104) Kiris A, Ozgocmen S, Kocakoc E, Ardicoglu O. Power Doppler assessment of overall disease activity in patients with rheumatoid arthritis. *J Clin Ultrasound* 2006 Jan;34:5-11.
- (105) Naredo E, Gamero F, Bonilla G, Uson J, Carmona L, Laffon A. Ultrasonographic assessment of inflammatory activity in rheumatoid arthritis: comparison of extended versus reduced joint evaluation. *Clin Exp Rheumatol* 2005 Nov;23:881-4.

- (107) Ellegaard K, Torp-Pedersen S, Jensen P S, Lund H, Danneskiold-Samsøe B, Bliddal H. The influence of physical activity, skin temperature and daily variation on the colour-Doppler measurements in patients with rheumatoid arthritis. *Ann.Rheum.Dis.* 66, 651. 7-7-2007.
Ref Type: Abstract
- (108) Ellegaard K, Torp-Pedersen S, Henriksen M, Lund H, Danneskiold-Samsøe B, Bliddal. Influence of Recent Exercise and Skin Temperature on US Doppler Measurements in Patients with Rheumatoid Arthritis - an Intervention Study. *Rheumatology (Oxford)*. 2009 Sep 18. [Epub ahead of print] PMID:197673388
- (109) Naredo E, Bijlsma JW. Becoming a musculoskeletal ultrasonographer. *Best Pract Res Clin Rheumatol* 2009 Apr;23:257-67.
- (110) Ellegaard K, Torp-Pedersen S, Stoltenberg M, Hansen A, Lorenzen T, Jensen D.V, et al. Standardized ultrasound examination of the wrist and metacarpophalangeal joints in patients with rheumatoid arthritis- A multicentre learning experience: Education in Ultrasound of Rheumatoid Arthritis- EURA. Apr 2009.
Ref Type: Unpublished Work
- (111) Brown AK, O'connor PJ, Wakefield RJ, Roberts TE, Karim Z, Emery P. Practice, training, and assessment among experts performing musculoskeletal ultrasonography: toward the development of an international consensus of educational standards for ultrasonography for rheumatologists. *Arthritis Rheum* 2004 Dec 15;51:1018-22.
- (112) Brown AK, O'connor PJ, Roberts TE, Wakefield RJ, Karim Z, Emery P. Recommendations for musculoskeletal ultrasonography by rheumatologists: setting global standards for best practice by expert consensus. *Arthritis Rheum* 2005 Feb 15;53:83-92.
- (113) Brown AK, O'connor PJ, Roberts TE, Wakefield RJ, Karim Z, Emery P. Ultrasonography for rheumatologists: the development of specific competency based educational outcomes. *Ann Rheum Dis* 2006 May;65:629-36.
- (114) Brown AK, Roberts TE, Wakefield RJ, Karim Z, Hensor E, O'connor PJ, et al. The challenges of integrating ultrasonography into routine rheumatology practice: addressing the needs of clinical rheumatologists. *Rheumatology (Oxford)* 2007 May;46:821-9.
- (115) Filippucci E, Unlu Z, Farina A, Grassi W. Sonographic training in rheumatology: a self teaching approach. *Ann Rheum Dis* 2003 Jun;62(6):565-7.
- (116) Naredo E, Bijlsma JW, Conaghan PG, Acebes C, Balint P, Hammer HB, et al. Recommendations for the content and conduct of EULAR Musculoskeletal Ultrasound Courses. *Ann Rheum Dis* 2008 Jul;67:1017-22.
- (117) Pineda C, Filippucci E, Chavez-Lopez M, Hernandez-Diaz C, Moya C, Ventura L, et al. Ultrasound in rheumatology. The Mexican experience. *Clin Exp Rheumatol* 2008 Sep;26:929-32.
- (118) Filippucci E, Meenagh G, Ciapetti A, Iagnocco A, Taggart A, Grassi W. E-learning in ultrasonography: a web based approach. *Ann Rheum Dis* 2007Jul;66:962-5
- (119) Taggart A, Filippucci E, Wright G, Bell A, Cairns A, Meenagh G, et al. Musculoskeletal ultrasound training in rheumatology: the Belfast experience. *Rheumatology (Oxford)* 2006 Jan;45:102-5.

- (120) D'Agostino MA, Maillefert JF, Said-Nahal R, Breban M, Ravaud P, Dougados M. Detection of small joint synovitis by ultrasonography: the learning curve of rheumatologists. *Ann Rheum Dis* 2004 Oct;63:1284-7.
- (121) Wakefield RJ, Freeston JE, Hensor EM, Bryer D, Quinn MA, Emery P. Delay in imaging versus clinical response: A rationale for prolonged treatment with anti-tumor necrosis factor medication in early rheumatoid arthritis. *Arthritis Rheum* 2007 Nov 29;5:1564-7.