MAGNETIC RESONANCE AND RADIOGRAPHY IN RHEUMATOID ARTHRITIS

MAGNETIC RESONANCE AND RADIOGRAPHY IN RHEUMATOID ARTHRITIS: INTERMODALITY COMPARISONS OF EROSION DETECTION

By

RUBEN TAVARES

A thesis

Submitted to the School of Graduate Studies

in Partial Fulfillment of the Requirements

for the Degree

Doctor of Philosophy

McMaster University

Doctor of Philosophy (2012) McMaster University
(Medical Sciences – Physiology and Pharmacology) Hamilton, Ontario

TITLE: Magnetic Resonance and Radiography in Rheumatoid Arthritis: Intermodality Comparisons of Erosion Detection

AUTHOR: Ruben Tavares, MBt, MSc, HBSc, CCRA

SUPERVISOR: Dr. J.D. Adachi

Number of Pages: *xi*, 296

ABSTRACT

BACKGROUND. Classically, radiography (x-ray) has been used to visualize the destructive effects of uncontrolled rheumatoid arthritis (RA) on articular bones. Both x-ray and magnetic resonance imaging (MRI) may be used to visualize erosive disease. The multi-slice and multi-planar characteristics of MRI provide greater detail per anatomy imaged than x-ray.

OBJECTIVE. The objective of this dissertation was to compare the relative merits of x-ray and MRI erosion detection.

METHODS. In Chapters 1 through 3, RA, its clinical management, and the role of diagnostic imaging were introduced. In Chapter 4, the overall objective was first investigated by evaluating the current state of knowledge using a systematic review. In Chapter 5, inter-rater reliability across four participating radiologists was investigated. In Chapter 6, reliability-adjusted evaluations were used to directly compare paired x-ray and MR images.

RESULTS. The systematic review indicated that x-ray has low sensitivity and high specificity for MRI erosions. The associations were dependent on RA symptom duration. The findings from the prospective studies conducted were consistent with the literature. In a patient-centred analysis, the proportion of patients with erosive disease detected on either modality was dependent on the anatomy compared. Despite similar proportions of patients with erosive disease detected in comparable diagnostic imaging sittings, the proportions were comprised of markedly different patients.

CONCLUSIONS. The literature comparing x-ray and MRI erosion detection was systematically reviewed. The dependence of the relative diagnostic test accuracy on symptom duration was highlighted. At the patient level of analysis, the greater number of anatomical sites examined on x-ray overcomes the superiority of MRI to detect erosions at the unit of measurement. The two modalities are complementary insofar that they detect erosive disease in distinct patients. Further investigation into the merit of imaging varied anatomy on MR and optimally accounting for inter-rater reliability in such comparisons is warranted.

ACKNOWLEDGEMENTS

This dissertation is the result of the commitment from many individuals. First and foremost, I thank Dr. Jonathan D. Adachi, my supervisor, for the opportunity to pursue this degree and whose unwavering work ethic remains my benchmark. I thank Dr. Colin E. Webber for his academic mentorship: the weekly meetings with him were the highlight of my experience in this program. I thank Dr. Karen A. Beattie, who always provided detailed feedback and shared her experiences to facilitate arduous challenges faced over the course of this program. Dr. Maggie Larche was an invaluable resource for joint count training. Her welcoming smile made it a joy to carry out my research every day. Her open door policy enabled me to resolve issues quickly and avoid delays to the study conduct. I thank Drs. Naveen Parasu, Karen Finlay, Erik Jurriaans, and Hao Wu. Without their expert diagnostic imaging evaluations, this work would have been impossible. I am grateful for Dr. Parasu's leadership among the radiologists. Without his keen interest in the project, it would have been extremely difficult to maintain the essential radiological component of this work. Dr. Hao Wu was my sounding board and motivator during difficult times. I hope I was able to reciprocate the active listening, patience, support, and respect that he showed me. I thank the many rheumatologists, who facilitated the recruitment of patients for the study. They included Drs. Adachi, Larche, William G. Bensen, Lawrence E. Hart, Raja S. Bobba, and Alfred A. Cividino. The office personnel of participating physicians were unbelievably helpful with respect to patient scheduling and providing advice on administrative matters. Ms. Christine Fyfe, Caitlin Steven, Mary Strain, Erika Arseneau and Dr. Steve Tytus helped to operationalize the ongoing study that fed data into this study. Ms. Fyfe and Steven have been unbreakable in their perseverance in successfully bringing this three-year study to a close. Dr. Tytus and Ms. Strain helped to develop a study participant scheduling database that facilated patient contact and limited attrition. Support from the operational team through the development of standard operating procedures and other systems was key in assuring the quality of the data collected and limiting patient attrition. Ms. Erica Nunes, Mr. Craig MacDougald, Jim Bowen, Ron Goeree, and Dr. Jean-Eric Tarride at the Programs for the Assessment of Technology in Health (PATH) Research Insitute helped develop a database from the 120+ page case report form. The involvement of PATH further raised the level of data quality. Much of the latter work did not make its way into this dissertation. It must be acknowledged that the work included would not have been possible outside of the infrastructure set in place for the much larger research program. I am indepted to the 191 patients who volunteered for the larger program and the 150+ who are currently on track to complete the two-year follow-up. I acknowledge the essential grant support from the Canadian Initiative for Outcomes in Rheumatology, Canadian Arthritis Network Rapid Impact Platform Program, and Ministry of Health and Long Term Care (the latter via PATH). I thank the Canadian Arthritis Network for a graduate studentship. Nearly $250 000 was raised to support the larger research program. Without the commitment and support from all of these individuals and institutions, the innovative findings from this work would not have been possible. Thank you to all.

TABLE OF CONTENTS

LIST OF ACRONYMS

-FS	Non-fat suppression
+FS	Fat suppression
3DGE	3D gradient echo
95% CI	95% confidence interval
αCCP	Anti-citrullinated cyclic peptide
ACR	American College of Rheumatology
AICAR	Aminoimidazole carboxamide ribonucleotide
ANOVA	Analysis of variance
ARA	American Rheumatism Association
AS	Ankylosing spondylitis
BRM	Biologic response modifier
CD##	Cluster of differentiation, refer to specific genes
CDAI	Clinical disease activity index
COX	Cytochrome C Oxidase
CPPD	Calcium pyrophosphate dihydrate deposition disease
CRP	C-reactive protein
CT	Computed tomography
DAS28	28-joint disease activity score
DC	Dendritic cell(s)
DIP	Distal interphalangeal (joints)
DMARD	Disease-modifying anti-rheumatic drug(s)
EIA	Early inflammatory arthritis
eMRI	extremity magnetic resonance imaging
EQ5D	EuroQol quality of life questionnaire
ESR	Erythrocyte sedimentation rate
ESSG	European Spondyloarthropathy Study Group
EULAR	European League Against Rheumatism
FSE	Fast-spin echo
GRADE	Grading of Recommendations, Assessment, Development and Evaluation
GM-CSF	Granulocyte marcophage colony-stimulation factor
HAQ	Health Assessment Questionnaire
HERA	Hydroxychloroquine in Early Rheumatoid Arthritis
HUI-III	Health Utility Index Mark III
HLA	Human Leukocyte Antigen
HLA-DRB	The ‘shared epitope’ genetic region linked to a subset of rheumatoid arthritis patients
IA	Inflammatory arthritis
ICAM-1	A gene implicated in expression of substrates involved in rheumatoid arthritis synovial pathogenesis
IBD	Inflammatory bowel disease
ICC	Intraclass Correlation Coefficient
IFN-γ	Interferon gamma

Ig	Immunoglobulin
IL	Interleukin
IQR	Interquartile range
IR	Inversion recovery
IU	International units
JSN	Joint space narrowing
MAPK	Mitogen-activated protein kinase(s)
MCP	Metacarpophalangeal
MDC	Minimum detectable change
MeSH	Medical subject headings
MMP	Matrix metalloproteinase(s)
MSK	Musculoskeletal
MTP	Metatarsophalangeal
MTX	Methotrexate
NSAID	Non-steroidal anti-inflammatory drug(s)
NF-κB	Nuclear factor kappa B
OMERACT	Outcome Measures in Rheumatology (formerly Outcome Measures for the Evaluation of Rheumatoid Arthritis Clinical Trials)
OR	Odds ratio
PATH	Programs for the Assessment of Technology in Health Research Institute
PICO	Population, intervention, comparator, outcome
PIP	Proximal interphalangeal (joint(s))
PMN	Polymorphonuclear neutrophils
PO	Per oral route of administration
PsA	Psoriatic arthritis
QUADAS	Quality Assessment of Diagnostic Accuracy Studies
RA	Rheumatoid arthritis
RADAI	Rheumatoid Arthritis Disease Activity Index
RADAR	Rapid Assessment of Disease Activity in Rheumatology
RAMRIS	Rheumatoid arthritis magnetic resonance score
RANK	Receptor activator of NF-κB
RAPID	Routine Assessment of Patient Index Data
ReA	Reactive arthritis
RF	Rheumatoid factor
SAARD	Slow-acting anti-rheumatic drug(s)
SC	Subcutaneous
SD	Standard deviation
SDAI	Simplified Disease Activity Index
SDC	Smallest detectable change
SDD	Smallest detectable difference
Se	Sensitivity
SJC	Swollen joint count
SLE	Systemic lupus erythematosus
Sp	Specificity

STARD	STAndards for Reporting Diagnostic accuracy
SpA	Spondylarthropath(y/ies)
STIR	short tau inversion recovery
T1w	T1-weighted
T2w	T2-weighted
T_H	Helper T cells
TJC	Tender joint count
TLR	Toll-like receptor(s)
TNF	Tumour necrosis factor
US	Ultrasonography
VCAM-1	A gene implicated in expression of substrates involved in rheumatoid arthritis synovial pathogenesis
vdHSS	van der Heijde-modified Sharp score

1. Introduction

1.1. Rheumatoid arthritis

1.1.1. Definition

Rheumatoid arthritis (RA) is an idiopathic, chronic inflammatory autoimmune disease characterized by polyarticular swelling and tenderness that causes functional disability, physical deformity, morbidity, and early mortality without appropriate, pharmacological intervention (1). Early in the disease course, synovitis drives functional disability. Uncontrolled synovitis and tenosynovitis begets generally irreversible tendon, ligament, cartilage and bone damage, which results in bone misalignment and disability (1). Fortunately, a number of treatment options are effective at controlling synovitis and inhibiting erosive bone damage. Caveats to treatment efficacy exist: the timing of treatment initiation is critical - anti-rheumatic therapy is most effective when initiated proximate to symptom onset; early treatment may be challenging for the large proportion of patients who present with insidious symptom onset, in whom the diagnosis is difficult to ascertain; heterogeneity in the response to therapy requires individually tailored treatment regimens; time is required to determine treatment efficacy.

The practical reality is that disease progression on diagnostic imaging remains an essential avoidance measure of therapeutic benefit. Evidence linking disease progression on diagnostic imaging to long-term functional disability is mainly derived from radiological evidence of erosions. The radiographic signal characterized by stark, juxta-articular, demarginated bone loss results from synovitis-mediated bone resorption. It is generally held that the MRI erosion signal represents the radiographic signal. Erosions

on MRI may be detected earlier in the disease course owing to the multiple cross-sectional segments of bone that can be imaged compared to the two-dimensional rendering that radiographs are limited to. The objective of this doctoral dissertation was to compare erosion detection on MRI and radiography.

1.1.2. Characteristics

The hallmarks of RA include symmetrical joint swelling and tenderness, morning stiffness, positive rheumatoid factor (RF), elevated acute phase reactants, and radiographic evidence of erosive bone loss. Symmetrical involvement of joint groups along the sagittal anatomical plane is common. The disease frequently involves the metacarpophalangeal (MCP), metatarsophalangeal (MTP), and/or radioulnar (wrist) joints initially. Stiffness in the morning or after prolonged periods of inactivity lasting more than 30 minutes (and moreso 60 minutes) is indicative of disease. Destructive articular bone damage (erosions) visualized on radiography, MRI, computed tomography (CT), or ultrasonography (US) develops over the first few years of disease. Disease symptoms are associated with functional decline that persist in the long-term due to resulting physical deformity, if uncontrolled. Rheumatoid arthritis is classically known as a crippling disease, visualized by bilateral hand deformity but extensively involving the lower extremities, resulting in wheelchair dependence.

Laboratory measures of systemic inflammation, such as C-reactive protein (CRP) or erythrocyte sedimentation rate (ESR) may be within normal ranges near symptom onset and become abnormally elevated with persistent disease (2,3). These acute phase

reactants are liver proteins, or surrogates thereof, that are produced in response to infection or inflammation. The principal liver proteins produced include fibrinogen, CRP, amyloid A protein, amyloid P protein, haptoglobin, ferritin, ceruloplasmin, and others. Whereas CRP is a direct measure of systemic inflammation, ESR is a surrogate measure, directly proportional to fibrinogen concentration. Neither test is a perfect measure of systemic inflammation. Although CRP is a direct measure, is sensitive to change and responsive to the presentation and withdrawal of the inflammatory stimulus, CRP levels may be confounded in patients with heart disease, especially those with a recent myocardial infarction. As a result of cardiac tissue damage, CRP is released into the circulation. In turn, ESR is a specimen-sensitive test and may be artificially elevated in patients with anemia (4).

Similarly, RF titre is frequently within the normal laboratory reference range near disease onset.[1] Up to 80% of the prevalent RA population has an abnormally elevated RF titre. A distinct 80% of the prevalent RA population has an abnormally elevated titre of anti-citrullinated cyclic peptide (αCCP) antibodies (5). The αCCP abnormality presents earlier than RF (6,7). Rheumatiod arthritis may be associated with low-grade anemia, fatigue, and sicca symptoms (1). Other extra-articular features, including full-blown secondary Sjogren's syndrome, rheumatoid nodules and vasculitis, are extra-articular manifestations of long-term disease (1). Early in the disease process functional disability

[1] The distinction between "disease" and "symptom" onset is not clearly differentiated in the literature and is problematic given that RA symptoms are not specific. In this thesis dissertation, the two expressions were used synonymously.

results from uncontrolled pain and synovitis while long-term disability is correlated with erosive joint damage (8,9).

An early diagnosis of RA is made clinically. Initially, inflammation must be distinguished from mechanical derangements and osteoarthritis (1). Concomitant prevalent non-inflammatory musculoskeletal (MSK) disorders, such as osteoarthritis and fibromyalgia, may hinder a diagnosis due to overlapping symptoms. Crystal-induced arthropathy (such as gout or calcium pyrophosphate dihydrate deposition disease) must also be ruled out (10,11). Clinical querying and assessment of joint swelling, tenderness, stiffness, pain with motion, and warmth indicate active inflammation and may help to rule out these non-inflammatory conditions (1). Early diagnosis may be hindered by the insidious onset of symptoms for the majority patients. Symptoms may wax and wane for prolonged periods prior to evolving into definite RA, remitting indefinitely, or differentiating into a non-RA diagnosis. In the presence of concomitant MSK conditions, the diagnosis of insidious onset RA may be obscured. Patient self-assessment of severe or prolonged stiffness in the morning or after rest is indicative of an inflammatory process (1,12-14). Episodes of morning stiffness enduring less than 30 minutes are more typical of osteoarthritis. Connective tissue disease diagnoses, such as systemic lupus erythematosus, polymyositis, and polymyalgia rheumatica, may have inflammatory involvement (11) and need to be ruled out. Likewise, palindromic rheumatism, viral arthritis, spondyloarthropathies and inflammatory bowel disease with peripheral arthritis manifestations must all be ruled out. In the subset of patients who present with acute polyarthritis, an RA diagnosis may be more readily made once an infectious etiology is

ruled out. Palindromic rheumatism is episodic, with episodes enduring in the order of days followed by complete remission and biannual relapse. Since 25-50% of palindromic rheumatism eventually evolves into RA, the disease presents a conundrum. Symptom chronicity greater than six weeks is sufficient to rule out viral arthritis and disease persistence greater than three months is predictive of chronic disease. Spondyloarthropathies are generally RF and αCCP seronegative, have non-symmetrical joint involvement, and frequently involve the axial skeleton. Despite challenges in discerning spondyloarthropathies from RA, they share some pharmacotherapeutic approaches in common. Although an initial positive clinical response to non-steroidal anti-inflammatory drugs (NSAID) is common for RA, it may delay the diagnosis and more appropriate therapy with disease-modifying anti-rheumatic drugs (DMARD). Generally, clinical, laboratory, and diagnostic imaging evidence may be utilized to ascertain prevalent disease; however, an incident diagnosis is often limited to clinical findings due to negative laboratory and diagnostic imaging findings very early in the disease course. Ultimately, like many rheumatic diseases, RA has a heterogeneous phenotype that is initially diagnosed clinically due to the unavailability of sensitive and specific laboratory and diagnostic imaging tests. Further investigations into the ability of imaging modalities to differentiate between inflammatory and self-limiting conditions may facilitate early, appropriate intervention for RA patients.

1.1.3. Classification

Rheumatoid arthritis has a heterogeneous phenotype. Classification criteria for RA have been developed and revised over the past 60 years in attempts to homogenize this heterogenous disease (15-17). The original 1958 criteria proposed by the American Rheumatism Association (ARA) were labeled as 'diagnostic' criteria that classified patients into 'classical', 'definite', 'probable', and 'possible' RA groups (15). The criteria were unfulfilling for a number of reasons: 1) there were several exclusions that inappropriately ruled out overlapping disease, e.g. with gout or systemic lupus erythematus; 2) the diagnostic performance of all classifications was unproven and suspect for the latter two categories; and 3) over time, invasive tests included in the criteria were considered inappropriate for implementation at the population level (18). Such limitations led to the development of the Rome (19) and New York (20) criteria for RA amenable to implementation across large cohorts. The Rome and New York criteria were not widely accepted and were succeeded by the 1987 revision by ARA (16). The latter criteria were developed using an evidence-based approach as opposed to the consensus and expert-opinion-based approaches it succeeded. Similar to the Rome and New York criteria, however, 1987 ARA revision was disseminated as an approach to define a reproducible subset of RA patients for clinical trials to enable inter-study comparisons of the safety and efficacy of interventions (16). The original criteria were labeled as diagnostic in nature. The performance of the 1987 criteria was acceptable for prevalent RA (16); however, it performed less-well for incident disease (21). Recently, new classification criteria were developed for the purposes of expediting early

intervention for RA (17). True to the 1987 revision, the latest attempt has two forms with unique classification properties. The most recent criteria included the patient-ascribed assessment of tenderness in the definition of joint involvement, diminished the importance of symptom longevity, and eliminated the diagnostic imaging and morning stiffness components (17). Also, the tree format necessitates positive serology for either an acute phase reactant, αCCP, or RF. Since the 1958 criteria, the dissemination of each revision has avoided claims of diagnostic utility.

In 2002, Emery *et al*. used an evidence-based approach to recommend that patients presenting with at least 3 swollen joints, a positive compression test of either the MCP or MTP joints, or at least 30 minutes of morning stiffness be referred to rheumatology for the assessment of RA (22). It was recently determined that this recommendation was strongly associated with rheumatologic opinion on both 'early RA' as an entity, and the recommendation to treat with a DMARD (including biologic response modifiers, BRMs) (23). In a prospective trial designed and executed by the author during the course of his PhD program, the Emery *et al*. recommendation was used as an all-encompassing criteria set for early and prevalent disease (see **Appendix A**). Exclusion criteria were used, but not limited to the purposes of excluding patients with diagnoses in the differential to RA.

1.1.4. Epidemiology

1.1.4.1. Burden of illness

Consuming 11.2% of the total societal healthcare costs, MSK conditions constitute the leading economic burden of illness in Canada (24). Arthritis affects 16% of the Canadian population and costs $6.4 billion annually (25). In the US, arthritis is the leading cause of disability (26). Up to 3% of the population suffers from the functional disability, physical disfigurement, stiffness, fatigue, pain, and quality of life declines caused by IA (24-27). Rheumatoid arthritis is the most prevalent and destructive form of inflammatory arthritis (IA) affecting 1% of the population and costing up to $19 000 per patient year (27). The disease causes long-term functional disability, early mortality, and an increased risk for several comorbidities (27-29). It reduces quality of life and increases work disability. The majority of short-term costs are indirectly incurred through work disability (25). In the long-term, the direct costs of hospitalizations due to comorbidities and joint replacement surgery carry its societal burden (30).

1.1.4.2. Incidence, prevalence, and demographic characteristics

In North America it is accepted that RA has a prevalence of 0.5 to 2.0% and incidence of 0.05-0.10% (1,31). Globally, the prevalence is commonly cited as 1% (27). First Nations and Inuit and populations of African descent have a higher prevalence of RA relative to Caucasians (27). Approximately two to three times more females than males are diagnosed with the disease and the mean age of RA onset is 52 years (1,27). The onset in males occurs later in life than in females (1). The disparity in incidence

between genders decreases with increasing age and is approximately equal among the elderly population (1).

Within three months of symptoms only 33% of RA patients are RF positive; within one year of disease 75% are RF positive (1); 80% of prevalent cases are RF positive. The role of smoking on the development of RA is contentious (32). Smoking is generally associated with an increased probability of RF-positivity. The inconclusive evidence around the independent effect of smoking on disease causality suggests that if one exists, its magnitude is small (32-33). Genetics, i.e. presence of the 'shared epitope' (HLA-DRB allotypes), has some diagnostic utility (3,34-39) but, only explains 13% of RA cases (2).

1.1.4.3. Prognostic factors

Prognostic factors refer to demographic, clinical, laboratory and diagnostic imaging indicators of declines in disease status, or disease progression. In 2008, the American College of Rheumatology (ACR; formerly the ARA) published RA treatment guidelines for DMARDs that stratified recommendations by prognostic factors (40). The prognostic factors included,

1) 28-joint Disease Activity Score (DAS28);
2) evidence of radiographic erosions;
3) RF or αCCP titre;
4) elevated acute phase reactants;
5) increasing age;
6) female sex;
7) cigarette smoking;

8) genetic predisposition (HLA DRB1 positivity); and,
9) increased RA-related physical disability as measured by the Health Assessment Questionnaire.

Six years earlier, the revised clinical management guidelines from the American College of Rheumatology emphasized age of disease onset, elevated acute phase reactants, RF seropositivity, and swelling of >20 total joints as indicators of poor prognosis, for which aggressive therapy was recommended (41). Other clinical management guidelines also specify elevated acute phase reactants, RF or αCCP seropositivity, radiographic evidence of erosive damage, elevated swollen joint count, presence of the shared epitope, and duration of cigarette smoking as a set of prognostic indicators" (42,43).

Within six weeks from symptom onset infectious arthritis, of viral or bacterial origins, is a common diagnosis in the differential (44). Further, spontaneous remission or remission induced by a short course of corticosteroids is common within three months from symptom onset (3). Generally, the diagnoses in the differential either have more favourable prognoses, or require a distinct therapeutic approach. Luckily, many of the peripheral inflammatory arthritides are recommended a similar pharmacotherapeutic approach and the diagnosis may be less important. Early inflammatory arthritis is considered to include early presentations of peripheral inflammatory arthritides, including RA, psoriatic arthritis, and undifferentiated arthritis (23). Intervention with one or more DMARD is the consensus first-line pharmacotherapeutic approach for RA patients (23).

Over the past 15 years there has been a movement towards early, intensive therapy contentiously characterized by treatment with a combination DMARD regimen within the 'Window of Opportunity' (i.e. three months from symptom onset) (45). The

challenges with early diagnosis render the implementation of an early, intensive treatment regimen within three months from symptom onset difficult without either 1) risking the over-treatment of patients with diagnoses in the differential; 2) treating undifferentiated IA rather than RA; or 3) limiting intervention to the subset of patients that present with acute onset of polyarthritis. For a sizeable proportion of incident RA, the establishment of a definitive diagnosis remains a challenge that results in heterogeneous classification.

Specific functional disability measurement tools exist. The Health Assessment Questionnaire (HAQ), a popular and widely validated, self-administered questionnaire based on a conceptual framework for function in arthritis patients. The HAQ score may reflect both permanent and reversible physical limitations. In contrast, radiography measures physical abnormalities associated with long-term physical and functional disability.

1.1.5. Pathophysiology

Rheumatoid arthritis is an idiopathic disease linking the immune and inflammatory systems (46). Although RA may exhibit extra-articular features, such as vasculitis, rheumatoid nodules and Sjorgen's syndrome, and may affect the spine (46), the disease primarily involves synovial articulations, including condyloid (e.g. wrist), hinge (e.g. humeroulnar), ball-and-socket (e.g. hip), saddle (e.g. metacarpophalangeal), and pivot (e.g. atlantoaxial) joints. Normal synovium is 1-3 cell layers thick and spans the joint cavity (47). In RA, synovial cell proliferation and inflammatory cell infiltration, in part through angiogenesis, increases the lining thickness to 10-15 cell layers (48,49).

The pathogenesis of the enigmatic autoimmune trigger is hypothesized from fundamental physiology. The cellular types implicated in the RA pathophysiology are numerous and include, dendritic cells, T cells, mast cells, synovial fibroblasts, monocytes and macrophages, B cells, neutrophils, and epithelial cells. The earliest pathogenic event may include triggering of dendritic cells (DC) through toll-like receptors (TLR) expressed on synovial cells by the elusive stimulus, or a combination of said stimulus and a self-antigen, either before or in parallel to T-cell involvement (46). Positive feedback mechanisms perpetuate the pro-inflammatory secretion of chemokines and cytokines by synovial fibroblasts, monocytes, macrophages, and B cells. The attachment of the proliferating pannus to the cartilage and bone surfaces is facilitated by the induction of ICAM-1, VCAM-1 to produce intercellular and cell-bone-matrix implicated selectins, cadherins, and integrins. TNF-α, IL-1 and IL-18 secreted by macrophages and dendritic cells induce the biosynthesis of aquaporins by fibroblasts leading the hydrarthrosis and bone damage. TNF-α also acts as an autocrine stimulator (50). TNF-α and IL-1 also induce the expression of the receptor activator of NF-κB (RANK) on macrophages, or osteoclast precursor cells. In conjunction with RANK ligand expression by synovial fibroblasts, T cells or stromal cells, macrophages differentiate into activated osteoclasts. Chrondocytes are activated and secrete cartilage-degrading matrix metalloproteinases (MMP). B cells are implicated in the propagation of the inflammatory process and their role is not well understood. The depletion of B cell populations, however, reduces inflammation and inhibits bone resorption.

Synovial fibroblasts promote chronic inflammation through the release of cytokines and chemokines, proangiogenic factors, and matrix-degrading enzymes (47). Monocytes are recruited to the synovium, where they differentiate into activated macrophages. B cells are recruited into the synovium where they are involved in rheumatoid factor antigen recognition, autoantibody proliferation, and are a potential source of cytokine and chemokine regulation. Cellular infiltration and proliferation results in tissue hyperplasia and hyperplasticity. The hyperplastic tissue forms a pannus, covering the cartilage and bony surfaces, exposing inflammatory and synovial cells to the site of cartilage and bone damage. Chronic inflammation in synovial joints eventually results in cartilage and bone destruction.

The obscure pathogenic antigen may initially activate the immune system by binding TLRs or CD14 to trigger a T-cell response (46). The resulting T-cell response may be polarized towards either T_H1 or T_H2 cell activity and decisively results in B cell activation. T_H1 cells are most strongly induced by interleukin (IL)-12; they release interferon-γ (IFN-γ) and tumour necrosis factor–β (TNF-β) and stimulate pro-inflammatory activities and certain humoral responses. In an appropriate epigenetic environment, T_H1 cells are activated by a self-antigen-mimicking, arthritogenic antigen presented by either dendritic cells, neutrophils, or B cells. In turn, T_H2 activity is induced by IL-4. These cells secrete IL-4, IL-5, IL-10, and IL-13; they have anti-inflammatory potential and promote other humoral responses, including the production of immunoglobulin (Ig) E (46). Through cell-cell interaction, such as CD11- and CD69-mediation, or by the release of by INF-γ, TNF-α, and IL-17, T_H1 cells activate

monocytes, macrophages, and synovial fibroblasts (46). The latter cells proliferate, forming the synovial pannus, and release TNF-α, IL-1 and IL-6 leading to chronic inflammation (46). Other cytokines and chemokines involved in the propagation of RA inflammation include IL-15, IL-18 and pro-angiogenic factors (46). Receptor-bound, these molecules trigger various signal transduction cascades, such as MAPK and RANK to induce pro-inflammatory, pro-bone resorption, and anti-bone forming processes (46).

The primary cellular composition of healthy synovium includes fibroblasts and macrophages. The physiological functions of synovial fibroblasts in a healthy state are the following: provide lubricants to the joint cavity and nutrients to cartilage; the biosynthesis of matrix components and degrading enzymes for matrix remodeling; and, contribute to tissue repair (47). In the disease state, RA synovial fibroblasts secrete cytokines and chemokines, proangiogenic factors and matrix-degrading enzymes to contribute to inflammation, angiogenesis and matrix degradation, respectively (47). Fibroblasts express HLA DR molecules resulting in a destructive phenotype independent of the local inflammatory environment (47).

Macrophages have the ability to activate T cells by antigen presentation (51). Monocytes, macrophage precursors, are recuited to the synovium where they differentiate into tissue-specific host defence cells. Macrophages are a major source of pro-inflammatory mediators, including cytokines, chemokines, and MMPs.

In healthy persons, neutrophils, sometimes referred to as polymorphonuclear neutrophils (PMN), constitute approximately 60% of white blood cells and provide the

primary line of defence against blood- or tissue-borne pathogens. Neutrophils are phagocytic: they eliminate pathogens by engulfing them and fusing the endocytosed phagosome with lysosomes. The lysosomic contents are acidic and contain digestive enzymes that degrade the pathogen.

Neutrophils exist in two phases: resting and activated. The resting phase ensures that their lysosomic contents do not damage host tissue in a pathogen free environment. Neutrophils are activated by a two-stage process, involving priming by bacterial products, and cytokines or chemokines, such as TNF-α, granulocyte marcophage colony-stimulation factor (GM-CSF), IL-8, and IFN-γ (51). Two priming mechanisms exist: first, intracellular granules with preformed receptors may be translocated to the cell surface; second, transcription factors may be induced to promote receptor and cytokine genes to enhance neutrophil function or lifespan (51). Neutrophils comprise the majority of cells in the synovial fluid of aspirated RA joints and are also present at the pannus/cartilage interface (52,53). Activated neutrophils carryout pro-inflammatory functions similar to macrophages, including antigen presentation, cytokine and chemokine release. Many of the cytokines and chemokines implicated in RA are potent modulators of neutrophilic activity (51). Interestingly, the modulatory effect of TNF-α is concentration dependent. At high concentrations, it induces apoptosis negating any cellular regulatory effects but potentially resulting in the release of its cytotoxic contents into the microenvironment contributing to disease-related tissue damage. At low TNF-α concentrations, the effect is biphasic: promoting apoptosis in a subset of cells but inhibiting apoptosis in the rest. In murine RA models, neutrophil depletion spares pannus

formation, synovial hyperplasia, joint swelling and cartilage destruction that otherwise prevails (54). Other animal studies inhibiting the synovial infiltration of neutrophils have reported similar results (55-58). Given their dynamic function, delayed apoptosis, and abundance, neutrophils may have been overlooked historically as important mediators of the chronicity and inter-articular migration of RA.

1.1.6. Summary

The pathophysiology of RA remains an area of intense research. The deregulated overproduction of proinflammatory cytokines, such as TNFα, IL-1, and IL-6, and disequilibrium with anti-inflammatory cytokines contribute to its pathophysiology (59). Our understanding of the disease processes has been sufficient to pursue therapeutic strategies targeting specific cytokines, chemokines, receptors, enzymes of the cells described above. In 2003, Smolen and Steiner summarized the ubiquitous therapeutic targets under investigation at the time. Nearly a decade later, the targets that have proven successful are limited to TNF-α, IL-1, IL-6, CD20, and CD28. Conventional DMARD are variably effective at controlling both inflammatory and structural disease progression; however, their mechanisms of action are less clearly understood than contemporary biologic therapies. To a lesser extent, COX(I) and COX(II) inhibitors have been effective for symptomatic control without any effect on destructive tissue damage. Corticosteroids, which have generalized anti-inflammatory effects, remain important options in the RA pharmacotherapeutic armamentarium. The role of DMARD, NSAID and corticosteroids in the clinical management of RA is summarized in Chapter 2.

1.2. References

1. O'Dell JR. Chapter 10. Rheumatoid arthritis: clinical aspects. *In* Koopman WJ, Boulware DW, Heudebert GR. Clinical Primer of Rheumatology. Philadelphia, Lippincott Williams and Wilkins, 2003. pp. 97-115.
2. Emery P. Prognosis in inflammatory arthritis: the value of HLA genotyping and oncological analogy. The Dunlop Dotteridge Lecture. J Rheumatol 1997;24:1436-42.
3. Green MJ, Marzo-Ortega H, McGonagle D, Wakefield R, Proudman S, Conaghan P, Gooi J, Emery P. Persistence of mild early in¯ammatory arthritis: the importance of disease duration, rheumatoid factor and the shared epitope. Arthritis Rheum 1999;42:2184-8.
4. Wener MH. Chapter 7: Diagnostic Laboratory Tests and Imaging. In Robbins L, Burckhardt CS, Hannan MT, deHoratius RJ (editors). Clinical Care in the Rheumatic Diseases. 2nd edition. Atlanta, Association of Rheumatology Health Professionals, 2001.
5. Nishimura K, Sugiyama D, Kogata Y, Tsuji G, Nakazawa T, Kawano S, Saigo K, Morinobu A, Koshiba M, Kuntz KM, Kamae I, Kumagai S. Meta-analysis: diagnostic accuracy of anti-cyclic citrullinated peptide antibody and rheumatoid factor for rheumatoid arthritis. Ann Intern Med 2007;146:797–808.
6. Jansen AL, van der Horst-Bruinsma I, van Schaardenburg D, van de Stadt RJ, de Koning MH, Dijkmans BA. Rheumatoid factor and antibodies to cyclic citrullinated Peptide differentiate rheumatoid arthritis from undifferentiated polyarthritis in patients with early arthritis. J Rheumatol 2002;29:2074–6.
7. Saraux A, Berthelot JM, Chales G, et al. Value of laboratory tests in early prediction of rheumatoid arthritis. Arthritis Rheum 2002;47:155–65.
8. Scott DL, Pugner K, Kaarela K, Doyle DV, Woolf A, Holmes J, Hieke K. The links between joint damage and disability in rheumatoid arthritis. Rheumatology 2000;39:122-32.
9. Scott DL, Smith C, Kingsley G. Joint damage and disability in rheumatoid arthritis: an updated systematic review. Clin Exp Rheumatol 2003;21(5 Suppl 31):S20-7.
10. Dore RK, Clements PJ, Fox RI, Furst DE, Kaplan H, Kitridou RC. Guidelines for Rheumatology Referral. Atlanta, American College of Rheumatology, 1996.
11. Holbrook AM (Chair) for Ontario Musculoskeletal Therapy Review Panel. Ontario Treatment Guidelines for Osteoarthritis, Rheumatoid Arthritis, and Acute Musculoskeletal Injury. Guideline. Toronto, Queen's Printer of Ontario, 2000.
12. Lineker S, Badley E, Charles C, Hart L, Streiner D. Defining morning stiffness in rheumatoid arthritis. J Rheumatol 1999;26:1052-7.
13. van der Helm-van Mil AH, le Cessie S, van Dongen H, Breedveld FC, Toes RE, Huizinga TW. A prediction rule for disease outcome in patients with recent-onset

undifferentiated arthritis: how to guide individual treatment decisions. Arthritis Rheum 2007;56:433-40.

14. van der Helm-van Mil AH, Detert J, le Cessie S, Filer A, Bastian H, Burmester GR, Huizinga TW, Raza K. Validation of a prediction rule for disease outcome in patients with recent onset undifferentiated arthritis: moving towards individualized treatment decision-making. Arthritis Rheum 2008;58:2241-7.
15. Ropes MW, Bennett GA, Cobb S, Jacox R, Jessar RA. Revision of diagnostic criteria for rheumatoid arthritis. Bull Rheum Dis 1958;9:175-6.
16. Arnett FC, Edworthy SM, Bloch DA, McShane DJ, Fries JF, Cooper NS, Healey LA, Kaplan SR, Liang MH, Luthra HS, Medsger TA Jr, Mitchell DM, Neustadt DH, Pinals RS, Schaller JG, Sharp JT, Wilder RL, Hunder GG. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. Arthritis Rheum 1988;31:315-24.
17. Aletaha D, Neogi T, Silman AJ, Funovits J, Felson DT, Bingham CO, 3rd, Birnbaum NS, Burmester GR, Bykerk VP, Cohen MD, Combe B, Costenbader KH, Dougados M, Emery P, Ferraccioli G, Hazes JM, Hobbs K, Huizinga TW, Kavanaugh A, Kay J, Kvien TK, Laing T, Mease P, Ménard HA, Moreland LW, Naden RL, Pincus T, Smolen JS, Stanislawska-Biernat E, Symmons D, Tak PP, Upchurch KS, Vencovský J, Wolfe F, Hawker G. 2010 Rheumatoid arthritis classification criteria: An American College of Rheumatology/European League Against Rheumatism Collaborative Initiative. Arthritis Rheum 2010;62:2569-81.
18. Symmons DPM. What is rheumatoid arthritis? Br Med Bull 1995;51:243-8.
19. Kellgren JH. Diagnostic criteria for population studies. Bull Rheum Dis 1962;13:291-2.
20. Bennett PH, Burch TA. New York symposium on population studies in the rheumatic diseases: new diagnostic criteria. Bull Rheum Dis 1967;17:453-8.
21. Dugowson CE, Nelson JL, Koepsell TD. Evaluation of the 1987 revised criteria for rheumatoid arthritis in a cohort of newly diagnosed female patients. Arthritis Rheum 1990;33:1042-6.
22. Emery P, Breedveld FC, Dougados M, Kalden JR, Schiff MH, Smolen JS. Early referral recommendation for newly diagnosed rheumatoid arthritis: evidence based development of a clinical guide. Ann Rheum Dis 2002;61:290-7.
23. Tavares R, Wells GA, Scarf J, Cagaanan R, Bykerk VP, Bell MJ. Classification of Rheumatologic Opinion on Early Inflammatory Arthritis: Harmonization of a Heterogeneous Standard. Arthritis Care Res 2010;62:1407-14.
24. Public Health Agency of Canada. Population Health Economics. Ottawa, Public Health Agency of Canada, 2010.
25. Public Health Agency of Canada. Life with Arthritis in Canada: A personal and public health challenge. Ottawa, Public Health Agency of Canada, 2010.

26. Centers for Disease Control and Prevention. National and state medical expenditures and lost earnings attributable to arthritis and other rheumatic conditions - United States, 2003. MMWR Morb Mortal Wkly Rep 2007:56:4-7.

27. Kvien TK. Epidemiology and burden of illness of rheumatoid arthritis. Pharmacoeconomics 2004;22(2 Suppl):1-12.

28. Ward MM, Leigh JP, Fries JF. Progression of functional disability in patients with rheumatoid arthritis. Associations with rheumatology subspecialty care. Arch Intern Med 1993;153:2229-37.

29. Scott DL, Smith C, Kingsley G. What are the consequences of early rheumatoid arthritis for the individual? Best Pract Res Clin Rheumatol 2005;19(1):117-36.

30. Fries JF, Williams CA, Morfeld D, Singh G, Sibley J. Reduction in long-term disability in patients with rheumatoid arthritis by disease-modifying antirheumatic drug-based treatment strategies. Arthritis Rheum 1996;39:616-22.

31. Centers for Disease Control and Prevention. Prevalence of arthritis-United States, 1997. Morb Mortal Wkly Rep 2001;50:334-6.

32. Symmons DP, Bankhead CR, Harrison BJ, Brennan P, Barrett EM, Scott DG, Silman AJ. Blood transfusion, smoking, and obesity as risk factors for the development of rheumatoid arthritis: results from a primary care-based incident case-control study in Norfolk, England. Arthritis Rheum 1997;40:1955-61.

33. Symmons DP. Epidemiology of rheumatoid arthritis: determinants of onset, persistence and outcome. Best Pract Res Clin Rheumatol 2002;16:707-22.

34. Visser H, le Cessie S, Vos K, Breedveld FC, Hazes JM. How to diagnose rheumatoid arthritis early: a prediction model for persistent (erosive) arthritis. Arthritis Rheum 2002;46:357-65.

35. Stockman A, Tait BD, Wolfe R, Brand CA, Rowley MJ, Varney MD, Buchbinder R, Muirden KD. Clinical, laboratory and genetic markers associated with erosions and remission in patients with early inflammatory arthritis: a prospective cohort study. Rheumatol Int 2006;26:500-9.

36. Nell VP, Machold KP, Stamm TA, Eberl G, Heinzl H, Uffmann M, Smolen JS, Steiner G. Autoantibody profiling as early diagnostic and prognostic tool for rheumatoid arthritis. Ann Rheum Dis 2005;64:1731-6.

37. van Gaalen FA, Linn-Rasker SP, van Venrooij WJ, de Jong BA, Breedveld FC, Verweij CL, Toes RE, Huizinga TW. Autoantibodies to cyclic citrullinated peptides predict progression to rheumatoid arthritis in patients with undifferentiated arthritis: a prospective cohort study. Arthritis Rheum 2004;50:709-15.

38. van Venrooij WJ, Hazes JM, Visser H. Anticitrullinated protein/peptide antibody and its role in the diagnosis and prognosis of early rheumatoid arthritis. Neth J Med 2002;60:383-8.

39. Jansen LM, van Schaardenburg D, van der Horst-Bruinsma I, van der Stadt RJ, de Koning MH, Dijkmans BA. The predictive value of anti-cyclic citrullinated peptide antibodies in early arthritis. J Rheumatol 2003;30:1691-5.

40. Saag KG, Teng GG, Patkar NM, Anuntiyo J, Finney C, Curtis JR, Paulus HE, Mudano A, Pisu M, Elkins-Melton M, Outman R, Allison JJ, Suarez Almazor M, Bridges SL Jr, Chatham WW, Hochberg M, MacLean C, Mikuls T, Moreland LW, O'Dell J, Turkiewicz AM, Furst DE, American College of Rheumatology. American College of Rheumatology 2008 recommendations for the use of nonbiologic and biologic disease-modifying antirheumatic drugs in rheumatoid arthritis. Arthritis Rheum 2008;59:762-84.

41. American College of Rheumatology Subcommittee on Rheumatoid Arthritis Guidelines. Guidelines for the management of rheumatoid arthritis: 2002 Update. Arthritis Rheum 2002;46:328-46.

42. Combe B, Landewe R, Lukas C, Bolosiu HD, Breedveld F, Dougados M, Emery P, Ferraccioli G, Hazes JM, Klareskog L, Machold K, Martin-Mola E, Nielsen H, Silman A, Smolen J, Yazici H. EULAR recommendations for the management of early arthritis: report of a task force of the European Standing Committee for International Clinical Studies Including Therapeutics. Ann Rheum Dis 2007;66:34-45.

43. Pham T, Gossec L, Fautrel B, Combe B, Flipo RM, Goupille P, Le Loët X, Mariette X, Puéchal X, Wendling D, Schaeverbeke T, Sibilia J, Sany J, Dougados M. Physical examination and laboratory tests in the management of patients with rheumatoid arthritis: development of recommendations for clinical practice based on published evidence and expert opinion. Joint Bone Spine 2005;72:222-8.

44. Quinn MA, Green MJ, Conaghan P, Emery P. How do you diagnose rheumatoid arthritis early? Best Pract Res Clin Rheumatol 2001;15:49-66.

45. Quinn MA, Emery P. Window of opportunity in early rheumatoid arthritis: possibility of altering the disease process with early intervention. Clin Exp Rheumatol 2003;21(5 Suppl 31):S154-7.

46. Smolen JS, Steiner G. Therapeutic strategies for rheumatoid arthritis. Nat Rev Drug Discov 2003;2:473-88.

47. Neumann E, Lefevre S, Zimmerman B, Gay S, Muller-Ladner U. Rheumatoid arthritis progression mediated by activated synovial fibroblasts. Trends Mol Med 2010;16:458-68.

48. Noss EH, Brenner MB. The role and therapeutic implications of fibroblast-like synoviocytes in inflammation and cartilage erosion in rheumatoid arthritis. Immunol Rev 2008;223:252-70.

49. Muller-Ladner U, Ospelt C, Gay S, Distler O, Pap T. Cells of the synovium in rheumatoid arthritis. Synovial fibroblasts. Arthritis Res Ther 2007;9:223.

50. Tak PP, Kalden JR. Advances in rheumatology: new targeted therapeutics. Arthritis Res Ther 2011;13(Suppl 1):S5.

51. Wright HL, Moots RJ, Bucknall RC, Edwards SW. Neutrophil function in inflammation and inflammatory diseases. Rheumatology 2010;49:1618-31.

52. Mohr W, Westerhellweg H, Wessinghage D. Polymorphonuclear granulocytes in rheumatic tissue destruction. III. an electron microscopic study of pmns at the pannus–cartilage junction in rheumatoid arthritis. Ann Rheum Dis 1981;40:396–9.

53. Wittkowski H, Foell D, af Klint E, De Rycke L, De Keyser F, Frosch M, Ulfgren AK, Roth J. Effects of intra-articular corticosteroids and anti-TNF therapy on neutrophil activation in rheumatoid arthritis. Ann Rheum Dis 2007;66(8):1020–5.

54. Wipke BT, Allen PM. Essential role of neutrophils in the initiation and progression of a murine model of rheumatoid arthritis. J Immunol 2001;167:1601–8.

55. Eyles JL, Hickey MJ, Norman MU, Croker BA, Roberts AW, Drake SF, James WG, Metcalf D, Campbell IK, Wicks IP. A key role for G-CSF-induced neutrophil production and trafficking during inflammatory arthritis. Blood 2008;112:5193–201.

56. Grespan R, Fukada SY, Lemos HP, Vieira SM, Napimoga MH, Teixeira MM, Fraser AR, Liew FY, McInnes IB, Cunha FQ. CXCR2-specific chemokines mediate leukotriene B4-dependent recruitment of neutrophils to inflamed joints in mice with antigen-induced arthritis. Arthritis Rheum 2008;58:2030–40.

57. Coelho FM, Pinho V, Amaral FA, Sachs D, Costa VV, Rodrigues DH, Vieira AT, Silva TA, Souza DG, Bertini R, Teixeira AL, Teixeira MM. The chemokinereceptors CXCR1/ CXCR2 modulate antigen-induced arthritis by regulating adhesion of neutrophils to the synovial microvasculature. Arthritis Rheum 2008;58:2329–37.

58. Grant EP, Picarella D, Burwell T, Delaney T, Croci A, Avitahl N, Humbles AA, Gutierrez-Ramos J-C, Briskin M, Gerard C, Coyle AJ. Essential role for the C5a receptor in regulating the effector phase of synovial infiltration and joint destruction in experimental arthritis. J Exp Med 2002;196:1461–71.

59. Nishimoto N. Interleukin-6 as a therapeutic target in candidate inflammatory diseases. Clin Pharmacol Ther. 2010;87:483-7.

2. Pharmacotherapy

2.1. Overview

The goal of RA disease management is to preserve long-term patient function. Diagnostic imaging evidence of disease progression, a surrogate for long-term functional disability is measured periodically to ensure treatment efficacy. In turn, in routine clinical management, measures of disease activity and function are used to determine treatment efficacy. Pharmacotherapeutic intervention is the primary means of effectively treating RA. Disease management strategies have been the focus of substantial research.

Disease management strategies in RA have evolved over the past two decades. Initially, the approach was to treat patients with non-steroidal anti-inflammatory drugs (NSAID), introduce disease-modifying anti-rheumatic drug (DMARD) when there was evidence of erosive damage, and use corticosteroids as a rescue therapy. Historically, RA was considered a benign disease (1,2) and the NSAID side effect profile was considered negligible. Epidemiologic data has since demonstrated the adverse mortality and morbidity consequences of RA (3,4). Light has been shed on the superior efficacy of DMARD and substantial gastrointestinal and cardiovascular risks associated with NSAID (5,6). As such, the so-called *treatment pyramid* approach is now obsolete. Early intervention with an intensive DMARD regimen has been shown to inhibit radiographic damage compared to do those treated solely with non-steroidal anti-rheumatic drugs (NSAIDs) (7).

Erosive damage is now considered an outcome to be avoided rather than a treatment trigger. Succeeding strategies, namely *sequential monotherapy*, *Sawtooth*,

combination therapy, and *inverted pyramid*, proposed one-size-fits-all approaches to DMARD treatment for RA patients (8). Sequential monotherapy refers to the sequential administration of DMARD monotherapy, one-by-one, until an efficacious intervention is found. Interventions are generally administered in order of increasing toxicity risk, which is roughly correlated with increasing potential efficacy. For many patients substantial time would lapse before an optimal DMARD therapy was arrived at, if at all. Many patients do not respond to monotherapy. The *Sawtooth* approach is a step-up approach, whereby if a response to monotherapy is not achieved, combination DMARD are tried. Combination therapy refers to the induction of DMARD therapy with multiple agents. The *Sawtooth* approach is at odds with clinical trials that demonstrate combination DMARD therapy to be superior to monotherapy. On the other hand, in the latter trials, a sizeable proportion of patients respond to monotherapy. Opposition to this strategy may cite the risk of overtreatment in patients that would otherwise respond to monotherapy. Subsequent strategies include the *management pyramid* and *targeted-treatment*. The *management pyramid* emphasizes the need for prompt DMARD intervention. Few would argue against this approach, which has been a major focus of research over the past decade. The *targeted-treatment* approach recommends the stratification of initial treatment based on disease severity. Despite opposition to this approach on the grounds of complexity, recommendations in recent disease management guidelines appear to be consistent with it (9). With respect to initial DMARD treatment decision-making, the *targeted-treatment* approach may best reflect current treatment decision-making by physicians in light of the heterogeneity in RA disease presentation.

Pharmacotherapeutic decision-making extends beyond the original treatment choice. Over the past 10-15 years, a number of trials began testing the efficacy of different transitional treatment decision-making approaches (10). Some studies investigated the effect of different criteria for treatment decision-making on downstream outcomes. Most of these trials have compared treatment decision-making based on a 28-joint disease activity score (DAS28) versus physician judgment. It was found that the use of formal disease activity measures to drive treatment decision-making is favoured over usual physician decision-making behaviour. The effects observed appear limited to the early disease onset population. In trials of prevalent patients, differences in outcome were not significant between patients treated using a targeted approach versus conventional physician-based treatment decision-making. Differences between targeted intervention strategies have not been directly compared. Others have tested different periods of assessment. The period of monitoring varies from trial to trial. The optimal period for monitoring needs to be reconciled with the lagtime required to ensure a treatment response. Others still have maintained both of the former and tested different pharmacological approaches.

The role of corticosteroids in RA management also warrants consideration. In early disease, Green *et al*. (1999) demonstrated that patients treated with corticosteroids were more likely to remit (11). In a recent follow-up trial, 1 in 10 patients were spared the development of RA when early IA was treated with corticosteroids (12).

The optimal approach to the clinical management of RA likely leverages many of these approaches to tailor treatment to the needs and expectations of individual patients.

Research into optimizing triage to rheumatologic care leverages the *management pyramid* approach (13-16). Recent guidelines recommend tailoring the initial DMARD regimen to the patient's prognostic profile (9). Disease activity monitoring and transitional treatment decision-making appears to be best achieved through the use of formal measure-based triggers (10). Treatment decision-making strategies based on maintaining patients below common disease activity composite thresholds work well to limit disease progression on diagnostic imaging. Despite this, 20% of patients in remission continue to progress on diagnostic imaging (17). As such, the assessment of disease progression on diagnostic imaging remains an important aspect of monitoring treatment efficacy. The role of glucocorticoids in early intervention and as a rescue or bridge therapy in patients with suboptimal response, and an effective agent to treat acute synovitis appear important. The use of NSAID for symptomatic relief of pain remains important as well.

The pharmacotherapeutic treatment options for RA are summarized in this chapter. The DMARD class is of primary importance and developments to the *biologics* subclass have substantially positively impacted the therapeutic options and clinical management complexity in recent years. Corticosteroids and NSAID remain important treatments for RA. These pharmacotherapeutic options are summarized as well.

2.2. Disease-modifying anti-rheumatic drugs

The DMARD therapeutic class is so-named due to the efficacy of these drugs to inhibit radiographic progression in RA (18-26). DMARD only mediate the autoimmune and inflammatory processes; they have no analgesic effect (27). These drugs are often effective at lowering circulating acute phase reactants. DMARDs must often be taken for

up to three months prior to observing a clinical response. In patients responsive to DMARD, the disease frequently relapses upon discontinuation of treatment.

The drug class is divided into subcategories including slow-acting anti-rheumatic drugs (SAARD, e.g. hydroxychloroquine, sulfasalazine), immunosuppressants (e.g. methotrexate (MTX), leflunomide), and biologic response modifiers (BRM) (e.g. etanercept, infliximab). The inhibitory effects of the hydroxychloroquine on radiographic progression are limited, if not questionable. Biologics, especially in combination with MTX, have a notably superior inhibitory effect on radiographic progression. A listing of DMARD is summarized in **Table 1** (8,28-32).

Table 1. Disease modifying anti-rheumatic drugs.

Generic Drug Name	Trade Name	Route of Administration	Maximum Recommended Daily Dose
Slow-Acting Anti-Rheumatic Drugs (SAARDs)			
Gold Compounds			
Auranofin	Ridaura®	PO	6 mg/day for 3 mth, then up to 9 mg/day for 3 mth (9 mg/day)
Aurothioglucose	Solganal®	IM	10 mg 1st wk, 25 mg 2nd and 3rd wk, then 50 mg/wk to 800 mg -1 g cumulative dose reached, then 25-50 mg q2-3w, then 50 mg q3-4w (50 mg/wk)
Gold Sodium Thiomalate	Myochrysine®	IM	10 mg 1st wk, 25 mg 2nd wk, then 25-50 mg/wk until 1 g cumulative dose, then 25-50 mg/wk q2-3w for 2-20 wk, then 25-50 mg/wk q3-4w
Anti-Malarials/Antibiotics			
Chloroquine	Aralen®	IM, PO	Off-label use
Hydroxychloroquine (sulfate)	Plaquenil®	PO	400-600 mg/day, increase to optimal response, reduce to maintenance dose of 200-400 mg/day after 4-12 wk
Minocycline	Minocin®	IV, PO	Off-label use
Other SAARDs			
Penicillamine	Cuprimine®, Depen®	PO	125-250 mg/day, increase at 1-3 mth intervals to optimal dose (1-1.5 g/day)
Sulfasalazine	Salazopyrin®, SAS™	PO	0.5-1g/day, increase to weekly to 2 g/day or optimal dose (3 g/day)
Cytotoxic, Immunosuppressants			
Azathioprine	Imuran®	PO	1 mg/kg/day, increase by 0.5 mg/kg/day until optimal response or max dose (2.5 mg/kg/day)
Cyclophosphamide	Cytoxan®	IV, PO	IV, 400-1800 mg/m^2/(1-5d) repeat q2-4w; PO, 50-100 mg/m^2 as continuous therapy or 400-1000 mg/m^2 q4-5d as intermittent therapy
Cyclosporin	Sandimmune®	PO	2.5 mg/kg/day tid, after 8 wk increase by 0.5-0.75 mg/kg/day until optimal response, increase again at 12 wk if necessary (4 mg/kg/day)
Leflunomide	Arava®	PO	100 mg/day for 3 days, followed by 10-20mg/day
Methotrexate	Rheumatrex®	PO, SC	7.5–25mg/w
Biologic Response Modifiers			
Anti-Tumour Necrosis Factor-alpha Agents			
Adalimumab	Humira®	SC	40 mg q2w
Certolizumab	Cimzia®	SC	400 mg q2w for 4w; 200 mg q2w thereafter
Etanercept	Enbrel®	SC	50 mg/wk
Golimumab	Simponi®	SC	50 mg/mth
Infliximab	Remicade®	IV	3 mg/kg at 0, 2, 6 wk up to 20 mg/kg q4-8w thereafter
Pharmaceutics against Other Therapeutic Targets			
Abatacept	Orencia®	IV	<12.5 mg/kg q2-4w (12.5 mg/kg)
Anakinra	Kineret®	SC	100 mg/day
Rituximab	Rituxan®	IV	1000 mg at 0,2 wk, additional courses q16-24w
Tocilizumab	Actemra®	SC	4-8 mg/kg q4w

PO = per oral; IM = intramuscular; IV = intravenous; SC = subcutaneous; mth = month; w or wk = week; q = every; d = day.

2.2.1. Methotrexate

Methotrexate is an analog of folic acid. Its polyglutamated metabolites inhibit aminoimidazole carboxamide ribonucleotide (AICAR)-transformylase (33), an enzyme involved in the synthesis of nucleic acids. The accumulation of AICAR-transformylase induces adenosine release. It is hypothesized the anti-inflammatory effects of MTX result from the downstream adenosine receptor activation in leukocytes resulting in lesser adhesion to endothelial cells, including synoviocytes (33). MTX is administered in doses ranging from 7.5 to 25 mg/wk orally (PO) or subcutaneously (SC). Folic acid supplementation is provided to limit gastrointestinal toxicity (27). Folic acid supplementation does not alter the efficacy of MTX (33). Parenteral administration of MTX also alleviates GI toxicity and may be more efficacious than oral administration.

Methotrexate is generally considered the optimal first-line monotherapy for RA given its superior efficacy relative to other single DMARD, low cost, acceptable safety profile, and additive efficacy in combination with virtually any secondary DMARD. Methotrexate monotherapy is the commonest comparator intervention in current clinical trials as placebo intervention is considered unethical. Virtually all contemporary drug candidates are tested with the co-administration of MTX in combination. MTX-biologic combinations provide greater inhibition of radiographic progression compared to MTX alone, an effect not as pronounced with biologic monotherapy.

At the doses used in RA, the anti-inflammatory and immunosuppressive properties of MTX are realized. MTX has cytotoxic effects that are dose-dependent. Patients are monitored routinely for liver and kidney function, and cytopenia (33).

Notably, impaired kidney function may impact the clearance of MTX leading to toxicity related to metabolite accumulation. Liver cirrhosis is uncommon and liver enzyme elevations usually return to normal by promptly withholding treatment and, if appropriate, retreating at a lower dose (33). It is recommended that patients on MTX therapy have laboratory test monitoring conducted on a monthly to bimonthly basis.

2.2.2. Leflunomide

Like MTX, a metabolite of leflunomide confers the active properties of the drug. Leflunomide inhibits an enzyme that mediates the synthesis of pyrimidine nucleic bases. It inhibits lymphocyte proliferation thereby exhibiting its efficacy in RA. Leflunomide confers an inhibitory effect on radiographic progression similar to MTX (34). It is administered orally with a loading dose of 100 mg/day for 3 days and a maintenance dose ranging from 10 to 20 mg/day. The drug carries a considerable risk profile primarily manifest through reversible alopecia, skin rash, stomatitis, diarrhea, and liver enzyme elevation. Leflunomide is teratogenic and requires up to 2 years to be completely cleared from circulation following discontinuation (27). As with methotrexate, leflunomide therapy warrants monthly or bimonthly laboratory test monitoring.

2.2.3. Sulfasalazine

Sulfasalazine is a two-drug-conjugate comprised of 5-aminosalicyclic acid and sulfapyridine. The salicylate affords anti-inflammatory effects while sulfapyridine metabolites have immunomodulatory effects (33). Sulfasalazine is a popular initial first-line DMARD in the United Kingdom (35). In combination with methotrexate and

hydroxychloroquine, the triple-combination therapy offers substantial therapeutic benefit over derivative two-drug combinations in patients naïve to DMARD combination therapy (36).

Sulfasalazine is administered orally at a dose of 500mg/d and raised to 2-3 g/d over 4-6 weeks (27). The principal side effect is gastrointestinal intolerance, which may resolve with dose attenuation. Rare hematologic serious adverse events include aplastic anemia, agranulocytosis, and hemolytic anemia, for which regular monitoring of laboratory test complete blood counts is recommended. Sulfa allergy is a contraindication to sulfasalazine treatment.

2.2.4. Antimalarials

Chloroquine and hydroxychloroquine are commonly used DMARD. Antimalarials have multiple effects on the immune system (33). Chloroquine is relatively more popular in Europe and may be both more efficacious and more toxic than hydroxychloroquine (37). In Canada, hydroxychloroquine is the anti-rheumatic drug most frequently used at DMARD treatment initiation (38). Interestingly, DMARD therapy is initiated prior to the establishment of a definitive diagnosis in Canada. The favourable benefit:risk ratio for hydroxychloroquine may be considered by physicians in early DMARD administration to provide patients with some treatment as the clinical work-up to a definitive diagnosis is undergone. The potential for overtreatment of patients with self-limiting disease resulting from treatment prior to RA diagnostic confirmation remains to be determined.

As discussed above, hydroxychloroquine, sulfasalazine, and methotrexate triple combination therapy is a very effective and well-tolerated drug regimen (36). In the same study it was demonstrated that hydroxychloroquine-methotrexate combination therapy is effective in a greater proportion of patients than sulfasalazine-methotrexate (36). The belief that the effect of hydroxychloroquine on RA is modest may stem from the results of the HERA study (39,40). This Canadian multi-centre study demonstrated small but significant effect sizes for hydroxychloroquine compared to placebo in an early RA sample over 36 weeks. These results were obtained despite a tendency to use more intra-articular corticosteroids in the placebo group ($p = 0.052$) (39). The anti-inflammatory and immunosuppressive effects of glucocorticoids are discussed later in this chapter. Many of the clinical effects were maintained at three years post-treatment (40).

Hydrochloroquine sulfate is administered orally and initiated at a dose ranging from 400-600 mg/d and reduced 200-400 mg/d after 4-6 weeks. Its use is associated with gastrointestinal intolerance, skin reactions, and mild central nervous systems-related adverse events (33). Generally, hydroxychloroquine is considered to have a very low risk profile. The primary concern with its use is ocular toxicity, for which ophthalmologic pre-screening and semi-annual-to-annual monitoring are recommended, depending on the reference (9,27).

2.2.5. Other conventional anti-rheumatic drugs

A host of other DMARD exist that are not commonly used in contemporary care. These drugs include gold compounds, cytotoxic immunosuppressant drugs (azathioprine, cyclophosphamide, cyclosporine, and penicillamine). These drugs are efficacious but

carry a substantial risk profile compared to the alternative DMARD discussed in this chapter.

Gold compounds, namely auranofin, gold sodium thiomalate, and aurothioglucose, were the "gold standard" remission-inducing agents used in the 1970s and 1980s (33). A draw to gold therapy is an apparent increase in the proportion of patients achieving remission over one year of treatment (41). The oral preparation (auranofin) has limited efficacy and the intramuscular options require regular office visits for administration and monitoring (33). Notably, gold sodium thiomalate is less-well tolerated than moderate dose MTX (41). Over the past few decades, the use of gold compounds has fallen out of favour. In the context of contemporary intensive clinical management strategies recommending high-dose methotrexate or combination therapy, the value of gold compounds may warrant reconsideration.

Immunospressant DMARD including azathioprine, cyclophosphamide, and cyclosporine, are efficacious in RA. Given contemporary therapeutic options, however, the adverse event profile of these medications renders their use exceptional. Exceptional circumstances include refractory disease to failed conventional therapeutic approaches and aggressive disease with life-threatening, extra-articular manifestations (33).

D-penicillamine has neither anti-inflammatory nor immunosuppressive effects (33). It contains a highly reactive sulfhydryl group with immunomodulatory effects on T- and B-cell lymphocytes (33). Penicillamine carries a risk profile similar to gold compounds that warrants careful monitoring. Like immunosuppressive agents, its use is

limited to patients unresponsive to conventional management who exhibit extra-articular disease manifestations (33).

2.2.6. Anti-tumour necrosis factor antagonists

Tumour necrosis factor alpha (TNFα) blockade has emerged as a major pharmacotherapeutic approach in RA clinical management. Over the past 12 years, five TNFα blockers have been approved internationally for major RA clinical indications. By earliest approval date, these include infliximab (Remicade®), etanercept (Enbrel®), adalimumab (Humira®), golulimumab (Symponi®), and certolizumab pegol (Cimzia®). In addition to inhibiting a novel target of the immune response, the α-TNFα therapeutics signaled the introduction of the first BRM or simply biologic agents approved for RA. Most α-TNFα therapeutics are monoclonal antibodies (mAb, e.g. infliximab, adalimumab, golimumab) or derivatives thereof (e.g. certolizumab pegol). Infliximab was first developed as a murine mAb and was later humanized to avoid xeno-immunoreactivity. Adalimumab and golimumab were developed as fully human mAb. Certolizumab pegol is a pegylated F_{ab} fragment of a humanized monoclonal antibody to TNFα. Pegylation refers to a common pharmaceutical technique of covalently binding polyethylene glycol polymer chains to a drug to increase its circulatory half-life. By comparison, etanercept is a fusion protein of soluble TNF receptor and constant portion of an antibody, F_c. All of these agents exert their therapeutic effects through binding TNFα.

These agents have all demonstrated efficacy in RA. Their clinical indications are agent-specific and all demonstrate superior efficacy when administered in conjunction with MTX. With exception, they are all clinically indicated for patients non-responsive to conventional DMARD. Some are indicated for RA treatment without MTX with lesser efficacy, especially when the concomitant medication is not tolerated. The indications usually include both the signs and symptoms of RA, as well and reducing radiographic evidence of disease progression.

Despite the known therapeutic target, the exact mechanism of action for the therapeutic effects of TNFα blockade has yet to be fully elucidated (42). It is thought that α-TNFα agents exert their effects primarily through binding and inhibiting soluble TNFα. The binding affinity of each agent differs and differential effects between specific agents call to question additional mechanistic effects including TNFα receptor inhibition or transmembrane TNFα induction (42). All α-TNFα agents bind transmembrane TNFα rendering the possibility of antibody-dependent and independent mechanisms of cytotoxicity.

The choice of initial α-TNFα agent is debatable. A recent meta-analysis suggests that adalimumab is most efficacious (43). These results, however, are not drawn from head-to-head studies, the optimal study design to determine interventional superiority. Further, the findings need to be reconciled with safety comparisons from similar analytical methods suggesting a greater risk for adalimumab and infliximab over etanercept (45). Moreover, patient preference on route of administration and frequency of dosing, factors that may affect adherence and hence efficacy, are important

considerations. Two for route of administration options exist: intravenous infusion (infliximab) or subcutaneous injection (all others). The frequency of dosing provides greater flexibility with twice weekly or weekly (etanercept), biweekly (adalimumab; certolizumab), monthly (certolizumab;golimumab), or bimonthly (infliximab). Anti-TNFα agents are generally expensive. Their costs range from $15000 to $25000 per year of treatment. Without insurance coverage, these medications are prohibitively expensive for most patients. In turn, most Canadian payers mandate documented evidence of treatment failure with more affordable conventional DMARD in combination prior to approving coverage for α-TNFα and other expensive biologic agents.

Due to pharmacological differences between specific agents, treatment failure with one agent does not necessitate failure on a second α-TNFα agent. Some small studies have demonstrated benefit from a second α-TNFα agent in patients non-responsive to a first α-TNFα intervention (46,47). In this population, there are alternative options (48-50). The relative efficacy of a second α-TNFα agent versus a biologic with a different therapeutic target remains largely unexplored. Emerging evidence suggests an advantage to attempting a secondary therapeutic target upon failure with one BRM (43).

2.2.7. Rituximab (Rituxan®)

Rituximab is a chimeric mAb against CD20-presenting cells. In combination with MTX, it is indicated for both the treatment of the signs and symptoms of RA and to reduce radiographic progression in patients non-responsive or intolerant to α-TNFα

agents. Effectively, rituximab reduces B cell counts and thereby inhibits the ill-defined B-cell-mediated inflammatory response. The drug was originally developed and approved in 1997 to treat non-Hodgkin's lymphomas. For RA, it is given as a course of two 1000mg intravenous infusions administered two weeks apart. The need for retreatment is evaluated 24 weeks after the initial course on the grounds of DAS28-ESR-defined relapse. Subsequent courses of treatment should be given at least 16 weeks after the initial one. The potential for infrequent dosing may have a patient preference advantage. The route of administration and length of time required for administration (at least 3 hours) may be disadvantageous. The medication is an additional option for patients unresponsive or otherwise intolerant to α-TNFα inhibitors.

2.2.8. Abatacept (Orencia®)

Abatacept is a selective T-cell costimulation modulator, which prevents the activation of T lymphocytes expressing the CD28 receptor (42). Similar in design to etanercept, abatacept is a fusion protein consisting of an extracellular domain of human cytotoxic T-lymphocyte-associated antigen 4 and a modified human IgG1 F_c (hinge, CH2, and CH3 domains) portion (51). As described in Chapter 1, activated T lymphocytes are recruited to the synovium in RA where they perpetuate the pathology of the disease. To become activated, T lymphocytes require two signals from antigen-presenting cells: a specific antigen and a co-stimulatory signal (51). In the absence of abatacept, one major co-stimulatory pathway involves the binding of CD80 and CD86 on antigen-presenting cells to CD28 on T lymphocytes (51). Abatacept inhibits this co-

stimulatory pathway by binding CD80 and CD86 (51). The apparent greater efficacy of abatacept in naïve T cells has prompted interest in the exploring the efficacy of abatacept in the early disease population (42).

Abatacept is indicated for RA patients who have an inadequate response to one or more DMARD, including α-TNFα agents (51). It may be administered as a monotherapy or in combination with another conventional DMARD, such as MTX. Concomitant treatment with α-TNFα agents should be avoided (51). As a first-line therapy, abatacept should be administered with MTX (51).

The drug is administered in doses ranging from 500mg (for a person <60 kg) to 1g (for a person >100 kg) every 2-4 weeks over the first two administrations and every 4 weeks thereafter (51). The weight-appropriate dose is administered by intravenous infusion over 30 minutes (51). From clinical trials, adverse events more common in the abatacept arm included headache, upper respiratory tract infection, nasopharyngitis, and nausea. It is recommended that patients treated with abatacept be up to date with their immunizations. Vigilance for hypersensitivity reactions is warranted.

2.2.9. Tocilizumab (Actemra®)

Tocilizumab is a humanized interleukin-(IL)-6 receptor mAb. It binds to both membrane-bound and soluble receptors to inhibit IL-6 function (30). As discussed in Chapter 1, IL-6 has ubiquitous roles in inflammation, immune response, bone metabolism, and cellular proliferation. Its role in the activation of osteoclasts and induction of MMP explains the efficacy of the tocilizumab on reducing radiographic

evidence of disease progression (42,52). The role of IL-6 in inflammatory processes such as anemia, fatigue, serum elevation in C-reactive protein and amyloid A help to explain other therapeutic effects of tocilizumab (52).

Administered in combination with MTX, other conventional DMARD, or as a monotherapy in patients intolerant to MTX, tocilizumab received international regulatory approval for treating RA signs and symptoms and reducing radiographic progression at 52 weeks in patients non-responsive to conventional DMARD or α-TNFα therapy (30). The medication is given every 4 weeks by intravenous infusion over one hour. The initial dose is 4 mg/kg and 8mg/kg on subsequent dosing. Tocilizumab has a considerable risk profile. General use in DMARD inadequate responders is cautioned against. Like α-TNFα therapy, IL-6-targeted therapy risks infection. Patients must be monitored for liver function tests, cytopenia, and lipidemia. Treatment needs to be evaluated on a case-by-base basis in consideration of disease and patient characteristics, and other therapeutic options (30).

2.2.10. Anakinra (Kineret®)

IL-1 blockade is another theoretically plausible mode of treating RA. Binding of the IL-1 to its receptor in many cells engages the IL-1R accessory protein and subsequently triggers
signalling and activation (53). Normally, an innate IL-1 receptor antagonist (IL-1Ra) competes with IL-1 for its receptor and blocks the signal transduction mechanisms (54). In IL-1Ra-deficient animals, a destructive arthritis phenotype results (55).

Anakinra is a recombinant, non-glycosylated version of the human IL-1Ra, identical except for the addition of a single N-terminal methionine residue (53). It is the only IL-1 blocking agent currently approved. Its effects in RA are inferior to α-TNFα therapy and other recently developed drugs. It is also associated with an increased risk for serious infection and anaphylactic reaction. Despite a generic indication, clinical use of this medication is limited.

2.3. Non-steroidal anti-Inflammatory drugs

The primary molecular targets of the NSAID drug class include cyclo-oxygenase (COX) enzyme inhibitors subclasses I and II (**Table 2**) (28,29,56). COX enzymes are involved in the release of prostaglandins. The primary function of COX-I enzymes is to regulate the release of prostaglandins involved with the protection of renal blood flow and gastric mucosa (56). COX-II enzymes produce prostaglandins to propagate the inflammatory response (56). As such, the primary function of the COX I and II subclasses differ. The NSAID drug class inhibits COX I and II enzymes.

These drugs are effective at alleviating pain and inflammatory symptoms in RA. They do not alter the underlying disease course (1,2,57). Historically, the erroneous belief that NSAID carry a benign risk profile combined with the under appreciation of the benefits limiting RA disease progression, favoured them as the first-line treatment (1,2). However, the side-effect profile is considerable and includes gastrointestinal and cardiovascular adverse events (58-75). Recently, product labels for all NSAID were revised to reflect the reassessment of the risk profile for this drug class (76,77).

Table 2. Common non-steroidal anti-inflammatory drug treatments.

Generic Drug Name	Common Trade Names	Usual Dosage	Maximum Recommended Daily Dose
Acetylsalicylic Acid	Aspirin®, others	3.6-5.4 g/day	-
Celecoxib	Celebrex®	100-200 mg bid	400 mg
Diclofenac	Voltaren®, Arthrotec® (diclofenac/misoprostol)	75 mg bid or 50 mg tid, 100 mg once/day sustained-release	150 mg
Fenoprofen	Naflon®	300-600 mg qid	3200 mg
Flurbiprofen	Ansaid®	100 mg bid or tid	300 mg
Ibuprofen	Various	400-800 mg qid	3200 mg
Indomethacin	Indocid®	25 mg tid to qid, 75 mg bid sustained-release	200 mg
Ketoprofen	Orudis®, others	50-75 mg qid, 200 mg qd sustained-release	300 mg
Meclofenamate	Meclomen®	50 mg tid or qid	300 mg
Meloxicam	Mobicox®	7.5-15 mg qid	15 mg
Nabumetone	Felafen®	1000-2000 mg/d in	2000 mg
Naproxen	Naprosyn®, Aleve®	250-750 mg tablet bid	1250 mg
Oxaprozin	Daypro®	1200 mg once/day	1200 mg
Piroxicam	Feldene®	20 mg once/day	20 mg
Sulindac	Clinoril®	150-200 mg bid	400 mg
Tolmetin	Tolectin®	400 mg tid	1600 mg

bid = twice daily; qd = per day; qid = 4 times daily; tid = 3 times daily.

2.4. Corticosteroids

Derived from native adrenal hormones, corticosteroids are effective at treating a variety of inflammatory conditions (56). Common corticosteroids used in RA care are listed in **Table 3** (28,29,56). In RA, they offer short-term anti-inflammatory and immunomodulatory properties (78). The efficacy of these drugs on inhibiting radiographic progression and remission when administered early is controversial (11,12,79). Evidence to support claims of radiographic progression inhibition is limited (11,79-81). Generally, corticosteroids are considered to have a considerable side effect profile when used on a long-term basis and the dosage must be tapered when coming off this class of drugs. To this point, intraarticular injection may more efficacious and better tolerated (82). Corticosteroids remain an important therapeutic option for RA (83-6).

Table 3. Common corticosteroids used in rheumatoid arthritis care.

Generic Drug Name	Common Trade Name	Common RA Routes of Administration	Usual Dosage*	Maximum Recommended Daily Dose
Methylprednisolone	Depo-Medrol®	IA	4-80 mg/joint	80 mg/joint q1-5w
Prednisolone	Inflamase®	PO	5-7.5 mg/day	7.5 mg/day
Prednisone	Winpred®	PO	≤ 7.5 mg/day	7.5 mg/day
Triamincolone	Kenalog®	IA	2.5-40 mg/joint	40 mg/joint q1-5w

*The maximum recommended dose of methylprednisolone for large joints (knees, ankles, shoulders) ranges from 20-80 mg. Lesser maximum doses for smaller joints ranging from 4-10 mg (for joints of the hand and chest) to 10-40 mg (for wrist and elbow) are recommended.
RA = rheumatoid arthritis; IA = intra-articular; PO = per oral; q = every; w = week.

2.5. Emerging therapies

As summarized by Smolen and Steiner (2003), a number of therapeutic targets have been inviestigated in clinical trials (87). Interventions against virtually all cells discussed in Chapter 1 have been investigated with multiple molecular targets in each case (87). Despite being published nearly 10 years ago, the molecular targets for medications approved since then are limited to TNFα, IL-1, IL-6, CD20, and CD28.

2.6. References

1. Scott DL. Pursuit of optimal outcomes in rheumatoid arthritis. Pharmacoeconomics 2004;22(2 Suppl):13-26.
2. Fries JF. Current treatment paradigms in rheumatoid arthritis. Rheumatology 2000;39(Suppl 1):30-5.
3. Kvien TK. Epidemiology and burden of illness of rheumatoid arthritis. Pharmacoeconomics 2004;22(2 Suppl):1-12.
4. Symmons DP. Epidemiology of rheumatoid arthritis: determinants of onset, persistence and outcome. Best Pract Res Clin Rheumatol 2002;16:707-22.
5. van der Heide A, Jacobs JW, Bijlsma JW, Heurkens AH, van Booma-Frankfort C, van der Veen MJ, et al. The effectiveness of early treatment with "second-line" antirheumatic drugs. A randomized, controlled trial. Ann Intern Med 1996;124:699-707.
6. Verstappen SM, Jacobs JW, Bijlsma JW, Heurkens AH, Booma-Frankfort C, Borg EJ, et al. Five-year followup of rheumatoid arthritis patients after early treatment with disease-modifying antirheumatic drugs versus treatment according to the pyramid approach in the first year. Arthritis Rheum 2003;48:1797-807.
7. Van Dongen H, van Aken J, Lard LR, Visser K, Ronday HK, Hulsmans HM, Speyer I, Westedt ML, Peeters AJ, Allaart CF, Toes RE, Breedveld FC, Huizinga TW. Efficacy of methotrexate treatment in patients with probable rheumatoid arthritis: a double-blind, randomized, placebo-controlled trial. Arthritis Rheum 2007;56:1424–32.
8. Tavares R. UNCOVER: Understanding Near-term Care Of Very Early Rheumatoid arthritis in a national, multicentre, randomly selected, historical inception cohort of newly diagnosed disease. Toronto, University of Toronto, 2007.
9. Saag KG, Teng GG, Patkar NM, Anuntiyo J, Finney C, Curtis JR, Paulus HE, Mudano A, Pisu M, Elkins-Melton M, Outman R, Allison JJ, Suarez Almazor M, Bridges SL Jr, Chatham WW, Hochberg M, MacLean C, Mikuls T, Moreland LW, O'Dell J, Turkiewicz AM, Furst DE, American College of Rheumatology. American College of Rheumatology 2008 recommendations for the use of nonbiologic and biologic disease-modifying antirheumatic drugs in rheumatoid arthritis. Arthritis Rheum 2008;59:762-84.
10. Knevel R, Schoels M, Huizinga TWJ, Aletaha D, Burmester GR, Combe B, Landewe RB, Smolen JS, Sokka T, van der Heijde DMFM. Current evidence for a strategic approach to the management of rheumatoid arthritis with disease-modifying antirheumatic drugs: a systematic literature review informing the EULAR recommendations for the management of rheumatoid arthritis. Ann Rheum Dis 2010;69:987-94.

11. Green M, Marzo-Ortega H, McGonagle D, Wakefield R, Proudman S, Conaghan P, Gooi J, Emery P. Persistence of mild, early inflammatory arthritis: the importance of disease duration, rheumatoid factor, and the shared epitope. Arthritis Rheum 1999;42:2184-8.

12. Verstappen SM, McCoy MJ, Roberts C, Dale NE, Hassell AB, Symmons DP; STIVEA Investigators. Beneficial effects of a 3-week course of intramuscular glucocorticoid injections in patients with very early inflammatory polyarthritis: results of the STIVEA trial. Ann Rheum Dis 2010;69:503-9.

13. Graydon SL, Thompson AE. Triage of referrals to an outpatient rheumatology clinic: analysis of referral information and triage. J Rheumatol 2008;35:1378-83.

14. Fitzgerald A, De Coster C, McMillan S, Naden R, Armstrong F, Barber A, Cunning L, Conner-Spady B, Hawker G, Lacaille D, Lane C, Mosher D, Rankin J, Sholter D, Noseworthy T. Relative urgency for referral from primary care to rheumatologists: the priority referral score. Arthritis Care Res 2011;63:231-9.

15. Bell MJ, Tavares R, Guillemin F, Bykerk VP, Tugwell P, Wells GA. Development of a self-administered early inflammatory arthritis detection tool. BMC Musculoskelet Disord 2010;11:50.

16. Tavares R, Wells GA, Bykerk VP, Guillemin F, Tugwell P, Bell MJ. Validation of a self-administered inflammatory arthritis detection tool for rheumatology triage. *Submitted*. 2012.

17. Brown AK, Conaghan PG, Karim Z, Quinn MA, Ikeda K, Peterfy CG, Hensor E, Wakefield RJ, O'Connor PJ, Emery P. An explanation for the apparent dissociation between clinical remission and continued structural deterioration in rheumatoid arthritis. Arthritis Rheum 2008;58:2958-67.

18. Clark P, Tugwell P, Bennet K, Bombardier C, Shea B, Wells G, Suarez-Almazor ME. Injectable gold for rheumatoid arthritis. Cochrane Database Syst Rev 1997;4.

19. Suarez-Almazor ME, Belseck E, Shea BJ, Tugwell P, Wells G. Sulfasalazine for treating rheumatoid arthritis. Cochrane Database Syst Rev 1998;2.

20. Suarez-Almazor ME, Belseck E, Shea BJ, Tugwell P, Wells G. Methotrexate for treating rheumatoid arthritis. Cochrane Database Syst Rev 1998;2.

21. Wells G, Haguenauer D, Shea B, Suarez-Almazor ME, Welch VA, Tugwell P. Cyclosporine for treating rheumatoid arthritis. Cochrane Database Syst Rev 1998;2.

22. Suarez-Almazor ME, Belseck E, Shea BJ, Tugwell P, Wells G. Cyclophosphamide for treating rheumatoid arthritis. Cochrane Database Syst Rev 2000;4.

23. Suarez-Almazor ME, Belseck E, Shea BJ, Homik JJ, Wells G, Tugwell P. Antimalarials for treating rheumatoid arthritis. Cochrane Database Syst Rev: Reviews 2000;4.

24. Suarez-Almazor ME, Spooner CH, Belseck E, Shea B. Auranofin versus placebo in rheumatoid arthritis. Cochrane Database Syst Rev 2000;2.

25. Suarez-Almazor ME, Spooner C, Belseck E. Azathioprine for treating rheumatoid arthritis. Cochrane Database Syst Rev 2000;4.

26. Osiri M, Shea B, Robinson V, Suarez-Almazor M, Strand V, Tugwell P, Wells G. Leflunomide for treating rheumatoid arthritis. Cochrane Database Syst Rev: Reviews 2002;3.

27. Gornisiewicz M, Moreland LW. Chapter 14: Rheumatoid Arthritis. *In* Robbins L, Burckhardt CS, Hannan MT, deHoratius RJ (editors). Clinical Care in the Rheumatic Diseases. 2nd Edition. Altanta, Association of Rheumatology Health Professionals, 2001.

28. Lacy CF, Armstrong LL, Goldman MP, Lance LL. Drug Information Handbook. 10th Anniversary ed. Hudson, OH: American Pharmaceutical Association; 2002.

29. CPS 2006: Compendium of Pharmaceuticals and Specialties. Ottawa, ON: Canadian Pharmacists Association; 2006.

30. Actemra® (tocilizumab) Product Monograph. Submission control no. 142808. Mississauga, Hoffmann-La Roche Limited, 2011.

31. Simponi® (golimumab) Product Monograph. Submission control no. 148073. Toronto, Janssen Inc., 2011.

32. Cimzia™ (certolizumab pegol) Product Monograph. Submission control no. 136802. Mississauga, UCB Canada Inc., 2010.

33. Chatham WW. Chapter 32: Monitoring of Patients on Antirheumatic Therapy. *In* Koopman WJ, Boulware DW, Heudebert GR (editors). Clinical Primer of Rheumatology. Philadelphia, Lippincott Wiliams and Wilkins, 2003.

34. Sharp JT, Strand V, Leung H, Hurley F, Loew-Friedrich I. Treatment with leflunomide slows radiographic progression of rheumatoid arthritis: results from three randomized controlled trials of leflunomide in patients with active rheumatoid arthritis. Leflunomide Rheumatoid Arthritis Investigators Group. Arthritis Rheum 2000;43:495-505.

35. Rachapalli SM, Williams R, Walsh DA, Young A, Kiely PD, Choy EH; Early Rheumatoid Arthritis Network. First-line DMARD choice in early rheumatoid arthritis--do prognostic factors play a role? Rheumatology 2010;49:1267-71.

36. O'Dell JR, Leff R, Paulsen G, Haire C, Mallek J, Eckhoff PJ, Fernandez A, Blakely K, Wees S, Stoner J, Hadley S, Felt J, Palmer W, Waytz P, Churchill M, Klassen L, Moore G. Treatment of rheumatoid arthritis with methotrexate and hydroxychloroquine, methotrexate and sulfasalazine, or a combination of the three medications: results of a two-year, randomized, double-blind, placebo-controlled trial. Arthritis Rheum. 2002;46:1164-70.

37. Felson DT, Anderson JJ, Meenan RF. The comparative efficacy and toxicity of second-line drugs in rheumatoid arthritis. Results of two metaanalyses. Arthritis Rheum. 1990;33:1449-61.

38. Tavares R, Pope JE, Tremblay J-C, Thorne C, Bykerk VP, Lazovskis J, Blocka KLN, Bell MJ, Lacaille D, Hitchon CA, Fitzgerald AA, Fidler WK, Bookman AAM, Henderson JM, Mosher DP, Sholter DE, Laupacis A, Boire G, Tomlinson G, Bombarder C. Early management of newly diagnosed rheumatoid arthritis by Canadian rheumatologists: a national, multicenter, retrospective cohort. J Rheumatol 2011;38:2342-5.

39. The HERA Study Group. A randomized trial of hydroxychloroquine in early rheumatoid arthritis: the HERA Study. Am J Med 1995;98:156-68.

40. Tsakonas E, Fitzgerald AA, Fitzcharles MA, Cividino A, Thorne JC, M'Seffar A, Joseph L, Bombardier C, Esdaile JM. Consequences of delayed therapy with second-line agents in rheumatoid arthritis: a 3 year followup on the hydroxychloroquine in early rheumatoid arthritis (HERA) study. J Rheumatol. 2000;27:623-9.

41. Rau R, Herborn G, Menninger H, Blechschmidt J. Comparison of intramuscular methotrexate and gold sodium thiomalate in the treatment of early erosive rheumatoid arthritis: 12 month data of a double-blind parallel study of 174 patients. Br J Rheumatol. 1997;36:345-52.

42. Tak PP, Kalden JR. Advances in rheumatology: new targeted therapeutics. Arthritis Res Ther. 2011;13(Suppl 1):S5.

43. Clinical and economic overview: biological response modifier agents for adults with rheumatoid arthritis. Ottawa, Canadian Agency for Drugs and Technologies in Health, 2010.

44. Alonso-Ruiz A, Pijoan JI, Ansuategui E, Urkaregi A, Calabozo M, Quintana A. Tumor necrosis factor alpha drugs in rheumatoid arthritis: systematic review and metaanalysis of efficacy and safety. BMC Musculoskelet Disord 2008;9:52.

45. Singh JA, Christensen R, Wells GA, Suarez-Almazor ME, Buchbinder R, Lopez-Olivo MA, Ghogomu ET, Tugwell P. A network meta-analysis of randomized controlled trials of biologics for rheumatoid arthritis: a Cochrane overview. CMAJ 2009;181:787-96.

46. Blom M, Kievit W, Fransen J, Kuper IH, den Broeder AA, De Gendt CM, Jansen TL, Brus HL, van de Laar MA, van Riel PL The reason for discontinuation of the first tumor necrosis factor (TNF) blocking agent does not infl uence the effect of a second TNF blocking agent in patients with rheumatoid arthritis. J Rheumatol 2009;36:2171-7.

47. van Vollenhoven R, Harju A, Brannemark S, Klareskog L. Treatment with infliximab (Remicade) when etanercept (Enbrel) has failed or vice versa: data from the STURE registry showing that switching tumour necrosis factor alpha blockers can make sense. Ann Rheum Dis 2003;62:1195-8.

48. Finckh A, Ciurea A, Brulhart L, Möller B, Walker UA, Courvoisier D, Kyburz D, Dudler J, Gabay C; Swiss Clinical Quality Management Programme for Rheumatoid Arthritis: Which subgroup of patients with rheumatoid arthritis benefits from switching to rituximab versus alternative antitumour necrosis factor (TNF) agents after previous failure of an anti-TNF agent? Ann Rheum Dis 2010;69:387-393.

49. Schiff M, Pritchard C, Huff stutter JE, Rodriguez-Valverde V, Durez P, Zhou X, Li T, Bahrt K, Kelly S, Le Bars M, Genovese MC. The 6-month safety and efficacy of abatacept in patients with rheumatoid arthritis who underwent a washout after anti-tumour necrosis factor therapy or were directly switched to abatacept: the ARRIVE trial. Ann Rheum Dis 2009;68:1708-14.

50. Emery P, Keystone E, Tony H, Cantagrel A, van Vollenhoven R, Sanchez A, Alecock E, Lee J, Kremer J. IL-6 receptor inhibition with tocilizumab improves treatment outcomes in patients with rheumatoid arthritis refractory to anti-tumour necrosis factor biologicals: results from a 24-week multicentre randomised placebo-controlled trial. RADIATE study. Ann Rheum Dis 2008;67:1516-1523.

51. Orencia™ (abatacept) Product Monograph. Submission control no. 128024. Montreal, Bristol-Myers Squibb Canada, 2010.

52. Nishimoto N. Interleukin-6 as a therapeutic target in candidate inflammatory diseases. Clin Pharmacol Ther 2010;87:483-7.

53. Kineret® (anakinra) Product Monograph. Submission control no. 142764. Dundas, C.R.I., 2010.

54. Arend, W. P., Malyak, M., Guthridge, C. J. & Gabay, C. Interleukin-1 receptor antagonist: role in biology. Annu Rev Immunol 1998;16:27–55.

55. Horai, R, Saijo S, Tanioka H, Nakae S, Sudo K, Okahara A, Ikuse T, Asano M, Iwakura Y. Development of chronic inflammatory arthropathy resembling rheumatoid arthritis in interleukin 1 receptor antagonist-deficient mice. J Exp Med 2000;191:313–20.

56. Beers MH, Berkow R. The Merck Manual of Diagnosis and Therapy. Whitehouse Station, NJ: Merck Research Laboratories, a division of Merck and Company Incorporated; 1999.

57. Quinn MA, Green MJ, Conaghan P, Emery P. How do you diagnose rheumatoid arthritis early? Best Pract Res Clin Rheumatol 2001;15:49-66.

58. Bensen W, Weaver A, Espinoza L, Zhao WW, Riley W, Paperiello B, Recker DP. Efficacy and safety of valdecoxib in treating the signs and symptoms of rheumatoid arthritis: a randomized, controlled comparison with placebo and naproxen. Rheumatology 2002;41:1008-16.

59. Bombardier C, Laine L, Reicin A, Shapiro D, Burgos-Vargas R, Davis B, Day R, Ferraz MB, Hawkey CJ, Hochberg MC, Kvien TK, Schnitzer TJ; VIGOR Study Group. Comparison of upper gastrointestinal toxicity of rofecoxib and naproxen in

patients with rheumatoid arthritis. VIGOR Study Group. N Engl J Med 2000;343:1520-8.

60. Caldwell JR, Roth SH. A double blind study comparing the efficacy and safety of enteric coated naproxen to naproxen in the management of NSAID intolerant patients with rheumatoid arthritis and osteoarthritis. Naproxen EC Study Group. J Rheumatol 1994;21:689-95.
61. Emery P, Kong SX, Ehrich EW, Watson DJ, Towheed TE. Dose-effect relationships of nonsteroidal anti-inflammatory drugs: a literature review. Clin Ther 2002;24:1225-91.
62. Deeks JJ, Smith LA, Bradley MD. Efficacy, tolerability, and upper gastrointestinal safety of celecoxib for treatment of osteoarthritis and rheumatoid arthritis: systematic review of randomised controlled trials. BMJ 2002;325:619-26.
63. Juni P, Nartey L, Reichenbach S, Sterchi R, Dieppe PA, Egger M. Risk of cardiovascular events and rofecoxib: cumulative meta-analysis Lancet 2004;364:2021-9.
64. Garner SE, Fidan DD, Frankish RR, Judd MG, Shea BJ, Towheed TE, Wells G, Tugwell P. Celecoxib for rheumatoid arthritis. Cochrane Database Syst Rev 2002;4:CD003831.
65. Garner SE, Fidan DD, Frankish RR, Judd MG, Towheed TE, Wells G, Tugwell P. Rofecoxib for rheumatoid arthritis. Cochrane Database Syst Rev 2005;1:CD003685.
66. Jajic Z, Malaise M, Nekam K, Koo E, Danko K, Kovacs M, Scarpignato C. Gastrointestinal safety of amtolmetin guacyl in comparison with celecoxib in patients with rheumatoid arthritis. Clin Exp Rheumatol 2005;23:809-18.
67. Krug H, Broadwell LK, Berry M, DeLapp R, Palmer RH, Mahowald M. Tolerability and efficacy of nabumetone and naproxen in the treatment of rheumatoid arthritis. Clin Ther 2000;22:40-52.
68. Martin-Mola E, Gijon-Banos J, Ansoleaga JJ. Aceclofenac in comparison to ketoprofen in the treatment of rheumatoid arthritis. Rheumatol Int 1995;15:111-6.
69. McKenna F. Diclofenac/misoprostol: the European clinical experience. J Rheumatol 1998;51(Suppl):21-30.
70. Pasero G, Marcolongo R, Serni U, Parnham MJ, Ferrer F. A multi-centre, double-blind comparative study of the efficacy and safety of aceclofenac and diclofenac in the treatment of rheumatoid arthritis. Curr Med Res Opin 1995;13:305-15.
71. Roth S, Bennett R, Caldron P, Hartman R, Mitchell C, Doucette M, Ekholm B, Goldlust B, Lee E, Wilson R. Reduced risk of NSAID gastropathy (GI mucosal toxicity) with nonacetylated salicylate (salsalate): an endoscopic study. Semin Arthritis Rheum 1990;19(4 Suppl 2):11-9.

72. Roth SH. Endoscopy-controlled study of the safety of nabumetone compared with naproxen in arthritis therapy. Am J Med 1987;83:25-30.

73. Silverstein FE, Faich G, Goldstein JL, Simon LS, Pincus T, Whelton A, Makuch R, Eisen G, Agrawal NM, Stenson WF, Burr AM, Zhao WW, Kent JD, Lefkowith JB, Verburg KM, Geis GS. Gastrointestinal toxicity with celecoxib vs nonsteroidal anti-inflammatory drugs for osteoarthritis and rheumatoid arthritis: the CLASS study: A randomized controlled trial. Celecoxib Long-term Arthritis Safety Study. JAMA 2000;284:1247-55.

74. Wojtulewski JA, Schattenkirchner M, Barcelo P, Le Loet X, Bevis PJ, Bluhmki E, Distel M. A six-month double-blind trial to compare the efficacy and safety of meloxicam 7.5 mg daily and naproxen 750 mg daily in patients with rheumatoid arthritis. Br J Rheumatol 1996;35(Suppl 1):22-8.

75. Furst DE, Kolba KS, Fleischmann R, Silverfield J, Greenwald M, Roth S, Hall DB, Roszko PJ, Meloxicam Rheumatoid Arthritis Investigators. Dose response and safety study of meloxicam up to 22.5 mg daily in rheumatoid arthritis: a 12 week multicenter, double blind, dose response study versus placebo and diclofenac. J Rheumatol 2002;29:436-46.

76. FDA News: FDA issues public health advisory on Vioxx as its manufacturer voluntarily withdraws the product. Available at: http://www.fda.gov/bbs/topics/news/2004/NEW01122.html. Accessed March 4, 2007.

77. FDA News: FDA announces series of changes to the class of marketed non-steroidal anti-inflammatory drugs (NSAIDs). Available at: http://www.fda.gov/bbs/topics/news/2005/NEW01171.html. Accessed March 4, 2007.

78. Smolen JS, Landewé R, Breedveld FC, Dougados M, Emery P, Gaujoux-Viala C, Gorter S, Knevel R, Nam J, Schoels M, Aletaha D, Buch M, Gossec L, Huizinga T, Bijlsma JW, Burmester G, Combe B, Cutolo M, Gabay C, Gomez-Reino J, Kouloumas M, Kvien TK, Martin-Mola E, McInnes I, Pavelka K, van Riel P, Scholte M, Scott DL, Sokka T, Valesini G, van Vollenhoven R, Winthrop KL, Wong J, Zink A, van der Heijde D. EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs. Ann Rheum Dis 2010;69:964-75. Erratum in: Ann Rheum Dis 2011;70:1519.

79. Criswell LA, Saag KG, Sems KM, Robinson V, Shea B, Wells G, Suarez-Almazor ME. Moderate-term, low-dose corticosteroids for rheumatoid arthritis. Cochrane Database Syst Rev 1998;3.

80. Combe B, Landewe R, Lukas C, Bolosiu HD, Breedveld F, Dougados M, Emery P, Ferraccioli G, Hazes JM, Klareskog L, Machold K, Martin-Mola E, Nielsen H, Silman A, Smolen J, Yazici H. EULAR recommendations for the management of early arthritis: report of a task force of the European Standing Committee for

International Clinical Studies Including Therapeutics (ESCISIT). Ann Rheum Dis 2007;66:34-45.

81. Landewe RB, Boers M, Verhoeven AC, Westhovens R, van de Laar MA, Markusse HM, van Denderen JC, Westedt ML, Peeters AJ, Dijkmans BA, Jacobs P, Boonen A, van der Heijde DM, van der Linden S. COBRA combination therapy in patients with early rheumatoid arthritis: long-term structural benefits of a brief intervention. Arthritis Rheum 2002;46:347-56.
82. Konai MS, Vilar Furtado RN, Dos Santos MF, Natour J. Monoarticular corticosteroid injection versus systemic administration in the treatment of rheumatoid arthritis patients: a randomized double-blind controlled study. Clin Exp Rheumatol 2009;27:214-21.
83. Kirwan JR, Bijlsma JWJ, Boers M, Shea BJ. Effects of glucocorticoids on radiological progression in rheumatoid arthritis. Cochrane Database Syst Rev 2007;1.
84. Laan RF, Jansen TL, van Riel PL. Glucocorticosteroids in the management of rheumatoid arthritis. Rheumatology 1999;38:6-12.
85. Wallen M, Gillies D. Intra-articular steroids and splints/rest for children with juvenile idiopathic arthritis and adults with rheumatoid arthritis. Cochrane Database Syst Rev 2006;1.
86. Gøtzsche PC, Johansen HK. Short-term low-dose corticosteroids vs placebo and nonsteroidal antiinflammatory drugs in rheumatoid arthritis. Cochrane Database Syst Rev 2005;1.
87. Smolen JS and Steiner G. Therapeutic strategies for rheumatoid arthritis. Nat Rev Drug Discov 2003;2:473-88.

3. Diagnostic imaging

3.1. Overview

Diagnostic imaging has a central role in the clinical management of RA, where the preservation of function (or limiting functional disability) is a principal aim. Radiographic evidence of disease progression is correlated with long-term functional disability in RA. Progression of both erosions and joint space narrowing are independently associated with long-term functional disability (1-3). Inhibition of radiographic progression is a key outcome measure in RA interventional clinical trials (4). As such, disease progression on radiography is an important surrogate outcome measure for long-term functional disability.

Radiography is the most commonly used imaging modality in RA clinical management (5): it is widely available, inexpensive, quick to execute, and effective. The therapeutic and health-economic impetus for identifying early RA and making objective treatment decisions has helped to stimulate research into the potential utility of other diagnostic imaging modalities as well. Multiplanar imaging techniques such as MRI, ultrasonography (US), scintigraphy, and computed tomography (CT) are candidate diagnostic imaging modalities currently being investigated to meet that demand. Several studies have demonstrated their greater sensitivity over conventional radiography for the detection of erosions (6-11). The increased sensitivity of these imaging modalities is attributed to the multiple slices and/or planes imaged compared to the single, superimposed, two-dimensional image with radiography. Whole body MRI, scintigraphy and CT may be prohibitively expensive to include as part of routine care. The advent of

extremity MRI (eMRI) renders this otherwise prohibitively costly modality of potential utility in RA management.

It is generally accepted that the erosion signal on MRI represents true cortical bone lesions (as verified for erosions on radiography). Edema and synovitis are also visualized on MRI. Together, the three features may help to improve the diagnostic specificity of the modality (12-16). Nonetheless, there are some challenges and limitations to the measurement of these features. Ultrasonography is a low cost alternative and may be best suited to visualize synovitis in early disease. Limitations to the joints that may be examined and low inter-acquisition reliability hamper the generalizability findings from this modality (17). Scintigraphy and CT are further limited by a questionable risk:benefit trade-off between radiation exposure and incremental value relative to the safer and effective aforementioned alternatives.

This chapter introduces the role of diagnostic imaging in the clinical management of RA. The roles of radiography and MRI are the focus. The relative merits of these two imaging modalities as diagnostic and prognostic tools are discussed. A description and biometric characterization of scoring schemes for the two modalities is also summarized. To begin, the relevance of radiography to the clinical management of RA is introduced.

3.2. Introduction

As a surrogate measure for long-term functional disability, radiography has long been utilized in the clinical management of RA patients to determine disease severity. In the early stages of RA, uncontrolled synovitis, a treatable and therefore reversible condition, is associated with pain, swelling, stiffness, and functional disability.

Uncontrolled, synovitis begets tendon, ligament, cartilage, and bone damage, which results in joint deformity and disability (18). Some studies suggest that the relationship between synovitis and erosion is weak and the two are largely independent processes (19). The majority of studies, however, support synovitis as a necessary precursor to the development of erosions (20-24). Certain features on diagnostic imaging, such as erosions and joint space narrowing, are practically irreversible and correlated with long-term disability. In conjunction with the view that radiography provides a source of objective, or *hard*, outcomes, the link between bone damage and long-term functional disability has diagnostic imaging positioned as a recommended method for evaluating disease status (2,25), prognostication (2), and interim progression monitoring (25).

3.3. Radiography

3.3.1. Overview

First discovered in 1895 by Wilhelm Conrad Roentgen (Röntgen), x-rays are electromagnetic radiation emitted by electrons. Hard x-rays, 0.10 to 0.01 nm wavelength, have the property of penetrating solid materials to varying degrees depending on the density of the object. The increased density of an object results in the increased attenuation of x-rays. X-rays that penetrate the imaged object are captured on a detector surface, or image receptor. Upon development of the receptor film, the exposed chemical surface darkens rendering a two-dimensional white-to-gray-scale image of the object against a black background.

Radiography has found widespread use in medical imaging. In RA, as described above, the potential deleterious effects of uncontrolled synovitis on bone are readily detected using radiography. The extent of erosive disease on radiography is correlated with long-term functional disability (2,26). Within two years from symptom onset, up to 75-90% of RA patients exhibit radiographic evidence of bone damage, i.e. bone erosions, or more simply erosions (27,28). Erosions have been detected in up to 97% of RA subjects in some cohorts (29). The majority of erosions are thought to develop over the first two years of disease (30). These findings may be cohort specific. At the population level, it is generally accepted that 20-50% of RA patients have erosive damage within the first three years of disease (18).

Disease progression on radiography may follow different patterns. In the clinical context, limiting radiographic evidence of disease progression translates into improved long-term functional outcomes. Overall, radiographic evidence of disease progression is only moderately correlated with of symptom duration (2). Following a small cohort of 109 patients over 30 years, Graudal *et al*. (1998) delineated five groups of patients with differential patterns of disease progression: non-progressors (<1%); slow onset with later exponential progression (39%); slow onset with acceleration and then deceleration of progression (20%); acute onset and rapid progression (11%); and, acute onset with slow progression (30%) (31). Over an eight-year timeframe across 114 early RA patients, Plant *et al*. (1998) demonstrated that 45% of RA patients progress linearly, 12% exhibit a lag in erosive progression, progression plateaus in 17%, and 26% have non-erosive disease (32). Despite the small numbers in each of these cohorts, the non-uniform pattern

of radiographic evidence of disease progression is consistent with data from one- and two-year follow-up clinical trials comparing the efficacy of DMARD (including biologic) interventions demonstrating non-uniform radiographic progression across patients. These trials generally demonstrate that radiographic evidence of disease progression occurs in 20-30% of trial participants over two years follow-up (33,34).

Care in the interpretation of these findings is important. The strength of clinical trials is internal validity not external validity. The external validity of the findings depends on whether or not differences between the sample and population exist that are associated with the outcome. Selection of clinical trial participants to maximize the outcome, or signal, (while limiting quiescence or non-outcomes) generally renders them as inappropriate sources of data for epidemiological benchmarking. Trial participants may have a greater propensity for erosive disease than the general RA population. In contrast, clinical trials have internal controls and outcomes are measured more rigorously than those collected from observational studies or from clinical practice. Trial outcomes are therefore less prone to bias and likely more accurate than otherwise derived. As such, the study populations of intervention-based clinical trials may have a selection bias towards patients with more severe disease and the rigour in the measurement of radiographic evidence of erosions may differ from that used in cohort settings. Generally, the external validity of the patterns reported in Graudal *et al*. and Plant *et al*. are difficult to ascertain; however, the heterogeneity of radiographic progression noted is consistent with suggestion of non-uniform progression from trial findings conducted over

the shorter follow-up periods. These findings emphasize the importance of identifying patients with a propensity to progress rapidly.

In 2002, Visser *et al*. published an important paper modeling predictors of erosive RA (35) in patients presenting with at least monoarthritis for a maximum duration of 2 years. The authors identified arthritis symptom duration greater than 6 weeks (and more so greater than 6 months), involvement in at least three joint groups, bilateral positive compression test of the metatarsophalangeal joints, rheumatoid factor titre greater than 5 IU/mL, and anti-citrullinated cyclic antibody (αCCP) titre greater than 92 IU/mL as important baseline predictors of the development of erosive disease (31). The antibody thresholds used were surprising. Current local normal laboratory reference ranges are less than 14-20 IU/mL for both of these tests (36-38). Similarly, the dichotomization of continuous variables (symptom and morning stiffness duration, and antibody titres) was noteworthy and puzzling. Despite the questionable decisions from a statistical modeling perspective (39), this work serves as an early rigorous effort to identify predictors of the development of radiographic evidence of erosive disease.

In reviewing radiographic progression studies, care must be taken to understand the disease features followed on imaging. Radiographic evidence of disease progression is most commonly measured by the extent of erosive disease. To a lesser extent, joint space narrowing, a surrogate measure for cartilage degradation, is also measured. Composites measuring both erosions and joint space narrowing are numerous and common. Despite the prominence of the former two features, periarticular osteopenia, ankylosis, and malalignment are additional features characteristic of RA that may be

visualized on radiography. Periarticular bone loss is often evident in the radiographs of RA patients. Osteopenia of the lower spine, as measured by bone mineral density using dual-energy x-ray absorptiometry, is also common in RA and more rapid in patients with worse prognosis (18). The prevalence and progression of osteopenia in RA is sufficient to recognize RA a risk factor for osteoporosis. Ankylosis is generally limited to long-standing, poorly controlled disease (18). Malalignment may result from tendon and ligament damage, and/or improper resetting following chronic inflammation. Collectively erosions, joint space narrowing, and periarticular osteopenia have high specificity for RA (18).

Radiographic evidence typically manifests in certain anatomical patterns. Erosive joint group involvement demonstrates considerable symmetry along the sagittal plane. These findings are aligned with the anatomical distribution of the clinical findings of joint swelling. Symmetrical involvement of joint swelling by joint group along the sagittal plane has long been identified as an important feature in disease classification (40,41). This feature may partially explain why the proportion of patients with disease progression detected on a single hand is similar to that detected bilaterally (42,43). Metatarsophalangeal joints (MTP) are among the first joint groups to develop erosions. Erosions also manifest early in metacarpophalangeal (MCP) and wrist joints of uncontrolled disease. Joint space narrowing is also common in these joints. Early radiographic changes to larger, clinically affected joints, such as the hip and knee, are more commonly limited to joint space narrowing (18). Erosive changes to larger joints occur later than in the hands, wrists and feet (14). Within joint groups, the suggestion

exists that joints more commonly used, e.g. MCP 2 and 3, are affected to a greater extent than MCP 4 and 5. This is presumed to be due to mechanical stress, although not conclusively demonstrated. Similarly, MTP 1 is more commonly affected than the lateral MTP joints. Generally, radiographic damage presents earlier in the feet than in the hands and wrists. The feet are thus an important anatomy to image especially in suspected or early established disease. Sometimes considered spine-sparing, cervical subluxation at C1-C2, and synovitis of ligaments and erosions of the odontoid process occur in RA (18).

Clinically, radiography is used for diagnostic confirmation and to assess the extent of disease (44). In follow-up, radiography is used to determine long-term treatment efficacy. Clinically, a trained radiologist evaluates radiographs qualitatively. Different radiologists are likely to be involved in assessing multiple radiographs of the same patient over time. For any one assessment, the radiologist may or may not have access to previous clinical impressions or images as there is little control over where x-ray requisitions are filled and centralized diagnostic imaging archives do not exist. Clinically, radiological renderings are susceptible to inter-rater variability. In research, inter-rater variability of semi-quantitative assessments is a noted phenomenon (45,46).

Composite measurement tools are used in rheumatology research to semi-quantitatively measure radiographic evidence of disease progression. Some common scales include van der Heijde-modified Sharp score (vdHSS), Sharp score, on which the former is based, Larsen score, and Genant score. Modifications of the Sharp and Larsen scores are common. Without exception, these tools call on the rater to make assessments of erosion and joint space narrowing. These tools are not used in clinical practice. They

are commonly used in research settings. An overview of common radiography scoring systems and an in-depth description of the van der Heijde-modified Sharp score is covered in the following sections.

3.3.2. Radiography scoring systems

Boini and Guillemin (2001) reviewed the characteristics of various radiography scoring methods used in RA (44). The anatomy imaged, reliability, and sensitivity to change was compared between scoring methods. Most scoring systems are based on the Sharp, Genant, and Larsen methods. Previous radiography scales, such as the Steinbrocker or Kellgren scales were global assessments of evidence of disease at the patient level (44,47). Whereas the Larsen score assigns a global assessment of joint damage (encompassing erosion and joint space narrowing) to each site assessed, Sharp- and Genant-based methods assign individual erosion and joint space narrowing scores to each joint or bone. Erosion and joint space narrowing are ubiquitously captured by all methods, except the Rau *et al*. (1998) modification of the Larsen score, termed the Ratingen score, which omits joint space narrowing (48). As noted above, soft tissue swelling, osteoporosis, subluxation and malalignment, and ankylosis are additional features on radiography associated with disease progression. No single score captures all of these features. Many of these latter features cannot be measured reliably and consistently across a large number of patients. This renders their standardized collection labour's love lost.

Generally, joints of the hands, wrists and feet are scored. All but the Sharp score include the feet. Modifications of the Sharp score, such as that by van der Heijde, include the feet. As described above, examining the feet becomes especially important for the detection of erosions in early disease. There are notable differences in the specific joints and anatomy considered, and the ordinal scales used for semi-quantification. The relative weight of erosion versus joint space narrowing and anatomy on the total score varies. Generally more weight is placed on erosions than joint space narrowing: the measurement tools developed by Kaye *et al*. (1987) (49) and modifications of the Genant method (50) are the exceptions. The upper extremities are weighted more heavily than the lower ones. The van der Heijde-modified Sharp score tends to be the most time-consuming. Paradoxically, the van der Heijde modification also appears to be most prevalent in the medical literature (51-53). Specific time requirements are rater-dependent as are the intra- and inter-rater reliability of the assessments. Training improves reliability and therefore sensitivity to change. Posteroanterior projections of the hands, wrists and feet have greatest utility; however, three projections are preferred. Supplementary projections include oblique (a.k.a. Norgaard, or "ball catcher's") and lateral (a.k.a. "okay" sign) views.

Sensitivity to change is measured by standardized response mean (SRM), adjusted SRM, smallest detectable difference (SDD), and smallest or minimum detectable change (MDC/SDC) (44). An SRM of 0.80 or greater is considered acceptable (44). Inter-rater reliability is assessed using the kappa and intra-class correlation coefficient (ICC) (44). Kappa and ICC values greater than 0.70 are commonly reported (44).

3.3.3. van der Heijde-modified Sharp score

The van der Heijde-modified Sharp score was originally published in 1989 (51) as a modification to the revised Sharp score published four years earlier (54) and derived from the original method published in 1971 (55). The Sharp method outlines the scoring of joints of the hands and wrists. It originally accounted for periosteal reaction, cortical thinning, osteoporosis, sclerosis, osteophyte formation, defects, cystic changes, erosions, joint space narrowing, and ankylosis. Periosteal reaction, cortical thinning, osteoporosis, sclerosis, and osteophyte formation were later omitted due to secondary relevance to RA disease or measurement challenges (47). All other features were captured in the resulting score as part of either erosion or joint space narrowing.

The major distinction between the van der Heijde modification and the revised Sharp score is the omission of the feet in the latter. In light of evidence demonstrating erosive disease manifests in the feet earlier than in the hands and wrists, van der Heijde modified the score to include the former anatomy. In addition, a few joints of the hands and wrists scored by the Sharp method were found to be difficult to resolve across multiple samples and were therefore dropped (52,53). One bone for erosion (os triquetrum) and three joints for joint space narrowing (radio-ulnar joint, lunar-triquetrum joint, first interphalangeal joint) were excluded from the van der Heijde modification (52,53). With these noted differences, the anatomy and scoring system used for the Sharp score and van der Heijde modification are described below.

Each of the following joints and bones of the hands and wrists are scored for erosions. Bilaterally, the five metacarpophalangeal, four proximal interphalangeal,

interphalangeal, metacarpal base, radius and ulnar bones, trapezium and trapezoid (as one unit; multangular), navicular, and lunate bone (1) are scored for erosions (51). The semi-quantitative ordinal scale ranges from 0 to 5 with the following descriptions for each individual score grade (52,53):

- 0 = normal; no defects;
- 1 = discrete, cortical break;
- 2 = lesion larger than a '1' not extending past the imaginary sagittal line dividing the bone in two halves;
- 3 = lesion larger than a '1' extending past the imaginary sagittal line dividing the bone in two halves;
- 5 = complete collapse of the bone.

A score of '4' is not assigned to any single lesion although it may be reached by multiple erosions on the same joint or bone. The joints of the hands and bones of the wrists are each assigned a maximum score of 5. Erosions are scored regardless of presumed underlying disease process, be it rheumatoid, osteoarthritis, or other (52,53). The van der Heijde modification includes a score for each of the five metatarsophalangeal and first interphalangeal joints of the foot, bilaterally. The system for scoring erosions in the feet differs slightly from that of the hands. Each joint of the feet is allotted a maximum score of 10 and maximum of 5 for each juxta-articulated bone (52,53).

Joint space narrowing is scored on a set of joints of the hands and wrists largely overlapping with those evaluated for erosions using an ordinal scale ranging from 0 to 4. Bilaterally, the joints assessed for joint space narrowing include the following: five metacarpophalangeal, four proximal interphalangeal, the third, fourth, and fifth

carpometacarpal, multangular-navicular, capitate-navicular-lunate, radiocarpal, five metatarsophalangeal, and first tarsal interphalangeal (51). The description for each joint space narrowing score is as follows:

- 0 = normal;
- 1 = focal or doubtful;
- 2 = generalized with > 50% of the original joint space left;
- 3 = generalized with < 50% of original joint space left or subluxation;
- 4 = bony ankylosis or complete luxation.

Here, "doubtful" is used when joint space narrowing is suspected but is not severe enough to score '2' (52,53). Notice that joint space narrowing in the feet is scored using the same convention as for the hands.

Using this method, 16 sites for each hand and 6 sites for each foot are scored for erosions. Bilaterally, a total erosion score of 280 is possible (16 sites/hand * 2 hands * 5 maximum score/site + 6 sites/foot * 2 feet * 10 maximum score/foot). In addition, joint space narrowing is scored on 15 sites per hand and 6 sites per foot resulting a maximum subtotal score of 168 ((15 sites/hand + 6 sites/foot) * 2 hands and feet * 4 maximum score/hand or foot). The total score for the assessment of erosions and joint space narrowing bilaterally for the hands, wrists and feet is 448.

A number of so-called "scoring rules" were proposed for the score. First, the assigned score was proposed to be unidirectional (52,53). That is, once a score is assigned to a bone or joint, a later score of lesser magnitude is not acceptable. Second, in temporal scenarios, the time sequence of the radiographs is to be known (52,53). These

scoring rules lack pragmatism and only serve to bias scores towards greater reliability between raters than if they are not used, resulting in an increased Type I-like error (concluding a change in score when one does not exist). The first rule does not account of the two-tailed uncertainty bound to any individual evaluation. That is, there is no reason to believe that the initial reading of an x-ray is absolutely accurate and precise. As such, a score of '1' assigned to one joint at timepoint 1 may be scored '0' at timepoint 2 strictly due to the variability in scoring. The first scoring rule ignores this possibility and by default, the minimum score at timepoint 2 would be '1'. Since the publication of these scoring rules, the determination that x-ray features of erosion and joint space narrowing are reversible for a minority of patients has provided further scrutiny to the imposition of this scoring rule. The second rule, 'known chronology', is subject to expectation bias, where later scores are artificially assigned greater scores due to preconceived expectations of disease progression over time (44). Practically, it is sometimes difficult to classify an x-ray feature from its continuous form to the ordinal form prescribed on the measurement scale. If the chronology of x-ray is known, raters may be inclined to assign a higher score than the previous one if the feature appears relatively larger, with consideration of neither the variability around any individual assessment nor agreement between the assigned score and its absolute definition. A less yet problematic alternative to the second scoring rule is to assess paired images (from the same patient) without explicitly knowing chronology.

As with those above, other suggestions have been made with the objective of improving the reliability of scoring. The use of multiple raters and averaged scoring is

claimed to improve the accuracy and precision of scoring (44). Sensitivity to change is improved by using paired images (unknown chronology) as compared to independent images. It is improved further still with known chronology. Many of these suggestions defy the clinical reality of radiological assessments. Clinically, x-rays are qualitatively assessed by a single radiologist who may or may not have access to previous images. Independent assessments by individual, trained radiologists provides the most clinically relevant assessment. Ultimately, scores are dependent on rater, rater training, film exposure, anatomical positioning, and the specific projections used. Although the van der Heijde-modified Sharp score has been widely adopted in RA research, use of the ascribed scoring rules has not, resulting in much herterogeneity of its application in the literature.

Given the heterogeneity in method of scoring x-ray, care is required in interpreting the biometric properties of these tools. In 1989, van der Heijde reported that the SDD of the van der Heijde-modified Sharp score was 10 (53). Smallest detectable difference is an assessment based on status score, i.e. individual image measurement error. As such, the chronology of images and the non-reversibility assumption are both irrelevant. Guillemin *et al*. (2005) reported an SDD of 10 for the vDHSS of the hands, which corresponds to an SDD of 16 assuming proportional variance between the scoring of hands and feet (56). This value was determined from a one-tailed test: SDD = $(1.645*SEM)^{1/2}$ whereas the two-tailed test and a different formula may be more appropriate (57). The score reported in Guillemin *et al*. (2005) may pertain to the smallest detectable change (SDC). The SDC makes use of paired data where the scoring rules may be applied. Generally, the SDC is smaller than the SDD. Two-tailed tests,

rather than one-tailed tests, are more commonly expressed for SDC and SDD values. The one-tailed test formula results in a smaller value. Despite these sources of variability, Bruynestyn *et al*. (2004) also report an SDD of approximately 15 for the vdHSS of the hands and feet (58).

Interestingly, the formulae for the calculation of SDD and SDC handle the composite score as if it were continuous data, not accounting for the uncertainty around each of the 86 individual joints or bones assessed. An alternative method of determining the SDD would be to determine the 95% confidence interval of the propagated error around each of the 86 individual joints or bones. Challenges to this approach exist as well, including the validity of error propagation assumption when dealing with ordinal data. To the author's knowledge no one has attempted to quantify imaging changes by this approach. In the next section, an introduction to the use of MRI in the clinical management of RA is provided.

3.4. Magnetic resonance Imaging

3.4.1. Overview

Developed over the latter half of the 20th century, MRI utilizes the magnetic moment of hydrogen atoms to differentiate between tissues of varying hydrogen density. The most prevalent sources of hydrogen in living systems include water and fat. In MRI, magnetic fields and radio frequency pulses are used to align and manipulate the magnetic moment of these hydrogen atoms, respectively. While the magnetic field of the MRI unit aligns the polarity of hydrogen atoms, radio frequency pulses are used to temporarily flip

the magnetic moment. Subsequently, the magnetic moment returns to its aligned, resting state and releases the energy that is collected by the MRI. The MRI converts the energy into a spatial presentation of hydrogen density. The radio frequency used can be adjusted to resolve either the water or fat signal. Based on differences in hydrogen density and the distinction between water and fat, MRI may be used to visualize and differentiate soft tissues.

The physics theory and engineering behind MRI are beyond the scope of this thesis. Of consequence to this work is the manipulation of MRI sequence parameters to optimally detect specific features of interest. Non-fat-saturated (-FS), T1-weighted (T1w) MRI sequences detect fatty bone. On T1w –FS MRI sequences, erosions are defined as dark lesions within bone that are observable in two planes and penetrate the cortical surface in at least one plane (59). Signal detection in a second plane is used to avoid mistakenly diagnosing enhancing tissue at the bone surface as an erosion (60). Equivalently, MRI images in one plane with isotropic voxels may be reconstructed in 3-dimensions and used to detect erosion (59). Fast spin echo (FSE) MRI sequences may be either T1w or T2-weighted (T2w). In contrast, gradient echo 3D (3DGE) images are T1w. Aqueous features, such as synovitis, effusion and edema, are detected using T2w sequences.

With T2w sequences, fat suppression is used to more clearly differentiate between the aqueous signal of interest and background fatty tissue. There are number of ways of achieving fat suppression, including but not limited to fat saturation and inversion recovery (IR) (61). The former is achieved by centring the radio frequency on the

resonance frequency for fat (which is of lower energy than water). Specifically, fat-suppressed (+FS) T2w and short tau IR (STIR) MRI sequences are commonly used to detect bone marrow edema. On either sequence, bone marrow edema is detected as a diffuse white-to-light-grayscale signal with ill-defined margins on a background of dark grayscale bone. The soft tissue and three-dimensional rendering capabilities of MRI expand the diagnostic imaging armamentarium beyond that available with x-ray.

The technology offers a safety advantage over other diagnostic imaging modalities: it does not use ionizing radiation as do x-ray and computed tomography. Nonetheless, pace-makers and other metal implants are generally considered contraindications to MRI. Depending on the ferrous content of a surgical implant, and the strength of the MRI magnetic field, many pose little or no risk to the patient undergoing MRI. Contrast agents are sometimes used to enhance the resolution of particular features on MRI. These agents may have associated risk profiles. Other non-life-threatening risks, such as claustrophobia, may be mediated by prophylaxis, or if applicable the use of extremity or walk-in MRI units. Generally, MRIs have a limited risk profile.

As applicable to RA, and as described above, diagnostic imaging change is a surrogate marker for long-term functional disability. Validation of this surrogate, however, is primarily limited to radiography. In turn, radiographic features only partially overlap with the MRI features visualized. Erosion, as measured on MRI, is thought to represent the same physical lesion as on radiography. The exception is that the former is thought to have a lower limit of detection owing to its multiplanar and multislice

capability. It is therefore expected to detect smaller lesions. Per bone imaged, MRI detects a greater number of lesions than does radiography (43,62,63).

A comprehensive review of the evidence comparing erosion detection on MRI and radiography is summarized in Chapter 4. Some studies report a 7- to 9-fold increase in erosion detection sensitivity of MRI over radiography (11,64). Some have noted that initial erosions observed on MRI appear on radiography 6 to 12 months later (62,65,66). Caveats to the increased sensitivity of MRI include the following:

1) correlations between diagnostic imaging progression and disease outcomes are almost exclusive to radiography and may not apply to MRI; and
2) decreased specificity of MRI erosions for RA disease as compared to its radiographic counterpart (12).

Both caveats may limit the value detecting pre-radiographic lesions. Regarding the latter, it has been suggested that the specificity of MRI erosions for RA may be as low as 35% (i.e. 2 of 3 healthy persons exhibit MRI evidence of erosions) (12). These data were taken from a study where the presence of erosions was measured on a single plane without isotropic voxels (12). Using a similar definition for MRI erosions to that above, Xie *et al*. (2008) noted the signal in the healthy population, more predominantly in the wrists of older controls, aged 49 to 74 (67). Other research, using the OMERACT RAMRIS definition suggests that MRI erosions in the healthy population are rare and limited to low-grade scores (68). One would expect the definition of erosion applied in the former two studies to result in a greater error in the detection of erosions due to the inability to differentiate partial volume artifacts (69). Klarlund *et al*. (1999) suggest that

as measured in a single plane, enhancing tissue at the bone surface may be mistakenly diagnosed as an erosion (60). The inclusion of a second imaging plane often depicts clear margins and help to mitigate partial volume effects (69). Reduction of slice thickness and interslice gap, improvement of inplane resolution (signal-to-noise ratio), and stabilization of the anatomy imaged may all contribute to minimizing partial volume artifacts.

Care must be taken to cross-reference the signal in two planes to ensure it is not confused with either synovitis, or effusion, both of which occur outside the bone within and including the synovial membrane. Pannus infiltration of erosions may also be confused as edema; however, this signal is of greater intensity than the diffuse, stippled appearance of edema. As described earlier, synovitis refers to inflamed synovial tissue and is indicative of disease activity. The fatty synovial tissue proliferates and is engorged with extracellular fluid when inflamed. As a result, MRI sequences sensitive to aqueous signal detect synovitis. The fatty content of the synovial membrane renders the signal intensity lower than effusion. Nonetheless, the delineation of synovitis from effusion on T2w +FS or STIR sequences may be challenging. The use of gadolinium-based contrast agents in T1w imaging is the current standard to detecting synovitis. The contrast agent is usually administered by bolus intravenous injection. Within minutes, the synovial tissue is perfused with contrast. The synovial membrane acts as a filter separating the blood and extra-articular matter from the intra-articular space. With additional time, the contrast agent diffuses across the synovial membrane into the intra-articular space. As such, the administration of contrast agent is time sensitive: if the MRI sequence is run too early, the tissue may be insufficiently perfused; if run too late, the resulting signal may be

contaminated with effusion. Other challenges to he use of contrast include cost and rare but serious adverse events, such as nephrogenic fibrosis syndrome (70,71). Fortunately, effusion is also closely related to active inflammation (72-74). Further, synovitis scoring with and without contrast is similar, even in studies biased to resolve a difference between the two (75). The use of T2w +FS is a viable alternative to T1w –FS ±Gd for detecting synovitis on MRI (75-7).

Most studies comparing the relative performance of MRI and radiography fail to account for the reliability of the assessments made. Reliability statistics are infrequently used to establish threshold scores for true positive diagnostic ascertainment. This is applicable to both MRI and x-ray. It may be used to account for inter-rater variability in assessments.

Initial reports on the inter-rater reliability of MRI erosion scoring using the Outcome Measures for Rheumatology (*previously* Outcome Measures for the Evaluation of Rheumatoid Arthritis Clinical Trials) (OMERACT) Rheumatoid Arthritis MRI Score (RAMRIS) report a fixed effects single measure ICC of 0.45-0.85 (26,78). In the context of intra-rater reliability, others report higher ICC (e.g. >0.95) values as expected (79). Notably, the fixed effects ICC is not generalizable to other raters (80). To Norman and Streiner's point that the ICC is equivalent to the kappa for binary data and the quadratic-weighted kappa for ordinal data, this holds true for the random effects ICC only (81). The random effects ICC is generalizable but seldom reported. Of consequence, the random effects ICC is almost always smaller than the fixed effects ICC. Although the

fixed effects ICC is perfectly valid and its reporting common, the generalizability of findings is lost on the reader.

A measure of agreement that provides more context than the ICC is the SDD. This measure refers to two standard deviations for a two-tailed test for a given measure. The standardized inter-rater SDD for bone erosions ranges from 24-42%, where a maximum actual score is used as the denominator (78,82). Lower SDD (3.3-12.4%) are observed in intra-rater reliability studies (79). An absolute inter-rater SDD (in scale units) for the total MCP erosion score for a single hand of 6.6 has been reported (82). An inter-rater SDD for the total wrist erosion score of a single hand of 4.8 has been reported (83). Care in the interpretation of these results is required as a number of factors may impact the measure. Generally, inter-rater variance is greater than intra-rater variance. Certain anatomy and features are more difficult than others to score. Finally, all reliability data are rater- and setting-specific.

The unit of analysis is also important. Use of the smallest unit of analysis and propagation of that error is expected to result in a larger SDD than if the score is determined from an aggregate score. Rarely is SDD reported per single bone or joint measured (i.e. the smallest unit of analysis). More commonly it is reported for a total composite score or feature examined. Between raters, the variance around absolute aggregate or composite scores may be dampened by a regression to the mean effect, where expectation bias renders some feature to be detected. On the other hand, if bias in scoring exists between raters, such that one rater is more aggressive in scoring over the

other, SDD determination from an aggregate score may have the impact of magnifying the variance between raters.

3.4.2. Diagnostic and prognostic measurement in RA

3.4.2.1. OMERACT RAMRIS

OMERACT is an international initiative by rheumatologists, methodologists, and other specialists interested in rheumatology with the goal of standardizing and endorsing the best clinical and research measurement tools for the care of these populations. The *Imaging in Inflammatory Arthritis* subcommittee developed RAMRIS in the early 2000s for the purpose of standardizing the semi-quantification of MRI to be used as an imaging outcome for this population. RAMRIS is an internationally recognized system for measuring RA disease changes on MRI (59,69,83-85). The literature is replete with clinimetric characterizations of the tools and validation for various purposes and extensions and subsets of the original population.

Three MRI features of prognostic relevance include synovitis, bone marrow edema, and erosions (12-16). Synovitis is a precursor to bone marrow edema (15,86). Synovitis and edema predict erosive bone damage (14,15,24). These three MRI features are captured in RAMRIS. The system is comprised of the following components:

1) A definition of the MRI sequences, planes of view, contrast agent requirements, and signal descriptions for each feature (59);
2) A scoring system (85);
3) An image atlas (84);
4) A guidance on pitfalls with the system (69); and,

5) Characterizations of the clinimetric properties of the system (83).

In the development of RAMRIS, a number of features detectable on MRI, including tenosynovitis and joint space narrowing, were determined to be of lesser diagnostic and prognostic utility than those described above.

The RAMRIS system utilizes a gadolinium-based contrast agent for the measurement of synovitis. This imposes a large operational cost (76). Alternatively, synovitis may be detected on MRI without a contrast agent using Fast Spin Echo (FSE), T2-weighted (T2w), fat-saturated (+FS) sequences (72-75). Concordance between the contrast-dependent and –independent methods is good to excellent (75) despite a deficit in tools to promote the standardization of the latter as available for the RAMRIS system (as listed above).

3.4.2.2. Other RA MRI scoring systems

Other, less popular RA MRI scoring systems also exist. Two examples of these include the Simplified Rheumatoid Arthritis Magnetic Resonance Imaging Score (SAMIS) (87) and compact MRI score (88). In this thesis, RA MRI measurement was specific to the RAMRIS method. Other systems were not considered.

3.5. Other diagnostic imaging modalities

As introduced above, other diagnostic and prognostic imaging modalities exist for the clinical management of RA. These included ultrasound (8), computed tomography (8,9,89), and bone scintigraphy (90). Other than ultrasound, which is commonly used, the latter two are rarely applied clinically. The focus of this thesis was to compare

erosion detection between MRI and radiography specifically. These investigations took the form of a systematic literature review (Chapter 4), the prospective evaluation of measurement reliability across four radiologists participating in this research program (Chapter 5), and an evaluation of the relative merits of MRI and x-ray erosion detection using prospective data (Chapter 6). A study to investigate incremental value of each modality over standard of care in the pharmacotherapeutic treatment decision-making was designed and executed by the author (see **Appendix A**). The former prospective studies were nested within the larger trial.

3.6. References

1. Scott DL, Smith C, Kingsley G. Joint damage and disability in rheumatoid arthritis: an updated systematic review. Clin Exp Rheumatol 2003;21(5 Suppl 31):S20-7.
2. Scott DL. Radiological progression in established rheumatoid arthritis. J Rheumatol 2004;69(Suppl):55-65.
3. Aletaha D, Funovits J, Smolen JS. Physical disability in rheumatoid arthritis is associated with cartilage damage rather than bone destruction. Ann Rheum Dis 2011:70:733-9.
4. United States Department of Health and Human Services, Food and Drug Administration. Guidance for industry: clinical development programs for drugs, devices and biological products for the treatment of rheumatoid arthritis. URL: www.fda.gov. *In* Paulus HE, Oh M, Sharp JT, Gold RH, Wong WK, Park GS, Bulpitt KJ, for the Western Consortium of Practicing Rheumatologists. Classifying structural joint damage in rheumatoid arthritis as progressive or nonprogressive using composite definition of joint radiographic change. A preliminary proposal. Arthritis Rheum 2004;50(4):1083-96.
5. Tavares R, Pope JE, Tremblay J-C, Thorne C, Bykerk VP, Lazovskis J, Blocka KLN, Bell MJ, Lacaille D, Hitchon CA, Fitzgerald AA, Fidler WK, Bookman AAM, Henderson JM, Mosher DP, Sholter DE, Laupacis A, Boire G, Tomlinson G, Bombardier C. Early management of newly diagnosed rheumatoid arthritis by Canadian rheumatologists: a national, multicenter, retrospective cohort. J Rheumatol 2011;38:2342-5.
6. Wakefield RJ, Gibbon WW, Conaghan PG, O'Connor P, McGonagle D, Pease C, Green MJ, Veale DJ, Isaacs JD, Emery P. The value of sonography in the detection of bone erosions in patients with rheumatoid arthritis: a comparison with conventional radiography. Arthritis Rheum 2000;43:2762–70.
7. Alasaarela E, Suramo I, Tervonen O, Lähde S, Takalo R, Hakala M. Evaluation of humeral head erosions in rheumatoid arthritis: a comparison of ultrasonography, magnetic resonance imaging, computed tomography and plain radiography. Br J Rheumatol 1998;37:1152–6.
8. Dohn UM, Ejbjerg BJ, Court-Payen M, Hasselquist M, Narvestad E, Szkudlarek M, Moller JM, Thomsen HS, Ostergaard M. Are bone erosions detected by magnetic resonance imaging and ultrasonography true erosions? A comparison with computed tomography in rheumatoid arthritis metacarpophalangeal joints. Arthritis Res Ther 2006;8:R110.
9. Dohn UM, Ejbjerg BJ, Hasselquist M, Narvestad E, Moller J, Thomsen HS, Ostergaard M. Detection of bone erosions in rheumatoid arthritis wrist joints with magnetic resonance imaging, computed tomography and radiography. Arthritis Res Ther 2008;10:R25.

10. Backhaus M, Kamradt T, Sandrock D, Loreck D, Fritz J, Wolf KJ, Raber H, Hamm B, Burmester G-R. Bollow M. Arthritis of the finger joints: a comprehensive approach comparing conventional radiography, scintigraphy, ultrasound, and contrast-enhanced magnetic resonance imaging. Arthritis Rheum 1999;42:1232-45.
11. Klarlund M, Ostergaard M, Jensen KE, Lysgard JM, Skjodt H, Lorenzen I, TIRA group. Magnetic resonance imaging, radiography and scintigraphy of the finger joints: one year of patients with early arthritis. Ann Rheum Dis 2000;59:521-8.
12. Olech E, Crues III JV, Yocum DE, Merrill JT. Bone marrow edema is the most specific finding for rheumatoid arthritis (RA) on noncontrast magnetic resonance imaging of the hands and wrists: a comparison of patients with RA and healthy controls. J Rheumatol 2010;37:265-74.
13. Ostergaard M, Hansen M, Stoltenberg M, Gideon P, Klaarlund M, Jensen KE, Lorenzen I. Magnetic resonance imaging-determined synovial membrane volume as a marker of disease activity and a predictor of progressive joint destruction in the wrists of patients with rheumatoid arthritis. Arthritis Rheum 1999;42:918-29.
14. Boyesen P, Haavardsholm EA, Ostergaard M, van der Heijde D, Sesseng S, Kvien TK. MRI in early rheumatoid arthritis: synovitis and bone marrow edema are independent predictors of subsequent radiographic progression. Ann Rheum Dis 2011;70:428-33.
15. McGonagle D, Conaghan PG, O'Connor P, Gibbon W, Green M, Wakefield R, Ridgway J, Emery P. The relationship between synovitis and bone changes in early untreated rheumatoid arthritis: a controlled magnetic resonance imaging study. Arthritis Rheum 1999;42:1706-11.
16. McQueen FM, Benton N, Perry D, Crabbe J, Robinson E, Yeoman S, McLean L, Stewart N. Bone edema scored on magnetic resonance imaging scans of the dominant carpus at presentation predicts radiographic joint damage of the hands and feet six years later in patients with rheumatoid arthritis. Arthritis Rheum 2003;48:1814-27.
17. Cheung PP, Dougados M, Gossec L. Reliability and ultrasonography to detect synovitis in rheumatoid arthritis: a systematic literature review of 35 studies (1415 patients). Arthritis Care Res 2010;62:323-34.
18. O'Dell JR. Chapter 10. Rheumatoid arthritis: clinical aspects. *In* Koopman WJ, Boulware DW, Heudebert GR. Clinical Primer of Rheumatology. Philadelphia, Lippincott Williams and Wilkins, 2003. pp. 97-115.
19. Kirwan JR. The relationship between synovitis and erosions in rheumatoid arthritis. Br J Rheumatol 1997;36:225–8.
20. McQueen FM, Stewart N, Crabbe J, Robinson E, Yeoman S, Tan PL, McLean L. Magnetic resonance imaging of the wrist in early rheumatoid arthritis reveals progression of erosions despite clinical improvement. Ann Rheum Dis 1999;58:156–63.

21. McGonagle D, Conaghan PG, O'Connor P, Gibbon W, Green M, Wakefield R, Ridgway J, Emery P. The relationship between synovitis and bone changes in early untreated rheumatoid arthritis: a controlled magnetic resonance imaging study. Arthritis Rheum 1999;42:1706–11.

22. Savnik A, Malmskov H, Thomsen HS, Graff LB, Nielsen H, Danneskiold-Samsøe B, Boesen J, Bliddal H. Magnetic resonance imaging of the wrist and finger joints in patients with inflammatory joint diseases. J Rheumatol 2001;28:2193–200.

23. Savnik A, Malmskov H, Thomsen HS, Graff LB, Nielsen H, Danneskiold-Samsøe B, Boesen J, Bliddal H. MRI of the wrist and finger joints in inflammatory joint diseases at 1-year interval: MRI features to predict bone erosions. Eur Radiol 2002;12:1203–10.

24. Conaghan PG, O'Connor P, McGonagle D, Astin P, Wakefield RJ, Gibbon WW, Quinn M, Karim Z, Green MJ, Proudman S, Isaacs J, Emery P. Elucidation of the relationship between synovitis and bone damage: a randomized magnetic resonance imaging study of individual joints in patients with early rheumatoid arthritis. Arthritis Rheum 2003;48:64–71.

25. Combe B, Landewe R, Lukas C, Bolosiu HD, Breedveld F, Dougados M, Emery P, Ferraccioli G, Hazes JMW, Klareskog L, Machold K, Martin-Mola E, Nielsen H, Silman A, Smolen J, Yazici H. EULAR recommendations for the management of early arthritis: report of a task force of the European Standing Committee for International Clinical Studies Including Therapeutics (ESCISIT). Ann Rheum Dis 2007;66:34-45.

26. Scott DL, Symmons DP, Coulton BL, Popert AJ. Long-term outcome of treating rheumatoid arthritis: results after 20 years. Lancet 1987;329:1108-10.

27. Ward MM, Leigh JP, Fries JF. Progression of functional disability in patients with rheumatoid arthritis. Associations with rheumatology subspecialty care. Arch Intern Med 1993;153:2229-37.

28. Scott DL, Smith C, Kingsley G. What are the consequences of early rheumatoid arthritis for the individual? Best Pract Res Clin Rheumatol 2005;19:117-36.

29. Pierre-Jerome C, Bekkelund SI, Mellgren SI, Torbergsen T, Husby G, Nordstrøm R. The rheumatoid wrist: bilateral MR analysis of the distribution of rheumatoid lesions in axial plan in a female population. Clin Rheumatol 1997;16:80–6.

30. van der Heijde DM. Joint erosions and patients with early rheumatoid arthritis. Br J Rheumatol 1995;34(Suppl 2):74–8.

31. Graudal Jurik AG, de Carvalho A, Graudal HK. Radiographic progression in rheumatoid arthritis: a long-term prospective study of 109 patients. Arthritis Rheum 1998;41:1470-80.

32. Plant MJ, Jones PW, Saklatvala J, Ollier WE, Dawes PT. Patterns of radiological progression in early rheumatoid arthritis: results of an 8 year prospective study. J Rheumatol 1998;25:417–26.

33. Kameda H, Kanbe K, Sato E, Ueki Y, Saito K, Nagaoka S, Hidaka T, Atsumi T, Tsukano M, Kasama T, Shiozawa S, Tanaka Y, Yamanaka H, Takeuchi T. Continuation of methotrexate resulted in better clinical and radiographic outcomes than discontinuation upon starting etanercept in patients with rheumatoid arthritis: 52-week results from the JESMR study. J Rheumatol 2011;38:1585-92.

34. Bathon J, Robles M, Ximenes AC, Nayiager S, Wollenhaupt J, Durez P, Gomez-Reino J, Grassi W, Haraoui B, Shergy W, Park S-H, Genant H, Peterfy C, Becker J-C, Covucci A, Moniz Reed D, Helfrick R, Westhovens R. Sustained disease remission and inhibition of radiographic progression in methotrexate-naïve patients with rheumatoid arthritis and poor prognostic factors treated with abatacept: 2-year outcomes. Ann Rheum Dis 2011;70:1949-56.

35. Visser H, le Cessie S, Vos K, Breedveld FC, Hazes JMW. How to diagnose rheumatoid arthritis early: a prediction model for persistent (erosive) arthritis. Arthritis Rheum 2002;46:357-65.

36. Lifelabs® Medical Laboratory Services, Toronto, ON. Per test results received in 2010.

37. CML Healthcare Inc. Mississauga, ON. Per test results received in 2010.

38. Gamma-Dynacare Medical Laboratories, Brampton, ON. Per test results received in 2010.

39. Royston P, Altman DG, Sauerbrei W. Dichotomizing continuous predictors in multiple regression: a bad idea. Stat Med 2006;25(1):127-41.

40. Ropes MW for A committee of the American Rheumatism Association. Diagnostic criteria for rheumatoid arthritis 1958 revision. Ann Rheum Dis 1959;18:49-53.

41. Arnett FC, Edworthy SM, Bloch DA, McShane DJ, Fries JF, Cooper NS, Healey LA, Kaplan SR, Liang MH, Luthra HS, Medsger Jr. TA, Mitchell DA, Neustadt DH, Pinals RS, Schaller JG, Sharp JT, Wilder RL, Hunder GG. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. Arthritis Rheum 1988;31:315-24.

42. Ejbjerg BJ, Vestergaard A, Jacobsen S, Thomsen HS, Ostergaard M. The smallest detectable difference and sensitivity to change of magnetic resonance imaging and radiographic scoring of structural joint damage in rheumatoid arthritis finger, wrist, and toe joints: a comparison of the OMERACT Rheumatoid Arthritis Magnetic Resonance Imaging Score applied to different joint combinaions and the Sharp/van der Heijde radiographic score. Arthritis Rheum 2005;52:2300-6.

43. Olech E, Freeston JE, Conaghan PG, Hensor EMA, Emery P, Yocum D. Using extremity magnetic resonance imaging to assess and monitor early rheumatoid

arthritis: the optimal joint combination to be scanned in clinical practice. J Rheumatol 2008;35:580-3.

44. Boini S, Guillemin F. Radiographic scoring methods as outcome measures in rheumatoid arthritis: properties and advantages. Ann Rheum Dis 2001;60:817-27.
45. Lassere MN, van der Heijde D, Johnson KR, Boers M, Edmonds J. Reliability of measures of disease activity and disease damage in rheumatoid arthritis: implications for smallest detectable difference, minimal clinically important difference, and analysis of treatment effects in randomized controlled trials. J Rheumatol 2001;28:892-903.
46. Lassere MN, Bird P. Measurements of rheumatoid arthritis disease activity and damage using magnetic resonance imaging. Truth and discrimination: does MRI make the grade? J Rheumatol 2001;28:1151-7.
47. Sokka T. Radiographic scoring in rheumatoid arthritis. A short introduction to the methods. Bull NYU Hosp Joint Dis 2008;66:166-8.
48. Rau R, Wassenberg S, Herborn G, Stucki G, Gebler A. A new method of scoring radiographic change in rheumatoid arthritis. J Rheumatol 1998;25:2094-107.
49. Kaye JJ, Nance EPJ, Callahan LF, Carroll FE, Winfield AC, Earthman WJ, Phillips KA, Fuchs HA, Pincus T. Observer variation in quantitative assessment of rheumatoid arthritis. Part II. A simplified scoring system. Invest Radiol 1987;22:41–6.
50. Genant HK. Methods of assessing radiographic change in rheumatoid arthritis. Am J Med 1983;75:35–47.
51. van der Heijde DM, van Riel PL, Nuver-Zwart IH, Gribnau FW, van de Putte LB. Effects of hydroxychloroquine and sulphasalazine on progression of joint damage in rheumatoid arthritis. Lancet 1989;1:1036-8.
52. van der Heijde D. How to read radiographs according to the Sharp/van der Heijde method. J Rheumatol 1999;26:743-5.
53. van der Heijde D. How to read radiographs according to the Sharp/van der Heijde method. J Rheumatol 2000;27:261-3.
54. Sharp JT, Young DY, Bluhm GB, Brook A, Brower AC, Corbett M, Decker JL, Genant HK, Gofton JP, Goodman N, Larsen A, Lidsky MD, Pussila P, Weinstein AS, Barbara N. Weissman BN. How many joints in the hands and wrists should be included in a score of radiologic abnormalities used to assess rheumatoid arthritis? Arthritis Rheum 1985; 28(12):1326-35.
55. Sharp JT, Lidsky MD, Collins LC, Moreland J. Methods of scoring the progression of radiologic changes in rheumatoid arthritis: correlation of radiologic, clinical and laboratory abnormalities. Arthritis Rheum 1971;14:706-20.

56. Guillemin F, Billot L, Boini S, Gerard N, Ødegaard S, Kvien TK. Reproducibility and sensitivity to change of 5 methods for scoring hand radiographic damage in patients with rheumatoid arthritis. J Rheumatol 2005;32:778-86.

57. Bruynesteyn K, Boers M, Kostense P, van der Linden S, van der Heijde D. Deciding on progression of joint damage in paired films of individual patients: smallest detectable difference or change. Ann Rheum Dis 2005;64:179-82.

58. Bruynesteyn K, van der Heijde D, Boers M, van der Linden S, Lassere M, van der Vleuten C. The Sharp/van der Heijde method out-performed the Larsen/Scott meothod on the individual patient level in assessing radiographs in early rheumatoid arthritis. J Clin Epidemiol 2004;57:502-12.

59. Ostergaard M, Peterfy C, Conaghan P, McQueen F, Bird P, Ejbjerg B, Shnier R, O'Connor P, Klarlund M, Emery P, Genant H, Lassere M, Edmonds J. OMERACT rheumatoid arthritis imaging studies. Core set of MRI acquisitions, joint pathology definitions, and the OMERACT RA-MRI Scoring System. J Rheumatol 2003;30:1385-6.

60. Klarlund M, Ostergaard M, Gideon P, Sorensen K, Jensen KE, Lorenzen I. Wrist and finger joint MR imaging in rheumatoid arthritis. Acta Radiologica 1999;40:400-9.

61. Delfaut EM, Beltran J, Johnson G, Rousseau J, Marchandise X, Cotten A. Fat suppression in MR imaging: techniques and pitfalls. Radiographics 1999;19:373-82.

62. McQueen FM, Benton N, Crabbe J, Robinson E, Yeoman S, McLean L, Stewart N. What is the fate of erosions in early rheumatoid arthritis? Tracking individual lesions using x rays and magnetic resonance imaging over the first two years of disease. Ann Rheum Dis 2001;60:859-68.

63. Brown AK, Conaghan PG, Karim Z, Quinn MA, Ikeda K, Peterfy CG, Hensor E, Wakefield RJ, O'Connor PJ, Emery P. An explanation for the apparent dissociation between clinical remission and continued structural deterioration in rheumatoid arthritis. Arthritis Rheum 2008;58:2958-67.

64. Lindegaard H, Vallo J, Horslov PK, Junker P, Ostergaard M. Low field dedicated magnetic resonance imaging in untreated rheumatoid arthritis of recent onset. Ann Rheum Dis 2001;60:770-6.

65. McQueen FM, Stewart N, Crabbe J, Robinson E, Yeoman S, Tan PLJ, McLean L. Magnetic resonance imaging of the wrist in early rheumatoid arthritis reveals a high prevalence of erosions at four months after symptom onset. Ann Rheum Dis 1998;57:350-6.

66. McQueen FM, Stewart N, Crabbe J, Robinson E, Yeoman S, Tan PLJ, McLean L. Magnetic resonance imaging of the wrist in early rheumatoid arthritis reveals progression of erosions despite clinical improvement. Ann Rheum Dis 1999;58:156-63.

67. Xie X, Webber CE, Adachi JD, O'Neill J, Inglis D, Bobba RS, Wu H. Quantitative, small bore, 1 Tesla, magnetic resonance imaging of the hands of patients with rheumatoid arthritis. Clin Exp Rheumatol 2008;26:860-5.

68. Ejbjerg B, Narvestad E, Rostrup E, Szkudlarek M, Jacobsen S, Thomsen HS, Ostergaard M. Magnetic resonance imaging of wrist and finger joints in healthy subjects occasionally shows changes resembling erosions and synovitis as seen in rheumatoid arthritis. Arthritis Rheum 2004;50:1097-106.

69. McQueen F, Østergaard M, Peterfy C, Lassere M, Ejbjerg B, Bird P, O'Connor P, Genant H, Shnier R, Emery P, Edmonds J, Conaghan P. Pitfalls in scoring MR images of rheumatoid arthritis wrist and metacarpophalangeal joints. Ann Rheum Dis 2005;64(Suppl 1):i48–i55.

70. Rydahl C, Thomsen HS, Marckmann P. High prevalence of nephrogenic systemic fibrosis in chronic renal failure patients exposed to gadodiamide, a gadolinium-containing magnetic resonance contrast agent. Invest Radiol 2008;43:141-4.

71. Shellock FG and Kanal E. Safety of magnetic resonance imaging contrast agents. J Magn Reson Imaging 1999;10:477–484.

72. Segami N, Nishimura M, Kaneyama K, Miyamaru M, Sato J, Murakami K-I. Does joint effusion on T2 magnetic resonance images reflect synovitis? Comparison of arthroscopic findings in internal derangements of the temporomandibular joint. Oral Surg Oral Med Oral Pathol Oral Radiol Endod 2001;92:341-5.

73. Segami N, Miyamaru M, Nishimura M, Suzuki T, Kaneyama K, Murakami K-I. Does joint effusion on T2 magnetic resonance images reflect synovitis? Part 2. Comparison of contrentration levels of preinflammatory cytokines and total preotein in synovial fluid of the temporomandibular joint with internal dergangements and osteoarthrosis. Oral Surg Oral Med Oral Pathol Oral Radiol Endod 2002;94:515-21.

74. Segami N, Suzuki T, Sato J, Miyamaru M, Nishimura M, Yoshimura H. Does joint effusion on T2 magnetic resonance images reflect synovitis? Part 3. Comparison of histologic findings of arthroscopically obtained synovium in internal derangements of the temporomandibular joint. Oral Surg Oral Med Oral Pathol Oral Radiol Endod 2003;95:761-6.

75. Østergaard M, Conaghan PG, O'Connor P, Szkudlarek M, Klarlund M, Emery P, Peterfy C, Genant H, McQueen FM, Bird P, Lassere M, Ejbjerg B. Reducing invasiveness, duration, and cost of magnetic resonance imaging in rheumatoid arthritis by omitting intravenous contrast injection - Does it change the assessment of inflammatory and destructive joint changes by the OMERACT RAMRIS? J Rheumatol 2009;36:1806-10. Erratum in: J Rheumatol. 2010;37:2198.

76. Østergaard M, Pedersen SJ, Dohn UM. Imaging in rheumatoid arthritis – status and recent advances for magnetic resonance imaging, ultrasonography, computed tomography and conventional radiography. Best Pract Res Clin Rheumatol 2008;22:1019–44.

77. McQueen FM. The MRI View of Synovitis and Tenosynovitis in Inflammatory Arthritis. Ann NY Acad Sci 2009;1154:21-34.

78. McQueen F, Lassere M, Edmonds J, Conaghan P, Peterfy C, Bird P, O'Connor P, Ejbjerg B, Klarlund M, Stewart N, Emery P, Shnier R, Genant H, Ostergaard M. OMERACT rheumatoid arthritis magnetic resonance imaging studies. Summary of OMERACT 6 MR imaging module. J Rheumatol 2003;30:1387-92.

79. Naraghi AM, White LM, Patel C, Tomlinson G, Keystone EC. Comparison of 1.0-T extremity MR and 1.5-T conventional high-field-strength MR in patients with rheumatoid arthritis. Radiology 2009;251:829-37.

80. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull 1979;86:420-8.

81. Norman GR, Streiner DL. Biostatistics: The Bare Essentials. Second Edition. Hamilton, B.C. Decker Inc., 2000.

82. Lassere M, McQueen F, Ostergaard M, Conaghan P, Shnier R, Peterfy C, Klarlund M, Bird P, O'Connor P, Stewart N, Emery P, Genant H, Edmonds J. OMERACT Rheumatoid Arthritis Magnetic Resonance Imaging Studies. Exercise 3: An international multicenter reliability study using the RA-MRI Score. J Rheumatol 2003;30:1366-75.

83. Haavardsholm EA, Ostergaard M, Ejbjerg BJ, Kvan NP, Uhlig TA, Lilleås FG, Kvien TK. Reliability and sensitivity to change of the OMERACT rheumatoid arthritis magnetic resonance imaging score in a multireader, longitudinal setting. Arthritis Rheum 2005;52:3860-7.

84. Conaghan P, Bird P, Ejbjerg B, O'Connor P, Peterfy C, McQueen F, Lassere M, Emery P, Shnier R, Edmonds J, Østergaard M. The EULAR–OMERACT rheumatoid arthritis MRI reference image atlas: the metacarpophalangeal joints. Ann Rheum Dis 2005;64(Suppl 1):i11–i21.

85. Østergaard M, Edmonds J, McQueen F, Peterfy C, Lassere M, Ejbjerg B, Bird P, Emery P, Genant H, Conaghan P. An introduction to the EULAR–OMERACT rheumatoid arthritis MRI reference image atlas. Ann Rheum Dis 2005;64(Suppl I):i3–i7.

86. Jimenez-Boj E, Redlich K, Turk B, Hanslik-Schnabel B, Wanivenhaus A, Chott A, Smolen JS, Schett G. Interaction between synovial inflammatory tissue and bone marrow in rheumatoid arthritis. J Immunol 2005;175:2579-88.

87. Cyteval C, Miquel A, Hoa D, Daures JP, Mariette X, Combe B. Rheumatoid arthritis of the hand: monitoring with a simplified MR imaging scoring method--preliminary assessment. Radiology 2010;256:863-9.

88. Suzuki T, Ito S, Handa S, Kose K, Okamoto Y, Minami M, Hayashi T, Goto D, Matsumoto I, Sumida T. A new low-field extremity magnetic resonance imaging and

proposed compact MRI score: evaluation of anti-tumor necrosis factor biologics on rheumatoid arthritis. Mod Rheumatol 2009;19:358-65.

89. Dohn UM, Ejbjerg BJ, Hasselquist M, Narvestad E, Court-Payen M, Szkudlarek M, Moller J, Thomsen H S, Ostergaard M. Rheumatoid arthritis bone erosion volumes on CT and MRI: reliability and correlations with erosion scores on CT, MRI and radiography. Ann Rheum Dis 2007;66:1388-92.
90. Duer A, Ostergaard M, Horslev-Petersen K, Vallo J. Magnetic resonance imaging and bone scintigraphy in the differential diagnosis of unclassified arthritis. Ann Rheum Dis 2008;67(1):48-51.

4. Systematic review

4.1. Introduction

From a pathophysiological perspective, bone erosion in RA refers to the loss of mineralized tissue at a juxta-articular site commonly involving a cortical bone lesion (1). In MRI, a sharply marginated bone lesion visible in two planes, with a cortical break in at least one plane constitutes a bone erosion. Equivalently, a three-dimensional rendering from a single plane using isotropic voxels may be used in MRI in place of the requirement for two planes (2). In MRI, erosions produce a devoid signal on T1-weighted (T1w) sequences. In x-ray, an erosion is depicted as a marginated break in otherwise smooth contours of cortical bone. Cortical bone produces little contrast in T1w MRI, a marginal break in the adjacent cancellous bone is used to detect erosions. Correct juxta-articular location is a requirement for both the x-ray and MRI features. Despite these subtle differences, the definition of the erosion signal is relatively consistent across the two diagnostic imaging modalities.

Reports indicate that MRI erosions are more prevalent than radiographic ones (3-6). MRI is expected to have a lower limit of detection for bone erosions than x-ray, resulting from its multi-sectional and multi-planar features (7). These features enable MRI to resolve small cortical lesions at one or more cross-sectional slices. The use of multiple planes helps to cross-reference the detected signal. Cross-referencing also helps to rule out imaging artifacts. It is also useful for accounting for out-of-plane features in any single plane. In contrast, in radiography, the entire anatomy of interest is superimposed on to a 2D image. This limits the detection of x-ray erosions to lesions

disrupting the otherwise smooth bone contour of the 2D silhouette projected onto the plain film. The use of multiple projections in radiography improves the chance of detecting erosions. Further, x-ray has the advantage of superior *in plane* resolution. Reports of an increased prevalence of MRI erosions suggest that the use of multiple projections and superior *in plane* resolution can only partially reconcile the fundamental differences in the limits of detection between these distinct approaches to measuring bone erosion.

In the preceding chapter MRI and x-ray were introduced as two prominent diagnostic imaging modalities used in rheumatology for the detection of bone erosions in RA. Radiographic evidence of erosive joint damage is prognostic of long-term functional disability (8). It is used in the clinical management of RA as a prognostic marker to determine disease severity and primary outcome in the assessment of long-term treatment efficacy (9). A primary aim of RA pharmacotherapy is the inhibition of radiographic erosive progression. Erosive progression models have been developed to determine predictors of erosive joint damage. Diagnostic imaging modalities with lower limits of detection for erosion may be of clinical utility in *targeted-treatment* approaches to limit radiographic progression. Magnetic resonance imaging may be one such modality. Pharmacotherapeutic management to pre-radiographic, MRI changes may help to inhibit radiographic evidence of disease progression and thereby improve long-term functional outcomes.

The evidence in the literature relating the association between MRI and X-ray detection of erosions was investigated in the current study. MRI is expected to have a

lower limit of detection for erosions than X-ray. In turn, all larger erosions detected on X-ray are expected to be detectable on MRI. Since RA is a progressive disease, erosions increase in size with time. It was expected that temporal associations between an MRI and x-ray erosion detection exist. Beyond a temporal dependence in the relative diagnostic test accuracy of the two modalities, additional sources of inter-study heterogeneity were expected. The primary objective of this systematic review was to summarize the sensitivity and specificity of x-ray for MRI erosions in RA. The secondary objective was to determine if the relative diagnostic test accuracy between the modalities was dependent on symptom duration and the time between imaging interventions. Generic research quality considerations and additional sources of inter-study heterogeneity specific to the research question were summarized and discussed.

4.2. Methods

4.2.1. Overview

The Cochrane Collaboration framework for conducting systematic reviews of diagnostic test accuracy was applied to this investigation (10). A protocol with explicit eligibility criteria was prepared *a priori*. A structured literature search strategy was applied across multiple citation indexes. Two reviewers sequentially screened the titles, abstracts and full articles of the identified citations. Discrepancies between the reviewers were discussed to the reconcile differences and a third reviewer was available to arbitrate the inclusion of discrepant citations. As part of the Cochrane Review Manager software utilized (version 5.1.5), items from the Quality Assessment of Diagnostic Accuracy

Studies (QUADAS) questionnaire were included. These items were used to assess the quality of included studies (11). Data from the included studies were entered in the Review Manager software.

4.2.2. Study eligibility and literature search strategy

In consultation with a McMaster University Research Librarian experienced with systematic reviews (12,13), a literature search strategy was designed using a modified population, intervention, comparator, outcome (PICO) framework. Medical Subject Heading (MeSH) terms were used and supplemented with keyword terms. The specific search strategy used was included in **Appendix B**.

Primary literature written in the English language was considered for study eligibility. Studies in which direct comparisons of MRI and x-ray erosion detection data could be extracted were included. Complete true positive, false positive, false negative, and true negative data were extractable from eligible studies. Eligibility was not restricted by any criteria potentially resulting in study heterogeneity, including the following:

1) Study design;
2) Quality assessment of diagnostic test accuracy studies appraisal items;
3) Reference standard or index test definition;
4) Time interval between reference standard and index test imaging intervention; and,
5) Patient symptom duration.

Multiple studies from same cohort were included provided that novel index test and reference standard comparative timepoints were reported. Secondary literature was excluded.

The search strategy was run on Wolters Kluwer Health Ovid SP MEDLINE® without Revisions 1996 to Present search engine. The same search strategy was then carried out on EMBASE 1998 to Present through the same search engine provider. The literature search was initially carried out in November 2008 and updated in April 2011. The reference lists of full articles reviewed were searched for additional eligible studies. Titles, abstracts and full articles of the literature search citations were sequentially screened by two reviewers to rule out ineligible articles.

Citations identified in the literature search were exported from OVID to Refworks™. Within Refworks™, citations were sorted into folders based on exclusionary phase: by title or abstract. Full articles were retrieved for the remaining citations. Citations were exported from Refworks™ to Review Manager™, version 5.1.5, where they were further sorted by study eligibility criteria. Citations identified from the reference list of full article reviews were imported into Review Manager™. All analyses were conducted using Review Manager™ following the structure of Cochrane Collaboration systematic reviews of diagnostic test accuracy.

4.2.3. Analysis

4.2.3.1. Assessment of methodological quality

The risk of bias in research needs to be considered in any critical appraisal of the literature in order to evaluate the validity of the results and strength of evidence. In a systematic review, research quality assessment may be used to weight the relative strength of evidence between compared studies and guide interpretation of its findings in the context of the research question. In the Cochrane Collaboration framework for systematic reviews of diagnostic test accuracy, QUADAS items are used to assess the quality of included articles (11). The original tool includes 14 items to assess the risk of bias in diagnostic accuracy studies (11). Of these, 11 are included in the Cochrane software. The QUADAS tool guidance was referred to to ensure correct interpretation (11).

The evaluation of bias in studies of the diagnostic test accuracy may be more subtle than in interventional studies. The items included in the QUADAS tool address the following aspects of studies for diagnostic test accuracy: external validity of the population and clinical setting (items 1, 2, and 12), appropriate selection of the reference standard (item 3), measurement bias of the reference standard or index test (items 4-11), and missing data (items 13 and 14). Items 2 (transparency of eligibility criteria), and 8 and 9 (transparency of reference standard and index test descriptions) were not included in the Cochrane framework and were therefore not considered here.

4.2.3.2. Statistical analysis and data synthesis

Given the hypothesis that MRI detects smaller erosions than x-ray, the former was used as the reference standard for the calculation of sensitivity and specificity. Three separate analyses were conducted. First, studies were organized by the time between MRI and x-ray imaging, such as baseline-vs-baseline comparisons, baseline MRI and year one x-ray, baseline MRI and year two x-ray, etc.. Across these stratifications, as the duration between initial MRI and follow-up x-ray increased, the sensitivity of x-ray erosions for MRI erosions was expected to increase. Specificity was expected to remain high across these stratifications; however, as the interval between the MRI and x-ray increased, the probability of the development of an erosion at a novel site not initially detected by MRI was also expected to increase. As such, specificity may decrease with increasing time interval between initial MRI and follow-up x-ray imaging. Within these stratifications, studies were sorted by increasing symptom duration. It was expected that the sensitivity of x-ray for the detection of MRI erosions would increase as the symptom duration increased. Specificity was expected to remain unchanged within strata. In a separate stratified analysis, studies were sorted by symptom duration only. It was unknown if symptom duration alone was of greater influence to the association between MRI and x-ray than the duration between imaging.

In these studies, it was expected that the sensitivity of x-ray for MRI erosions would increase with increasing symptom duration. Within cross-sectional comparisons, the specificity of x-ray for MRI erosions was expected to remain high. As explained above, the specificity was expected to decrease as the time interval between MRI and

follow-up x-ray increased. A hybrid variable of symptom duration and follow-up time interval was created as both data were expected to positively affect the association between initial MRI and follow-up x-ray detection of erosions. Finally, within longitudinal studies, the temporal association between initial MRI erosion and follow-up x-ray was determined. Here too, sensitivity was expected to increase while specificity was expected to decrease. Consistency in the temporal association between initial MRI and downstream x-ray would support the study hypothesis.

4.2.3.3. Investigations of heterogeneity

The studies included in systematic reviews of diagnostic test accuracy are expected to have more heterogeneity than those of interventional studies (14). In interventional studies, tests of heterogeneity and inconsistency in results are sometimes interpreted as providing insight to potential publication bias. In systematic review of interventional studies, tests for the asymmetry of funnel plots may be conducted to test for heterogeneity between the results of included studies. Such tests for heterogeneity possess low statistical power (14). The accuracy of the funnel plot asymmetry test deteriorates with increasing magnitude of odds ratios for the association measured (14). In interventional systematic reviews where effect sizes approach an odds ratio of 1, these tests find some limited application. In contrast, the large odds ratios expected from studies of diagnostic test accuracy render these tests of no value. For systematic reviews of diagnostic test accuracy, it is instead recommended that the sources of hypothesized heterogeneity be explored (14). Multivariable methods of doing so were beyond the scope of this project. Instead, heterogeneity was explored on a bivariate basis by

stratifying the results as described above. Data were categorized or stratified by the variables in question and trends in sensitivity and specificity examined.

4.3. Results

4.3.1. Results of the search

The literature search identified 2657 citations, of which 16 journal articles were included in the review (**Figure 1**). A total of 2184 citations were ruled out by screening the title. An additional 322 were ruled out upon reviewing the abstract. The remaining 151 citations required full article review to determine eligibility. Of these, 87 did not meet at least one inclusion criterion: 18 citations were secondary literature, 14 citations pertained to the wrong disease, and 55 did not include data on either x-ray (n=39) or MRI (n=16). An additional 48 articles fulfilled exclusion criteria. Of these, 43 contained insufficient data for extraction and 5 were either duplicate citations or unique citations containing redundant data with an included article. The remaining 16 articles were included in the review.

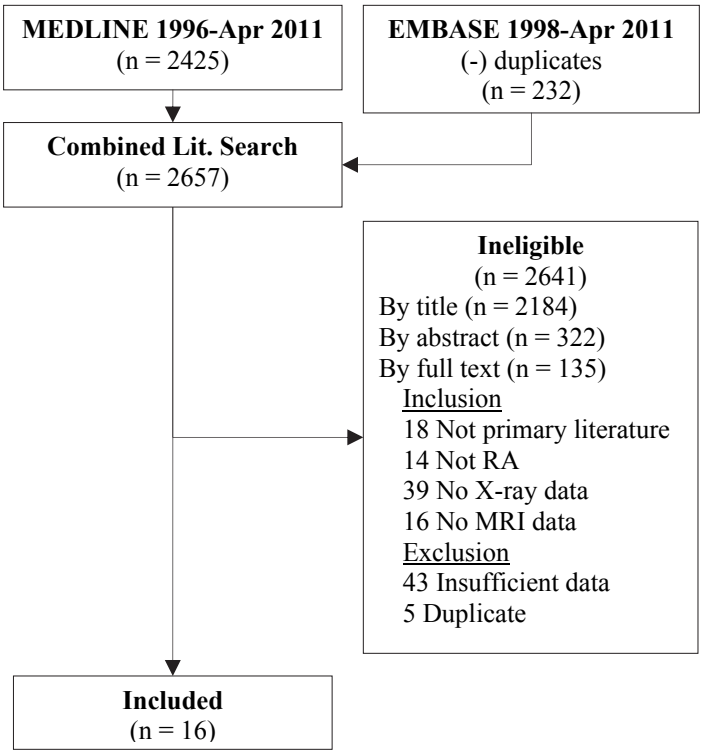


Figure 1. Systematic review flow diagram.

4.3.2. Methodological quality of included studies

The majority of the quality items were satisfied across the included articles (**Figure 2**). Some corollaries and exceptions to this general finding exist and were reported below. The avoidance of partial verification and accounting for missing data were exceptions not satisfied by a considerable number of articles. The quality assessment was evaluated liberally and may have been biased towards a higher quality rating. Evaluations of methodological quality were also summarized by study and QUADAS questionnaire items (**Figure 3**).

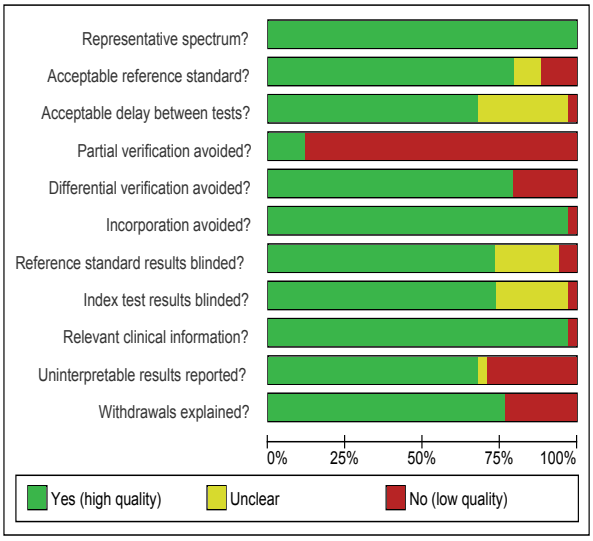


Figure 2. Percentage of included studies fulfilling QUADAS methodological quality items (11).

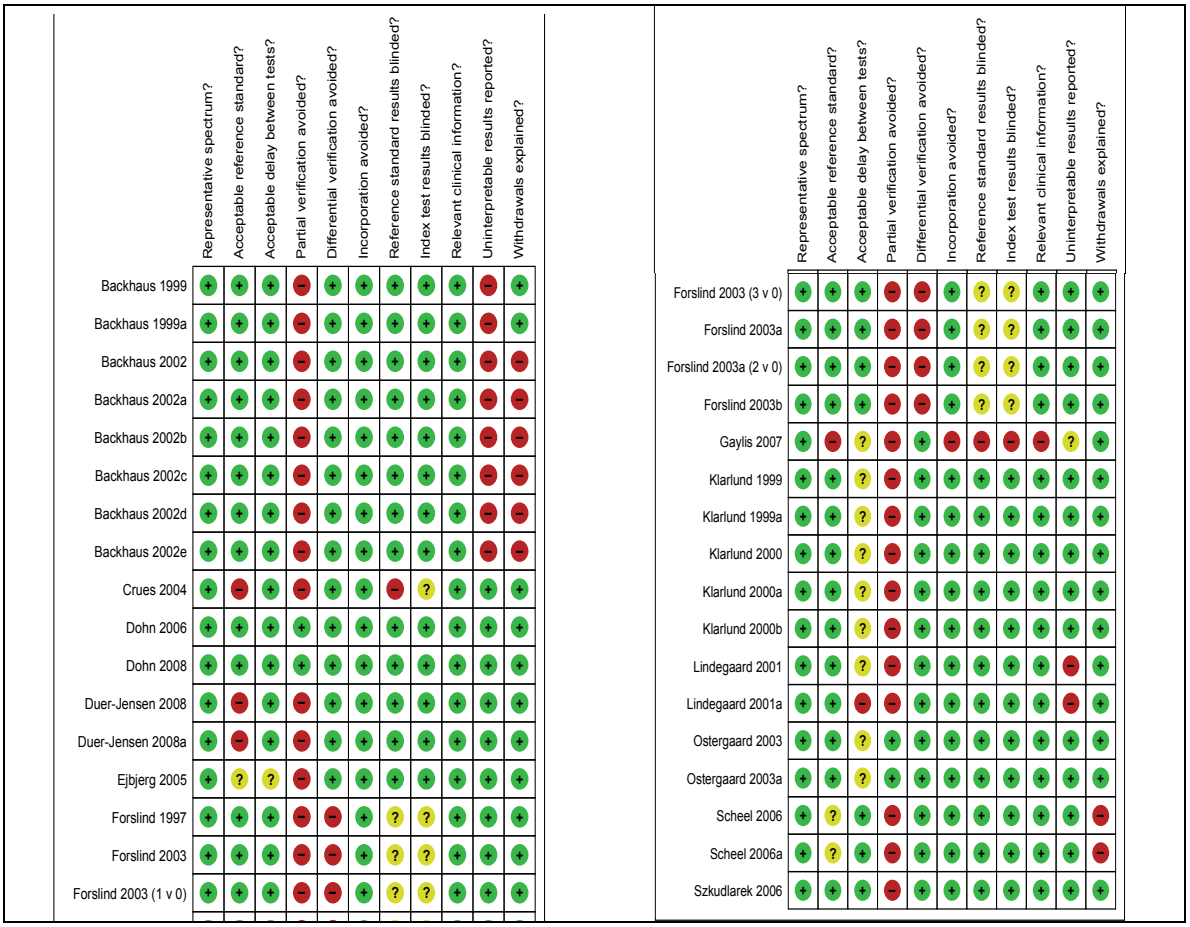


Figure 3. QUADAS methodological quality assessment of included studies (11).

Item 1, representative spectrum, was evaluated quite liberally relative to the QUADAS user guidance (11). The guidance recommends that any study including normal controls be negatively evaluated. Normal, or healthy, participants are theoretically expected to have negative findings for both the reference standard and index test, thereby artificially increasing the association between the two modalities. A handful of included studies incorporated a minority of healthy controls (6,15-20). Other studies included a minority of non-RA IA and non-IA conditions (5,21-3). For two reasons QUADAS item 1 was favourably evaluated for these articles. First, the proportion of non-disease participants was usually small compared to the number of disease participants included in the analysis. Even among the disease population, the majority of observations were expected to have negative findings given the use of the joint quadrant, bone, or joint as the unit of analysis. As such, the impact of including a minority of normal patients with a lesser expected number of erosions on the determination of relative accuracy was expected to be small.

Item 2, acceptable reference standard, was evaluated based on the fulfillment of the Outcome Measures in Rheumatology (OMERACT) RA MRI score (RAMRIS) criteria for the measurement of erosions on MRI (2). Some notable exceptions to the use of an acceptable reference standard were noted (17,23,24). In two studies, the erosion definition included T1-weighted signal loss in one plane only and required a corresponding STIR signal (17,22). In the latter, T1-weighted imaging was limited to a single plane and multiplanar reconstruction was conducted using images with unequal (non-isotropic) voxels (23).

Item 3, acceptable delay between tests, refers to the time interval between administration of the index test and the reference standard. Ideally, the two imaging modalities would be tested at the same time to negate the chance of disease progression between imaging affecting the results. In clinical research, the ideal is often not practical and the delay between the two interventions occurs. The delay between the two interventions was rarely controlled and for 30% was not reported and therefore *'unclear*'. This item was evaluated quite liberally since clinical trial and longitudinal cohort data suggest that a minority of patients progress on an annual basis, approximately 20%. As such, the likelihood of change over a shorter period of time is likely to be less than 20%. In this review, delays less than one month were accepted and delays greater than that were more subjectively evaluated, usually as unacceptable.

Item 4, partial verification, was the most problematic quality measure across the included studies. Partial verification bias refers to the scenario where the reference standard and index tests are carried out on neither all nor a random sample of all units of analysis. Partial verification may bias the estimates of overall diagnostic accuracy if data are available for one imaging modality and not the other. If a correlation exists between uninterpretable observations and the presence of erosions, then removing these observations may bias the analysis. In approximately 18% of the included articles (n=3), the number of units of analysis evaluated (be it joint quadrants, bones or joints) did not represent the expected number indicated in the methods (15,16,25). Observations were more frequently discarded due to missing MRI data than x-ray data, frequently related to the limited MRI field of view.

Here, the distinction between partial verification and unreported, uninterpretable results was difficult to differentiate. As a result, the approach was taken in the quality assessment to not penalize an article multiple times for the same issue. Therefore, if the number of observations in the results did not match those expected from the methods section and the study was evaluated to suffer from partial verification, this evidence was not used to gauge uninterpretable results.

Item 5, differential verification, refers to using a different reference standard depending on the result of the index test. In some articles the MRI device used may have varied between patients or over the longitudinal course of the study. Differential verification bias may also affect the results in studies where reference standard and index test results are unblinded and subject to incorporation bias.

Item 6, incorporation bias, refers to the independence of the reference standard and index test. MRI and x-ray are independent imaging modalities. Because erosions are uncommon at the bone or joint level, studies may be designed to increase the erosion signal. One way of doing this is to image only certain joints known to be more commonly affected: this is always done. A additional strategy is to include only patients with x-ray and/or MRI evidence of erosion in at least one joint. This selection process introduces a bias that may artificially increase the association between the two modalities. For example, if a positive signal on the reference standard is a selection criterion, the number of false positives is diminished, driving up the specificity of the index test. In contrast, studies of participants with predominantly negative reference standards drive results towards low sensitivity and high specificity.

Items 7 and 8, blinding of the reference standard and index tests, respectively, are related to 5 and 6. Blinding is a technique used to restrict knowledge of the reference standard results from persons involved in the measurement of the index test and vice-versa. It ensures the independence of the measurement of the reference standard and index test. Multiple levels of blinding are possible, such as participant, investigator/rater, analysis, etc. In this review, the blind of interest pertained to the rater(s) of the images. Lack of a blind was expected to introduce analytical bias affecting items 5, differential verification, and 6, incorporation bias. Its absence was most blatant in Gaylis *et al*. (2007) where, "3 erosions were visible on x-ray retrospectively after viewing the MRI image, and 2 erosions were visible on MRI retrospectively after seeing the x-rays." In this example, the lack of a blind led to both differential verification and incorporation bias, where the subset of dataset with positive reference standard or index test results was reinterpreted. The affected units of analysis were re-examined with greater scrutiny than those devoid of signal in the first analysis. Greater or lesser rater scrutiny may be imposed to artificially modify the association between the two modalities.

Item 9, presentation of relevant clinical information, is difficult to critically evaluate in the context of the research question. Knowledge of the clinical context of the imaging requisition may affect the diagnostic test result, especially when interpretation is required (11). The use of clinical information may introduce expectation bias. In the context of the current research question, where the diagnostic imaging modalities were compared for the detection of the erosion signal without commentary on clinical correlates or presumed causation, the relevance of clinical information may have been

less important. For example, erosions are not expected in healthy subjects and therefore signals consistent with the definition of erosion may be dismissed on clinical grounds or rationalized against in consideration of potential pitfalls with the modality (26). This could result in an artificially low signal in both the reference standard and index tests, resulting in high specificity. By comparison, in the setting of research, where the raters are required to adhere to a specific protocol beyond the scope of their clinical standard of care, the raters are likely to be aware of the disease context, which may result in more frequent reporting of signal and increase sensitivity due to chance agreement and lower specificity due to reduced measurement reliability.

Whiting *et al*. suggest that the relevance of clinical information to the diagnostic interventions being studied be determined *a priori*, and the evaluation of compliance with this quality assessment item be assessed in that context. Others have suggested that the provision of clinical data introduces *'clinical review bias'* (27-30). Clinically, the data available to a radiologist for interpretation may be dependent on the referring physician's and radiologist's standard of care. Different referring physicians may provide differing amounts of clinical information to help guide the radiologist's assessment. In turn, the radiologist may or may not dictate a requirement for clinical correlates to evaluate the diagnostic imaging test. In this review, this item was evaluated liberally, on the basis of a systematic presentation of some clinical data expected to be presented equivalently to the raters across the two imaging modalities. All but one study was evaluated as satisfying the criterion. The one exception, Gaylis *et al*. 2007, presented minimal disease characteristics for the screened sample and not the included participants (24). As

discussed above, in the context of the research question, the relevance and effect of this item on the integrity of the study results is unclear.

Items 10 and 11 pertain to data management biases associated with improper handling of uninterpretable data and study withdrawals, respectively. Uninterpretable results are common to studies of diagnostic tests and are frequently discarded from the analysis with little consideration. The introduction of bias with this practice depends on the correlation between uninterpretable results and disease status, and disproportionate distribution between the reference standard and the index test. Theoretically, a random distribution of uninterpretable results has no affect on the accuracy of the results (11). As described above, MRI data were frequently missing from many studies, due to limited field of view. The disproportionality between uninterpretable MRI vs x-ray data may result in a biased assessment. In the current review, approximately 30% of the included studies did not appropriately report uninterpretable results. Interestingly, the quality assessment guidance only calls for transparency of such data and not the adoption of a strategy to manage it. Study design and statistical techniques may be used to manage missing data issues. These were rarely considered in the articles reviewed.

Overall, the articles included in the analysis were of acceptable quality as determined using a lenient interpretation of the generic QUADAS quality items included in the Cochrane Collaboration diagnostic test accuracy systematic review software (11,14). The quality evaluation conducted here may have been biased towards higher quality than what others would assess.

Other Quality Assessment Considerations. A number of additional quality considerations specific to the research question were considered as potential sources of risk to study quality and heterogeneity in the results. In no particular order, the issues identified were the following:

- Symptom duration
- Anatomy investigated
- MRI magnetic strength (as a surrogate for signal to noise ratio)
- Erosion definition
- The number of x-ray projections used
- Prescribed systems used to score MRI and x-ray erosions
- Inter-rater reliability assessment
- Number of raters
- Unit of analysis

The Cochrane Collaboration recommends that sources of expected heterogeneity between studies be explored to estimate their impact on the results. In this study such an exploration was limited to symptom duration and the time interval between initial MRI and follow-up x-ray assessment as described in the findings below. The prevalence in the included studies of other potential sources heterogeneity were reported in the findings below and discussed in the subsequent section.

4.3.3. Findings

Study Characteristics. The characteristics of the included studies were summarized in **Table 1**. The 16 articles included 34 datasets and 10 953 paired observations directly comparing bone erosion detection on MRI and radiography. Of

these, 26 cross-sectional datasets, including 9648 observations, compared the two modalities at the same timepoint. Eight datasets characterized temporal associations between initial MRI detection of erosion and subsequent x-ray follow-up (n=1305). Of these, two datasets including 295 observations compared baseline MRI data with one-year radiography data; three datasets including 709 observations compared baseline MRI data with two-year radiography data; an additional three datasets including between 23 and 150 observations compared baseline with three-, five-, and seven-year radiography data.

Four longitudinal cohorts reported in 7 articles included 18 paired datasets with 4189 observations. Collectively, these were used to investigate temporal patterns in the association between initial MRI and follow-up x-ray. The first of these longitudinal cohorts, covered in three separate papers (21,22,31), included 7 datasets with 3162 paired observations. Intra-study follow-up included cross-sectional comparisons at multiple symptom durations as well as data comparing baseline MRI with two- and seven-year x-ray data. The symptom durations covered by cross-sectional data extracted from this cohort ranged from 1.8 to 14 years. The second longitudinal cohort was covered by a single citation (25). It included two datasets with 285 paired observations. Similarly, a third cohort included two datasets with a total of 544 paired observations (5). The final cohort included seven datasets and 198 paired observations (32,33).

Table 1. Characteristics of included studies.

Study (Origin)	Clinical Features and Settings	Study Design	Reference Standard	Index Test	Anatomy	No. of Datasets
Backhaus 1999 (Germany) (21)	Group I - Larsen grades 0-1: 1987 ACR RA (n=17); ESSG SpA (n=6); SLE (n=2); UIA (n=7); 24 XX; median (range) age, 39.5 (19-64) years; symptom duration, 1.8 (0.2-16.2) years; NSAID, 50%; steroids, 34%; DMARD, 12-24%; CRP 14 (0.4-102.6) mg/dL; ESR 20 (2-67) mm/h Group II - Larsen grades 2+: 1987 ACR RA (n=19); ESSG SpA (n=9); 16 XX; median (range) age, 55 (28-84) years; symptom duration, 9.25 (0.6-26.1) years; CRP, 9.2 (1.7-73.1) mg/dL; ESR, 23 (2-70) mm/h; DMARD, 43-79%; steroids, 50%; NSAID, 64%	Cross-sectional, paired blinded control	1 radiologist (other than x-ray rater); 0.2T Magnetom Open (Siemens, Erlangen, Germany); OMERACT compliance of erosion definition unclear	1 radiologist (other than MRI rater); conventional radiography optimized to visualize trabeculae, joints and soft tissue of the hands; 2 projections (PA; L); Larsen scoring system	W MCP, PIP, DIP	2
Backhaus 2002 (Germany) (22)	From Backhaus 1999; Group I - Larsen grades 0-1: 1987 ACR RA (n=16); ESSG SpA (n=7); SLE (n=2); UIA (n=3); 20 XX; median (range) age, 41 (20-66) years; symptom duration, 3.5 (2-18.4) years; CRP 3.7 (1.0-72.8) mg/dL; ESR 15 (2-81) mm/h; RF+, 36%; HLA-DRB1+, 57%; at baseline, DMARD, 46-57%; NSAID, 0%; corticosteroids, 46% Group II - Larsen grades 2+: 1987 ACR RA (n=15); ESSG SpA (n=6); 10 XX; median (range) age, 58 (35-86) years; symptom duration, 9.5 (2.7-27.5) years; CRP, 3.9 (1.0-28.8) mg/dL; ESR, 28 (4-81) mm/h; RF+, 67%; HLA-DRB1+, 48%; DMARD, 76-96%; corticosteroids, 67%; NSAID, 14%	Prospective, paired blinded control (temporal blind unclear)	1 radiologist (other than x-ray rater); 0.2T Magnetom Open (Siemens, Erlangen, Germany); OMERACT compliance of erosion definition unclear	1 radiologist (other than MRI rater); conventional radiography; 2 projections (PA; L); Larsen scoring system	W MCP, PIP, DIP	6
Crues 2004 (USA) (23)	n=132; 95% 1987 ACR; 5% joint symptoms with psoriasis; 101 XX; mean (range) age, 62 (32-88) years	Cross-sectional, paired unblinded control; recruitment strategy NS	2 MSK radiologists (per x-ray); 0.2T MagneVu 1000 (MagneVu, Carlsbad, CA, USA); Erosion definition not OMERACT compliant	2 MSK radiologists (per MRI); conventional radiography; 3 projections (PA,L,O)	B MCP 2-3, wrist	1

Study (Origin)	Clinical Features and Settings	Study Design	Reference Standard	Index Test	Anatomy	No. of Datasets
Dohn 2006 (Denmark) (15)	1987 ACR RA (n=17) with ≥1 radiographically occult MCP MRI erosion/healthy controls (n=4); 13/3 XX; median (range) age, 52 (33-78)/35.5 (34-57) years; RA symptom duration 8 (4-22) years; RF+, 82%	Cross-sectional, paired blinded control; recruited from pre-existing MRI study	1 MSK radiologist (other than x-ray rater); 0.6T Philips Panorama unit (Philips Medical Systems; Helsinki, Finland); OMERACT-compliant erosion definition	1 MSK radiologist (other than MRI rater); conventional radiography; 2 projections (PA,O); Philips Digital Diagnost unit (Philips Medical Systems; Hamburg, Germany)	S MCP	1
Dohn 2008 (Denmark) (16)	1987 ACR RA (n=17) with ≥1 radiographically occult MCP MRI erosion/healthy controls (n=4); 13/3 XX; median (range) age, 51 (33-78)/36 (34-57) years; symptom duration, 8 (4-22) years; RF+, 82%	Cross-sectional, paired blinded control; recruited from pre-existing MRI study	1 radiologist; 0.6T Philips Panorama unit (Philips Medical Systems; Helsinki, Finland); OMERACT-compliant erosion definition;	1 MSK radiologist other than MRI rater); conventional radiography; 2 projections (PA,O); Philips Digital Diagnost unit (Philips Medical Systems; Hamburg, Germany); van der Heijde-modified Sharp Score	S Wrist	1
Duer-Jensen 2008 (Denmark) (17)	1987 ACR RA (n=15)/healthy controls (n=4); 10/2 XX; median (range) age, 58 (25-79)/56 (35-67) years; symptom duration, 7.5 (1-33) years; SJC/28, 2.5 (0-9), TJC/28, 0.5 (0-20), DAS28, 2.7 (0-5.4), RF+, 60%; CRP <8mg/dL, 67%; 40% MTX combination therapy; MTX monotherapy, 26.6%; non-MTX DMARD monotherapy, 20%	Cross-sectional, paired blinded control	1 rheumatologist; 0.2T Artoscan MRI unit (Esaote Biomedica, Genoa, Italy); erosion definition not OMERACT compliant	1 MSK radiologist; conventional radiography; 2 projections (PA,O)	MCP, Wrist	2
Ejbjerg 2005 (Denmark) (18)	1987 ACR RA patients (n=37); healthy controls (n=28); XX 34/20; median (range) age 52 (33-84) / 38 (24-57); disease duration, 5 (1-37)/0 (0-0); TJC, 8 (0-22)/0 (0-0); SJC, 6 (0-16)/0 (0-0); AMS, 15 (0-120)/0 (0-0); HAQ, 0.875 (0-2.38)/0 (0-0); CRP, 10 (8-111)/8 (8-16) g/dL; RF+, 73%/3.5%; DAS28-CRP, 4.36 (2.0-6.9)/NA; DMARD, 70%/NA; Biologic therapy, 13%/NA; corticosteroids, 16%/NA; NSAID, 35%/NA; analgesics, 29%/NA	Cross-sectional; paired, modality-blinded control	1 radiologist (per x-ray rater); 1T Siemens Impact high field MRI unit; OMERACT-compliant erosion definition	1 radiologist (per MRI rater); conventional radiography; 2 projections (PA,O)	MCP, Wrist	1

Study (Origin)	Clinical Features and Settings	Study Design	Reference Standard	Index Test	Anatomy	No. of Datasets
Forslind 1997 (Sweden) (32)	1987 ACR RA (n=30); 23 XX; median (range) age, 59 (24-80) years, disease duration, 8 (2-14) months	Cross-sectional, paired control (blinding unclear)	2 radiologists; 1T Siemens Magnetom Impact; OMERACT-compliant erosion definition	1 radiologist; conventional radiography; 3 projections (NOS)	Knee, MTP 5	1
Forslind 2003 (Sweden) (33)	Subset of Forslind 1999; 1987 ACR RA (n=23); 18 XX; median (range), baseline age, 56 (28-78) years; disease duration, 6 (1-12) months. Characteristics and treatment at baseline/1y/3y: CRP (g/L), 13 (4-121)/4 (4-87)/4 (4-77); DAS28, 4.85 (2.30-6.39)/2.89 (0.77-5.4)/2.55 (0.97-5.51); HAQ, 0.8 (0.0-2.5)/0.3 (0.0-1.4)/0.19 (0.0-2.4); DMARD, 0/35/48%; steroid, 26/70/65%	Prospective, paired unblinded control (temporal blind unclear)	2 radiologists; 1T Siemens Magnetom Impact; OMERACT-compliant erosion definition	1 radiologist (per MRI rater); conventional radiography; 3 projections (PA,O,L)	MTP 5	6
Gaylis 2007 (USA) (24)	1987 ACR RA (n=31); 1y infliximab study compliers with imaging data	Retrospective cohort, paired unblinded control	2 radiologists (per x-ray raters); 0.2T MagneVu 1000 (Carlsbad, CA, USA); erosion definition not OMERACT-compliant	2 radiologists (per MRI raters); conventional radiography; 2 projections (PA,O,L)	MCP 2-3	1
Klarlund 1999 (Denmark) (19)	1987 ACR RA (n=41)/healthy controls (n=3); median(range) age 61(24-83)/31(24-33) years	Cross-sectional, paired blinded control	1 radiologist; 1.0T Siemens Impact /1.5T SP63 Siemens Magnetom MR units; OMERACT-compliant erosion definition	1 radiologist (per MRI rater); conventional radiography; 2 projections (PA,O)	MCP, Wrist	2
Klarlund 2000 (Denmark) (5)	Subset of Klarlund 1999; 1987 ACR RA (n=21)/new RA (n=5)/UIA (n=8); n XX, 11/5/7; median (range) age, 50 (20-82) years; disease duration, 0.25 (0-1.83) years; RF+, 47%; SJC/28, 6 (0-18); TJC/28, 15 (0-24); ESR, 15 (3-42) mm/h; sCRP, 95 (95-543) nmol/L; DMARD, 24%; steroid, 29%; NSAID, 68%	Prospective, paired modality blinded and temporally unblinded control	1 radiologist (per x-ray rater); 1.0T Siemens Magnetom Impact Unit (Erlangen, Germany); OMERACT-compliant erosion definition	1 radiologist (per MRI rater); conventional radiography; 2 projections (PA,O); Larsen scoring system	MCP, PIP	3
Lindegaard 2001 (Denmark) (6)	1987 ACR RA (n=25)/healthy controls (n=3); median (range) age, 55 (20-69)/46 (34-55) years; XX, 18/1; disease duration, 0.33 (0.15 - 0.99) years	Cross-sectional, paired blinded control	1 radiologist; 0.2T Artoscan (Esaote Biomedica, Italy); OMERACT-compliant erosion definition	1 MSK radiologist; conventional radiography; 2 projections (PA,O)	MCP, Wrist	2

Study (Origin)	Clinical Features and Settings	Study Design	Reference Standard	Index Test	Anatomy	No. of Datasets
Ostergaard 2003 (Denmark) (25)	1987 ACR RA (n=10); median (range) age, 62 (20-76) years, disease duration, 1.5 (0.5-9) years; DMARD, 100%	Prospective; paired, blinded control (temporal blinding unclear)	Multiple raters; MRI unit NOS; OMERACT-compliant erosion definition	Multiple raters; conventional radiography; 1 projection (PA)	Wrist	2
Scheel 2006 (Germany) (31)	Subset of Backhaus 1999; Larsen grades 0-1; 1987 ACR RA (n=16)	Prospective, paired modality and temporally blinded control	2 radiologists; 0.2T Magnetom Open (Siemens, Erlangen, Germany)/0.2 Tesla Esaote (Esaote C-Scan, Genoa, Italy) MRI units; erosion definition not OMERACT-compliant	1 radiologist (other than MRI rater); conventional radiography; 2 projections (PA,L); Larsen scoring system	MCP, PIP	2
Szkudlarek 2006 (Denmark) (20)	1987 ACR RA with ≥3 swollen or tender MCP or PIP joints (n=40)/healthy controls (n=20); n XX, 32/16; median (range) age, 58 (23-79)/52 (27-79) years; disease duration 5 (0-20) years; DMARD, 87.5%	Cross-sectional, paired blinded control	1 radiologist; 1.0 T Siemens Impact MR unit (Siemens, Erlangen, Germany); OMERACT-compliant erosion definition	1 radiologist; conventional radiography; 2 projections (PA,O)	MCP, PIP	1

ACR, 1987 American College of Rheumatology (formerly American Rheumatism Association) classification criteria for RA; B, bilateral; (s)CRP, (serum) C-reactive protein; DAS28, 28-joint disease activity score; DIP, distal interphalangeal (joints); DMARD, disease-modifying anti-rheumatic drug(s); ESR, erythrocyte sedimentation rate; ESSG, European Spondyloarthropathy Study Group; HAQ, Health Assessment Questionnaire; HLA-DRB1+, positive genetic test for RA shared epitope; L, lateral; MCP, metacarpophalangeal (joints); MRI, magnetic resonance imaging; MSK, musculoskeletal; MTP, metatarsophalangeal (joints); MTX, methotrexate; NA, reported as not applicable; NOS, not otherwise specified; NSAID, non-steroidal anti-inflammatory drug(s); O, oblique; OMERACT, Outcome Measures in Rheumatology; PA, posteroanterior; PIP, proximal interphalangeal (joints); RA, rheumatoid arthritis; RF+, rheumatoid factor positivity; S, select; SJC/28, 28-point swollen joint count; SLE, systemic lupus erythematosus; SpA, spondylarthropathy; SP63, Siemens MRI model number; TJC/28, 28-point tender joint count; UIA, undifferentiated inflammatory arthritis; USA, United States of America; W, worst-effected; XX, female.

Overall, the average symptom duration for patients in the included datasets ranged from three months (5) to 14 years (31). The joints studied varied (**Figure 4**). Data were infrequently stratified by the joint groups studied. The majority of studies included investigations of the metacarpophalangeal joints (MCP) (67%). Other joints commonly investigated in these datasets included the proximal interphalangeal (PIP) (41%), wrist (38%), distal interphalangeal (DIP) (23%), metatarsophalangeal (MTP) (15%), and knee (3%). The unit of analysis varied markedly between studies. More than half (61.8%) compared the modalities at the joint level; 20.6% looked at agreement at the bone level; 14.7% narrowed the comparisons to the joint-quadrant level; and, one study (2.9%) compared the modalities at the joint-group level (e.g. MCP 2-5, wrist, of MTP 1-5). Studies that compared the modalities at the patient level were excluded from the study (34-6). The strengths of MRI magnetic field used ranged from 0.2T for 47.1% of studies and 1T for an additional 40.7%. Magnetic fields strengths of 0.6T and 1.5T were used in a minority of studies (5.8% and 1.5%, respectively). The strength of the magnetic field was not mentioned in 5.9% of datasets. In 17.6% of included studies the definition of MRI erosion was not compliant with the OMERACT definition; in an additional 23.5%, the definition was not specified. The majority of studies (94.1%) used at least two x-ray projections. A formal scoring system was used to assess x-ray erosions in 38.2% of studies. The effect of symptom duration on the research question and temporal association between MRI and x-ray erosion detection were investigated as part of the exploration of potential sources of heterogeneity in relative diagnostic test accuracy across the included studies. Theoretical considerations to the potential impact of the

other characteristics reported here on study result heterogeneity were addressed in the discussion section.

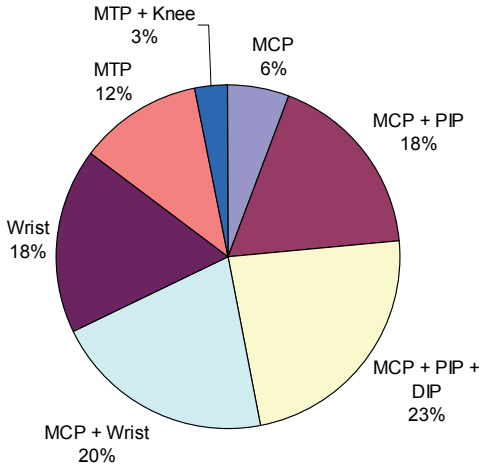


Figure 4. Pie diagram of the anatomy evaluated for erosions across the included studies.

4.3.3.1. Primary Objective

Across studies, the results were characterized by a low sensitivity and high specificity of x-ray for MRI erosions (**Figure 5**). As expected for systematic reviews of diagnostic test accuracy (14), heterogeneity existed in the test results. The mean sensitivity weighted by study sample size was 0.28 (95% CI: 0.28-0.29) and ranged from 0.00 (95% CI: 0.00-0.04) to 0.87 (95% CI: 0.60-0.98). Specificity results approximated unity with a weighted mean of 0.97 (95% CI: 0.97-0.97) and ranged from 0.38 (95% CI: 0.14-0.68) to 1.00 (95% CI: 0.99-1.00). There is low validity to the pooled result for studies with substantial heterogeneity (14). The pooled result was provided for summary purposes only. Heterogeneity associated with symptom duration and the time between initial MRI and follow-up x-ray assessment were investigated as part of the secondary objectives.

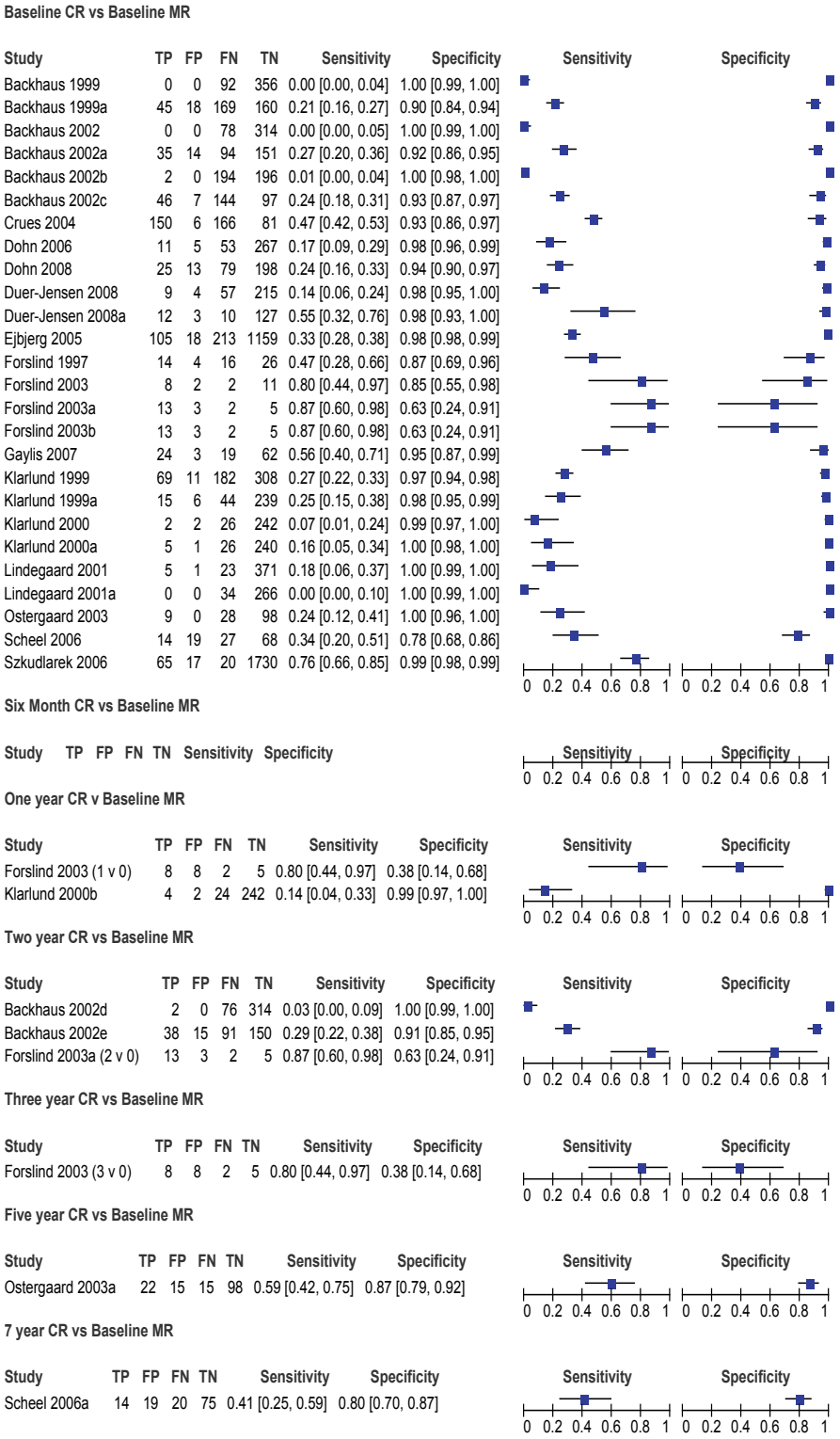


Figure 5. Sensitivity and specificity forest plots for unique paired study comparisions.

4.3.3.2. Secondary objectives

Symptom Duration. With increasing symptom duration, sensitivity increased and specificity decreased (**Figure 6**). The sensitivity of x-ray for MRI erosions appeared to increase with increasing symptom duration. Excluding two articles with extreme values, the sensitivity ranged from 0.07 (95% CI: 0.01-0.24) for patients with an average symptom duration of 0.25 years, to 0.34 (95% CI: 0.20-0.51) for patients with 14 years symptom duration. Collectively, the outliers accounted for 29% of the data in this analysis (n=2616). One outlier, Backhaus *et al*. (2002), reported a very low sensitivity of 0.01 (95% CI: 0.00-0.03) from 784 observations in patients with an average of 3.5 years of symptoms. The second outlier, Szkudlarek *et al*. (2006) reported a sensitivity of 0.76 (95% CI: 0.66-0.85) from 1832 observations for patients with an average of 5 years of symptoms. Excluding both articles resulted in the more conservative pattern of increasing sensitivity with increasing symptom duration.

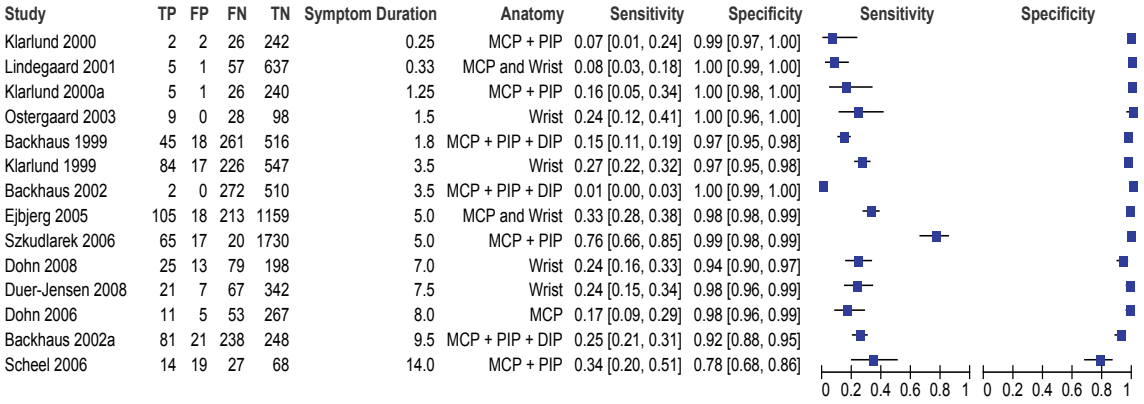


Figure 6. Sensitivity and specificity forest plots of included studies sorted by average symptom duration.

In turn, specificity decreased slightly with increasing symptom duration. The specificity ranged from approximately 0.99 (95% CI: 0.97-1.00) to 0.92 (95% CI: 0.88-0.95) for symptom durations ranging from 0.25 to 9.5 years, respectively. A single dataset by Scheel *et al*. (2006), including 128 observations, reported an extreme specificity of 0.78 (95% CI: 0.68-0.86) for a symptom duration of 14 years. The magnitude of the increase in sensitivity over symptom duration appeared greater than the decrease in specificity over the same symptom duration.

In this exploration of the impact of symptom duration on the heterogeneity of study results, the cross-sectional comparisons were sorted by this variable to determine if any patterns across the studies emerged. Here, multiple datasets with the same average symptom duration within the same cohort and citation were pooled. Datasets for temporal associations were excluded from this analysis but were considered in a separate analysis below (n=1305). In addition, two articles were excluded due to missing symptom duration (23,24). Two additional articles were excluded for other reasons: their markedly divergent results relative to other studies; their focus on the knee and/or MTP 5, anatomy with limited clinical relevance; and small numbers involved resulting in large 95% confidence intervals around the sensitivity and specificity estimates (33,34). The latter cross-sectional datasets represented a small minority of the total data (33,34). Regardless, the latter datasets were reported both with the overall results (primary objective, **Figure 5**) as well as with the intra-study temporal association analysis (below).

Temporality. The studies included in this review reported x-rays conducted cross-sectionally, as well as at 1-, 2-, 3-, 5-, and 7-years after the initial MRI (**Figure 7**). Few

datasets existed to investigate inter-study temporality. The large majority of data was cross-sectional in nature. Two studies investigated temporal findings at 1- and 2-years, respectively. Single studies existed for the 3-, 5- and 7-year follow-up studies. For cross-sectional, and 1- and 2-year temporality studies, inter-study heterogeneity was marked. Due to marked inter-study heterogeneity, intra-study temporal trends between MRI and x-ray were explored.

Intra-study results suggested that with increasing time between initial MRI and follow-up x-ray, the sensitivity of the latter for the former increased and the specificity decreased (**Figure 7**). One half of the studies included in the review (n=8) represented data from four unique longitudinal cohorts. The first cohort included four cross-sectional and three temporal datasets (21,22,31). Across these data, the sensitivity of x-ray for MRI erosions increased from approximately 0.14 (95% CI: 0.11-0.19) for patients with an average 1.8 years of symptoms to 0.34 (95% CI: 0.20-0.51) for the patients at 14 years of symptoms. Among patients with no x-ray erosions at baseline, the sensitivity of x-ray for initial MRI erosions slightly increased from 0.01 (95% CI: 0.00-0.03) at 3.5 years of symptoms to 0.03 (95% CI: 0.00-0.09) for x-rays taken two years thereafter. Over this period, specificity remained 1.00 (95% CI: 0.99-1.00). At 9.5 years of symptoms, among patients with x-ray erosions at baseline, the sensitivity of x-ray for MRI erosions was 0.25 (95% CI: 0.21-0.31) and increased to 0.29 (95% CI: 0.22-0.38) for x-rays taken two years later. The corresponding specificities for these data were 0.92 (95% CI: 0.88-0.95) and 0.91 (95% CI: 0.85-0.95), respectively. For patients with an average of seven years of disease symptoms, the sensitivity of x-ray taken seven years after the initial MR

imaging was 0.41 (95% CI: 0.25-0.59) and specificity was 0.80 (95% CI: 0.70-0.87). Ostergaard *et al*. (2003) studied 10 patients with a median 1.5 years of RA symptoms (25). X-rays had a sensitivity of 0.24 (95% CI: 0.12-0.41) and a specificity of 1.00 (95% CI: 0.96-1.00) for erosions on MRI imaged at the same time. X-rays taken five years thereafter had a sensitivity of 0.59 (95% CI: 0.42-0.75) and a specificity of 0.87 (95% CI: 0.79-0.92) for baseline MR erosions. Over a one-year interval, Klarlund *et al*. (2000) and Klarlund *et al*. (1999) demonstrated an increase in sensitivity from 0.07 (95% CI: 0.01-0.24) to 0.14 (95% CI: 0.04-0.33) while the specificity remained at 0.99 (95% CI: 0.97-1.00) (21,22). Forslind *et al*. (1997) initially found a sensitivity of 0.47 (95% CI: 0.28-0.66) and a specificity of 0.87 (95% CI: 0.69-0.96) for x-ray detection of MRI erosions in the knee and MTP 5 of 30 patients with an average of 0.67 years of symptoms (32). Six years later, in the longitudinal study reported in Forslind *et al*. (2003), a subset of this cohort was followed over an additional three years (33). In the follow-up report imaging was limited to MTP 5 in 23 of the original 30 patients. In the latter study, there were no clear patterns of change in either sensitivity or specificity. Between the Forslind *et al*. 1997 and Forslind *et al*. 2003 papers, there was an increase in sensitivity to approximately 0.80 (95% CI: 0.60-0.98) while specificity dropped to between 0.38 (95% CI: 0.14-0.68) and 0.63 (95% CI: 0.24-0.91) (32,33). The cohort reported in Forslind *et al*. (1997) and Forslind *et al*. (2003) was of limited value (32,33). It included a negligible number of observations and clinically limited anatomy. It was included in the analysis for completeness of reporting.

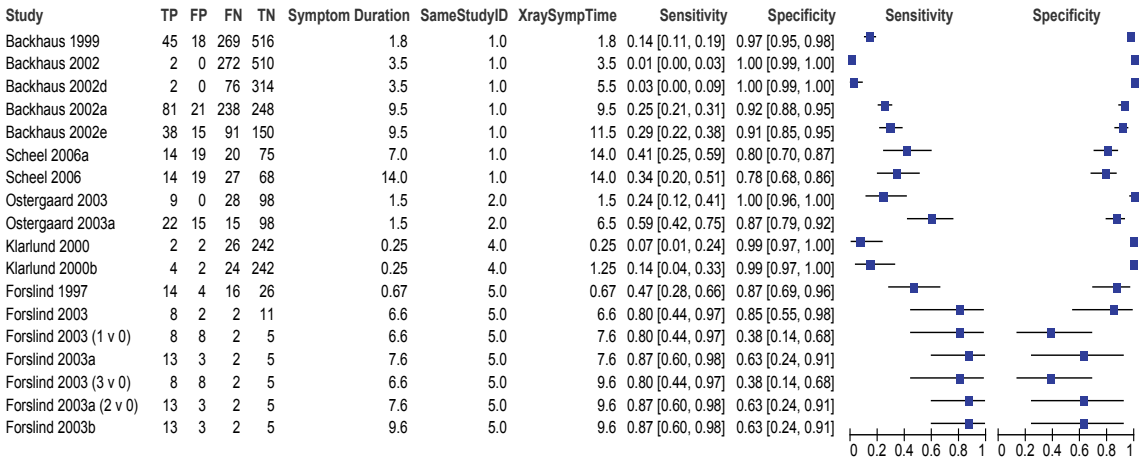


Figure 7. Sensitivity and specificity forest plots of included studies with internal longitudinal follow-up. Intra-study patterns (SameStudyID) of the temporal association between initial MRI and follow-up x-ray detection of erosons sorted by study, and a combination of symptom duration and the time between baseline MRI and follow-up x-ray (XraySympTime).

4.4. Discussion

4.4.1. Summary of main results

Included in this systematic review were 16 eligible studies selected from a literature search that identified 2657 citations. The 16 studies reported on 34 unique datasets and included 10 953 paired observations. Heterogeneity in study quality and characteristics existed. Using a lenient approach, the methodological quality of the included studies was evaluated to be acceptable, primarily lacking in the avoidance of partial verification and the reporting of uninterpretable data. Studies varied by average symptom duration, anatomy investigated, MRI magnetic strength, definition of erosion, number of x-ray projections used, use of prescribed scoring systems, number of raters, and unit of analysis. No studies accounted for inter-rater reliability in the comparison of the association between x-ray and MRI detection of erosions.

Overall, sensitivity was generally low and specificity high. The patterns of increasing sensitivity, and to a lesser extent decreasing specificity, both with increasing symptom duration emerged from this review. Among a subset of longitudinal studies, the following patterns emerged: 1) sensitivity increased with increasing interval between initial MRI and subsequent x-ray; 2) sensitivity increased with increasing symptom duration; 3) specificity decreased with increasing interval between initial MRI and subsequent x-ray; and 4) specificity decreased with increasing symptom duration.

4.4.2. Strengths and weakness of the review

To investigate the association between MRI and x-ray detection of erosion in RA, an extensive literature search of two principal medical research citation indexes was conducted. Studies directly comparing MRI and x-ray were examined as opposed to indirect comparisons. Indirect comparisons are considered a lower level of evidence but are commonly used in studies of diagnostic test accuracy (37). Where available, multiple datasets were extracted from the same study. Care was taken not to include redundant cohort data reported across multiple publications. Some sources of study heterogeneity hypothesized to effect the association were identified, explored, and reported. Other study characteristics potentially affecting the research question were discussed as well. Inter- and intra-study patterns illustrating the temporal association between x-ray and MRI detection of erosions were summarized. Critical appraisal of the literature was conducted using the QUADAS questionnaire, a methodological quality assessment tool for studies of diagnostic test accuracy (11).

There were several limitations to this review. Study heterogeneity precluded the validity of deriving pooled figures to summarize the association between MRI and x-ray detection of erosions. As such, only patterns across the data could be described and reported. A pooled result would have been more convincing but was not appropriate for these data.

The principal sources of study heterogeneity explored, namely symptom duration and time interval between initial MRI and follow-up x-ray, would have been more appropriately investigated at the individual participant level of analysis as opposed to the study or dataset level. This would have required access to raw data and likely the use of multivariable modeling and simulation approaches. These approaches were not pursued.

Quality appraisal was conducted by a single individual and was admittedly lenient. If the guidelines for the use of the QUADAS questionnaire were followed more strictly, the included studies may have been interpreted to be of lower quality. One of the weaknesses across the included studies not specifically captured by the generic QUADAS questionnaire was the different units of analysis used. In this study we aimed to categorize anatomy into units of analysis, within which the binary presence of an erosive signal can be determined and confidently presumed to refer to the same anatomical location for the two modalities (i.e. geo-anatomical cross-referencing). Studies that compared the detection of erosions across the two modalities using the whole person as a unit of analysis were excluded since the geo-anatomical cross-referencing lacked validity for the research question. However, several other units of analysis were included, each carrying a varied degree of validity. Included studies compared the detection of erosion

at the anatomical region, joint, bone, or joint quadrant (in order of increasing precision of the anatomy compared and hence increasing validity of the comparison). Results across varying units of analysis were compared together, despite a lack of validity to do so.

Likewise, different studies examined different joints and joint combinations. Evidence suggests that different joints, e.g. MCP 2,3 and MTP 1, are commonly affected in RA. Further, the MRIs with limited field of view are sometimes unable to appropriately image the carpal bones. The latter point appeared to be the major shortcoming as evaluated using the generic QUADAS-based quality appraisal.

The study compared the relative detection of erosions by x-ray against MRI as the reference standard. In the case of erosion detection, computed tomography may be superior to MRI (16). Both appear to lower the limit of detection of the clinically relevant reference standard: x-ray. X-ray, not MRI nor CT, has been correlated with long-term physical and functional disability (8). Inferring the x-ray correlations onto other diagnostic imaging modalities ought to be cautioned against.

The literature search was limited to articles from 1996 to present. Findings from the literature search suggest the existence of relevant literature published before 1996. A search of the reference lists of articles selected for full review was informally conducted by single reviewer and no additional articles were selected. The study may have benefited from the expansion of the dates searched to ensure the systematic consideration of articles published prior to 1996. Further, the literature search was limited to the English language. Conference proceedings were not considered due to the perceived inability to appraise the study quality from abstracts.

4.4.3. Applicability of findings to clinical practice

Despite its limitations, this systematic review suggests that MRI erosions develop into x-ray erosions with time. MRI erosions precede x-ray erosions. With the RA clinical management impetus to limit radiographic disease progression, MRI may find utility in identifying and monitoring pre-radiographic erosions. Given the low sensitivity of x-ray for MRI erosions, the clinical relevance of MRI erosions may be questionable. It may be more prudent to use MRI to identify patients with progressive disease, who may be at greater risk of radiographic erosion and long-term functional disability. An appropriate interval for monitoring MRI progression has yet to be determined. For x-ray, annual monitoring of disease progression has been recommended and more frequently in early disease (38). Likewise, the most appropriate anatomy to be monitored remains to be determined. Olech *et al*. (2008) suggest a 16% increase in the detection of patients with erosive disease on MRI by imaging bilateral hands compared to a single hand and only an additional 3% increased detection by including both feet (39). In turn, for detecting disease progression, monitoring a single hand was comparable to monitoring both hands and feet (39). In the RA subset with radiographic erosions, the relative merits of MRI and radiographic monitoring have yet to be determined.

4.5. Authors' conclusions

4.5.1. Implications for practice

Erosions are detectable on MRI prior to x-ray. With increasing symptom duration, erosions detected on MRI are increasingly detectable on x-ray. As the time

between initial MRI and follow-up increases, the erosions detected on MRI become increasingly detectable on x-ray. The decreasing specificity of x-ray for MRI erosions as the time between initial MRI and follow-up x-ray increases suggests that erosive progression is not limited to joints initially detected as affected on MRI. Many MRI erosions did not progress to x-ray erosions. It is possible that many MRI erosions are slow- or non-progressing or artifacts resulting for not accounting for the reliability of erosion detection. In the pre-radiographic context MRI may help clinicians identify patients at risk for developing radiographic erosions. Given the low sensitivity of x-ray for MRI erosions, disease progression on MRI may be a more appropriate predictor of radiographic disease than imaging results from a single timepoint. The most appropriate interval for monitoring has yet to be determined and may vary by patient. In patients with radiographic evidence of disease, the relative merits of monitoring disease progression using x-ray or MRI have yet to be determined.

4.5.2. Implications for research

Further research is required to better characterize the relationship between MRI and x-ray detection of erosions. This study suggests that symptom duration affects the association and should be more tightly controlled in the future studies. Likewise, the time interval between MRI and x-ray ought to be more tightly controlled. In longitudinal studies, image evaluation ought to made as independent as possible. Blinding by patient, modality, and temporality may help to keep image evaluations independent. Reliability, or the reproducibility of rater evaluations, ought to be accounted for in the assessments. Reliability concepts, such as the smallest detectable difference, ought to be applied at the

unit of analysis level to discern true evaluations from those within predetermined confidence bounds of normal. Greater transparency on partial verification, and the handling of uninterpretable data and study withdrawals is required to minimize the introduction of bias in the analysis. Studies of diagnostic test accuracy ought to heed methodological quality considerations captured in the QUADAS questionnaire, as well as others specific to MRI and x-ray measurement. The Standards for Reporting of Diagnostic Accuracy recommendations provide a foundation of points to consider in the preparation of research in this area and should be followed (27).

4.6. References

1. Schett G. Erosive arthritis. Arthritis Res Ther 2007;9(Suppl 1):S2.
2. Ostergaard M, Peterfy C, Conaghan P, McQueen F, Bird P, Ejbjerg B, Shnier R, O'Connor P, Klarlund M, Emery P, Genant H, Lassere M, Edmonds J. OMERACT rheumatoid arthritis magnetic resonance imaging studies. Core set of MRI acquisitions, joint pathology definitions, and the OMERACT RA-MRI scoring system. J Rheumatol 2003;30:1385-6.
3. Foley ND, Stack JP, Ryan M, et al. Magnetic resonance imaging in the assessment of rheumatoid arthritis: a comparison with plain film radiographs. Br J Rheumatol 1999;30:101-6.
4. Jorgensen C, Cytevel C, Anaya JM, Baron MP, Lamarque JL, Sany J. Sensitivity of magnetic resonance imaging of the wristi in very early rheumatoid arthritis. Clin Exp Rheumatol 1993;11:163-8.
5. Klarlund M, Ostergaard M, Jensen KE, Lysgard JM, Skjodt H, Lorenzen I, TIRA group. Magnetic resonance imaging, radiography and scintigraphy of the finger joints: one year of patients with early arthritis. Ann Rheum Dis 2000;59:521-8.
6. Lindegaard H, Vallo J, Horslov PK, Junker P, Ostergaard M. Low field dedicated magnetic resonance imaging in untreated rheumatoid arthritis of recent onset. Ann Rheum Dis 2001;60:770-6.
7. McQueen F, Lassere M, Edmonds J, Conaghan P, Peterfy C, Bird P, et al. OMERACT rheumatoid arthritis magnetic resonance imaging studies. Summary of OMERACT 6 MR imaging module. J Rheumatol 2003;30:1387-92.
8. Scott DL, Smith C, Kingsley G. Joint damage and disability in rheumatoid arthritis: an updated systematic review. Clin Exp Rheumatol 2003;21(5 Suppl 31):S20-7.
9. United States Department of Health and Human Services, Food and Drug Administration. Guidance for industry: clinical development programs for drugs, devices and biological products for the treatment of rheumatoid arthritis. URL: www.fda.gov. *in* Paulus HE, Oh M, Sharp JT, Gold RH, Wong WK, Park GS, Bulpitt KJ, for the Western Consortium of Practicing Rheumatologists. Classifying structural joint damage in rheumatoid arthritis as progressive or nonprogressive using composite definition of joint radiographic change. A preliminary proposal. Arthritis Rheum 2004;50(4):1083-96.
10. Deeks JJ, Bossuyt PM, Gatsonis C (editors), Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 1.0. The Cochrane Collaboration, 2010. Available from: http://srdta.Cochrane.org/.
11. Whiting P, Ruties AWS, Reitsma JB, Bossuyt PMM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies on diagnostic accuracy included in systematic reviews. BMC Med Res Meth 2003;3:25.

12. Devereaux PJ, Choi PT, Lacchetti C, Weaver B, Schünemann HJ, Haines T, Lavis JN, Grant BJ, Haslam DR, Bhandari M, Sullivan T, Cook DJ, Walter SD, Meade M, Khan H, Bhatnagar N, Guyatt GH. A systematic review and meta-analysis of studies comparing mortality rates of private for-profit and private not-for-profit hospitals. CMAJ 2002;166:1399-406.

13. Levy M, Heels-Ansdell D, Hiralal R, Bhandari M, Guyatt G, Yusuf S, Cook D, Villar JC, McQueen M, McFalls E, Filipovic M, Schünemann H, Sear J, Foex P, Lim W, Landesberg G, Godet G, Poldermans D, Bursi F, Kertai MD, Bhatnagar N, Devereaux PJ. Prognostic value of troponin and creatine kinase muscle and brain isoenzyme measurement after noncardiac surgery: a systematic review and meta-analysis. Anesthesiology 2011;114:796-806.

14. Macaskill P, Gatsonis C, Deeks JJ, Harbord RM, Takwoingi Y. Chapter 10: Analysing and Presenting Results. In: Deeks JJ, Bossuyt PM, Gatsonis C (editors), Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 1.0. The Cochrane Collaboration, 2010. Available from: http://srdta.Cochrane.org/.

15. Dohn UM, Ejbjerg BJ, Court-Payen M, Hasselquist M, Narvestad E, Szkudlarek M, Moller JM, Thomsen HS, Ostergaard M. Are bone erosions detected by magnetic resonance imaging and ultrasonography true erosions? A comparison with computed tomography in rheumatoid arthritis metacarpophalangeal joints. Arthritis Res Ther 2006;8:R110.

16. Dohn UM, Ejbjerg BJ, Hasselquist M, Narvestad E, Moller J, Thomsen HS, Ostergaard M. Detection of bone erosions in rheumatoid arthritis wrist joints with magnetic resonance imaging, computed tomography and radiography. Arthritis Res Ther 2008;10:R25.

17. Duer-Jensen A, Vestergaard A, Dohn UM, Ejbjerg B, Hetland ML, Albrecht-Beste E, Ostergaard M. Detection of rheumatoid arthritis bone erosions by two different dedicated extremity MRI units and conventional radiography. Ann Rheum Dis 2008;67:998-1003.

18. Ejbjerg BJ, Narvestad E, Jacobsen S, Thomsen HS, Ostergaard M. Optimised, low cost, low field dedicated extremity MRI is highly specific and sensitive for synovitis and bone erosions in rheumatoid arthritis wrist and finger joints: comparison with conventional high field MRI and radiography. Ann Rheum Dis 2005;64:1280-7.

19. Klarlund M, Ostergaard M, Gideon P, Sorensen K, Jensen KE, Lorenzen I. Wrist and finger joint MR imaging in rheumatoid arthritis. Acta Radiologica 1999;40:400-9.

20. Szkudlarek M, Klarlund M, Narvestad E, Court-Payen M, Strandberg C, Jensen KE, Thomsen HS, Ostergaard M. Ultrasonography of the metacarpophalangeal and proximal interphalangeal joints in rheumatoid arthritis: a comparison with magnetic resonance imaging, conventional radiography and clinical examination. Arthritis Res Ther 2006;8:R52.

21. Backhaus M, Kamradt T, Sandrock D, Loreck D, Fritz J, Wolf KJ, Raber H, Hamm B, Burmester G-R. Bollow M. Arthritis of the finger joints: a comprehensive approach comparing conventional radiography, scintigraphy, ultrasound, and contrast-enhanced magnetic resonance imaging. Arthritis Rheum 1999;42:1232-45.
22. Backhaus M, Burmester GR, Sandrock D, Loreck D, Hess D, Scholz A, Blind S, Hamm B, Bollow M. Prospective two year follow up study comparing novel and conventional imaging procedures with arthritic finger joints. Ann Rheum Dis 2002;61:895-905.
23. Crues JV, Shellock FG, Dardashti S, James TW, Troum OM. Identification of wrist and metacarpophalangeal joint erosions using a portable magnetici resonance imagines system compared to conventional radiographs. J Rheumatol 2004;31:676-85.
24. Gaylis NB, Needell SD, Rudensky D. Comparison of in-office magnetic resonance imaging versus conventional radiography in detecting changes in erosions after one year of infliximab therapy in patients with rheumatoid arthritis. Mod Rheumatol 2007;17:273-8.
25. Ostergaard M, Hansen M, Stoltenberg M, Jensen KE, Szkudlarek M, Pedersen-Zbinden B, Lorenzen I. New radiographic bone erosions in the wrists of patients with rheumatoid arthritis are detectable with magnetic resonance imaging a median of two years earlier. Arthritis Rheum 2003;48:2128-31.
26. McQueen F, Østergaard M, Peterfy C, Lassere M, Ejbjerg B, Bird P, O'Connor P, Genant H, Shnier R, Emery P, Edmonds J, Conaghan P. Pitfalls in scoring MR images of rheumatoid arthritis wrist and metacarpophalangeal joints. Ann Rheum Dis 2005;64(Suppl I):i48–i55.
27. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, Moher D, Renie D, de Vet HCW, Lijmer JG. The STARD statement of reporting studies of diagnostic accuracy: explanation and elaboration. Clin Chem 2003;49:7-18.
28. Begg CB. Biases in the assessment of diagnostic tests. Stat Med 1987;6:411-23.
29. Doubilet P, Herman PG. Interepretation of radiographs: effect of clinical history. Am J Roentgenol 1981;137:1055-8.
30. Berbaum KS, Franken EA Jr, Dorfman DD, Barloon T, Ell SR, Lu CH, Smith W, Abu-Yousef MM. Tentative diagnoses facilitate the detection of diverse lesions in chest radiographs. Invest Radiol 1986;21:532-9.
31. Scheel AK, Hermann K-GA, Ohrndorf S, Werner C, Schirmer C, Detert J, Bollow M, Hamm B, Muller GA, Burmester GR, Backhaus M. Prospective 7 year follow up study comparing radiography, ultrasonography, and magnetic resonance imaging in rheumatoid arthritis finger joints. Ann Rheum Dis 2006;65:595-600.

32. Forslind K, Larsson EM, Johansson A, Svensson B. Detection of joint pathology by magnetic resonance imaging in patients with early rheumatoid arthritis. Br J Rheumatol 1997;36:683-8.

33. Forslind K, Johanson A, Larsson EM, Svensson B. Magnetic resonance imaging of the fifth metatarsophalangeal joint compared with conventional radiography in patients with early rheumatoid arthritis. Scand J Rheumatol 2003;32:131-7.

34. McQueen FM, Stewart N, Crabbe J, Robinson E, Yeoman S, Tan PLJ, McLean L. Magnetic resonance imaging of the wrist in early rheumatoid arthritis reveals a high prevalence of erosions at four months after symptom onset. Ann Rheum Dis 1998;57:350-6.

35. McQueen FM, Stewart N, Crabbe J, Robinson E, Yeoman S, Tan PLJ, McLean L. Magnetic resonance imaging of the wrist in early rheumatoid arthritis reveals progression of erosions despite clinical improvement. Ann Rheum Dis 1999;58:156-63.

36. McQueen FM, Benton N, Crabbe J, Robinson E, Yeoman S, McLean L, Stewart N. What is the fate of erosions in early rheumatoid arthritis? Tracking individual lesions using x rays and magnetic resonance imaging over the first two years of disease. Ann Rheum Dis 2001;60:859-68.

37. Brazzelli M, Sandercock PAG, Chappell FM, Celani MG, Righetti E, Arestis N, Wardlaw JM, Deeks JJ. Magnetic resonance imaging versus computed tomography for detection of acute vascular lesions in patients presenting with stroke symptoms. Cochrane Database Syst Rev 2009;4:CD007424.

38. Combe B, Landewe R, Lukas C, Bolosiu HD, Breedveld F, Dougados M, Emery P, Ferraccioli G, Hazes JMW, Klareskog L, Machold K, Martin-Mola E, Nielsen H, Silman A, Smolen J, Yazici H. EULAR recommendations for the management of early arthritis: report of a task force of the European Standing Committee for International Clinical Studies Including Therapeutics (ESCISIT). Ann Rheum Dis 2007;66:34-45.

39. Olech E, Freeston JE, Conaghan PG, Hensor EMA, Emery P, Yocum D. Using extremity magnetic resonance imaging to assess and monitor early rheumatoid arthritis: the optimal joint combination to be scanned in clinical practice. J Rheumatol 2008;35:580-3.

5. Inter-rater reliability of MRI and x-ray erosion detection

5.1. Introduction

In Chapter 4, the literature investigating the relationship between erosion detection on MRI and x-ray in RA was summarized in the form of a systematic review. One of the findings from that literature review was that few, if any, of the investigations accounted for the reliability of the evaluations carried out in each study. This is surprising in light of the abundance of literature investigating the reliability of the measures used in the evaluation of MRI and x-ray findings in this clinical setting (1,2).

As described in Chapter 3, the prominent, internationally recognized system for measuring RA disease changes on MRI is the Outcome Measures in Rheumatology (OMERACT) RA MRI score (RAMRIS) (3-7). The three MRI features of prognostic relevance measured on RAMRIS include synovitis, bone marrow edema, and erosions (8,9). Synovitis refers to inflammation of the synovial tissue of articular joints and is a precursor to bone marrow edema (10). Edema is an MRI-specific feature characterized by the presence of a diffuse water signal in the normally fatty cancellous bone (11). It may represent inflammatory infiltrate. Synovitis and edema each predict erosive bone damage (8,12,13). The three MRI features are reversible to varying degrees. Synovitis is a transient feature characteristic of active inflammation and is reversible with effective anti-rheumatic drug intervention. To a lesser extent, and over a longer time interval, effective intervention can reverse edema. Although MRI and x-ray evidence of erosion healing exists, it is a rare event. Practically, erosions are progressive, irreversible features

detectable on diagnostic imaging that are correlated with long-term functional disability (14,15). The RAMRIS system is comprised of the following components:

1. A definition of the MRI sequences, planes of view, contrast agent requirements, and signal descriptions for each feature (3);
2. A scoring system (4);
3. An image atlas (5);
4. A guidance on pitfalls with the system (6); and,
5. Characterizations of the reliability and validation of the system (7,16).

The RAMRIS system utilizes a gadolinium-based contrast agent for the measurement of synovitis. This imposes a large operational cost (17) and a limited safety risk (18,19). Synovitis may also be detected on MRI without a contrast agent using Fast Spin Echo (FSE), T2-weighted (T2w), fat-saturated (+FS) sequences (20,21). Concordance between the contrast-dependent and -independent methods is good to excellent (21) despite a deficit in tools to promote the standardization of the latter, as available for the RAMRIS system (as listed above). Reliability of the tool has been investigated for the overall and component feature subscores (7,16,22-24). These studies have reported intra- and inter-reliability of status and change scores as well as other performance properties partially related to reliability. To the author's knowledge the tool's reliability at the unit of measurement has never been studied. A deficit in the literature also exists on the reliability and validity of contrast-independent measurement of synovitis on MRI.

In turn, x-ray measurement tools were also introduced in Chapter 3. Erosive lesions, a hallmark of RA, are classically and still most commonly detected using radiography. Radiographic erosions correlate with long-term functional disability

(14,15). In addition to erosions, conventional radiography scoring tools measure joint space narrowing, a surrogate measure of cartilage thinning. The reliability of other radiographic features commonly attributed to RA, such as periarticular osteopenia, ankylosis and malalignment, are questionable and therefore do not figure prominently in measurement tools. Perhaps the most popular conventional radiographic damage measure is the van der Heijde-modified Sharp score (vdHSS) (25-27). The tool measures erosions and joint space narrowing (JSN) at 86 sites in the hands, wrists and forefeet. The unit of measurement is either the joint or bone. For both features, numbered, ordinal scales are used to characterize the severity of the feature at the unit of measurement. The ordinal scores at each unit of measurement are summed into an overall score, or feature-specific sub-score (26,27). The reliability of both status and change scores at the overall and feature-specific subscore levels of analysis are frequently reported (1,25,28-31). Reliability at the unit of measurement has been identified as an important, underappreciated consideration for this instrument (32). The author is unaware of any published literature to report the reliability of the measure at the unit of measurement.

Reliability data are rater- and setting-specific (33). With an interest in advancing the current state of knowledge on the relationship between erosion detection on MRI and x-ray, a necessary prerequisite was to determine the reliability of the evaluations conducted by local raters. It was equally prudent to leverage existing international recommendations for studies of the diagnostic test accuracy (34,35), methodological quality assessment criteria (36), and other question-specific issues with a potential to introduce bias into the investigation. Many of these were discussed in the Chapter 4.

Where possible, in Chapters 5 and 6 those identified points were leveraged and accounted for in the investigation of the association between erosion detection on MRI and x-ray using a local cohort of early inflammatory arthritis and established RA. The objective of the current chapter was to determine the reliability of MRI and x-ray evaluations by local study radiologists using a contrast-free modification of the OMERACT RAMRIS and the van der Heijde-modified Sharp score respectively.

5.2. Methods

This inter-rater reliability investigation was conducted in conjunction with a clinical trial designed to compare the effect of biannual peripheral magnetic resonance imaging, radiography and standard of care on pharmacotherapeutic augmentation in inflammatory arthritis as described in Chapter 7. Included in this section are methodological details and protocol elaborations related specifically to the inter-rater reliability investigation.

5.2.1. Study design

Four radiologists each scored MRI scans of MCP 2-5 of 19 patients and x-rays of both hands, wrists, and forefeet of 9 patients. A sample size of 9 for each modality was estimated from comparable studes in the literature (1,7,16,22,24,25,28-31). For MRI, the sample size was increased to 19 due to the coincidental availability of data for the 4 radiologists. Post-hoc, all derived ICC estimates were significantly different than 0 indicating adequate statistical power. Raters were trained and provided a run-in period to gain experience with the scoring systems. Evaluations for centrally prepared, completely

anonymized image sets were visualized using Merge eFilm software. Evaluations were captured on standardized forms designed to limit recording error.

5.2.2. Patients

A subset of 19 patients enrolled in the clinical trial with at least one suspected erosion on MRI were randomly selected to have the baseline MR images of the MCP 2-5 evaluated by each of the four participating radiologists. A partially overlapping subset of 9 participants were randomly selected to have baseline x-rays of their hands, wrists and forefeet or six-month x-rays of their hands and wrists evaluated by each participating radiologist.

5.2.3. Raters

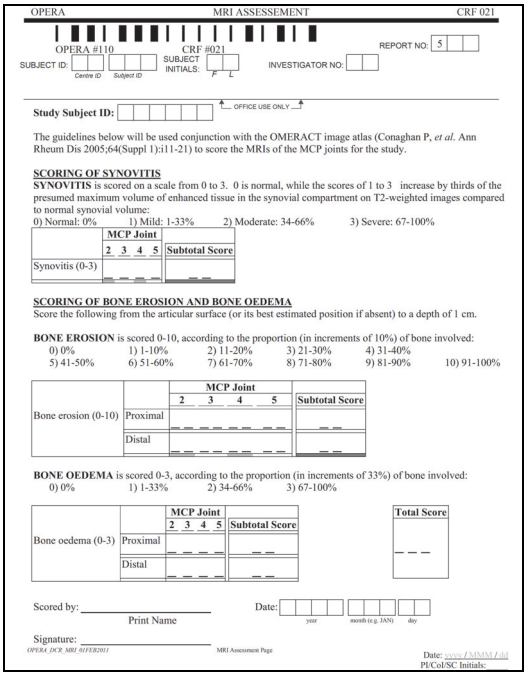
Four radiologists participated in the study. The radiologists had different levels of radiology specialty training and experience. Three of the four were musculoskeletal fellowship trained and had minimal familiarity with either the OMERACT RAMRIS or vdHSS measurement tools used in the study. The fourth radiologist was not fellowship trained and had nearly a decade of experience evaluating x-ray and MRI diagnostic imaging in rheumatology, including familiarity with evaluating both the vdHSS and the RAMRIS.

5.2.4. Measurement

Each radiologist evaluated the MRI and x-ray image sets from each participant. All raters attended an introductory presentation on the scoring systems. They were also

provided reference articles for each scoring system (3-6,25-27). A high-quality printed copy of the RAMRIS image atlas was also provided to each rater (5). High quality reference illustrations of published vdHSS example scores were supplied by Dr. van der Heijde and printed and distributed to each of the raters (26,27). A run-in period was used to give the radiologists an opportunity to become familiar with the scoring systems.

Evaluations were captured on scoring sheets designed to promote reliability in scoring as illustrated in **Figures 1** and **2**. Both scoring sheets include an anonymized study subject identifier (study subject ID), summary definitions of each feature, the scoring scale, fields to capture rater identification, and the date of the evaluation. Specifically, the scoring sheet for RAMRIS included a citation for the reference image atlas although the raters had equivalent high quality prints of this reference (5). The RAMRIS scoring sheet was directly adopted from the published example form (4). To the author's knowledge, neither reference radiographic images nor scoring forms exist for the vdHSS. These were developed specifically for the study and designed to limit error in the interpretation of the scoring system and anatomy of interest at the unit of measurement. The scoring fields were superimposed directly over the images of the anatomy to be scored. This was expected to reduce the interpretative and transcription error. In addition to providing the raters with high quality prints of example scored illustrations, lower quality examples of the same illustrations were contained within the vdHSS forms (26,27).



OPERA MRI ASSESSEMENT CRF 021

OPERA #110 CRF #021 REPORT NO: 5

SUBJECT ID: SUBJECT INITIALS: INVESTIGATOR NO:

Centre ID Subject ID F L

OFFICE USE ONLY

Study Subject ID:

The guidelines below will be used conjunction with the OMERACT image atlas (Conaghan P, *et al.* Ann Rheum Dis 2005;64(Suppl 1):i11-21) to score the MRIs of the MCP joints for the study.

SCORING OF SYNOVITIS

SYNOVITIS is scored on a scale from 0 to 3. 0 is normal, while the scores of 1 to 3 increase by thirds of the presumed maximum volume of enhanced tissue in the synovial compartment on T2-weighted images compared to normal synovial volume:

0) Normal: 0% 1) Mild: 1-33% 2) Moderate: 34-66% 3) Severe: 67-100%

	MCP Joint 2 3 4 5	Subtotal Score
Synovitis (0-3)	_ _ _ _	_ _

SCORING OF BONE EROSION AND BONE OEDEMA

Score the following from the articular surface (or its best estimated position if absent) to a depth of 1 cm.

BONE EROSION is scored 0-10, according to the proportion (in increments of 10%) of bone involved:

0) 0% 1) 1-10% 2) 11-20% 3) 21-30% 4) 31-40%
5) 41-50% 6) 51-60% 7) 61-70% 8) 71-80% 9) 81-90% 10) 91-100%

		MCP Joint 2 3 4 5	Subtotal Score
Bone erosion (0-10)	Proximal	_ _ _ _	_ _
	Distal	_ _ _ _	_ _

BONE OEDEMA is scored 0-3, according to the proportion (in increments of 33%) of bone involved:

0) 0% 1) 1-33% 2) 34-66% 3) 67-100%

		MCP Joint 2 3 4 5	Subtotal Score	Total Score
Bone oedema (0-3)	Proximal	_ _ _ _	_ _	_ _ _
	Distal	_ _ _ _	_ _	

Scored by: _____ Date: year month (e.g. JAN) day

Print Name

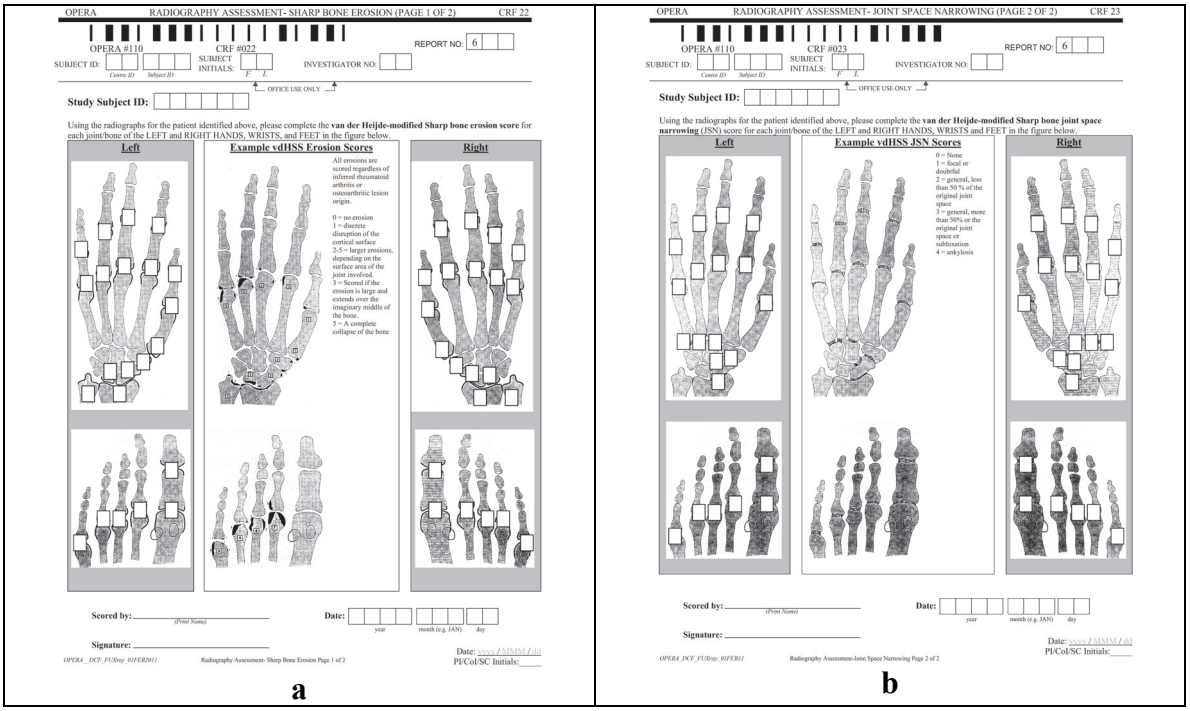
Signature: _____

OPERA_DCF_MRI_01FEB2011 MRI Assessment Page

Date: yyyy / MMM / dd

PI/Col/SC Initials:

Figure 1. OMERACT RAMRIS evaluation form.



OPERA RADIOGRAPHY ASSESSMENT- SHARP BONE EROSION (PAGE 1 OF 2) CRF 22

OPERA #110 CRF #022 REPORT NO: 6

SUBJECT ID: SUBJECT INITIALS: INVESTIGATOR NO:

Centre ID Subject ID F L

OFFICE USE ONLY

Study Subject ID:

Using the radiographs for the patient identified above, please complete the **van der Heijde-modified Sharp bone erosion score** for each joint/bone of the LEFT and RIGHT HANDS, WRISTS, and FEET in the figure below.

Left **Example vdHSS Erosion Scores** **Right**

All erosions are scored regardless of inferred rheumatoid arthritis or osteoarthritic lesion origin.

0 = no erosion
1 = discrete disruption of the cortical surface
2-5 = larger erosions, depending on the surface area of the joint involved.
3 = Scored if the erosion is large and extends over the imaginary middle of the bone.
5 = A complete collapse of the bone

Scored by: _____ Date: year month (e.g. JAN) day

(Print Name)

Signature: _____

OPERA_DCF_FUDrop_01FEB2011 Radiography Assessment- Sharp Bone Erosion Page 1 of 2

Date: yyyy / MMM / dd

PI/Col/SC Initials:

a

OPERA RADIOGRAPHY ASSESSMENT- JOINT SPACE NARROWING (PAGE 2 OF 2) CRF 23

OPERA #110 CRF #023 REPORT NO: 6

SUBJECT ID: SUBJECT INITIALS: INVESTIGATOR NO:

Centre ID Subject ID F L

OFFICE USE ONLY

Study Subject ID:

Using the radiographs for the patient identified above, please complete the **van der Heijde-modified Sharp bone joint space narrowing** (JSN) score for each joint/bone of the LEFT and RIGHT HANDS, WRISTS and FEET in the figure below.

Left **Example vdHSS JSN Scores** **Right**

0 = None
1 = focal or doubtful
2 = general, less than 50 % of the original joint space
3 = general, more than 50% or the original joint space or subluxation
4 = ankylosis

Scored by: _____ Date: year month (e.g. JAN) day

(Print Name)

Signature: _____

OPERA_DCF_FUDrop_01FEB11 Radiography Assessment-Joint Space Narrowing Page 2 of 2

Date: yyyy / MMM / dd

PI/Col/SC Initials:

b

Figure 2. van der Heijde-modified Sharp score evaluation forms for **a**) erosion and **b**) joint space narrowing.

5.2.5. Diagnostic imaging

Conventional radiography of the hands, wrists, and forefeet were conducted. Posteroanterior, oblique, and lateral projections of the hands were acquired together with those of the wrists. Posteroanterior and lateral views of forefeet were also obtained. Electronic radiography images formatted as DICOM formatted files were anonymized and distributed to the radiologists for scoring. Technical specifications of the x-ray source were summarized in **Appendix C**.

Magnetic resonance imaging was conducted using the following sequences on a 1T extremity scanner (OrthOne, GE Healthcare): erosion, 1) coronal view, 3 dimensional gradient echo (3DGE), non-fat saturated (-FS), non-isotropic voxels, 1 mm slice thickness, 0 mm interslice gap, 2) axial view, fast spin echo (FSE) T1-weighted (T1w) – FS, 2 mm slice thickness, 0 mm interslice gap; synovitis, axial view, FSE T2w +FS, 2 mm slice thickness, 0 mm interslice gap; edema, coronal view, FSE short-tau inversion recovery (STIR), 2 mm slice thickness, 0 mm interslice gap. Technical specifications of the MRI settings were captured in **Appendix C**.

The MRI sequences used in this study largely overlap with those used in previously published studies (37-39). The sequences used across those studies were merged, redundant sequences dropped, and an axial T1w sequence in the secondary plane was added to bring the sequences in compliance with the OMERACT requirements for the erosion detection. The sequence order was optimized to minimize acquisition time, comprised of the MRI sequence times, inter-sequence orientation set-up time, and

hardware-related, inter-sequence delays (4). The study radiologists reviewed the redundant sequences available and selected the optimal ones for the study.[2]

The majority of patients were scanned using a 100 mm diameter coil. In exceptional patients with large hands, or a physical deformity precluding the use of the 100 mm diameter coil, a 110 mm diameter coil was used. The sum of the total sequences time was approximately 21 minutes. With *in situ* number of slice optimization, the total run time was readily reduced to 25 minutes or less. With patient preparation time, a single set of MCP 2-5 joints could be scanned in less than 30 minutes. MRI technicians were trained and adhered to the standard operating procedures to promote consistent image acquisition.

As with the radiographic images, the MRI image sets were acquired as DICOM files. Each image set was anonymized with a unique identifier to prevent cross-referencing image scoring between raters and thereby promote the independence of the evaluations.

5.2.6. Data management

The anonymized patient evaluation forms were faxed into an iDataFax database prepared by the McMaster University and St. Joseph's Healthcare Hamilton, Programs for the Assessment of Technology in Health (PATH) Research Institute (Hamilton, ON, Canada) in consultation with the author. This ensured that the data were entered at the

[2] The in plane resolution of the coronal sequence 3DGE –FS was deemed to be of lesser quality than an alternative coronal FSE T1w –FS sequence but was selected at the request of a study investigator given its application to a tangential project. At the request of the radiologists, an exploratory 3DGE +FS sequence of 6 minutes in duration was kept in the MRI sequence protocol.

unit of measurement, enabling analyses to be conducted at multiple levels: overall scores; component feature subscores; and, the unit of measurement. Data were quality assured in the accordance with good clinical data management practice recommendations (40). Data were exported into SAS® version 9.2 (Cary, NC) for further cleaning and analysis.[3]

5.2.7. Statistical Analysis

Study participant characteristics were reported using the mean and standard deviation. The Shrout and Fliess random effects intra-class correlation coefficients (rICC) (41) were determined at the overall, component feature subscore, and feature unit of measurement levels using the %INTRACC SAS® macro (42). The F-test *p* values were determined to report the level of significance of the ICC. The absolute smallest detectable difference was determined (43) and reported as a percentage of the maximum actual score as well as a percentage of the maximum measure value (1).

5.3. Results

The demographic and disease characteristics of the study participants included in the evaluation of the RAMRIS and vdHSS score inter-rater reliability assessments among the four study radiologists were determined (**Table 1**). All participants were clinically diagnosed with rheumatoid arthritis. A total of 19 participants were included in the inter-rater reliability assessment of the RAMRIS of the MCP 2-5 of a single hand, a tool with 20 assessment points (8 erosion; 8 edema, and 4 synovitis). In contrast, 9 participants were selected for the inter-rater reliability assessment of the vdHSS. The vdHSS of the

[3] Notably, one rater reversed the entry of MRI proximal and distal joint scores. This recording error was verified with the rater and data entries were corrected.

bilateral hands (including wrists) and feet includes 86 units of measurement (44 erosion; 42 joint space narrowing).

Table 1. Study participant characteristics.

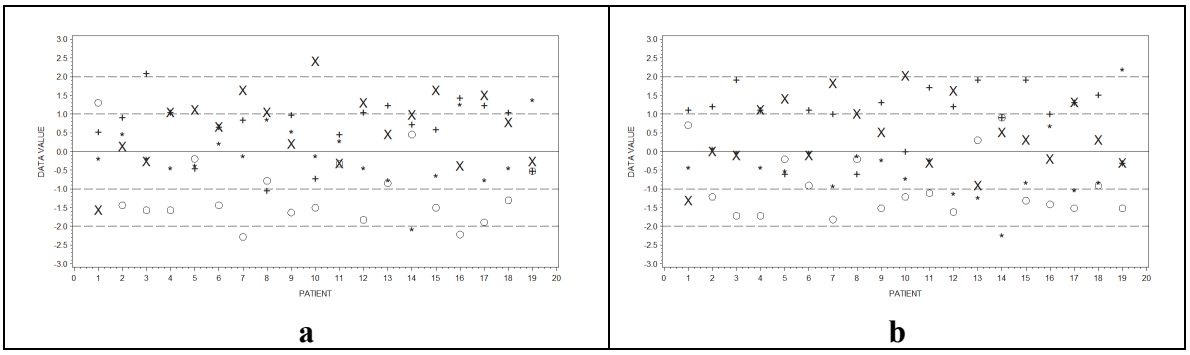
Baseline Characteristic	Value	
	MRI patients (n= 19)	X-ray patients (n=9)
Age, mean (SD)	57.0 (14.7)	59.3 (12.5)
Gender, male n (%)	4 (21.1)	1 (11.1)
Ethnicity, %		
Caucasian	14 (73.7)	7 (77.8)
South Asian	2 (10.5)	0 (0.0)
Not reported	3 (15.8)	2 (22.2)
Smoking status, yes n (%)	10 (52.6)	3 (33.3)
Symptom Duration years, mean (SD)	6.8 (6.4)	7.6 (7.3)
Evidence of Erosions, yes n (%)	9 (47.4)	3 (33.3)
ESR (mm/h), mean (SD)	20.0 (18.2)	22.4 (23.7)
C-reactive Protein (mg/L), mean (SD)	17.9 (36.7)	24.5 (55.6)
Rheumatoid Factor Positive, %	14 (73.7)	7 (77.7)
Anti-nuclear Antibody Positivity, n (%)	7 (36.8)	3 (33.3)
Treating Physician, n (%)		
1	15 (78.9)	6 (66.7)
2	2 (10.5)	1 (11.1)
3	0 (0.0)	1 (11.1)
8	2 (10.5)	1 (11.1)
Tender Joint Count, mean (SD)	4.6 (6.5)	6.1 (8.1)
Swollen Joint Count, mean (SD)	8.2 (5.1)	8.3 (3.5)
DAS28 score, mean (SD)	3.9 (1.6)	4.0 (2.6)
HAQ score, mean (SD)	1.6 (1.4)	1.8 (1.7)
RADAI score, mean (SD)	22.6 (11.8)	21.7 (15.7)
SDAI score, mean (SD)	18.9 (16.6)	9.6 (2.1)
CDAI score, mean (SD)	49.1 (44.1)	44.4 (53.4)
RAPID3 score, mean (SD)	20.2 (8.5)	21.6 (9.0)
RAPID4 score, mean (SD)	20.8 (8.3)	22.1 (10.2)
RAPID5 score, mean (SD)	51.2 (34.6)	50.0 (46.1)
EQ5D score, mean (SD)	0.8 (0.2)	0.8 (0.2)
HUI-III score, mean (SD)	0.7 (0.9)	0.6 (0.7)
OMERACT RAMRIS score, mean (SD)*	11.7 (7.8)	10.4 (9.5)
vdHSS, mean (SD)†	24.6 (38.3)	32.7 (34.6)

CDAI = clinical disease activity index; DAS28 = 28-joint disease activity score; ESR = erythrocyte sedimentation rate; HAQ = health assessment questionnaire; EQ5D = EuroQol quality of life questionnaire; HUI-III = Health Utility Index Mark III; OMERACT = Outcome Measure in Rheumatology (formerly Outcome Measures in Rheumatoid Arthritis Clinical Trials); RAMRIS = Rheumatoid Arthritis Magnetic Resonance Imaging Score; RAPID3-5 = Routine Assessment of Patient Index Data versions 3 to 5; SDAI = Simplified Disease Activity Index; vdHSS = van der Heijde-modified Sharp score.

* RAMRIS measured in dominant hand of study participants.

† vdHSS measured in hands, wrists, and feet of study participants.

The rICC for the measurement of the RAMRIS across the four study radiologists were determined (**Figure 3**). As opposed to the fixed effects ICC, the rICC was reported as it is more conservative and generalizable (41). In the plots, each radiologist was represented by a specific symbol (+, *, x, o). The ICC for overall scores were generally greater than for subscores. The rater represented by the "o" symbol rated all features most conservatively, providing the minimum of the four evaluations for 14 overall scores, 14 erosion, 15 edema, and 13 synovitis subscores. The radiologists represented by the symbols "x" and "+" provided the greatest overall scores for 10 and 9 of the 19 participants, respectively. The radiologists represented by the symbols "+" and "x" rated the greatest erosion subscore for 12 and 7 of the 19 participants, respectively (including two ties for the greatest erosion subscore). Similarly, "+" and "x" rated the greatest edema scores for 10 and 8 of the 19 participants, respectively. Radiologists represented by the symbols "x" and "*" provided evaluations with the highest synovitis scores for 10 and 9 of the 19 participants, respectively. All evaluations were within four standard deviations from the mean rating. At the individual unit of measurement (plots not shown), the erosion rICC was 0.66; edema rICC was 0.54; and, synovitis rICC was 0.35.



a **b**

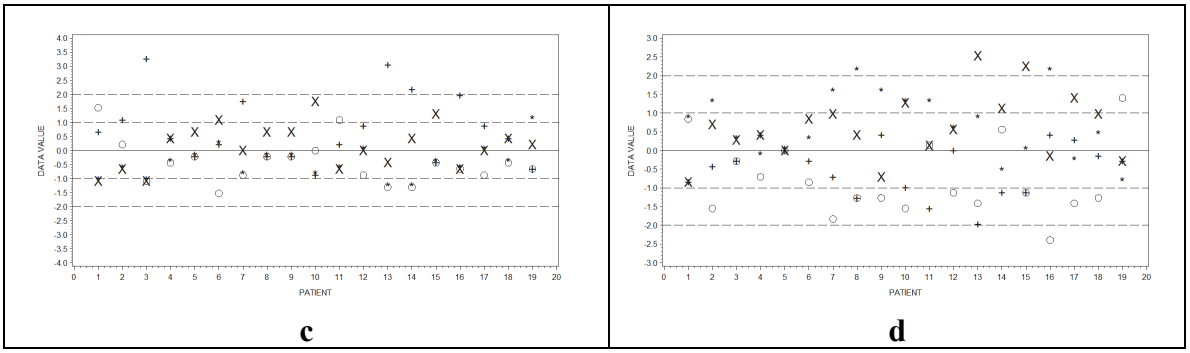


Figure 3a-d. Intra-class correlation coefficient plots for the measurement of the OMERACT RAMRIS of the MCP 2-5 for 19 rheumatoid arthritis patients by four study radiologists. Each symbol (+, *, x, o) represented a specific radiologist. **a**) Overall score, rICC = 0.54; **b**) Erosion subscore, rICC = 0.42; **c**) Edema subscore, rICC = 0.54; **d**) Synovitis subscore, rICC = 0.29.

The ICC plots for the measurement of the overall vdHSS, and erosion and joint space narrowing subscores were also determined (**Figure 4**). The symbols representing each radiologist were consistent with **Figure 3**. Interestingly, the ICC for the JSN subscore was of greater magnitude than the vdHSS overall score, which, in turn, was of greater magnitude than the erosion subscore. For hands only (n=9), the overall rICC was 0.56, the erosion subscore rICC was 0.34, JSN subscore, rICC was 0.78. The erosion at the unit of measurement rICC was 0.59. The radiologist represented by the symbol "+" provided evaluations with the greatest overall score, and erosion and joint space narrowing subscores for 6, 6, and 5 of 7 participants with available hands and feet scores, respectively. In turn, the radiologist represented by the symbol "*" provided evaluations with the lowest overall score and erosion and joint space narrowing subscores for 6, 5, and 6 of the 7 participants with available hands and feet scores, respectively. Despite the observed patterns, all evaluations were within three standard deviations of the mean rating.

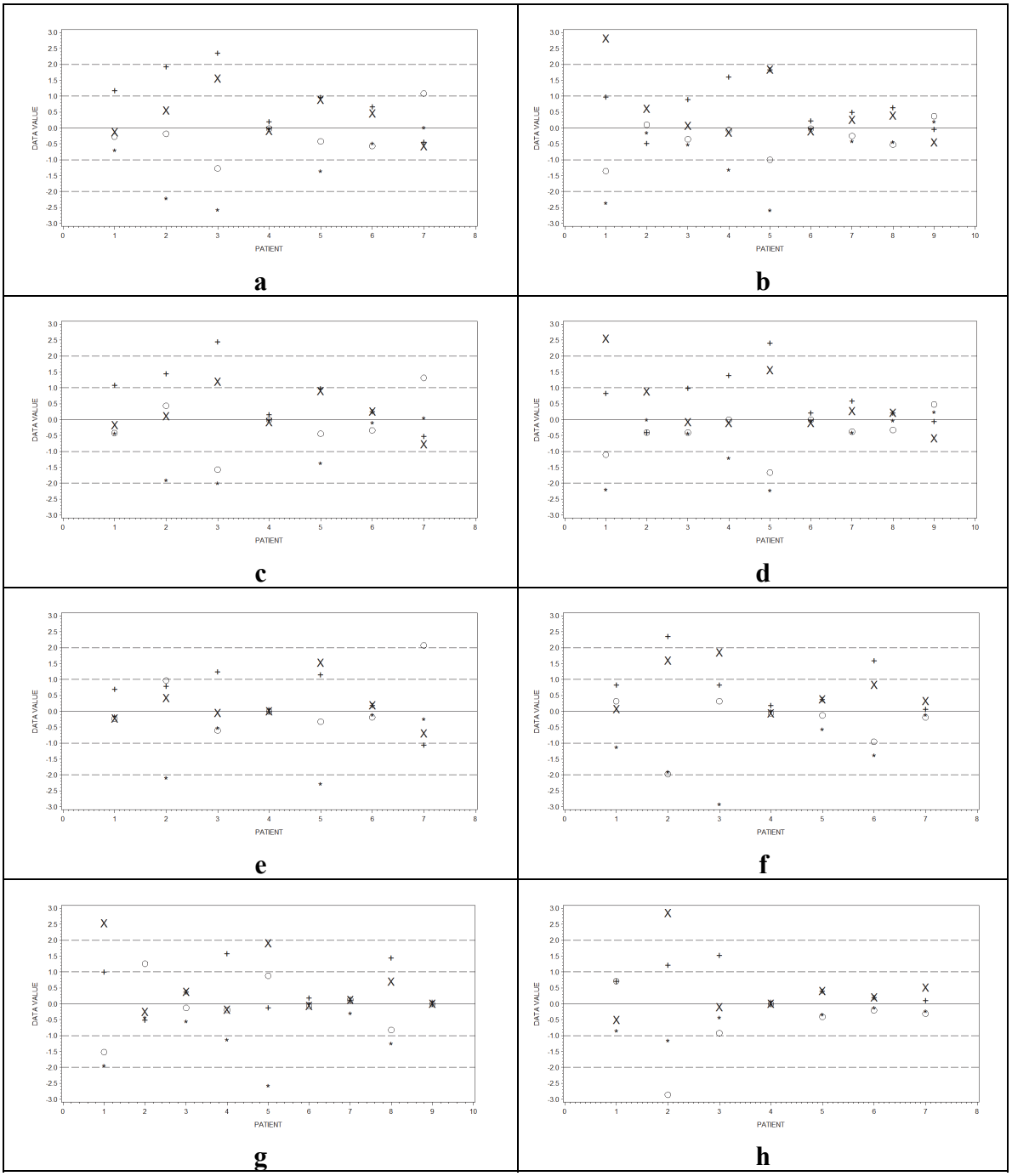


Figure 2a-h. Intra-class correlation coefficients for the measurement of the van der Heijde-modified Sharp score for seven (nine for hands only subset) rheumatoid arthritis patients by 4 study radiologists. Each symbol (+, *, x, o) represented a specific radiologist. **a**) Overall, rICC = 0.57; subscores for **b**) erosions and joint space narrowing, hands only, rICC=0.56; **c**) erosions, hand and feet, rICC = 0.52; **d**) erosions, hands only, rICC = 0.34; **e**) erosions, feet only, rICC = 0.76; **f**) joint space narrowing, hands and feet,

rICC = 0.75; **g**) joint space narrowing, hands only, rICC = 0.78; **h**) joint space narrowing, feet only, rICC = 0.60.

The smallest detectable differences for RAMRIS and vdHSS across the four study radiologists were calculated (**Table 2**). The overall RAMRIS for MCP 2-5 unilaterally has a maximum score of 116. The smallest detectable difference (SDD), interval units (calculated) for the RAMRIS evaluation of these joints was 11 (10.63). For the RAMRIS erosion, edema, and synovitis subscores, the SDD was 7 (6.87), 4 (3.18), and 5 (4.93), respectively. At the smallest unit of the analysis, at the individual bone, the SDD for erosion was 2 (1.48) and for edema was 1 (0.88). At the individual joint the SDD for synovitis was 2 (1.60). The maximum vdHSS for hands and feet is 448. Across the four study radiologists, the overall SDD for the vdHSS was 39 (38.21). The SDD for the erosion subscore for the hands and feet was 34 (33.17) and 26 (25.89) for the hands. The SDD for the joint space narrowing (JSN) subscore for hands and feet was 11 (10.88) and 11 (10.97) for the hands. At the individual joint and or bone level, the erosion SDD as 2 (1.86) and for JSN was 2 (1.29).

Table 2. Intra-class correlation coefficients and Smallest detectable differences for the measurement of OMERACT RAMRIS and van der Heijde-modified Sharp scores by four study radiologists.

Measure	**rICC**	**SDD***	**% SDD / Max score**	**% SDD / Scale Max**
OMERACT RAMRIS				
Overall, n = 19	0.54	11 (10.63)	37	10
Component Subscore, n=19				
Erosion	0.42	7 (6.87)	39	9
Edema	0.54	4 (3.18)	50	17
Synovitis	0.29	5 (4.93)	45	42
Unit of Measurement				
Erosion, n=152	0.66	2 (1.48)	22	20
Edema, n=152	0.54	1 (0.88)	33	33

Measure	rICC	SDD*	% SDD / Max score	% SDD / Scale Max
Synovitis, n=76	0.35	2 (1.60)	66	66
vdHSS, hands and feet				
Overall, n=7	0.57	39 (38.21)	45	9
Component Subscore, n=7				
Erosion	0.52	34 (33.17)	56	12
JSN	0.75	11 (10.88)	31	7
vdHSS, hands only				
Overall, n=9	0.56	34 (33.33)	41	13
Component Subscore, n=9				
Erosion	0.34	26 (25.89)	53	16
JSN	0.78	11 (10.97)	31	9
Unit of Measurement				
Erosion, n=372	0.59	2 (1.86)	22	20
JSN, n=354	0.67	2 (1.29)	50	50

*Rounded (actual).
JSN = joint space narrowing; OMERACT = Outcome Measure in Rheumatology (formerly Outcome Measures in Rheumatoid Arthritis Clinical Trials); RAMRIS = Rheumatoid Arthritis Magnetic Resonance Imaging Score; rICC = random effects intraclass correlation coefficient; SDD = smallest detectable difference; vdHSS = van der Heijde-modified Sharp score.

5.4. Discussion

In this chapter, the reliability of semi-quantitatively measuring MRI and x-ray data from RA patients across four participating radiologists was investigated. The RAMRIS was used to evaluate MRI of the MCP 2-5 and vdHSS was used to evaluate x-rays of the hands, wrists and forefeet. As reported by others, the ICC and SDD for the overall and specific feature subscores were determined to evaluate the rater- and setting-specific reliability of the two measures (22,24,28-31,43). Given that these tools are composite measures comprised of several individual evaluations, the ICC and SDD were also determined at the level of the smallest possible unit of the analysis, the unit of measurement.

The rICC reported were generally in the good to moderate range (22,44). The ICC plots demonstrated a tendency towards bias between raters in the evaluations: certain

raters tended to provide evaluations with greater or lesser scores. Despite this tendency for bias, most individual rater evaluations fell within 2 standard deviations of the mean for the unit of analysis.

These ICC results translated into the following SDD findings. It was determined that the SDD for the RAMRIS of the MCP 2-5 of a single hand was 11. For the component features, the SDD for erosion, edema, and synovitis were 7, 4, and 5 respectively. At the unit of measurement, the SDD for erosion, edema, and synovitis were 2, 1 and 2, respectively. For the vdHSS of the hands, wrists and feet, the SDD was 39. For the component erosion and JSN subscores, the SDD were 34 and 11, respectively. At the unit of measurement, the SDD for erosion and joint space narrowing were 2 and 2, respectively.

These findings impact the interpretation of diagnostic imaging evaluations using these measurement tools. First and foremost, the findings dictate the extent to which any individual evaluation from any participating radiologist may be considered abnormal. Second, the analyses provide context to the appropriate use of the Shrout and Fleiss fICC and rICC. Finally, these findings call to question the conventional approach to determining the reliability of composite measures.

The SDD is the conventional threshold for determining differences in status scores accounting for the variability within or between raters evaluations, as applicable (45). It is directly derived from the Bland and Altman ICC plots (46) and the standard deviation or mean squared error term from an analysis of variance (ANOVA) statistical test (43,46). Only differences greater than the SDD can be considered statistically

significant. In the setting of multiple raters, inter-rater reliability is of greater relevance than intra-rater reliability considering the variability between raters exceeds the variability within individual raters (33). Clinically, inter-rater reliability is a more practically relevant consideration than intra-rater reliability since the evaluating radiologist is seldom controlled for across the radiology laboratories frequented by patients and the radiologists who interpret the films within the same laboratory. Similarly, given its greater clinical relevance, status scores were investigated as opposed to change scores. Here, status scores refer to the independent evaluations of diagnostic imaging from a single time point. In turn, change scores refer to the dependent (simultaneous) evaluation of diagnostic imaging from two different timepoints. In these analyses, the SDD was rounded to the next whole number considering the ordinal or discrete, interval nature of the two scales investigated. By doing so, differences of the magnitude of the SDD reported here may be considered statistically significant.

For RAMRIS, overall scores varying by 11 or more can be considered significantly different. From a diagnostic perspective, only overall scores 11 or greater can be considered abnormal. As such, in serial independent evaluations, only RAMRIS differences of 11 or more can be considered to represent true disease progression or regression (45).

Notably, the scoring of synovitis using the RAMRIS was subject to the lowest reliability among the three MRI component features. In this study, T2w +FS MRI sequences were used to evaluate synovitis. The RAMRIS method strictly measured synovitis using T1w –FS with and without a gadolinium contrast agent. The low

reliability may have been partially attributed to the lack of an image atlas for scoring synovitis on T2w +FS MRI sequences compared to the OMERACT standard. Despite this limitation, T2w +FS is an accepted MRI sequence for the measurement of the synovitis (47). In an investigation biased to demonstrate a difference between these two methods, good to excellent agreement resulted (21). In the current study, participating raters were supplied with the contrast-dependent image atlas as part of their training and instructed to infer that the magnitude of the T1w –FS +Gd signal corresponded to the T2w +FS signal of interest. As a tangential project, a consensus meeting was held to derive an image atlas for scoring synovitis using T2w +FS axial sequences (**Appendix D**). The effect of the derived image atlas on the reliability of the synovitis scoring has yet to be determined and is the focus of an accepted peer-reviewed grant conceptualized, designed, and written by the author (**Appendix E**). In addition, other causes of the relatively low reliability in scoring synovitis exist, such as possible technical limitations to the use of T2w +FS for the measurement of the synovitis, an unclear definition of the synovitis signal on this MRI sequence, and the potential confounding and/or contaminating effect of the effusion signal on this sequence. These alternative explanations will be explored as part of the investigation into the development of the T2w +FS MRI synovitis atlas.

The study findings were compared to those reported by other authors. Lassere *et al*. (1999) reported a SDD of 15.5 for the vdHSS of the hands, wrists and feet for the mean of any two trained raters (48). Lassere *et al*. (2003) reported on the reliability of an earlier version of RAMRIS, which utilized different scale for the measurement of edema

(22). An overall RAMRIS score was not calculated. Among six raters in that study, the SDD for erosion across 10 sets of MCP 2-5 was 6.6. Expressed as a percentage of the maximum erosion subscore, the %SDD was 27% (22). The data reported here are outside the range of the reliability reported by others for these instruments (1,7,16,24,25,28-31).

Despite the similarities and differences between these findings and those of others, this analysis calls to question fundamental aspects of how reliability of composite measures is determined. The data presented demonstrate that the reliability of composite measures evaluated at a level other than the unit of measurement may be artificial and incorrect if acceptable reliability at the smallest unit of analysis is not first established.

To date, this point has not been fully appreciated in the literature, at least not in the context of the reliability of the two measures investigated here. In a related research, Lukas *et al*. (2009) investigated absolute and nearly complete agreement in change scores measured using vdHSS over the course of the first year in a three-arm clinical trial comparing the effect of anti-rheumatic therapy on radiographic progression (32). They reported that absolute and nearly complete agreement across 4 total ratings by 2 raters occurred in 7.4% of joints (52/706) evaluated to have a change by any one rater (32). In their paper, the authors did not focus specifically on reliability. It is presumed by the reader that complete agreement was reached with the remaining 6549 joints evaluated to demonstrate no progression. Including the joints with no progression, absolute agreement in change scores occurred for 90.4% (6561/7255) of joints investigated. Perhaps due to the results where a positive signal was detected by at least one rater, the authors conceded "poor absolute agreement". Change scores, however, are not directly comparable to

status scores. In the current investigation on status scores conducted by 4 different raters, the absolute agreement would be expected to be worse than that of change scores tested across fewer raters. In this study, 3 out of 4 raters agreed on 60.8%, and 4 out of 4 raters agreed on 48.4% of 372 joints evaluated for erosion on x-ray. In turn, 3 out of 4 raters agreed on 50.7%, and 4 out of 4 raters agreed on 12.5% of the 152 bones evaluated for erosion of MRI.

In the case of the vdHSS, the validity of the conventional approach to measuring reliability is further complicated by the ordinal nature of the scale used to evaluate each unit of measurement. First and foremost, the erosion scale was devised such that the score at the unit of measurement results from a summation of all erosive signals detected in the joint or bone. As a result, there are multiple distinct evaluations that can lead to the same score even at the unit of measurement. Consider a hypothetical example: in evaluating the left MCP 2, the first rater scores 3 by observing 1 discrete disruption of the cortical surface of the phalangeal base and 1 larger erosion that does not extend over the imaginary middle of the metacarpal head. The second rater, evaluating the same joint, interprets the larger erosion as a 3 and no discrete cortical disruptions. In this example, the two evaluations result in perfect agreement at the unit of measurement when, in reality, only partial agreement exists. As such, even at the unit of measurement, the true reliability in erosion scoring is limited. Nonetheless, the closer one can come to cross-referencing the anatomical features contributing to the derived score, the closer reliability at that unit of analysis will reflect the true reliability. Notwithstanding these considerations, at the vdHSS unit of measurement, the ordinal nature of the scale leads to

the interpretation that higher scores result in greater abnormality. In assessing the reliability of both the vdHSS and RAMRIS instruments, the assessment must be carried out at the smallest available unit of measurement to best estimate the true reliability.

It has been recognized in the literature that reliability improves with the use of averaging the evaluations of multiple raters (41). Averaging has the direct effect of lowering the SDD by a factor of $k^{-0.5}$, where k = the number of evaluations averaged. The application of this theory developed for continuous, single measurement scales also breaks down in application to composite measures and measures utilizing non-continuous scales. As discussed above, the true reliability of composite measures must be made at the unit of measurement. Beyond this, the parametric average (mean) of ordinal data is meaningless. The parametric averaging of the discrete interval is of lesser concern. To this end, non-parametric averaging at the unit of measurement is not helpful as the data are not additive and therefore cannot be used to derive an average overall score.

The raters in this study were radiologists either specialized in MSK imaging or were experienced with both the x-ray and MSK scoring systems employed. As described above, they were provided with tutorial session on the scoring systems and were provided the resources published to guide their scoring. They were also provided a "*run-in*" period to gain experience the scoring systems. Despite this training, the smallest detectable difference for 4 of the 5 features scored was greater than the rating scale unit. A few approaches used in mutual exclusivity or in combination may be used to improve the reliability results reported. Discrepant evaluations may be arbitrated. This approach is utilized in systematic reviews to resolve discrepancies in literature selection and data

extraction. It is noteworthy that in systematic reviews, both the reliability and accuracy of incorporating arbitration may be determined given the irrefutable source data. Accuracy may not be as easily ascertainable in diagnostic imaging applications where the true reference standard assessment may be more difficult to ascertain. Arbitration would increase the cost of imaging due to the use of multiple raters. Further, the resulting arbitration would constitute a single evaluation as opposed to multiple and *k* would be reduced to 1. A second approach is to count only evaluations at the unit of measurement greater than the SDD. This approach may undermine the scales of the instrument and may limit its sensitivity to change or responsiveness. Third, rater training and testing may be conducted until SDD at the unit of measurement becomes less than a scale unit as exemplified by the OMERACT RAMRIS edema unit of measurement SDD of 0.88 and rounded to 1. Since the SDD is biased due to the floor or basement effect (22), especially in situations where the features measured are uncommon relative to the number of measurements, this approach may also be problematic. Nonetheless, reducing the SDD at the unit of measurement to a value of smaller magnitude than the scale units would allow one to confidently use the conventional approach to reliability testing. Fourth, automated measures may be developed and validated against the reference standard with the goal of replacing the standard. Unfortunately, the performance of a test is only valid within the limits of the reference against which it was tested. Inferences outside the tested range are generally unacceptable. This limitation, however, does not forego the use of other standards such as cadaver, phantoms, or histopathology, which may be used to validate or

calibrate automated measurements both inside and outside the detection limits of rater evaluations.

The principal limitations to this study are those pertaining to the statistics used to assess the reliability of the tools investigated. For the purposes of comparison, the conventional reliability statistics were calculated for the data collected. Specifically in relation to the tools investigated, the limitations with these statistics were discussed and illustrated by reporting the corresponding results for the dataset. Some potential novel approaches for addressing inter-rater reliability were discussed. These and other approaches that may be more appropriate for investigating the reliability of composite measures should be the focus of future research.

Beyond these principal limitations, semi-quantitative scoring methods are not used clinically. Although this limits the direct clinical relevance of the study findings, the study results raise questions about the reliability of the diagnostic imaging results reported clinically. Time requirements have been often cited as the reason limiting the clinical use of semi-quantitative measures (47,49). To bridge the gap between research and clinical use, automated methods of quantifying diagnostic imaging have been developed for x-ray joint space narrowing (50,51), and MRI synovitis (52,53) and erosions (38,39,54-57). Despite such facilitating tools, all of these semi-quantitative techniques are dependent on the accurate detection of the features of interest by radiologists and generally result in more work for the limited radiologist resource.

5.5. Conclusions

In summary, this chapter investigated the inter-rater reliability for scoring the RAMRIS and vdHSS. Using a conventional analysis to investigate this, moderate to good fICC values were obtained for the overall and feature-specific subscores. The reliability measures corresponded to SDD values towards the upper range of those reported by others. Notably, however, reliability statistics are rater- and setting-specific. Given, the precedent in the literature, the derived overall SDD values enable one to differentiate abnormalities from the variability in ratings between the participating radiologists. Unlike other literature reports, the SDD was determined at the unit of measurement to be able to quantify the reliability of individual evaluations at each joint or bone. For this group of raters, the scales were not true for 4 of the 5 features measured (MRI: synovitis, erosion; x-ray: erosion; JSN). This investigation shed light on the challenges to the validity of measuring the reliability of composite measures at the overall or subscore levels without accounting for error at the unit of measurement. The findings from this chapter were applied in Chapter 6 to investigate the relationship between MRI and x-ray erosion detection using prospective data.

5.6. References

1. Lassere MN, van der Heijde D, Johnson KR, Boers M, Edmonds J. Reliability of measures of disease activity and disease damage in rheumatoid arthritis: implications for smallest detectable difference, minimal clinically important difference, and analysis of treatment effects in randomized controlled trials. J Rheumatol 2001;28:892-903.
2. Lassere MN, Bird P. Measurements of rheumatoid arthritis disease activity and damage using magnetic resonance imaging. Truth and discrimination: does MRI make the grade? J Rheumatol 2001;28:1151-7.
3. Østergaard M, Peterfy C, Conaghan P, McQueen F, Bird P, Ejbjerg B, et al. OMERACT rheumatoid arthritis magnetic resonance imaging studies. Core set of MRI acquisitions, joint pathology definitions, and the OMERACT RA-MRI scoring system. J Rheumatol 2003;30:1385-6.
4. Østergaard M, Edmonds J, McQueen F, Peterfy C, Lassere M, Ejbjerg B, et al. An introduction to the EULAR–OMERACT rheumatoid arthritis MRI reference image atlas. Ann Rheum Dis 2005;64(Suppl I):i3–i7.
5. Conaghan P, Bird P, Ejbjerg B, O'Connor P, Peterfy C, McQueen F, et al. The EULAR-OMERACT rheumatoid arthritis MRI reference image atlas: the metacarpophalangeal joints. Ann Rheum Dis 2005;64(Suppl I):i11–i21.
6. McQueen F, Østergaard M, Peterfy C, Lassere M, Ejbjerg B, Bird P, et al. Pitfalls in scoring MR images of rheumatoid arthritis wrist and metacarpophalangeal joints. Ann Rheum Dis 2005;64(Suppl I):i48–i55.
7. Haavardsholm EA, Østergaard M, Ejbjerg BJ, Kvan NP, Uhlig TA, Lilleås FG, et al. Reliability and sensitivity to change of the OMERACT rheumatoid arthritis magnetic resonance imaging score in a multireader, longitudinal setting. Arthritis Rheum 2005; 52:3860–3867.
8. McGonagle D, Conaghan PG, O'Connor P, Gibbon W, Green M, Wakefield R, et al. The relationship between synovitis and bone changes in early untreated rheumatoid arthritis: a controlled magnetic resonance imaging study. Arthritis Rheum 1999;42:1706-11.
9. McQueen FM, Benton N, Perry D, Crabbe J, Robinson E, Yeoman S, et al. Bone edema scored on magnetic resonance imaging scans of the dominant carpus at presentation predicts radiographic joint damage of the hands and feet six years later in patients with rheumatoid arthritis. Arthritis Rheum 2003;48:1814-27.
10. Jiminez-Boj E, Redlich K, Turk B, Hanslik-Schnabel B, Wanivenhaus A, Chott A, Smolen JS, Schett G. nteraction between synovial inflammatory tissue and bone marrow in rheumatoid arthritis. J Immunol 2005;175:2579-88.
11. Schett G. Bone Marrow Edema. Ann N Y Acad Sci 2009;1154:35-40.

12. Conaghan PG, O'Connor P, McGonagle D, Astin P, Wakefield RJ, Gibbon WW, Quinn M, Karim Z, Green MJ, Proudman S, Isaacs J, Emery P. Elucidation of the relationship between synovitis and bone damage. Ann Rheum Dis 2003;48(1):64-71.
13. Boyesen P, Haavardsholm EA, Ostergaard M, van der Heijde D, Sesseng S, Kvien TK. MRI in early rheumatoid arthritis: synovitis and bone marrow edema are independent predictors of subsequent radiographic progression. Ann Rheum Dis 2011;70:428-33.
14. Scott DL, Smith C, Kingsley G. Joint damage and disability in rheumatoid arthritis: an updated systematic review. Clin Exp Rheumatol 2003;21(5 Suppl 31):S20-7.
15. Scott DL. Radiological progression in established rheumatoid arthritis. J Rheumatol 2004;69(Suppl):55-65.
16. Conaghan P, Lassere M, Ostergaard M, Peterfy C, McQueen F, O'Connor P, Bird P, Ejbjerg B, Klarlund M, Shnier R, Genant H, Emery P, Edmonds J. OMERACT Rheumatoid Arthritis Magnetic Resonance Imaging Studies. Exercise 4: An international multicenter longitudinal study using the RA-MRI Score. J Rheumatol 2003;30:1376-9.
17. Østergaard M, Pedersen SJ, Dohn UM. Imaging in rheumatoid arthritis – status and recent advances for magnetic resonance imaging, ultrasonography, computed tomography and conventional radiography. Best Pract Res Clin Rheumatol 2008;22:1019–44.
18. Rydahl C, Thomsen HS, Marckmann P. High prevalence of nephrogenic systemic fibrosis in chronic renal failure patients exposed to gadodiamide, a gadolinium-containing magnetic resonance contrast agent. Invest Radiol 2008;43(2):141-4.
19. Shellock FG and Kanal E. Safety of magnetic resonance imaging contrast agents. J Magn Reson Imaging 1999;10:477–84.
20. Segami N, Nishimura M, Kaneyama K, Miyamaru M, Sato J, Murakami K-I. Does joint effusion on T2 magnetic resonance images reflect synovitis? Comparison of arthroscopic findings in internal derangements of the temporomandibular joint. Oral Surg Oral Med Oral Pathol Oral Radiol Endod 2001;92:341-5.
21. Østergaard M, Conaghan PG, O'Connor P, Szkudlarek M, Klarlund M, Emery P, et al. Reducing invasiveness, duration, and cost of magnetic resonance imaging in rheumatoid arthritis by omitting intravenous contrast injection - Does it change the assessment of inflammatory and destructive joint changes by the OMERACT RAMRIS? J Rheumatol 2009;36:1806-10. Erratum in: J Rheumatol 2010;37:2198.
22. Lassere M, McQueen F, Ostergaard M, Conaghan P, Shnier R, Peterfy C, Klarlund M, Bird P, O'Connor P, Stewart N, Emery P, Genant H, Edmonds J. OMERACT Rheumatoid Arthritis Magnetic Resonance Imaging Studies. Exercise 3: An international multicenter reliability study using the RA-MRI Score. J Rheumatol 2003;30:1366-75.

23. McQueen F, Lassere M, Edmonds J, Conaghan P, Peterfy C, Bird P, O'Connor P, EjBjerg B, Klarlund M, Stewart N, Emery P, Shnier R, Genant H, Ostergaard M. OMERACT rheumatoid arthritis magnetic resonance imaging studies. Summary of OMERACT 6 MR imaging module. J Rheumatol 2003;30:1387-92.

24. Conaghan PG, Ejbjerg B, Lassere M, Bird P, Peterfy P, Emery P, McQueen F, Haavardsholm E, O'Connor P, Edmonds J, Genant H, Ostergaard M. A multicenter reliability study of extremity-magnetic resonance imaging in the longitudinal evaluation of rheumatoid arthritis. J Rheumatol 2007;34:857-8.

25. van der Heijde DM, van Riel PL, Nuver-Zwart IH, Gribnau FW, van de Putte LB. Effects of hydroxychloroquine and sulphasalazine on progression of joint damage in rheumatoid arthritis. Lancet 1989;1:1036-8.

26. van der Heijde D. How to read radiographs according to the Sharp/van der Heijde method. J Rheumatol 1999;26:743-5.

27. van der Heijde D. How to read radiographs according to the Sharp/van der Heijde method. J Rheumatol 2000;27:261-3.

28. Boers M, Verhoeven AC, Markusse HM, van de Laar MAFJ, Weshovens R, van Denderen JC, van Zeben D, Dilkmans BAC, Peeters AJ, Jacobs P, van den Brink HR, Schouten HJA, van der Heijde DMRM, Boonen A, van der Linden S. Randomised comparison of combined step-down prednisolone, methotrexate and sulphasalazine with sulphasalazine alone in early rheumatoid arthritis. Lancet 1999;350:309-18.

29. Klareskog L, van der Heijde D, de Jager JP, Gough A, Kalden J, Malaise M, Mola EM, Pavelka K, Sany J, Settas L, Wajdula J, Pedersen R, Fatenejad S, Sanda M, for the TEMPO (Trial of Etanercept and Methotrexate with Radiographic Patient Outcomes) study investigators. Lancet 2004;363:675-81.

30. St Clair EW, van der Heijde DM, Smolen JS, Maini RN, Bathon JM, Emery P, Keystone E, Schiff M, Kalden JR, Wang B, Dewoody K, Weiss R, Baker D; Active-Controlled Study of Patients Receiving Infliximab for the Treatment of Rheumatoid Arthritis of Early Onset Study Group. Combination of infliximab and methotrexate therapy for early rheumatoid arthritis: a randomized, controlled trial. Arthritis Rheum 2004;50:3432-43.

31. Lassere MN, van der Heijde D, Johnson K, Bruynesteyn K, Molenaar E, Boonen A, Verhoeven A, Emery P, Boers M. Robustness and generalizability of smallest detectable difference in radiological progression. J Rheumatol. 2001;28:911-3.

32. Lukas C, Landewé R, Fatenejad S, van der Heijde D. Subtle changes in individual joints result in both positive and negative change scores in a patient: results from a clinical trial in patients with rheumatoid arthritis. Ann Rheum Dis 2009;68:1691-5.

33. Norman GR, Streiner DL. Biostatistics: The Bare Essentials. Second Edition. Hamilton, B.C. Decker Inc., 2000.

34. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, Lijmer JG, Moher D, Rennie D, de Vet HCW, for the STARD Group. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. Ann Intern Med 2003;138:40-4.

35. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, Moher D, Rennie D, de Vet HCW, Lijmer JG. The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. Clin Chem 2003;49:7-18.

36. Whiting P, Rutjes AWS, Reitsma JB, Bossuyt PMM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. BMC Med Res Methodol 2003;3:25.

37. Xie X, Webber CE, Adachi JD, O'Neill J, Inglis D, Bobba RS, Wu H. Quantitative, small bore, 1 Tesla, magnetic resonance imaging of the hands of patients with rheumatoid arthritis. Clin Exp Rheumatol 2008;26:860-5.

38. Emond PD, Choi A, O'Neill J, Xie J, Adachi R, Gordon CL. The development of EERA: Software for assessing rheumatic joint erosions. Can Assoc Rad J 2009;60:63-8.

39. Emond PD, Inglis D, Choi A, Tricta J, Adachi JD, Gordon CL. Volume measurement of bone erosions in magnetic resonance images of patients with rheumatoid arthritis. Magn Reson Med 2012;67:814-23.

40. Society for Clinical Data Management. Good clinical data management practices, version 4.0. Milwaukee (Wisconsin): Society for Clinical Data Management, 2005.

41. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychological Bulletin 1979;86:420-8.

42. Hamer RM. Sample 25031: Compute six intraclass correlation measures. Chapel Hill: Robert M. Hamer, 1990. Available: http://support.sas.com/kb/25/031.html. Accessed 5 Apr 2011.

43. Bruynesteyn K, Boers M, Kostense P, van der Linden S, van der Heijde D. Deciding on progression of joint damage in paired films of individual patients: smallest detectable difference or change. Ann Rheum Dis 2005;64:179-82.

44. Landis R, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159-74.

45. Bruynesteyn K, van der Heijde D, Boers M, van der Linden S, Lassere M, van der Vleuten C. The Sharp/van der Heijde method out-performed the Larsen/Scott method on the individual patient level in assessing radiographs in early rheumatoid arthritis. J Clin Epidemiol 2004;57:502-12.

46. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;1:307-10.

47. McQueen FM. The MRI View of Synovitis and Tenosynovitis in Inflammatory Arthritis. Ann NY Acad Sci 2009;1154:21-34.

48. Lassere M, Boers M, van der Heijde D, Boonen A, Edmonds J, Saudan A, Verhoeven AC. Smallest detectable difference in radiological progression. J Rheumatol 1999;26:731-9.

49. 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;22:1109-26.

50. Duryea J, Jiang Y, Zakharevich M, Genant HK. Neural network based algorithm to quantify joint space width in joints of the hand for arthritis assessment. Med Phys 2000;27:1185-94.

51. Finckh A, de Pablo P, Katz JN, Neumann G, Lu Y, Wolfe F, Duryea J. Performance of an automated computer-based scoring method to assess joint space narrowing in rheumatoid arthritis: a longitudinal study. Arthritis Rheum 2006;54:1444-50.

52. Savnik A, Bliddal H, Nyengaard JR, Thomsen HS. MRI of the arthritic finger joints: synovial membrame volume determination, a manual vs a stereologic method. Eur Radiol 2002;12:94-8.

53. Ostergaard M, Hansen M, Stoltenberg M, Gideon P, Klaarlund M, Jensen KE, Lorenzen I. Magnetic resonance imaging-determined synovial membrame volume as a marker of disease activity and a predictor of progressive joint destruction in the wrists of patients with rheumatoid arthritis. Arthritis Rheum 1999;42:918-29.

54. Dohn UM, Ejbjerg BJ, Hasselquist M, Narvestad E, Court-Paven M, Szkudlarek M, Moller J, Thomsen HS, Ostergaard M. Rheumatoid arthritis bone erosion volumes in CT and MRI: reliability and correlations with erosion scores on CT, MRI and radiography. Ann Rheum Dis 2007;66:1388-92.

55. Bird P, Joshua F, Lassere M, Shnier R, Edmonds J. Training and calibration improve inter-reader reliability of joint damage assessment using magnetic resonance imaging scoring and computerized erosion volume measurement. J Rheumatol 2005;32:1452-8.

56. Bird P, Ejbjerg B, McQueen F, Ostergaard M, Lassere M, Edmonds J. OMERACT rheumatoid arthritis magnetic resonance imaging studies. Exercise 5: an international multicenter reliability study using computerized MRI erosion volume measurements. J Rheumatol 2003;30:1380-4.

57. OsiriX - DICOM Viewer. Available at: http://www.osirix-viewer.com/. Accessed 10 Oct 2010.

6. Reliability-adjusted comparison of MRI and x-ray erosion detection

6.1. Introduction

In Chapter 4, a systematic review of the literature directly comparing erosion detection in rheumatoid arthritis (RA) by MRI and x-ray was reported. In these analyses paired at the joint, bone, or bone quadrant unit of measurement, these studies generally demonstrated a low sensitivity and high specificity of x-ray for the detection of MRI erosions (1-16). The low sensitivity results support the hypothesis that the MRI has a lower limit of detection for erosion than does x-ray. The specificity results approaching unity suggest that all erosions detected on x-ray are detected on MRI. These results also indicate that smaller erosions and erosions at an earlier disease stage are detected on MRI. Further, the role of disease duration was a positive effect modifier of the sensitivity of x-ray detection of MRI erosions was also suggested.

Across the studies, the results were marked by heterogeneity. The methodological quality of the studies varied (17-20). Study-specific considerations that could have affected the outcome also varied. The latter include, but are not limited to, MRI magnet strength, MRI and x-ray erosion signal definition, number of x-ray projections, anatomy imaged, imaging hardware, and rater qualification, experience, and training. Beyond these points, the specificity of MRI erosions for RA is also questionable (21-24).

Despite a considerable amount of literature on the reliability of MRI and x-ray semi-quantitative measures used in RA (25-32), none of the studies comparing erosion detection across the two modalities directly accounted for measurement reliability. Most of the studies included in the review evaluated diagnostic imaging using a single rater

(1,2,4-7,11-14,16). The remainder of the studies reported consensus between two raters to classify erosion detection (3, 8-10,15). The smallest detectable difference (SDD) is a measure used to quantify measurement reliability using the original units of the scale (32). It is based on Bland and Altman's *'limits of agreement'* approach (33) and defines the 95% confidence interval bound for measurement error. The SDD threshold is used to differentiate true differences in scores from measurement error (34). The approach has been used to investigate the reliability of RAMRIS and vdHSS at the overall score or component subscore levels (25-32). In the original reporting of the *'limits of agreement'* approach, however, it was illustrated using a single continuous data variable (peak expiratory flow rate) (33). The RAMRIS and vdHSS measures, on the other hand, are composites of aggregated scores made at multiple anatomical evaluation sites, so-called units of measurement. Reliability assessment of a composite or aggregate measure at the overall score level does not account for reliability at the unit of measurement (35). Formulae exist that may be used to propagate the error to the overall score level (36-37).

In this chapter, the primary objective was to determine the relative diagnostic test accuracy of MRI and x-ray for erosion detection in RA accounting for inter-rater reliability. To compare the relative performance of the two technologies, the primary objective was investigated by pairing data at the most precise anatomical level possible (herein referred to as *'geo-anatomical cross-referencing'*). To also provide clinical context, erosion detection between the modalities was also investigated at the patient level. For the latter, paired differences in proportions of patients with erosive disease were determined by comparing anatomical joint groups commonly imaged by each

modality. The effect of disease duration on erosion detection by each imaging modality was also explored in an effort to explain the trend that was revealed in the systematic review.

6.2. Methods

This study was nested within a clinical trial originally designed to compare the effect of biannual peripheral MRI, radiography and standard of care on anti-rheumatic pharmacotherapeutic treatment decision-making in inflammatory arthritis (Chapter 7). Participant eligibility criteria, diagnostic imaging equipment specifications, and other general methodological considerations are reported in Chapter 7. The current study utilized available baseline data to investigate the preset research question. This section details methodological considerations specific to the investigation pursued in this chapter.

6.2.1. Study design

A paired, cross-sectional study of RA patients with a range of symptom duration was conducted. For each participant, MRI scans of the bilateral MCP 2-5 joints were acquired as were x-ray of the hands, wrists and feet.

6.2.2. Patients

A total of 65 patients who had both baseline MRI and x-ray data available were included. A subset of 57 participants had complete sets of imaging data for both hands of MRI and x-ray of the hands, wrists and feet. Images were randomly assigned to raters after study enrolment in the larger trial was completed. As such, included participants

represented a random sample of the 191 patients enrolled in the trial who complied with baseline diagnostic imaging requirements.

6.2.3. Raters

The four radiologists who contributed to the inter-rater reliability data reported in Chapter 5 evaluated the images included here. The characteristics of the raters previously described also apply to the current investigation. Three of the four radiologists were MSK fellowship trained. The fourth was not MSK-fellowship trained but had extensive experience with MRI and x-ray imaging in rheumatology.

6.2.4. Measurement

As described in Chapters 5 and **Appendix A**, the OMERACT RAMRIS and vdHSS systems were used to semi-quantitatively evaluate the MRI and x-ray images, respectively. Evaluations for each anatomical region were captured. Where multiple evaluations for the same image set were available, the most recent one was used in the study to standardize the selection method. This was determined prior to inspecting or analyzing the data. Image sets assigned to raters were anonymized, effectively blinding the raters to the patients' imaging data from the opposite modality. Likewise, raters were blinded to specific participant clinical and laboratory characteristics except for knowledge that the sample was comprised of IA and predominant clinically diagnosed RA.

6.2.5. Diagnostic imaging

The erosion signal on MRI was consistent with the OMERACT RAMRIS definition (38). The cross-referenced devoid signal in cancellous bone had to be present on both T1-weighted planes imaged (coronal and axial used here), with a cortical break detectable on at least one plane (38). The general diagnostic imaging specifications used in this investigation were reported in Chapters 5 and 7. To summarize, bilateral MRI scans of the MCP 2-5 joints were acquired. A unilateral MCP joint set was acquired per sitting and images of both sets were acquired in consecutive sittings. The RAMRIS system was originally developed to evaluate the MCP 2-5 and wrist joints (38). In this study, imaging was limited to the MCP joints for a number of reasons. The two primary reasons included the following: imaging was restricted to one anatomical area in consideration of the impact of imaging time on patient study compliance; the decision to image the MCP instead of the wrist joints was based on the collective experience of our research group (perceived better reliability; data transferability to concomitant projects; and the availability of OMERACT-compliant MRI sequences). The MRI unit had a magnetic strength of 1.0T (GE Healthcare, ONI 1.0T OrthOne, Wilmington, MA). A 100 mm coil was most frequently used. A 110 mm coil was used for patients with larger hands. T1-weighted images in coronal and axial planes were acquired. Also included were axial T2-weighted and coronal STIR MRI sequences, for synovitis and edema detection, respectively, and an exploratory fat-saturated 3D gradient echo (T1-wieghted) sequence for cartilage visualization. All raters used Merge eFilm version 1.9 to view

both MRI and x-ray image sets. MRI technical specifications were summarized in **Appendix C**.

On x-ray, the conventional erosion definition was applied and three imaging projections were used: posteroanterior (PA), oblique and lateral. Bilateral hands and wrists were scanned in one sitting and feet were imaged in a second, consecutive sitting, or vice versa. DICOM electronic files were collected, anonymized, and distributed to the raters to evaluate on Merge eFilm version 1.9 using vdHSS. X-ray technical specifications were summarized in **Appendix C**.

To address the primary objective, the anatomy imaged on x-ray corresponded to that imaged on MRI. Clinically, however, the hands, wrists and forefeet are commonly imaged on x-ray (39). The 2007 European League Against Rheumatism guidelines for the management of early RA recommend radiographic imaging of the hands, wrists and feet every 6 to 12 months for the first few years of disease (40). These small joint groups tend to be affected earlier than larger joint groups, such as the knees, shoulders and elbows (41). Among small joints, erosions may present earlier in the forefeet than in the hands or wrists (42). For these reasons, x-ray of the hands, wrists and feet were collected. The bilateral hands and wrists can be imaged in a single sitting, and have specific Ministry of Health and Long Term Care service billings (43). Likewise, the bilateral feet may be imaged in a single sitting and have a separate service billing (43). As part of the secondary objective, inter-modality erosion detection was compared at the patient level using data collected from clinically relevant anatomical comparisons taking

into consideration of clinical practice (39), clinical management guidelines, and billable services (43).

6.2.6. Data management

Data management considerations relevant to this investigation were summarized in Chapters 5 and **Appendix A**. Most importantly, diagnostic imaging data collected was entered at the units of measurement for each of the scoring systems. This enabled reliability to be determined at the unit of measurement. It also had implications on the precision to which inter-modality, geo-anatomical cross-referencing could be made.

6.2.7. Statistical analysis

Study participant characteristics were reported using non-parametric measures of central tendency. Most interval data followed skewed distributions. A 2x2 contingency table was derived to compare paired evaluations at the joint level of analysis. Anatomical cross-referencing for data pairing was limited to the most precise unit of measurement common to both imaging modalities: the joint used in the vdHSS scale. In turn, the analysis was limited to the MCP 2-5 joints, the anatomy imaged by MR. MRI data collected at the distal and proximal aspects of the MCP joint were combined. Interval data (RAMRIS) and ordinal data (vdHSS) were then classified by erosion presence (yes/no).

Evaluations were adjusted for the SDD values, as calculated in Chapter 5. To do this, at the scale unit of measurement, the raw evaluation score was divided by the SDD and rounded down to the next whole number. Only after SDD adjustment at the unit of

measurement native to the scale were data then classified by erosion presence. Data paired at the patient-joint level were used to populate the 2x2 contingency table comparing the binary signals of erosion detection on x-ray as the test against MRI as the reference standard. From this data arrangement, the following properties of diagnostic test accuracy were determined: odds ratio, sensitivity, specificity, and accuracy. Reliability-adjusted and unadjusted, evaluations were compared.

A number of analyses were then conducted to provide additional clinical context to the study results. First, the joints most commonly affected by erosive lesions on both imaging modalities were determined. Here, differences in the frequency of erosion detection between raw and SDD-adjusted values were plotted and inter-modality comparisons were made per joint. Second, inter-modality differences in the proportion of patients with erosive disease detected were compared using different combinations of clinically relevant anatomy imaged. McNemar's test of paired proportions with an exact correction for the binomial probability distribution was used to test for differences in the proportion of patients with erosive disease detected across different anatomical groupings imaged by each modality. Cohen's kappa was used to explore agreement on the patients with erosive disease detected by either imaging modality. Finally, to further investigate the trends in Chapter 4, depicting a symptom duration-dependent increase in sensitivity of x-ray for MRI erosions, scatter plots and Spearman's correlation coefficients were determined for erosion subscores on RAMRIS and vdHSS, respectively.

All analyses were conducted on SAS/STAT version 9.2 (Cary, NC). All statistical tests were conducted using a 5% level of significance.

6.3. Results

A total of 488 joints from 122 hands of 65 patients were included in this analysis. Demographic and disease characteristics for these patients were collected (**Table 1**). Patients were a median (IQR) 59 (49, 66) years of age, predominantly female (83.1%), and caucasian (≥61.5%). At least 24 of the 65 patients (36.9%) had radiographic evidence of erosions diagnosed outside the study setting by an independent radiologist based on the same radiographic images. Study participants had 4.3 (2.6, 7.0) years of symptoms on the date of MRI imaging. The median time (range) between x-ray and MRI imaging was 4.5 (-78, 61) days. The OMERACT RAMRIS for the 122 individual MCP 2-5 joint sets imaged was of 9 (5, 14) and the corresponding vdHSS for these patients was 14 (0, 33). Spearman's rho for the correlation between the OMERACT RAMRIS and vdHSS of individual hands was 0.31 (p=0.0004).

Table 1. Study participant characteristics, n = 65.

Baseline Characteristic	Value
Age, median (IQR)(range)	59.0 (49.0, 66.0) (36.0, 83.0)
Gender, male n (%)	11 (16.9)
Ethnicity, n (%)	
Caucasian	40 (61.5)
Arab/West Asian	3 (4.6)
Black	2 (3.1)
South Asian	1 (1.5)
Other	2 (3,1)
Missing	17 (26.2)
Smoking status, yes n (%)	25 (38.5)
Missing	12 (18.5)
Symptom Duration, median years (IQR)(range)	4.3 (2.6-7.0)(1.0-24.5)
Radiographic Evidence of Erosions, n (%)	24 (36.9)
Missing	7 (10.8)
Erythrocyte sedimentation rate (mm/h), median (IQR)(range)	22.0 (10.0-32.0) (0.0-69.0)
C-reactive Protein (mg/L), median (IQR)(range)	6.1 (3.0-21.9) (0.5-138.0)
Rheumatoid Factor Positivity, n (%)	46 (70.8)
Anti-nuclear Antibody Positivity, n (%)	28 (43.1)
Treating Physician, n (%)	
1	38 (58.5)

Baseline Characteristic	Value
2	7 (10.8)
3	13 (20.0)
5	2 (3.1)
8	5 (7.7)
Tender Joint Count, median (IQR)(range)	5.0 (1, 13) (0, 25)
Swollen Joint Count, median (IQR)(range)	10.0 (5, 13) (0, 22)
DAS28, median (IQR)(range)	4.5 (3.3-5.7) (0.8-9.2)
HAQ, median (IQR)(range)	1.5 (0.5-3.2) (0.0-5.3)
RADAI, median (IQR)(range)	23.8 (13.5-33.2) (0.8-61.9)
SDAI, median (IQR)(range)	35.0 (18.6-52.7) (0.2-115.7)
CDAI, median (IQR)(range)	62.3 (32.7-91.6) (2.2-155.5)
RAPID3, median (IQR)(range)	22.1 (15.7-30.6) (0.0-50.6)
RAPID4, median (IQR)(range)	24.4 (17.5-33.6) (1.5-56.0)
RAPID5, median (IQR)(range)	71.1 (44.0-102.8) (2.8-180.8)
HUI-III, median (IQR)(range)	0.8 (0.4-0.9) (-0.4-1.00)
OMERACT RAMRIS, median (IQR)(range)*	9.0 (5, 14) (0, 30)
vdHSS, median (IQR)(range)†	14.0 (4, 33) (0, 110)

CDAI = clinical disease activity index; DAS28 = 28-joint disease activity score; ESR = erythrocyte sedimentation rate; HAQ = health assessment questionnaire; EQ5D = EuroQol quality of life questionnaire; HUI-III = Health Utility Index Mark III; OMERACT = Outcome Measure in Rheumatology (formerly Outcome Measures in Rheumatoid Arthritis Clinical Trials); RAMRIS = Rheumatoid Arthritis Magnetic Resonance Imaging Score; RAPID3-5 = Routine Assessment of Patient Index Data versions 3 to 5; SDAI = Simplified Disease Activity Index; vdHSS = van der Heijde-modified Sharp score.
* OMERACT RAMRIS measured in dominant hand of study participants.
† vdHSS measured in hands, wrists, and feet of study participants.

From imaging data available across all 65 participants, the frequency distributions of erosion scores per joint were determined (**Figure 1**). The RAMRIS erosion raw evaluations were characterized by 45.8% units of measurement with signal (**Figure 1a**). In contrast, for vdHSS, 18.6% of raw evaluations had an erosion signal (**Figure 1b**).

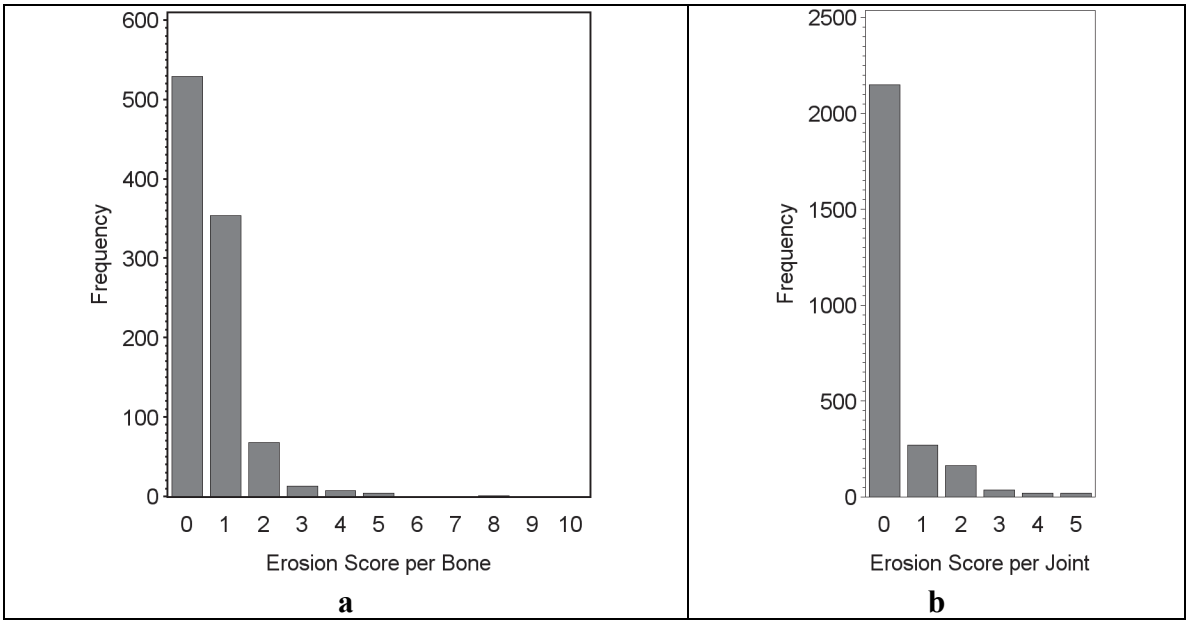


Figure 1. Frequency distribution of erosion scores at the unit of measurement for **a**) OMERACT RAMRIS, and **b**) vdHSS. OMERACT, Outcome Measures for Rheumatology (formerly Outcome Measures for the Evaluation of Rheumatoid Arthritis Clinical Trials); RAMRIS, Rheumatoid Arthritis Magnetic Resonance Imaging Score; vdHSS, van der Heijde-modified Sharp Score.

Comparing the paired data, the principal findings of this investigation are reported in **Table 2**. From the raw evaluations, 331 MCP 2-5 joints were assessed as having erosions on MRI (67.8%). In turn, 134 erosions were detected in the same joints on x-ray (27.5%). The odds ratio (OR) (95% CI) for the association between x-ray and MRI erosion detection was 1.8 (1.2-2.9). The Se (SD) was 0.31 (0.03) and Sp (SD) was 0.80 (0.03). The accuracy of erosion detection between the two modalities by raw evaluation was 0.47. Utilizing the reliability data from Chapter 5, and thereby limiting erosion detection to features of at least the magnitude of the SDD, the number of joints evaluated to have erosions on MRI decreased by 72.5% to 91 (18.6%). The number of erosions on x-ray decreased by 67.2% to 44 (9.8%). The OR, Sp, and accuracy increased to 3.2 (1.5-

6.1), 0.93 (0.01), and 0.79, respectively, while the Se decreased to 0.19 (0.04). At the joint level of analysis, 2.1-fold the erosions detected on x-ray were detected on MRI. Spearman's rho for correlation between RAMRIS and vdHSS raw erosion scores for common joints at the joint level of analysis was 0.28 (p=0.002).

Table 2. MCP 2-5 bone erosions detected on radiography and MRI with MRI as the reference standard (n=65; n_{hands}=122; n_{Joints}=488).

Property	**Evaluation**	
	Raw	**SDD-adjusted**
Odds Ratio	1.8 (1.2-2.9)	3.2 (1.5-6.1)
True Positive	103	17
False Positive	31	27
False Negative	228	74
True Negative	126	370
Sensitivity	0.31±0.03	0.19±0.04
Specificity	0.80±0.03	0.93±0.01
Accuracy	0.47	0.79

SDD_{mri}=2. SDD_{xray}=2.
SDD = smallest detectable difference.

Figure 2 illustrates the frequency of raw and SDD-adjusted erosions by unit of measurement. On MRI, of the MCP 2-5 bones (**Figure 2a**), erosions most frequently appeared on the metacarpal heads of MCP 2 and 3 joints. The frequency of erosions at these two sites dropped by 62.0% from 92 to 35 for MCP 3 and by 71.7% from 92 to 26 for MCP 2, when adjusted for the SDD. Across all distal bones, or phalangeal bases, the frequency of erosions dropped from raw counts ranging from 27 to 39 (22.1-32.0%) to the SDD-adjusted frequencies of 0 to 1 (0.0-0.8%). In x-rays of the hands, wrists and feet (**Figure 2b**), erosions were most frequently detected at MCP 2 for raw, 44 (36.1%), and at scaphoid for SDD-adjusted, 20 (16.4%), evaluations. The second most common site of erosions on x-ray was PIP3 with 17 (13.9%) SDD-adjusted erosions. At sites common to

both imaging modalities using SDD-adjusted figures (**Figure 2c**), MRI detected 2.6- to 8.0-fold the number of erosions detected on x-ray.

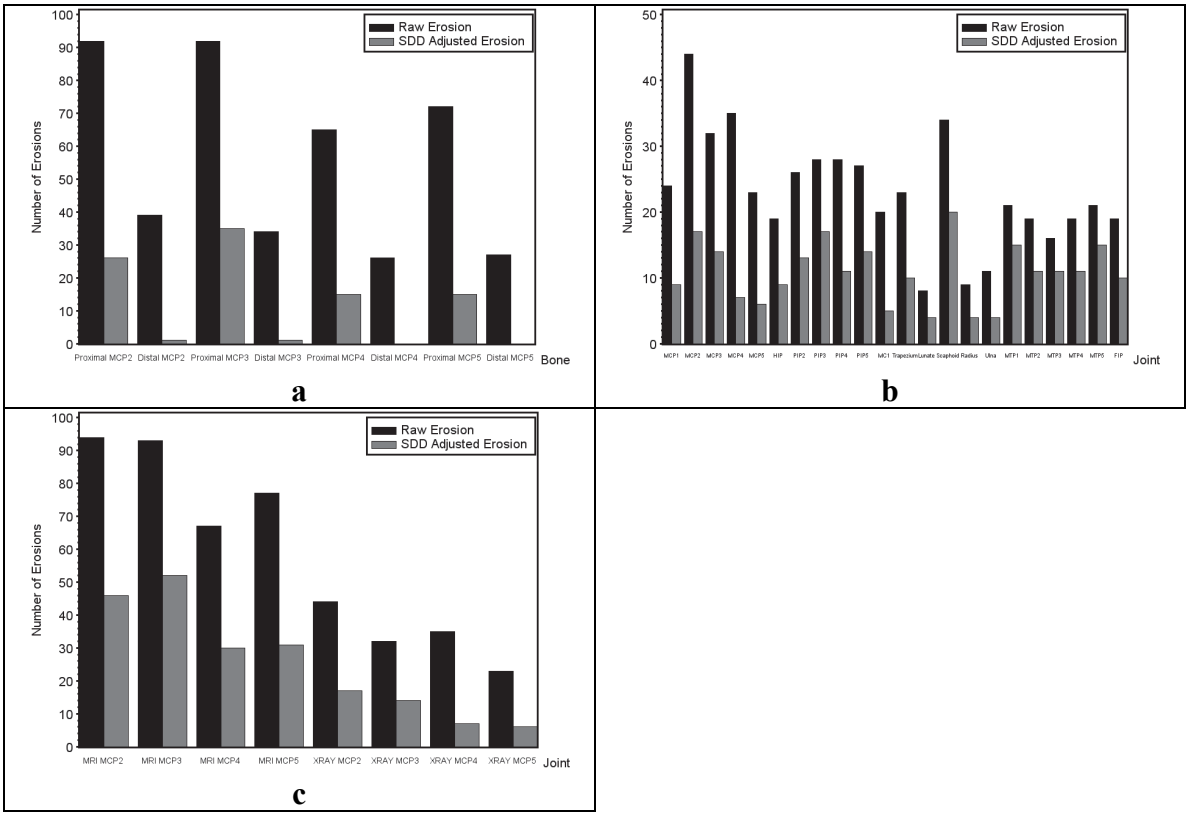


Figure 2. Raw and smallest detectable difference-adjusted frequency of erosion detection by unit of measurement. For **a**) MRI, **b**) x-ray, and **c**) joints common to both imaging modalities.

Cumulative frequency distributions of RAMRIS and vdHSS erosion subscores by unilateral MCP 2-5 joints were are illustrated in **Figure 3**. These distributions illustrate the effect of SDD adjustment at the unit of measurement on the frequency of erosion detection and erosion subscores. From raw evaluations, 108 sets of MCP 2-5 joints (88.5%) had evidence of erosive lesions on MRI (**Figure 3a**). The median (IQR) erosion score was 4 (2-7) and a maximum score of 15 was evaluated. Adjusting for SDD, the

number of joint sets with erosive disease decreased by 45.4% to 59, (48.4% overall). The average SDD-adjusted RAMRIS erosion subscore for a single hand dropped to 0 (0-2) with a maximum of erosion score of 12. In turn, 65 MCP 2-5 joint sets were evaluated to have evidence of erosive disease (53.3%) on x-ray by raw evaluation (**Figure 3b**). The median raw vdHSS erosion subscore for these joints was 1 (0-2) with a maximum score of 12. The SDD-adjusted erosion detection using vdHSS assessment for the paired hand and wrist decreased by 56.9% to 28, or 23.0% of included joint sets. The SDD-adjusted, average vdHSS erosion subscore for these joint sets was 0 (0-0) with a maximum score of 10. Overall, 2.1-fold the MCP 2-5 joint sets with erosion on x-ray had erosions on MRI.

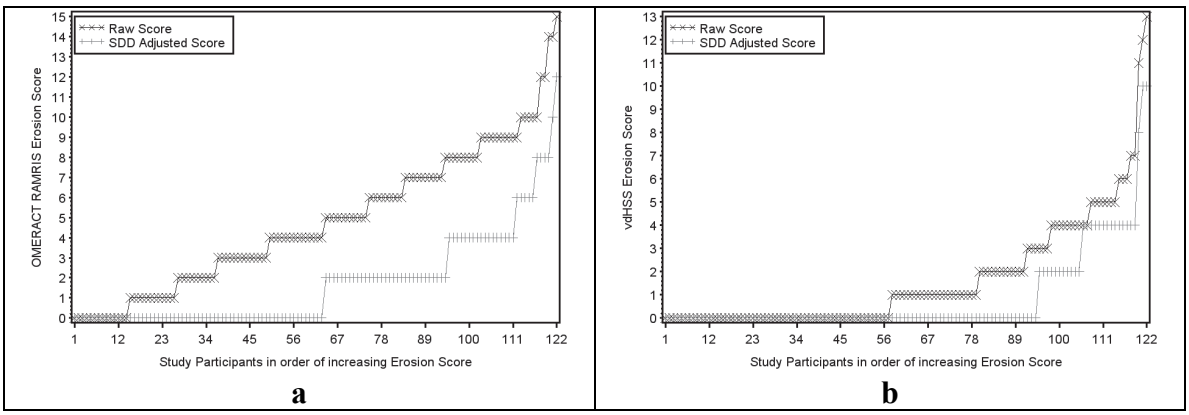


Figure 3. Cumulative probability plot of raw and unit of measurement SDD-adjusted erosion subscores for 122 2nd to 5th metacarpophalangeal joint sets on **a**) MRI using OMERACT RAMRIS and **b**) x-ray using vdHSS. OMERACT, Outcome Measures for Rheumatology (formerly Outcome Measures for the Evaluation of Rheumatoid Arthritis Clinical Trials); RAMRIS, Rheumatoid Arthritis Magnetic Resonance Imaging Score; SDD, smallest detectable difference; vdHSS, van der Heijde-modified Sharp Score.

Patient-centred analyses comparing erosion detection between MRI and x-ray were conducted (**Figure 4**). The frequency of bilateral sets of MCP 2-5 joints with SDD-adjusted erosions on MRI was 37 (64.9%) (**Figure 4a**). Limited to MCP 2-5 joint sets of

the dominant hand, erosions were detected on 23 of 57 hands (40.4%) (**Figure 4b**). On x-ray of the hands, wrists and feet for the same 57 patients who underwent bilateral MRI of the MCP joints, the frequency of SDD-adjusted erosions detected was 35 (61.4%) (**Figure 4c**). The evaluation of the x-ray of the feet resulted in the detection of 23 patients with erosions (40.4%) (**Figure 4d**). When the x-ray evaluation was limited to the hands and wrists, the frequency of patients with SDD-adjusted erosions was 31 (54.4%) (**Figure 4e**). The number of patients with erosions detected on MRI of MCP 2-5 bilaterally was 1.1-fold for the number detected on x-ray of the hands, wrists and feet: 12 patients had erosions exclusively on MRI and 10 had erosion exclusively on x-ray (McNemar's test, $p = 0.83$). Cohen's κ between this comparison was 0.17 ± 0.13 ($p=0.16$). The number of erosions detected on x-ray of the hands, wrists and feet was 1.5-fold the number on unilateral MRI of the MCP 2-5 joints; 5 patients had erosions exclusively on MRI and 17 had erosions exclusively on x-ray (McNemar's test, $p = 0.02$; $\kappa = 0.26 \pm 0.12$, $p = 0.03$). Comparing unilateral MRI to x-rays of the bilateral feet resulted in the detection of the same number of erosions detected on either modality: 10 patients with erosive disease exclusive to MRI; 10 patients with erosive disease exclusive to x-ray (McNemar's test, $p = 1.00$). Comparing unilateral MRI to x-rays of the hands and wrists, 1.3-fold the number of erosions detected on MRI were detected on x-ray (McNemar's test, $p = 0.12$; $\kappa = 0.31 \pm 0.12$ $p = 0.02$).

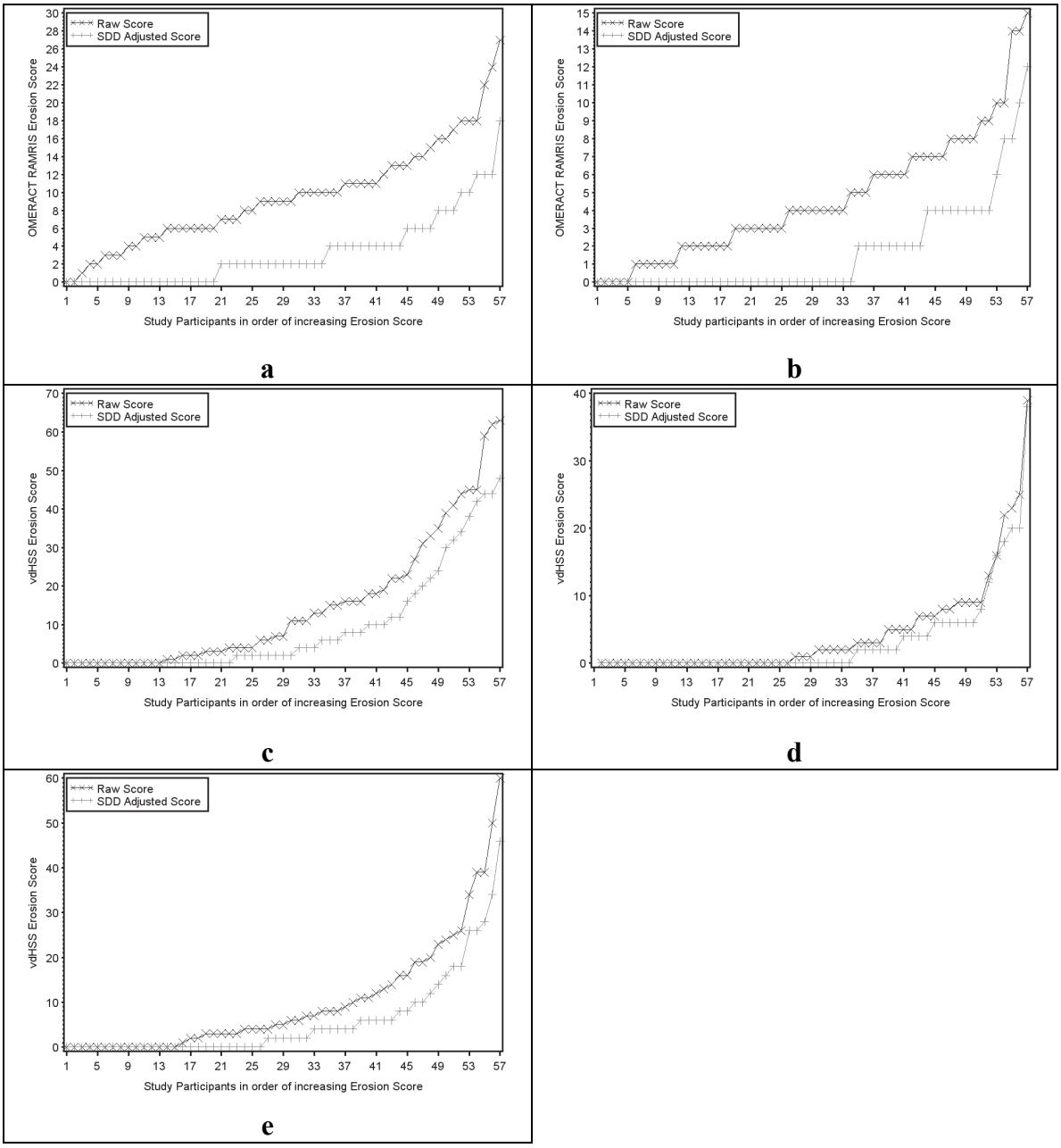


Figure 4. Cumulative probability plot of raw and unit of measument SDD-adjusted OMERACT RAMRIS and vdHSS erosion subscores at a patient level of analysis. **a**) MRI using OMERACT RAMRIS of bilateral MCP 2-5 joint sets. **b**) OMERACT RAMRIS of dominant hand MCP 2-5 joints. **c**) vdHSS of the hands, wrists and feet. **d**) vdHSS of the feet. **e**) vdHSS of the hands and wrists.
OMERACT, Outcome Measures for Rheumatology (formerly Outcome Measures for the Evaluation of Rheumatoid Arthritis Clinical Trials); RAMRIS, Rheumatoid Arthritis Magnetic Resonance Imaging Score; SDD, smallest detectable difference; vdHSS, van der Heijde-modified Sharp Score.

Given the results from Chapter 4, the potential modifying effect of symptom duration on the association between x-ray and MRI erosion detection was explored. The bivariate associations between raw erosion scores on either imaging modality and symptom duration were determined (**Figure 5**). A large amount of scatter in the data for both MRI and x-ray suggests the involvement of other variables in the association between symptom duration and erosion subscore. Adjusting for the SDD had little impact on the association between the erosion subscore and symptom duration, Spearman's rho of 0.10 (p=0.26). The correlation between SDD-adjusted vdHSS erosion score and symptom duration increased to 0.37 (p<0.0001).

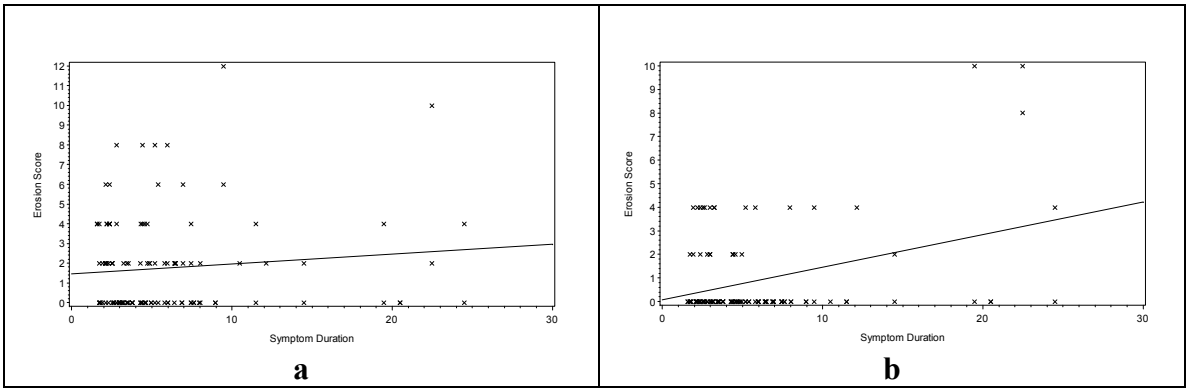


Figure 5. Correlation between unit of measurement-SDD-adjusted individual hand erosion subscore and symptom duration for **a**) OMERACT RAMRIS, Spearman rho = 0.10 (p=0.26), and **b**) van der Heijde-modified Sharp score, Spearman rho = 0.37 (p<0.0001). OMERACT, Outcome Measures for Rheumatsology (formerly Outcome Measures for the Evaluation of Rheumatoid Arthritis Clinical Trials); RAMRIS, Rheumatoid Arthritis Magnetic Resonance Imaging Score; SDD, smallest detectable difference; vdHSS, van der Heijde-modified Sharp score.

6.4. Discussion

The relative merits of MRI and x-ray for the detection of bone erosions in RA were compared using a prospectively collected, cross-sectional, paired dataset of 488 MCP 2-5 joints from 122 hands of 65 patients, 57 of whom had complete imaging data. Building on the findings from Chapter 5, erosion detection was adjusted for measurement reliability. Adjustment resulted in increases to the strength of association, specificity and accuracy, and decreases in sensitivity and frequency of erosion detection. It was determined that the association between erosion detection on the MRI and x-ray had an OR of 3.2 (1.5-6.1), a Se of 0.19 (0.04), Sp of 0.93 (0.01), and accuracy of 0.79. Adjustment also lowered the number of erosions detected per MCP 2-5 joint from 67.8% to 18.6% on MRI and 27.5% to 9.8%% on x-ray. Per common joint imaged, 2.6- to 8.0-fold the erosions detected on x-ray were detected on MRI. By anatomical grouping, comparing MCP 2-5 joint sets, adjustment resulted in MRI detection of 2.1-fold the erosive disease detected on x-ray. At the patient level of analysis, bilateral MRI of the MCP 2-5 joints resulted in the detection of erosive disease in 1.1-fold the number detected on x-ray of the hands, wrists and feet. Limiting MRI to the dominant MCP 2-5 joints, the proportion of patients with erosive disease was 66% of the frequency detected on x-ray of the hands, wrists, and feet; the same frequency detected on x-rays of the feet alone; and 1.3-fold the frequency detected on x-rays of the hands and wrists. A significant, moderate positive Spearman correlation existed between x-ray erosion subscore and symptom duration of 0.37 (p<0.0001). The corresponding weak correlation between MRI and symptom duration of 0.10 (p=0.26) was non-significant. These data

corroborate the findings from Chapter 4 that suggest the association between erosion detection on x-ray and MRI is dependent on symptom duration and should be accounted for in future studies. Technologically, the findings demonstrate the superiority of MRI to detect erosions per joint imaged. Practically, the relative performance of the two imaging modalities is highly dependent on the anatomy imaged.

6.4.1. Literature comparison

As summarized in Chapter 4, a number of studies have investigated the relative merits of MRI and x-ray for erosion detection (1-16). The findings reported in this chapter were compared to the diagnostic performance reported by others. In the current study, considerable effort was directed at understanding and integrating inter-rater reliability into the evaluation of the association between the two imaging modalities. Alternative approaches to handling measurement error exist and were discussed. Beyond inter-rater reliability, a number of other sources of heterogeneity between the current and comparable studies exist and were discussed as well. These discussions were placed in context with recommendations for investigations on diagnostic tests.

6.4.1.1. Relative diagnostic performance

Diagnostic test properties are specific to clinical indication and setting (44). Whereas the clinical indication relates the test signal to the reference standard, the setting dictates the prevalence of the signal in persons with and without the disease. Diagnostic tests may be indicated for disease screening, diagnosis, identification of physiological derangements, prognosis, and monitoring of disease progression or treatment response

(20). For diagnosis, the aim of a test is to accurately differentiate disease from non-disease.

Several properties exist to characterize diagnostic test performance including, Se, Sp, positive and negative predictive values, and receiver operating characteristic area under the curve, positive and negative likelihood ratio, OR, and accuracy (20,45-9). The diagnostic test properties focused on were OR, Se, Sp and accuracy. Of these, Se and Sp are most commonly reported. The other properties may be derived directly from the same data.

The Se findings from the literature are low but are heterogeneous, ranging from less than 0.20 (1-2,4-5,12-3) to greater than 0.50 (4,9-10,16). These data indicate that only a fraction of erosions detected on MRI were detected on x-ray. The Sp of x-ray for the detection of MRI erosions frequently approached unity (1-7,10-14,16). The high Sp suggests that most erosions detected on x-ray are also detected on MRI. Across comparable studies, there appeared to be a trend of decreased Sp with increased Se. The trend is consistent with the findings in this chapter before and after reliability error correction. Upon reliability adjustment in the current study, the Se dropped from 0.31 (0.03) to 0.19 (0.04) and the Sp increased from 0.80 (0.03) to 0.93 (0.01). The drop in Se may be partially explained by the reduction of chance agreement resulting from the reduction in measurement error. Similarly, the measurement error in the raw evaluations contributed to the low Sp reported. The adjusted Sp may also be artificially elevated due to the low prevalence of erosions versus non-erosions detected on x-ray. By definition, the resulting small proportion of false positive signals (x-ray, yes; MRI, no) relative to

true negative (x-ray, no; MRI, no) results in a high specificity despite a modest positive predictive value of 0.39. Nonetheless, these data generally support the hypothesis that MRI has a lower limit of detection than x-ray and therefore may be used to detect smaller erosions, or erosions at an earlier stage of disease. And despite some analytical limitations, the data corroborate the general findings that erosions detected on x-ray are nearly 100% restricted to erosions that appear on MRI.

6.4.1.2. Inter-rater reliability adjustment

The primary analysis of this chapter was contingent on correction for measurement error using the SDD at the unit of measurement. This is not the only available approach to account for inter-rater reliability. In the comparable literature, a consensus between two raters was the preferred alternative (3,7-8,10,15) albeit implemented in a minority of studies. In contrast to the sensitivity-lowering effect exhibited in the current study, the consensus approach resulted in some of the highest sensitivity values reported in the literature (3,7-8,10,15). These studies reported mean sensitivity values of at least 0.34 (0.20-0.51). Theoretically, advantages and limitations exist to both approaches (50). In the context of the many sources of heterogeneity encountered in the literature, it is difficult to compare the two approaches based on the existing data. Theoretically, both approaches would be expected to improve both the internal and external validity of the comparison.

In the current study, each radiologist's evaluation was independently obtained: all image sets were anonymized to guard against intra- and inter-rater and -modality interpretative contamination (51). The image sets included here were evaluated as part of

a larger pool of similar image sets to further dilute the possibility of contamination. Inter-rater reliability at the unit of measurement was investigated using Bland and Altman's *'limits of agreement'* approach, from which the SDD is directly derived (33). The SDD concept is becoming increasingly common as a method of accounting for reliability of OMERACT RAMRIS and vdHSS (25-32). In this investigation it was applied as a criterion to differentiate true scores from measurement error (32,34). As illustrated in Bruynesteyn *et al*. (2005), to date the SDD has been determined at the overall composite score or component feature subscore levels, effectively ignoring measurement error at the unit of measurement and handling these composite overall or component subscores as single, interval variables (34). It may be a significant detriment to the validity of data in the literature. The *'limits of agreement'* concept was introduced as a method to investigate agreement between interval measures, such as the continuous variable exemplified, peak expiratory flow rate (33). Neither RAMRIS, vdHSS, nor the comprising subscores are interval data variables. Instead they are composites of aggregated interval or ordinal scales. These measures are composed of more than one component feature ('composite'). Each feature is evaluated at multiple units of measurement and individual scores pooled ('aggregate'). Both RAMRIS and vdHSS sum the unweighted score at each unit of measurement and across components, as applicable, to produce the overall score.[4]

It is incorrect to estimate the reliability of either a composite or aggregate measure at the level of the overall score (35). Doing so fails to account for the measurement error

[4] The vdHSS erosion subscore scale for feet is evaluated on a scale for 0 to 10, compared to the hand scales that range for 0 to 5, which effectively doubles the weight of the feature at these joints.

associated with individual measures (at the unit of measurement). The errors at the unit of measurement must be propagated to determine the true reliability of the composite (36-7). Otherwise, the inter-rater reliability ought to be accounted for at the unit of measurement where the *'limits of agreement'* approach more appropriately applies. In this chapter, the latter approach was adopted.

For the sake of comparison, the effect of calculating SDD at the erosion subscore level is compared to that at the unit of measurement. For MRI, SDD-adjustment at the erosion subscore level (erosion subscore SDD of 11 from Chapter 5 applied to **Figure 3a**) meant that only 5 (4.1%) of 122 unilateral MCP 2-5 joint sets imaged would have erosions. Contrast that with the application of the SDD at the unit-of-measurement: this resulted in the identification of 59 (48.4%) MCP 2-5 joint sets with erosive findings. For x-ray findings measured using the vdHSS, the relative results were similar. The vdHSS erosion subscore SDD for the hands, wrists and feet was 34 (**Chapter 5, Table 2**). Applied to the cumulative erosive score for this anatomy (**Figure 4c**) resulted in the identification of 9 (15.8%) of 57 patients with erosive disease. Using the unit of measurement SDD, 35 (61.4%) of 57 patients were determined to have erosive disease. Limited to the hands and wrists, a vdHSS erosion subscore SDD of 26 resulted in the identification of 5 (8.8%) of 57 patients with erosions. In contrast, applied at the unit of measurement, erosions were detected in 31 (54.4%)of 57 patients.

These data illustrate that SDD adjustment at the subscore level may be insensitive to changes at individual joints. Further, the SDD calculations at the aggregate score level reported here account only for variations in the overall score and not the variance of the

individual measurements. If the error around individual measurements were accounted for, the SDD at the aggregate score would be expected to be greater still.

6.4.1.3. Sources of heterogeneity

Aside from the methods employed to account for measurement error, the literature investigating this research question varies in several other ways as described in Chapter 4. Some sources of heterogeneity that exist in the literature include the following: number of plain film projections; strength of the MRI magnet; definition of the MRI erosion signal; anatomy investigated; unit of measurement compared; and, failure to account for symptom duration. A greater number of x-ray projections is expected to improve the accuracy of the erosion detection. The conventional maximum of three projections was used in the current study. The increased strength of the magnet, and more specifically a greater ratio between magnetic strength and MRI bore size, results in greater resolution and increased accuracy of erosion detection. Bore size, however, is rarely reported in the literature. Use of multiple MRI planes is expected to improve the measurement error by reducing the number of erosion-like artifacts detected (11). Anatomy plays a role in the assessment. Anatomical peculiarities may alter the relative capacity of either imaging modality to resolve the erosion signal. Less specific geo-anatomical cross-referencing is expected to buffer measurement error and therefore result in an increased association between the two modalities. Finally, the association between the two imaging modalities for erosion detection is dependent on symptom duration. In retrospect, this seems intuitive. However, it was first suggested in Chapter 4. It does not appear to be accounted for in the literature. Symptom duration reflects the duration of

uncontrolled or poorly controlled synovitis. Synovitis begets erosive bone damage and erosive lesions progress in size with time leading to their early detection on MRI and later detection on x-ray.

In our study, a 1.0T magnet and 100 mm coil were used. Others have found that the performance of this system is comparable to that of 1.5T whole body MRI units for the detection of erosion, edema and synovitis (52). Likewise, a number of other studies have compared magnetic strength as low as 0.2T to conventional whole body high field standard 1.5T MRI units (53-6). Without exception, the studies report comparable findings with the caveat that none were designed as equivalence studies. Notably, two studies captured in Chapter 4, utilizing at 0.2T MRI, consensus image evaluation and comparing overlapping anatomy with that investigated here reported some the highest Se of x-ray for MRI erosions (3,10). Coincidentally, neither of these studies used an OMERACT-compliant definition of erosion. One required a signal on a STIR sequence in addition to the correct devoid signal on the T1-weighted sequence (3). Both studies evaluated the presence of erosions from images acquired from a single plane. (3,10). The STIR signal included in the former erosion definition may represent either effusion or pannus infiltration of the erosive cavity similar to the T1-weighted post-gadolinium enhancement signal (38). In the RAMRIS definition (38), and as defined in all remaining studies included in Chapter 4, erosions may exist in the absence of a T2-weighted signal. The absence of pannus tissue in the erosion may have implications on bone healing: healing occurs more prominently in hands devoid of inflammation associated with a T2-

weighted signal (57). The inclusion of a T2-weighted signal in the erosion definition would lower the number of MRI erosions detected resulting in a greater sensitivity.

As suggested from the findings in Chapter 4, the sensitivity of x-ray for MRI erosions increased with symptom duration. The correlational findings between symptom duration and erosive progression on each modality supported this hypothesis. A significant, moderate correlation between vdHSS erosion subscore and symptom duration was determined. The weak correlation between individual hand MCP 2-5 RAMRIS erosion subscore and symptom duration was not significant. These findings support the hypothesis that as symptom duration increases, erosions increase in size, increasingly surpassing the x-ray limit of detection. The plateau of erosion detection on MRI may suggest a ceiling effect on the size of many erosions, a varying susceptibility of sites to erosions, and/or a differential protective effect of anti-rheumatic treatment to the development of new erosions relative to the progression of existing erosions. These hypotheses warrant further investigation. Data also indicate that most erosions are small. In contrast, the number of erosions detected on x-ray increases as symptom duration increases.

The difference in the limit of detection between the two modalities is such that many of the erosion-susceptible sites appear affected on MRI prior to detection on x-ray. Restated, initially a prominent floor effect exists for x-ray detection for early disease. As disease duration increases, erosion size increases resulting in detection on x-ray. Since the joints examined over time are fixed and erosions are evaluated on a binary basis (as opposed to a scale that accounts for erosion size), the comparison imposes a ceiling effect

on MRI erosion detection earlier than with x-ray. As such, with increased disease duration, the sites susceptible to erosion are more quickly saturated on MRI than on x-ray. This allows for a greater number of erosions to be detected on x-ray with time. This does not explain why existing erosions do not appear to progress with symptom duration though. Further research is required to investigate these observations.

6.4.2. Clinical relevance

The clinical relevance of the results is that the relative merits of the two diagnostic imaging modalities may be dependent on disease stage, measurement scale, and the number of joints imaged. Despite the importance to validity of precise, geo-anatomical cross-referencing, extrapolation of the study findings to the patient level has the greatest clinical relevance. A benefit of applying reliability concepts to the measurement of erosion is the lowering of the rate of over-diagnosis. Study diagnostic imaging results were compared to independent clinical radiology assessments for the same patient timepoints. In these comparisons, adjustment for inter-rater reliability at the unit of measurement reduced the percentage of patients with erosive disease by approximately 20-50% percent compared to raw study evaluations. At the patient-level, 2.1-fold the patients with x-ray evidence of erosions in the clinical setting were determined to have erosion in study setting from raw evaluations (McNemar's test, $p<0.0001$; $\kappa = 0.24 \pm 0.08$, $p = 0.004$). In contrast, 1.7-fold the patients with evidence of erosions in the clinical setting were detected in the study via SDD-adjusted assessment of erosions on x-ray (McNemar's test, $p = 0.004$; $\kappa = 0.19 \pm 0.11$, $p = 0.07$). These exploratory results were limited by the unavailability of reliability data from the clinical

setting. They suggest a disparity in erosion detection between the clinical and study settings. The disparity cannot be fully accounted for by SDD adjustment. Clinically, evidence of erosions contributes to the assessment of disease severity. This may influence the treating physician to recommend more intensive pharmacotherapy. From a regulatory perspective, diagnostic imaging evidence of erosive disease is used to support government reimbursement for expensive anti-rheumatic agents. Over-diagnosis of erosions may lead to an unintentional increase in drug expenditure.

In the context of the sensitivity of detecting erosive disease, the study provides a rationale for maximizing the number of anatomical regions imaged both on MRI and x-ray. In this chapter, it was demonstrated that MRI of bilateral MCP 2-5 joints resulted in the detection of 1.6-fold the proportion of patients with erosive disease detected on unilateral imaging of the same joints. Consistent with this, imaging the hands, wrists and feet on x-ray, resulted in the detection of erosive disease in 1.5 the proportion of RA patients detected from plain films the bilateral feet and 1.2-fold the proportion detected from plain films of the bilateral hands and wrists. Other studies reported a greater advantage to evaluating disease progression over different anatomical regions as opposed to bilateral imaging of the same anatomy (25-6). For MRI, the other anatomical site most commonly imaged, and RAMRIS applicable, is the wrist.

With inter-modality comparisons, the proportion of patients with erosive disease detected by bilateral MRI of the MCP joints was comparable to that detected on x-ray of the hands, wrists and feet. X-ray of the hands and wrists detected a greater proportion of patients with erosions than unilateral MRI of the MCP 2-5 joints. Unilateral MRI of the

MCP 2-5 joints detected a greater proportion of patients with erosions than x-ray of the feet. In all cases, however, many patients had erosions detected exclusively on MRI or x-ray. From the anatomy imaged in this study, the data suggest an additive value to the use of both modalities.

Specificity is another important diagnostic test property. It quantifies the extent to which false positives are avoided. In this chapter it was determined that the Sp of x-ray for MRI erosions approached unity. Limitations to this result are discussed below; however, of greater clinical relevance is the Sp of MRI erosions for RA. This is important considering our understanding of the long-term consequences of erosive disease on physical and functional disability stem from studies using x-ray data. Only a fraction of MRI erosions ever appear on x-ray: a point worth considering in the interpretation of the low sensitivity of x-ray for MRI erosions.

The specificity of MRI erosions for RA is uncertain, if not controversial (21-4). Studies indicate that the false positive rate may vary from less than 2% (21) to more than 5% of normal bones (22). At the patient level, this represents up to 65% of the healthy, non-RA patients with at least one erosion-like lesion on MRI of bilateral MCP 2-5 and wrist joints (22). Others have also reported erosions in a substantial proportion of healthy persons (23-4). Xie *et al*. (2008) reported non-zero mean RAMRIS erosion subscores for MCP 2-5 and wrist joint of healthy older controls of 62±7 years of age that were absent from younger control group of 25±5 years of age (23). A moderate positive correlation between erosion score and age was also reported by Olech *et al*. (2010) (22). These studies did not account for measurement error; despite providing some form of reliability

results. Assuming the SDD values reported in Chapter 5 applied to the Olech *et al*. (2010) study (22), 35.0% of healthy controls and 60% of RA patients would have at least one erosion of the bilateral MCP 2-5 and wrists. The authors reasonably hypothesized that limiting MRI erosion detection to a single anatomical plane and blinding to diagnosis increased the prevalence of erosions in their study (22). To the authors' credit, however, it ought to have been emphasized that although the former limitation was a source of bias towards greater erosion detection in their results, the latter provision safeguarded against a second bias, 'clinical review bias' (18,58-60). The OMERACT definition of erosion requires that the MRI T1-weighted devoid signal in cancellous bone be detected on two planes, or on a 3D reconstruction of a single plane with isometric voxels, and a cortical break observed on at least one plane (38). In a pre-RAMRIS paper, Klarlund *et al*. (1999) suggested that the second plane helped "to avoid … enhancing tissue at bone surfaces [from being] mistaken for bone erosions". Clinically, erosion detection on MRI may offer superior detection of erosive disease than x-ray. Further investigation into the specificity and other diagnostic test properties may be required to fully assess the clinical value of erosions exclusively detected on MRI.

The moderate positive correlation between symptom duration and x-ray erosions but not with MRI suggests a greater disparity in favour of MRI would exist in earlier disease. As such, in early inflammatory arthritis, or undiffierentiated arthritis, MRI detection of erosion may have a greater role than x-ray. Nonetheless, the isolated role of erosions has been investigated here. Each imaging modality may be used to evaluate additional distinct features of diagnostic relevance, including joint space narrowing,

edema (exclusive to MRI), and synovitis or soft tissue swelling. Similarly, the utility of the diagnostic imaging ought to be evaluated in the context of other clinical measures integrated into the care for RA patients. Interestingly, the recent classification criteria for RA dismiss the relevance of diagnostic imaging evidence in early RA detection (61) despite the longtime presence of x-ray evidence in earlier iterations (62-3). The relative merits of the imaging modalities ought to be evaluated in the context of the scientific, clinical, and economic considerations.

As a prognostic tool, early data on radiographic progression measured using vdHSS suggested that annual progression measured in the hands and wrists, 2.3-3.2% of maximum score, was comparable to that in the hands, wrists and feet, 2.1-3.2% (42). Here, where the absolute detection of erosions is secondary to the detection of erosive progression, the advantages of the minimizing the number of joints imaged, while accurately measuring progression, may outweigh the disadvantages of not monitoring all joints initially evaluated. A balanced consideration of all performance properties of diagnostic tests should be considered in determining the optimal combination of joint groups to image. These may vary depending on the specific clinical indication for the diagnostic test.

6.4.3. Limitations

There were a number of limitations to the current study. Some pertain to the inferences that may be made from these data, while others relate to the placement of this work in relation to existing gaps in our current knowledge of the role of diagnostic imaging in the clinical management of RA. Regarding the primary objective, the major

limitation was the failure to account for the effect of symptom duration. Unfortunately, this is a limitation of all studies that have investigated this research question, or at least those included in Chapter 4. In this body of work, symptom duration was identified as an effect modifier of the relationship between x-ray and MRI (64). This finding sets the expectation for the differential diagnostic test performance properties of imaging modalities at different stages of disease. Moving forward, symptom duration must be accounted for in evaluations of the relative diagnostic test accuracy of MRI and x-ray detection of erosions.

The use of rigorously investigated instruments for the semi-quantitative measurement of erosions was a strength in many regards, however, it also limited the precision to which geo-anatomical cross-referencing of the erosion signal could be made. Specifically, the vdHSS system for scoring x-rays limited cross-referencing to the joint level of analysis. Although cross-referencing at the joint level of analysis has been done by others (1-2,8-10,12,15), others still have improved the precision of cross-referencing to the bone (4,6-7,11,14), and better still, to the bone-quadrant unit of measurement (5,11-3,16). Some studies compared diagnostic test accuracy between these two imaging modalities at the anatomical joint group (3) or the patient unit of measurement (65-6). The latter comparison was not appropriate to assess diagnostic test accuracy of erosion detection, despite its acknowledged relevance to clinical decision-making.

The use of vdHSS imposed an additional unforeseen limitation. The vdHSS allows for multiple individual features to be summed into one aggregate score within each unit of measurement. RAMRIS does as well. However, vdHSS uses an ordinal

scale; whereas, RAMRIS uses an interval scale. This imposes a limitation to the validity the inter-rater reliability assessment of the vdHSS. To illustrate the issue, the ordinal nature of the scale ensures that the summed score of three features scored '1' is closer to a single evaluation of a '3' than is one '1' or two '1' evaluations; however, it may not be closer than a single '2' evaluation. Such is a property of ordinal scales: the difference between categories is not equal. Further, there may be clear differences in the reliability of assessing a single '2' score compared to two '1' scores that cannot be differentiated using the current scoring sheet. The current investigation of inter-rater reliability did not account this for limitation of utilizing vdHSS. Future research needs to investigate if this limitation is practically relevant, or relegated to a theoretical note.

As alluded to above, the specificity of x-ray for MRI erosions may be misleading. After correcting for measurement reliability, very few erosions were detected on x-ray and many joints were negative for both MRI and x-ray erosions. As such, the high specificity may be an artifact of this imbalance in TP, FP, FN, and TN cases. This artifact speaks to the importance of considering all diagnostic properties and not just the most common sensitivity and specificity. Here the positive predictive value of 0.39 indicating that only 39% of x-ray erosions were detected on MRI. Future research should consider methods of appropriately balancing the 2x2 contingency table and accounting for the vdHSS scale artifact of multiple erosion evaluations accounting for the evaluation at the unit of measurement. Despite the artifact and modest positive predictive value, the moderate odds ratio for the SDD-adjusted association, supports the conclusions made.

Data from Chapter 4 suggest that the association between MRI and x-ray erosions may be dependent on the anatomy investigated. The performance of diagnostic imaging devices and raters on the detection of erosions on the different anatomy may vary from that determined for MCP 2-5. Other anatomy commonly affected by erosions in the early course of disease includes wrist and MTP joints. Due to the expected inverse relationship between time demands on participants and study compliance and other practical reasons, wrist MRI were not examined. The RAMRIS system does not support MTP evaluations. Approximately 30 minutes are required to scan the unilateral MCP 2-5 of a study participant compared to less than 5 minutes to perform three projection x-rays of the hands, wrists, and feet. Conversely, the time required to score vdHSS is likely greater than RAMRIS due to the greater number of assessments points (units of measurement). Despite this limitation, the key findings from this study were expected to apply to other anatomy: measurement reliability and symptom duration serve important roles in quantifying the association between erosion detection on MRI and x-ray.

This study was not designed to and cannot be used to measure the absolute diagnostic performance of either modality. To this end, the STAndards for Reporting Diagnostic accuracy (STARD) and the Grading of Recommendations, Assessment, Development and Evaluation (GRADE) criteria for evaluating diagnostic test recommendations describe methodological considerations and their impact on the diagnostic test properties (17-20). In this study, the specific focus was to compare the relative utility of MRI and x-ray to detect erosions in RA. For this purpose, it was reasonable to assemble a sample strictly comprised of RA patients. At the unit of

measurement, there was a sufficient mix of expected true- and non-cases. Similarly, at the patient level of analysis, not all patients were expected to have erosions, especially not a sample of patients selected in terms of liberal disease activity eligibility criteria.

A much larger sample size would have been required to investigate factors contributing to variability in scores using multivariable analysis of covariance. A major finding from the current study was the determination that inter-rater reliability accounted for approximately a 20-50% over-diagnosis of erosions. The paired design and anonymization of image sets helped to control much of the variability in ratings other than the different raters themselves. Yet, potential sources for the variability included MSK fellowship training, years of experience, hardware for viewing images, MSK patient caseload, duration of pre-study experience with the measurement tools, and intra-study temporal dependence. In order to account for the variability in assessments at the unit of measurement, a sample size of 10-times the total number of independent variables (the number of units of measurement) is recommended (67). An investigation into these sources of variability among the scores reported may help to increase the reliability of future work.

The study design and sample composition (predominantly RA) imposed limitations on the multivariable investigations that could be studied with this dataset. As described earlier, the applications of diagnostic tests are varied. The specific indication of interest would determine the research question and hence the study design, sample size and variables collected (covariables, confounders and effect modifiers). For example, erosive progression on x-ray has long been viewed as an outcome to be avoided in RA. It

prompted a classic study to investigate the predictors of erosive RA among patients newly referred to rheumatology (68). In the Visser *et al*. (2002) study, clinical and laboratory assessments at presentation to rheumatology were used to predict the following outcomes for two years: 1) persistence of an inflammatory arthritis, and 2) erosive disease. For the former outcome, baseline erosive disease was identified as a predictor of persistence. For the latter, erosive disease was used as an outcome. The predictors of erosive disease included RF and anti-CCP positivity, positive compression test of the bilateral MTP, at least 1h or morning stiffness, and swelling in at least 3 joints. As a precursor to x-ray erosions, it may be relevant to determine the predictors of MRI erosions as a strategy to avoid erosive damage on x-ray. Alternatively, the investigation of early changes in clinical and laboratory measures on erosive progression with time may be intriguing. It may also be interesting to investigate MRI as a prognostic for x-ray erosive disease as an outcome, or to compare the performance of RAMRIS-based algorithms to the new RA classification criteria (61) as predictors of persistent disease in new referrals to rheumatology with suspected inflammatory arthritis. When used as an outcome, it should not be forgotten that the clinical relevance of radiographic erosions is their correlation with long-term functional disability. The further removed the surrogate is from the clinically relevant outcome, there is an increased risk of dissociation. These are all important questions. Relative to the current dataset, these can not be pursued without 1) longitudinal follow-up to verify the temporality between risk factor and outcome, 2) a sample selected to reflect the indication for which predictors of erosive disease are clinically relevant, and 3) a much larger sample size.

6.4.4. Future considerations

Perhaps most importantly, the study sheds light on the most appropriate method of applying SDD to aggregate and composite scores. Applied in this manner, the SDD concept needs to be further developed. In contrast to pre-existing studies implementing the SDD at the level of aggregated data, this study accounted for measurement reliability at the unit of measurement. Only changes beyond Bland and Altman's *'limits of agreement'* at the unit of measurement were counted as true signal, or abnormality (33). A more in-depth investigation into the limitations of applying this approach to ordinal scales, such as the vdHSS, requires further attention. In the current study, it was also assumed that the magnitude of the SDD was constant over the range of the two scales investigated. This may be inappropriate, especially when dealing with ordinal scales, where the magnitude of signal required for adjacent points varies. In this study, the strong positive skew in the distribution of vdHSS data (towards 0 scores) suggests that the SDD values are more applicable to lower end scores. In turn, a large sample, or the sampling of patients with a greater disease duration would be required to investigate this. Investigations into the most appropriate application of the SDD at the unit of measurement for aggregate and composite measures are needed and are applicable for all composite measures, not only OMERACT RAMRIS and vdHSS. Future reports should consider adjusting for measurement reliability when reporting diagnostic test performance properties. Given, the effect on the magnitude to the results reported here, reliability adjustment should perhaps be mandatory. Additional measures may be pursued to improve the quality of future diagnostic imaging accuracy studies. To this

end, the STARD and GRADE recommendations should be followed, where applicable (17-20). Consideration should be given to adapting the current systems for scoring diagnostic imaging to enable a more precise cross-referencing of features detected between raters and modalities. Such changes may overcome some of the limitations experienced here.

The proposed investigations will enable clinicians and researchers to further understand the implications of studies of diagnostic test accuracy. For example, our limited understanding of the prevalence of erosions in non-RA patients is based on raw evaluations, usually by a single rater. The study findings indicate that measurement error is pronounced even between MSK-fellowship trained radiologists. The application of these concepts to the epidemiology of erosion detection will help to ascertain the absolute test accuracy of diagnostic imaging.

6.5. Conclusions

This study demonstrated that MRI detects a greater proportion of erosions than x-ray in the MCP 2-5 joints of RA patients. It detects a similar proportion of RA patients with erosive disease when bilateral MCP 2-5 joints imaged on MRI are compared to the x-rays of the hands, wrists, and feet. The *'limits of agreement'* concept was applied using the SDD to account for inter-rater reliability and used to help elucidate the association between the detection of erosions on MRI and x-ray. When applied, the OR, Sp, and accuracy of the measurement increased while the Se decreased. Applying the reliability concept to the measurement of these features resulted in a moderate significant association between erosion detection on the two modalities. The Sp approached unity,

indicating that nearly all negative evaluations on x-ray were also negative on MRI, and the accuracy of the evaluations nearly doubled. The findings corroborate earlier studies and further support the hypothesis that the association is dependent on symptom duration. The study findings support the hypothesis that MRI has a lower limit of detection for erosions owing to its multiple slices and planes imaged compared to the single, superimposed, two-dimensional image with radiography.

6.6. References

1. Backhaus M, Kamradt T, Sandrock D, Loreck D, Fritz J, Wolf KJ, Raber H, Hamm B, Burmester G-R. Bollow M. Arthritis of the finger joints: a comprehensive approach comparing conventional radiography, scintigraphy, ultrasound, and contrast-enhanced magnetic resonance imaging. Arthritis Rheum 1999;42:1232-45.
2. Backhaus M, Burmester GR, Sandrock D, Loreck D, Hess D, Scholz A, Blind S, Hamm B, Bollow M. Prospective two year follow up study comparing novel and conventional imaging procedures with arthritic finger joints. Ann Rheum Dis 2002;61:895-905.
3. Crues JV, Shellock FG, Dardashti S, James TW, Troum OM. Identification of wrist and metacarpophalangeal joint erosions using a portable magnetici resonance imagines system compared to conventional radiographs. J Rheumatol 2004;31:676-85.
4. Duer-Jensen A, Vestergaard A, Dohn UM, Ejbjerg B, Hetland ML, Albrecht-Beste E, Ostergaard M. Detection of rheumatoid arthritis bone erosions by two different dedicated extremity MRI units and conventional radiography. Ann Rheum Dis 2008;67:998-1003.
5. Dohn UM, Ejbjerg BJ, Court-Payen M, Hasselquist M, Narvestad E, Szkudlarek M, Moller JM, Thomsen HS, Ostergaard M. Are bone erosions detected by magnetic resonance imaging and ultrasonography true erosions? A comparison with computed tomography in rheumatoid arthritis metacarpophalangeal joints. Arthritis Res Ther 2006;8:R110.
6. Dohn UM, Ejbjerg BJ, Hasselquist M, Narvestad E, Moller J, Thomsen HS, Ostergaard M. Detection of bone erosions in rheumatoid arthritis wrist joints with magnetic resonance imaging, computed tomography and radiography. Arthritis Res Ther 2008;10:R25.
7. Ejbjerg BJ, Narvestad E, Jacobsen S, Thomsen HS, Ostergaard M. Optimised, low cost, low field dedicated extremity MRI is highly specific and sensitive for synovitis and bone erosions in rheumatoid arthritis wrist and finger joints: comparison with conventional high field MRI and radiography. Ann Rheum Dis 2005;64:1280-7.
8. Forslind K, Larsson EM, Johansson A, Svensson B. Detection of joint pathology by magnetic resonance imaging in patients with early rheumatoid arthritis. Br J Rheumatol 1997;36:683-8.
9. Forslind K, Johanson A, Larsson EM, Svensson B. Magnetic resonance imaging of the fifth metatarsophalangeal joint compared with conventional radiography in patients with early rheumatoid arthritis. Scand J Rheumatol 2003;32:131-7.
10. Gaylis NB, Needell SD, Rudensky D. Comparison of in-office magnetic resonance imaging versus conventional radiography in detecting changes in erosions after one

year of infliximab therapy in patients with rheumatoid arthritis. Mod Rheumatol 2007;17:273-8.

11. Klarlund M, Ostergaard M, Gideon P, Sorensen K, Jensen KE, Lorenzen I. Wrist and finger joint MR imaging in rheumatoid arthritis. Acta Radiologica 1999;40:400-9.
12. Klarlund M, Ostergaard M, Jensen KE, Lysgard Madsen J, Skjodt H, Lorenzen I, and the TIRA group. Magnetic resonance imaging, radiography, and scintigraphy of the finger joints: one year follow up of patients with early arthritis. Ann Rheum Dis 2000;59:521-8.
13. Lindegaard H, Vallo J, Horslev-Petersen K, Junker P, Ostergaard M. Low field dedicated magnetic resonance imaging in untreated rheumatoid arthritis of recent onset. Ann Rheum Dis 2001;60:770-6.
14. Ostergaard M, Hansen M, Stoltenberg M, Jensen KE, Szkudlarek M, Pedersen-Zbinden B, Lorenzen I. New radiographic bone erosions in the wrists of patients with rheumatoid arthritis are detectable with magnetic resonance imaging a median of two years earlier. Arthritis Rheum 2003;48:2128-31.
15. Scheel AK, Hermann K-GA, Ohrndorf S, Werner C, Schirmer C, Detert J, Bollow M, Hamm B, Muller GA, Burmester GR, Backhaus M. Prospective 7 year follow up study comparing radiography, ultrasonography, and magnetic resonance imaging in rheumatoid arthritis finger joints. Ann Rheum Dis 2006;65:595-600.
16. Szkudlarek M, Klarlund M, Narvestad E, Court-Payen M, Strandberg C, Jensen KE, Thomsen HS, Ostergaard M. Ultrasonography of the metacarpophalangeal and proximal interphalangeal joints in rheumatoid arthritis: a comparison with magnetic resonance imaging, conventional radiography and clinical examination. Arthritis Res Ther 2006;8:R52.
17. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glazziou PP, Irwig LM, Lijmer JG, Moher D, Rennie D, de Vet HCW, for the STARD Group. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. Ann Intern Med 2003;138:40-4
18. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, Moher D, Renie D, de Vet HCW, Lijmer JG. The STARD statement of reporting studies of diagnostic accuracy: explanation and elaboration. Clin Chem 2003;49:7-18.
19. Schunemann HJ, Oxman AD, Brozek J, Glasziou P, Jaeschke R, Vist GE, Williams Jr. JW, Kunz R, Craig J, Montori VM, Bossuyt P, Guyatt G. GRADE: grading quality of evidence and strength of recommendations for diagnostic tests and strategies. BMJ 2008;336:1106-10.
20. Schunemann HJ, Oxman AD, Brozek J, Glazziou P, Bossuyt P, Chang S, Muti P, Jaeschke R, Guyatt GH. GRADE: assessing the quality of evidence for diagnostic recommendations. EBM 2008;13:162-3.

21. Ejbjerg B, Narvestad E, Rostrup E, Szkudlarek M, Jacobsen S, Thomsen HS, Ostergaard M. Magnetic resonance imaging of wrist and finger joints in healthy subjects occasionally shows changes resembling erosions and synovitis as seen in rheumatoid arthritis. Arthritis Rheum 2004;50:1097-106.

22. Olech E, Crues III JV, Yocum DE, Merrill JT. Bone marrow edema is the most specific finding for rheumatoid arthritis (RA) on noncontrast magnetic resonance imaging of the hands and wrists: a comparison of patients with RA and healthy controls. J Rheumatol 2010;37:265-74.

23. Xie X, Webber CE, Adachi JD, O'Neill J, Inglis D, Bobba RS, Wu H. Quantitative, small bore, 1 Tesla, magnetic resonance imaging of the hands of patients with rheumatoid arthritis. Clin Exp Rheumatol 2008;26:860-5.

24. Tonolli-Serabian I, Poet JL, Dufour M, Carasset S, Mattei JP, Roux H. Magnetic resonance imaging of the writst in rheumatoid arthritis: comparison with other inflammatory joint diseases and control subjects. Clin Rheumatol 1996;15:137-42.

25. Olech E, Freeston JE, Conaghan PG, Hensor EMA, Emery P, Yocum D. Using extremity magnetic resonance imaging to assess and monitor early rheumatoid arthritis: the optimal joint combination to be scanned in clinical practice. J Rheumatol 2008;35:580-3.

26. Ejbjerg BJ, Vestergaard A, Jacobsen S, Thomsen HS, Ostergaard M. The smallest detectable difference and sensitivity to change of magnetic resonance imaging and radiographic scoring of structural joint damage in rheumatoid arthritis finger, wrist, and toe joints: a comparison of the OMERACT Rheumatoid Arthritis Magnetic Resonance Imaging Score applied to different joint combinaions and the Sharp/van der Heijde radiographic score. Arthritis Rheum 2005;52:2300-6.

27. Jawaid WB, Crosbie D, Sholton J, Reid DM, Stewart A. Use of digital radiogrammetry in the assessment of joint damage in rheumatoid arthritis. Ann Rheum Dis 2006;65:459-64.

28. Sharp JT, Wolfe F, Lassere M, Boers M, van der Heijde D, Larsen A, Paulus H, Rau R, Strand V. Variability of precision in scoring radiographic abnormalities in rheumatoid arthritis by experienced readers. J Rheumatol 2004;31:1062-72.

29. van der Heijde D, Lassere M, Edmonds J, Kirwan J, Strand V, Boers M. Minimal clinically important difference in plain films in RA: group discussions, conclusions, and recommendations. OMERACT Imaging Task Force. J Rheumatol 2001;28:914–7.

30. Lassere M, van der Heijde D, Johnson KR, Boers M, Edmonds J. Reliability of measures of disease activity and disease damage in rheumatoid arthritis: implications for smallest detectable difference, minimal clinically important difference, and analysis of treatment effects in randomized controlled trials. J Rheumatol 2001;28:892-903.

31. Lassere MN, van der Heijde D, Johnson K, Bruynesteyn K, Molenaar E, Boonen A, Verhoeven A, Emery P, Boers M. Robustness and generalizability of smallest detectable difference in radiological progression. J Rheumatol 2001;28:911-3.

32. Lassere M, Boers M, van der Heijde D, Boonen A, Edmonds J, Saudan A, Verhoeven AC. Smallest detectable difference in radiological progression. J Rheumatol 1999;26:731-9.

33. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;1:307-10.

34. Bruynesteyn K, Boers M, Kostense P, van der Linden S, van der Heijde D. Deciding on progression of joint damage in paired films of individual patients: smallest detectable difference or change. Ann Rheum Dis 2005;64:179-82.

35. Robins et al. Handbook of Research Methods in Personality Psychology. New York: The Guilfod Press, 2009. p. 515.

36. He Q. Estimating the reliability of composite scores. Coventry, Qualifications and Curriculum Authority, 2010. Report no.: Ofqual/10/4703.

37. Bevington PR, Robinson DK. Data reduction and error analysis of the physical sciences. 3rd edition. Boston, McGraw-Hill, 2003.

38. Ostergaard M, Peterfy C, Conaghan P, McQueen F, Bird P, Ejbjerg B, Schnier R, O'Connor P, Klarlund M, Emery P, Genant H, Lassere M, Edmonds J. OMERACT Rheumatoid Arthritis Magnetic Resonance Imaging Studies. Core set of MRI acquisitions, joint pathology definitions, and the OMERACT RA-MRI scoring system. J Rheumatol 2003;30:1385-6.

39. Tavares R, Pope JE, Tremblay J-C, Thorne C, Bykerk VP, Lazovskis J, Blocka KLN, Bell MJ, Lacaille D, Hitchon CA, Fitzgerald AA, Fidler WK, Bookman AAM, Henderson JM, Mosher DP, Sholter DE, Laupacis A, Boire G, Tomlinson G, Bombarder C. Early management of newly diagnosed rheumatoid arthritis by Canadian rheumatologists: a national, multicenter, retrospective cohort. J Rheumatol 2011;38:2342-5.

40. Combe B, Landewe R, Lukas C, Bolosiu HD, Breedveld F, Dougados M, Emery P, Ferraccioli G, Hazes JMW, Klareskog L, Machold K, Martin-Mola E, Nielsen H, Silman A, Smolen J, Yazici H. EULAR recommendations for the management of early arthritis: report of a task force of the European Standing Committee for International Clinical Studies Including Therapeutics (ESCISIT). Ann Rheum Dis 2007;66:34-45.

41. O'Dell JR. Chapter 10. Rheumatoid arthritis: clinical aspects. *In* Koopman WJ, Boulware DW, Heudebert GR. Clinical Primer of Rheumatology. Philadelphia, Lippincott Williams and Wilkins, 2003. pp. 97-115.

42. van der Heijde DMFM, van Leeuwen M, van Riel PLCM, van de Putte LBA. Radiographic progression on radiographs of hand and feet during the first 3 years of

rheumatoid arthritis measured according to Sharp's method (van der Heijde modification). J Rheumatol 1995;22:1792-6.

43. Schedule of Benefits. Physician services under the Health Insurance Act (September 1, 2011). Toronto, Ontario Ministry of Health and Long Term Care, 2011.
44. Bleeker SE, Moll HA, Steyerberg EW. External validation is necessary in prediction research: A clinical example. J Clin Epidemiol 2003;56:826–32.
45. Altman DG, Bland JM. Diagnostic tests 1: sensitivity and specificity. BMJ 1994;308:1552.
46. Altman DG, Bland JM. Diagnostic tests 2: predictive values. BMJ 1994;309:102.
47. Altman DG, Bland JM. Diagnostic tests 3: receiver operating characteristic plots. BMJ 1994;309:188.
48. Deeks JJ, Altman DG. Diagnostic tests 4: likelihood ratios. BMJ 2004;329:168-9.
49. Glas AS, Lijmer JG, Prins MH, Bonsel GJ, Bossuyt PMM. The diagnostic odds ratio: a single indicator of test performance. J Clin Epidemiol 2003;56:1129-35.
50. Stemler SE. A comparison of consensus, consistency, and measurement approaches to estimating interrater reliability. Practical Assessment, Research and Evaluation. 2004;9. Accessed April 19, 2012, http://PAREonline.net/getvn.asp?v=9&n=4.
51. Streiner DL, Norman GR. PDQ Epidemiology. 2nd edition. Hamilton, BC Decker Inc., 1998. p. 60.
52. Naraghi AM, White LM, Patel C, Tomlinson G, Keystone EC. Comparison of 1.0-T extremity MR and 1.5-T conventional high-field-strength MR in patients with rheumatoid arthritis. Radiology 2009;251:829-37.
53. Savnik A, Malmskov H, Thomsen HS, Bretlau T, Graff LB, Nielsen H, Danneskiold-Samsoe B, Boesen J, Bliddal H. MRI of the arthritic small joints: comparison of extremity MRI (0.2 T) vs high-field MRI (1.5 T). Eur Radiol 2001;11:1030-8
54. Conaghan PG, Ejbjerg B, Lassere M, Bird P, Peterfy C, Emery P, McQueen F, Haavardsholm E, O'Connor, P, Edmonds J, Genant H, Ostergaard M. A multicenter reliability study of extremity-magnetic resonance imaging in the longitudinal evaluation of rheumatoid arthritis. J Rheumatol 2007;34:857-8.
55. Freeston JE, Conaghan PG, Dass S, Vital E, Hensor EMA, Stewart S, Emery P. Does extremity-MRI improve erosion detection in severely damaged joints? A study of long-standing rheumatoid arthritis using three imaging modalities. Ann Rheum Dis 2007;66:1538-40.
56. Suzuki T, Ito S, Handa S, Kose K, Okamato Y, Minami M, Sugihara M, Horikoshi M, Tsuboi H, Hayashi T, Goto D, Matsumoto I, Sumida T. New low-field extremity MRI, compacTscan: comparison with whole-body 1.5 T conventional MRI. Mod Rheumatol 2010;20:331-6.

57. Lukas C, van der Heijde D, Fatenajad S, Landewe R. Repair of erosions occurs almost exclusively in damaged joints without swelling. Ann Rheum Dis 2010;69(5):851-5.

58. Begg CB. Biases in the assessment of diagnostic tests. Stat Med 1987;6:411-23

59. Doubilet P, Herman PG. Interepretation of radiographs: effect of clinical history. AJR Am J Roentgenol 1981;137:1055-8.

60. Berbaum KS, Franken EA Jr, Dorfman DD, Barloon T, Ell SR, Lu CH, Smith W, Abu-Yousef MM. Tentative diagnoses facilitate the detection of diverse lesions in chest radiographs. Invest Radiol 1986;21:532-9.

61. Aletaha D, Neogi T, Silman AJ, Funovits J, Felson DT, Bingham III CO, Birnbaum NS, Burmester GR, Bykerk VP, Cohen MD, Combe B, Costenbader KH, Dougados M, Emery P, Ferraccioli G, Hazes JMW, Hobbs K, Huizinga TWJ, Kavanaugh A, Kay J, Kvien TK, Laing T, Pincus T, Smoeln JS, Stanislawska-Biernat E, Symmons D, Tak PP, Upchurch KS, Vencovsky J, Wolfe F, Hawker G. 2010 Rheumatoid Arthritis Classification Criteria: An American College of Rheumatology/European League Against Rheumatism collaborative initiative. Arthritis Rheum 2010;62:2569-81.

62. Arnett FC, Edworthy SM, Bloch DA, McShane DJ, Fries JF, Cooper NS, Healey LA, Kaplan SR, Liang MH, Luthra HS, Medsger Jr. TA, Mitchell DA, Neustadt DH, Pinals RS, Schaller JG, Sharp JT, Wilder RL, Hunder GG. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. Arthritis Rheum 1988;31:315-24.

63. Ropes MW for A committee of the American Rheumatism Association. Diagnostic criteria for rheumatoid arthritis 1958 revision. Ann Rheum Dis 1959;18:49-53.

64. Kelsey JL, Whittemore AS, Evans AS, Thompson WD. Methods in Observational Epidemiology. 2nd edition. New York, Oxford University Press, 1996. p. 15.

65. McQueen FM, Stewart N, Crabbe J, Robinson E, Yeoman S, Tan PLJ, McLean L. Magnetic resonance imaging of the wrist in early rheumatoid arthritis reveals a high prevalence of erosions at four months after symptom onset. Ann Rheum Dis 1998;57:350-6.

66. McQueen FM, Benton N, Crabbe J, Robinson E, Yeoman S, McLean L, Steward N. What is the fate of erosions in early rheumatoid arthritis? Tracking individual lesions using x-rays and magnetic resonance imaging over the first two years of disease. Ann Rheum Dis 2001;60:859-68.

67. Norman GR, Streiner DL. Biostatistics: The Bare Essentials. Hamilton, B.C. Decker, Inc., 2000. p. 153.

68. Visser H, le Cessie S, Vos K, Breedveld FC, Hazes JMW. How to diagnose rheumatoid arthritis early: a prediction model for persistent (erosive) arthritis. Arthritis Rheum 2002;46:357-65.

7. Integrated discussion and conclusions

7.1. Bringing it all together

The primary objective of this dissertation was to determine the relative merits of x-ray and MRI for the detection of erosions in RA. To establish the rationale for these investigations, earlier chapters introduced RA (Chapter 1), and reviewed the pharmacotherapeutic (Chapter 2) and diagnostic imaging (Chapter 3) approaches to its clinical management. The first investigation was a systematic review of the literature to determine the association between MRI and x-ray erosion detection (Chapter 4). The second investigation aimed to prospectively determine the inter-rater reliability of MRI and x-ray of disease progression in RA by 4 study radiologists (Chapter 5). Finally, the objective of the third investigation was to prospectively compare MRI and x-ray erosion detection in RA (Chapter 6).

In Chapter 1, the definition, epidemiology, burden of disease, and pathophysiology of RA were introduced. Rheumatoid arthritis is an idiopathic autoimmune disorder characterized by articular inflammation, stiffness and pain. Although primarily manifest in synovial joints, RA is a systemic inflammatory disease with substantial comorbidity, including an increased risk for cardiovascular disease, malignancy, and early mortality. The disease frequently has an immediate impact on functional disability. In its early course, functional disability is driven by uncontrolled synovitis (1). Prolonged, uncontrolled synovitis results in ligament and tendon damage resulting in the misalignment of bones and deformity (1). In the long-term, prolonged synovitis triggers the pathological resorption of bone, known as erosions (2,3). The

extent of erosive joint damage is correlated with long-term functional disability (4,5). Erosions are a principal component of radiographic measurement of disease progression. In clinical trials, radiographic evidence of disease progression is a primary avoidance outcome.

In Chapter 2, the pharmaceutical armamentarium for RA disease management was summarized. Pharmacotherapeutic interventions for RA aim to inhibit radiographic evidence of disease progression and disease activity. Disease-modifying anti-rheumatic drugs are differentially effective at controlling radiographic progression and disease activity. Recent pharmaceutical development for RA has been near-exclusively limited to the biologic response modifiers subclass of DMARD therapeutic agents. Whereas the mechanisms of action of many older DMARDs are not well understood, newer agents target specific cytokines, chemokines, receptors, or enzymes of the cells involved in the pathophysiology of the disease. Over the past two decades, other non-DMARD drug classes have assumed secondary, yet important roles in the clinical management of RA. Of these, non-steroidal anti-inflammatory drugs (NSAID) are effective at controlling symptomatic pain and acute inflammation but are ineffective at controlling radiographic progression. Corticosteroids have long been an important drug class in rheumatic care due to their fast-acting, generalized anti-inflammatory effects. Despite a long history as a therapeutic option, corticosteroids are used with caution due to their significant long-term side effects.

In Chapter 3, diagnostic imaging modalities utilized in RA care were discussed. The introduction focused on x-ray and MRI. Ultrasonography is also commonly used to

detect erosions, effusions, synovitis and synovial vascularisation in synovial joints but, was not the focus of this dissertation. A description of the x-ray and MRI signals detected and biometric characterization of common scoring systems for each were summarized in this chapter. The vdHSS and RAMRIS are common and well-characterized semi-quantitative scoring systems used to evaluate disease progression on x-ray and MRI, respectively. Both measures include evaluations of erosive disease. The vdHSS also evaluates JSN, a surrogate for cartilage degradation. In turn, beyond erosions, RAMRIS measures bone marrow edema and synovitis. The relative merits of these scoring systems were discussed.

In Chapter 4, a systematic review of the literature was carried out to investigate the relative diagnostic test accuracy of x-ray and MRI. A total of 16 studies with 34 unique datasets comprising a total of 10 953 paired observations was identified. The results demonstrated that x-ray had a low Se and high Sp for erosions detected on MRI. Heterogeneity in the study results precluded a meta-analysis. Study characteristics and compliance with general recommendations for the methodological quality of studies of diagnostic test accuracy (6-8) were also heterogeneous. Study-specific sources of heterogeneity included symptom duration, anatomical site investigated, MRI magnetic strength, erosion signal definition, number of x-ray projections used, use of prescribed scoring systems, number of raters, and unit of analysis. The methodological quality was leniently evaluated to be acceptable. The studies were primarily lacking in the avoidance of partial verification and the reporting of uninterpretable data (6-8). A trend of increasing Se and decreasing Sp with increasing symptom duration emerged from the

review. The effects of other variables were discussed. No studies accounted for measurement reliability.

In Chapter 5, the reliabilities of MRI and x-ray evaluations by 4 participating raters using a contrast-free modification of the RAMRIS and the vdHSS were determined, respectively. For the reliability assessment of RAMRIS, imaging data from 19 RA participants were evaluated. Per set of MCP 2-5 joints evaluated on RAMRIS, 20 assessments points (8 erosion; 8 edema; and 4 synovitis) were scored. Therefore, 76 evaluations (19 scores*4 raters) were used to evaluate the reliability of RAMRIS overall score and each of the component feature subscores. At the unit of measurement, a total of 304-608 evaluations were used to determine the reliability of the individual evaluations of erosion, edema, and synovitis. In turn, the reliability of the vdHSS was determined using data from 9 RA participants. The vdHSS includes a total of 86 assessment points (44 erosion; 42 JSN). As such, a total of 36 evaluations (9 scores * 4 raters) were used to determine the reliability of vdHSS overall score and each component feature subscore. At the unit of measurement, 1584 assessments were used to determine the reliability of individual evaluations of erosion and 1476 were used to determine the reliability of JSN. Random effects ICC values for inter-rater agreement were in the moderate range (9,10). Expressed in terms of the SDD, the findings dictated the extent to which an individual evaluation from a single rater could be considered abnormal. The SDD was determined a few different ways. First, for consistency with the approach reported in the literature (11), it was determined at the composite score level or component feature subscore level without accounting for error at the unit of

measurement. The SDD was also determined at the unit of measurement to account for inter-rater reliability of individual assessments. The RAMRIS SDD for a unilateral set of MCP 2-5 joints were 19 (overall), 11 (erosion), 5 (edema), and 7 (synovitis). At the unit of measurement, the RAMRIS SDD were 2 (erosion), 1 (edema), and 2 (synovitis). The vdHSS SDD for the hands, wrists and feet for the overall score of component feature subscores were 38 (overall), 13 (erosion), and 10 (JSN). Limited to the hands and wrists, the vdHSS SDD were 13 (overall), 11 (erosion), and 9 (JSN). At the unit of measurement, the vdHSS SDD were 2 (erosion) and 2 (JSN). The validity of measuring the reliability of composite or aggregate scores at levels other than the unit of measurement was challenged (12-15).

In Chapter 6, the relative diagnostic test accuracy of x-ray and MRI for erosion detection in RA was determined accounting for inter-rater reliability of individual joint assessments in a prospectively collected dataset. In this investigation, to compare the relative performance of the two technologies, the primary objective was investigated by pairing data at the most precise anatomical level common to both imaging modalities, referred to as geo-anatomical cross-referencing in the dissertation. To provide clinical context, the utility of either modality to detect erosive disease at the patient level was conducted. At a joint level of analysis of 488 MCP 2-5 joints, MRI detected 2.1-fold the erosions detected in x-ray. The relative diagnostic test accuracy of x-ray for MRI erosions using unit-of-measurement-SDD corrected values was characterized by the following properties: Se, 0.19 (0.04); Sp, 0.93 (0.01); accuracy, 0.79. The vdHSS erosion subscore for the MCP 2-5 joints was positively correlated with symptom duration, 0.37 (*p*

< 0.0001). The correlation between RAMRIS erosion subscore of unilateral MCP 2-5 joints and symptom duration was not significant, 0.10 ($p = 0.26$). Adjustment for SDD at the unit of measurement had the effect of lowering Se, and increasing OR, Sp, and accuracy. At the patient level of analysis, there was no significant difference in the proportion of erosive disease detected between bilateral MRI of the MCP 2-5 joints and x-ray of the hands, wrists and feet ($p = 0.83$), although 12 patients with erosive disease were uniquely identified by MRI and 10 were uniquely identified by x-ray. The comparison of unilateral MRI compared to the hands, wrists and feet favoured x-ray ($p = 0.02$). Comparisons of unilateral MRI to x-ray of the hands and wrists ($p = 0.12$), or feet alone ($p = 1.00$), were not significantly different given the available sample size. In the case of the former comparison, the results trended towards significance in favour of increased detection by MRI.

7.2. Conclusions

A number of conclusions were made from this body of work. Some of the findings were novel; others adopted concepts from other fields of research to challenge conventions held in the field. In this thesis, a systematic review of the relative diagnostic test accuracy of x-ray and MRI detection of erosion was conducted (Chapter 4). The review represented a novel initiative to systematically summarize and synthesize the collective findings of the literature in the area. In Chapter 5, the reliabilities of RAMRIS and vdHSS were reported. In addition to reporting reliability statistics for the overall composite and component feature subscores, as reported in the literature, they were also reported at the unit of measurement. Importantly, the validity of reporting the reliability

of composite measures at the overall score level without accounting for it at the unit of measurement was challenged. The challenge was supported by measurement theory reported in education (12), psychology (13-14), and medicine (15) backed by pioneering work by Spearman and Cronbach (14). The SDD at the unit of measurement was used to adjust the raw evaluations in Chapter 6, where the relative diagnostic test accuracy of MRI and x-ray was investigated. To the author's knowledge, others have not studied this. The potential for reliability concerns at the unit of measurement of the vdHSS has been recognized (16). Similar to the work of Ejbjerg et al. (2005) (17), to bring clinical context to the results reported, the relative merits of x-ray and MRI to detect absolute changes were reported. Unlike the former study, which applied the SDD to disease progression, here comparisons of unit-of-measurement-SDD-adjusted status scores were made. Below, the conclusions from this dissertation were summarized in point-form by chapter.

7.2.1. Systematic review

- X-ray detection of MRI erosions was characterized by low Se and high Sp
- Across studies, the results were markedly heterogeneous and precluded the meta-analytic pooling of data
- The methodological quality of the studies was markedly heterogeneous, principally lacking in the avoidance of partial verification and reporting of uninterpretable data (6-8)

- A number of study-specific considerations were hypothesized to affect the relative diagnostic test accuracy of x-ray and MRI erosion detection
- A trend of increased Se and, to a lesser extent, decreased Sp with increased symptom duration was noted

7.2.2. Inter-rater reliability of MRI and x-ray erosion detection

- Using the approach reported in comparable studies to investigate inter-rater reliability of vdHSS and RAMRIS, moderate to good rICC values were obtained
- The validity of evaluating the reliability of a composite or aggregate measure using the variance of the summed score alone was challenged. The measurement error associated with comprising evaluations must also be accounted for. This is especially important when the unit of measurement SDD exceeds a single unit of the scale.
- The reliability of the vdHSS erosions and JSN at the unit of measurement across 4 study radiologists were determined 2 (greater than the smallest scale unit)
- The reliability of RAMRIS erosions, edema, and synovitis at the unit of measurement across 4 study radiologists were determined 2, 1, and 2 (greater than the smallest scale unit for erosion and synovitis)

7.2.3. Reliability-adjusted comparison of MRI and x-ray erosion detection

- MRI detected a greater proportion of erosions than x-ray in the MCP 2-5 joints

- SDD-adjustment had the effect of decreasing the Se and increasing OR, Sp and accuracy of the association between x-ray and MRI erosion detection at the joint level
- The proportion of patients with erosive disease detected by MRI of the bilateral MCP 2-5 joints was comparable to x-ray of the hands, wrists and feet
- The proportion of patients with erosive disease detected by unilateral MRI of the MCP 2-5 joints was inferior to x-ray of the hands, wrists and feet, and comparable to the proportion detected by x-ray of either the hands and wrists, or feet alone.
- The study findings supported the hypothesis that MRI has a lower limit of detection for erosions owing to its multiple slices and planes imaged compared to the single, superimposed, two-dimensional image characterizing x-ray.
- The technological superiority of MRI over x-ray for detecting individual erosions per geo-anatomically cross-referenced unit of assessment was offset by the greater number of units of the assessment imaged on x-ray per clinically relevant image acquisition.

7.3. Clinical relevance

This scientific merit and clinical relevance of this dissertation is considerable. The scientific merit spans from the rigorous review of the literature in this area (Chapter 4) to the recognition of the special consideration required to apply the *"limits of agreement"* approach to evaluating the reliability of composite measures. These investigations represent novel contributions to the field of diagnostic imaging for RA care. The clinical relevance is also noteworthy as summarized below.

In Chapter 4, a trend between symptom duration and the relative diagnostic test accuracy of x-ray for MRI erosion was identified. In Chapter 6, the moderate positive correlation of 0.37 between vdHSS erosion subscores and symptom duration was also demonstrated ($p<0.0001$). The slight positive correlation of 0.10 between RAMRIS erosion subscores and symptom duration was not significant ($p=0.26$). Together, these findings suggest that MRI may be more sensitive than x-ray to the detection of early erosive disease. MRI may find utility in identifying and monitoring pre-radiographic erosions. In early IA, or undiffierentiated disease, MRI detection of erosions may have an even greater role than x-ray. This suggestion needs to be balanced with the uncertainty regarding the absolute specificity of MRI erosions for RA (18-21). Researchers have reported mixed results in this regard. Further, in the context that x-ray erosions are surrogates of long-term functional disability and not all MRI erosions progress into x-ray erosions, care needs to be taken in inferring x-ray outcomes onto MRI. Nonetheless, the data support further research into the clinical utility of MRI as a diagnostic and prognostic in the early stages of disease. This research does not preclude the relevance of MRI erosion detection at later stages of disease; however, a greater disparity in erosion detection favouring MRI than determined in the current study may be expected at early stages of disease. Relative diagnostic test accuracy is dependent on disease stage investigated, but also measurement system used, and number and variety of anatomical groups imaged.

The systematic review conducted in Chapter 4 compared the relative merits of x-ray and MRI for erosion detection at varying levels of analysis, e.g. joint group, joint,

bone, or bone quadrant. Despite the scientific importance of precise geo-anatomical cross-referencing, the relative performance of x-ray and MRI at the patient level has the greatest clinical relevance. In Chapter 6 it was demonstrated that the application of unit of measurement SDD-adjusted results reduced the percentage of patients with erosive disease by approximately 20-33% compared to raw evaluations. Alone, the clinical context of these results is limited. To gauge the clinical relevance, the SDD-adjusted x-ray results were compared to paired, non-study, radiological interpretations of the same images for these patients. It was determined that despite a decrease in the proportion of erosive disease detected relative to raw evaluations, unit of measurement-SDD-adjusted erosion detection still detected a greater proportion of patients with erosive disease compared to the radiological investigations in the clinical setting ($p = 0.004$). Of 24 patients with erosive disease detected on the same x-rays by different radiologists in the clinical setting, 18 were also detected in the study setting using SDD-adjustment. Although unit of measurement-SDD adjustment decreased the overall proportion of erosive disease detected, the relative accuracy of SDD-adjusted and raw study evaluations for clinical-based findings was equivalent. The comparison was exploratory, limited, and warranted further investigation.

In the context of the Se of detecting erosive disease, the study indirectly provided a rationale for further investigating the relative performance of x-ray and MRI where the number of different anatomical regions imaged was maximized. Others have noted an increased Se to erosion detection when different anatomical regions are imaged as opposed to the same anatomy bilaterally (17,22). The other anatomical site applicable to

the RAMRIS system is the wrist. In Chapter 6, it was demonstrated that bilateral MRI of MCP 2-5 joints resulted in the detection of a greater proportion of patients with erosive disease detected on unilateral MRI of the same joints ($p<0.0001$). The difference in the proportion detected by bilateral MRI and x-ray of the hands, wrists and feet was not significant ($p= 0.83$). Despite the existence of a number of patients with erosions specific to MRI or x-ray, any incremental benefits over standard of care by increasing the number of anatomical regions imaged, and/or the utilization of both MRI and x-ray need to be balanced with consideration of the increased cost and limited access to MRI.

The optimal combination of joint groups to image may vary depending on the specific clinical indication. For example, as a prognostic tool, the rate of annual progression of the vdHSS as measured from the hands and wrists was 2.3-3.2% of the maximum score compared to 2.1-3.2% as measured at the hands, wrists and feet (23). Similarly, on MRI, the proportions of RA patients with progressive disease were similar if progression on the RAMRIS was assessed through unilateral or bilateral evaluation of the MCP 2-5 joints (22). As a prognostic measure, the absolute detection of erosions is secondary to the detection of erosive progression. The advantages of minimizing the number of joints imaged while accurately measuring progression may outweigh the disadvantages of not ascertaining absolute progression at all joints commonly affected at the early stages of disease.

In Chapter 6, it was determined that the Sp of x-ray for MRI erosions was very high, 0.93 (0.01). The high Sp corroborates the findings reported in the studies included in Chapter 4. The findings generally indicate that, with little exception, negative

evaluations on MRI are also negative on x-ray. There are some limitations to this interpretation as described in Chapter 6 and summarized in the section below. Also, the implications of these findings need to be interpreted with consideration that only a fraction of MRI erosions ever progress into x-ray erosions and the absolute Sp of the MRI erosions for RA remains uncertain (18-21).

The role of additional distinct features of diagnostic relevance, including JSN, edema, and synovitis (or soft tissue swelling on x-ray) need to be considered to fully assess the clinical utility of these two diagnostic imaging modalities. Similarly, the utility of the diagnostic imaging ought to be evaluated in tandem with clinical measures integrated into the care for RA patients. Interestingly, the recent classification criteria for RA dismissed the relevance of diagnostic imaging evidence in early disease detection (24) despite the longtime presence of x-ray evidence in earlier iterations (25-6). The relative merits of the imaging modalities ought to be evaluated in the context of scientific, clinical, and economic considerations.

7.4. Limitations

As with all studies, there were a number of limitations to this body of work. The limitations were described in the discussion sections of the individual chapters and summarized below.

7.4.1. Systematic review

There were several limitations to the systematic review conducted. Study heterogeneity precluded the validity of conducting a meta-analysis of the results. As

such, potential sources of heterogeneity were discussed and patterns with relative diagnostic test performance and symptom duration were explored. The effect of disease duration on diagnostic test accuracy is best accounted for at the individual participant level as opposed to the study or dataset level. This requires access to raw data and the use of multivariable modeling and simulation approaches. These approaches were outside of the scope the literature review conducted here and were not pursued.

Quality appraisal was conducted by a single, lenient individual. If methodological quality was assessed by a second reviewer or applied more strictly, the included studies may have been interpreted to be of lower quality. Further, study-specific considerations not accounted for in the generic quality assessment tool (6-8) were expected to impact study quality. For example, the level of at which geo-anatomical cross-referencing was carried out affects the accuracy of the comparison: the more precise the cross-referencing, the more accurate the relative diagnostic test accuracy. Results across units of analysis were compared together, despite a limited validity to do so.

Further, the anatomical sites compared across studies varied. Evidence suggests that certain joints, e.g. MCP 2 and 3 and MTP 2 and 5, are most commonly affected in RA. Therefore, the relative diagnostic test accuracy at different joints may vary.

The study compared the relative detection of erosions using MRI as the reference standard. In the case of erosion detection, computed tomography (CT) may be superior to MRI (27). Clinically, however, both MRI and CT appear to lower the limit of detection of the clinically relevant reference standard, x-ray. X-ray, not MRI nor CT, is correlated with long-term physical and functional disability in RA. Care is therefore

required in inferring the radiographic correlations with clinically relevant outcomes onto other diagnostic imaging modalities.

The literature search was limited to articles from 1996 to present. Supplemental reading suggests the existence of relevant literature published before 1996. Although the search of the reference lists of included studies yielded no additional articles of interest, the formal literature search could have encompassed a broader timeframe. Other limitations to the literature search included the following: only English language studies were considered; conference proceedings were not considered due to inability to evaluate study quality from abstracts; and the search was limited to two, albeit prominent, citation indexes

7.4.2. Inter-rater reliability of MRI and x-ray erosion detection

In Chapter 5, the principal limitations were those pertaining to the statistics used to assess the reliability of the tools investigated. For the purposes of comparison, overall reliability statistics were calculated as reported in the literature. The composite or aggregated scores were interpreted as single variables and the measurement error of individual comprising evaluations were not considered. There may be an opportunity to optimize the approach used in future research. Nonetheless, for the primarily analysis, reliability at the unit of measurement was conducted. As such, the primary results of the study were not affected by this limitation.

In the case of the vdHSS, the ordinal nature of the scale at the unit of measurement limits the validity of applying the *"limits of agreement"* approach to measuring reliability. As exemplified in the discussion of Chapter 6, the same score at

the unit of measurement on the vdHSS may be achieved through the summation of different combinations of features that do not comprise the same erosive volume, or that would represent markedly different pathology. As such, even at the unit of measurement, the assessment of reliability of the vdHSS is limited. Notwithstanding this limitation, the assessment of reliability at the unit of measurement is relatively more accurate than its determination from the composite or aggregate score, while assuming the latter to be a single, interval datum.

Beyond these principal limitations, semi-quantitative scoring methods are not used clinically. This limits the direct clinical relevance of the study findings. Time requirements have been cited as the reason limiting the clinical use of semi-quantitative measures (32). Automated methods of quantifying x-ray JSN (28-9), MRI synovitis (30-1), and erosions (32-7) may one day help to bridge this gap. Currently, automated methods remain dependent on the accurate detection of the features by radiologists, resulting in more work for the limited radiology resource.

Reliability is a context- and rater-specific concept (38). Despite this, the rater- and study-specific nature of reliability data, expressed in terms of the composite score or component feature subscore SDD (erroneously), the results reported may have been more variable than those reported elsewhere (9,39-46). Subtleties in methodology may account for this. In the current study, image anonymization and the use of individual status scores were obtained to ensure evaluations were made as independent as possible without contamination from other images or raters. Nonetheless further consideration of methods

to improve the reliability of the study raters without affecting the accuracy may be required.

7.4.3. Reliability-adjusted comparison of MRI and x-ray erosion detection

In Chapter 6, because the unit of measurement-SDD derived in Chapter 5 was used to adjust the raw scores, the majority of the limitations summarized in the section above also applied here. Additional limitations specific to this chapter existed.

A major limitation was the inability to account for symptom duration. In Chapter 4, the trend was observed that x-ray Se for MRI erosions increased and Sp decreased with increasing symptom duration. The Chapter 6 dataset was not amenable to investigating this phenomenon. Instead, correlations between x-ray and MRI erosion subscores with symptom duration were explored. The findings suggested that symptom duration is an effect modifier of the relative diagnostic test accuracy of x-ray and MRI erosion detection.

The systems used to evaluate erosions on either modality limited the precision to which cross-referencing could be made. Here, cross-referencing was limited to the joint level of analysis, the smallest unit of measurement for vdHSS. Others have cross-referenced at the bone (47-51) or bone-quadrant level of analysis (50,52,53-5).

The study was limited to the investigation of the MCP 2-5 joints by MRI. As discussed in Chapter 4, the association between MRI and x-ray erosions may depend on the anatomy investigated. Despite RAMRIS supporting the evaluation of wrist joints,

due to the negative impact of time demands on participant compliance, MRI of these joints were not conducted. At baseline and two years, bilateral MCP 2-5 joint scans were imaged, instead of two different anatomical sites, for practicality: availability of RAMRIS-compliant MRI sequences for the MCP joints; applicability of MCP joint data to concomitant projects within the research group; and, perceived greater reliability to the measurement of the MCP joints.

The Sp of x-ray for MRI erosions may be misleading. The high Sp may be an artifact of the large imbalance in true positive and true negative units of analysis. Although the Sp and Se were focused on here to render the data comparable to the results reported by others, all diagnostic properties need to be considered when evaluating diagnostic test performance (56-60).

In this study, the specific focus was to compare the relative utility of MRI and x-ray to detect erosions in RA. This study was not designed to and cannot be used to measure the absolute diagnostic performance of either modality. To this end, the STAndards for Reporting Diagnostic accuracy (STARD) and the Grading of Recommendations, Assessment, Development and Evaluation (GRADE) criteria for evaluating diagnostic test recommendations describe methodological considerations and their impact on the diagnostic test properties (6-8,61-2). For the intended purpose of this study, the study population was appropriate.

A major finding from the current study was the determination that inter-rater reliability accounted for approximately a 20-33% of the variability in the measurement of erosions on either modality. Potential sources for the variability included MSK

fellowship training, years of experience, hardware for viewing images, MSK patient caseload, duration of pre-study experience with the measurement tools, intra-study temporal dependence, or other yet unforeseen sources. A sample size of 10-times the total number of independent variables (the number of units of measurement) is recommended to investigate the factors potentially contributing to the variance in the association between x-ray detection of MRI erosions (63). Such investigations may help to increase the reliability of future work.

7.5. Future considerations

Follow-up projects to the investigations studied here are evident, one of which, the trial reported in **Appendix A**, is ongoing. Future considerations for this research were summarized below.

Further research is required to better characterize the relationship between MRI and x-ray detection of erosions. From Chapter 4, it was suggested that symptom duration is an effect modifier of the relative diagnostic test accuracy of x-ray and MRI detection of erosions and should be more tightly controlled in the future studies. Likewise, the time interval between MRI and x-ray ought to be more tightly controlled. Several additional study-specific considerations were identified as potentially affecting the relative diagnostic test accuracy. These specific considerations are in addition to the general methodological considerations for studies of diagnostic test accuracy (6-8,61-2). It would be valuable to unequivocally determine the impact of these variables on the comparative performance of each modality. In the least, future studies should consider these issues

and report how they were managed. Most of these general considerations were applied to this body of work.

From the reliability investigation in Chapter 5, future research should be aimed at better understanding and addressing sources of inter-rater reliability. Certainly, a move towards the development of an automated, consistent system of scoring images would be advantageous. As reported earlier, several automated methods of quantifying these diagnostic imaging have been developed or are currently under development (28-37). This field of research should consider work towards validating an alternative mechanism of categorically identifying the abnormalities of interest, an aspect currently relegated to the radiologist. In doing so, subsequent developments in feature quantification would have greater clinical relevance. In the effort to do this, a more sophisticated analysis of the factors contributing the variability between raters may be required. In this regard, two interesting hypotheses need to be tested: the dependence of erosion detection on 1) the detection of concomitant erosion in the same in individual, and 2) the geo-anatomical localization of erosions. Given the study-specific nature of reliability properties, such work will need to be validated against a broad range of radiologists, or otherwise a robust radiological standard, to ensure external validity.

This research sheds light on the method of determining SDD for composite and aggregate measures. The approach is not novel, *per se,* and is referenced in education, psychology, clinical research literature (12-5). The principle is simple: the measurement error of comprising tests must be accounted for to determine the reliability of the overall composite score. The reliability of vdHSS and RAMRIS may need to be re-investigated

using this approach. Also, by further breaking down the units of measurement to the joint quadrant (or more specific), issues with the vdHSS described above may be lessened. The drawback of this latter consideration is the increased workload on the rater. Further, in the current study, it was also assumed that the magnitude of the unit of measurement-SDD was constant over the range of the two scales investigated. This may be inappropriate, especially when dealing with ordinal scales, where the magnitude of signal varies for adjacent points. In this study, the strongly positive skew in the distribution of data (towards 0 scores) may suggest that the SDD values are more applicable to lower end scores. Future work may sample patients with more severe disease to measure reliability along the range of the score scale. In the meantime, future reports should consider the advantages and disadvantages of adjusting for measurement reliability on reported outcomes.

The findings from this research were integrated into a prospective investigation of the utility of diagnostic imaging on pharmacotherapeutic treatment decision-making for early IA and RA over two years (**Appendix A**). This study represents an initial investigation. The work will contribute to the knowledge of how diagnostic imaging may be leveraged to optimize rheumatologic care. Future work will investigate the optimization of this intervention strategy, including, but not limited to, identifying population subsets at greatest risk, adjusting the interval over which imaging is conducted, and testing in conjunction with DMARD therapy-controlled, disease-activity based, intensive clinical management strategies.

7.6. Summary

Rheumatoid arthritis is an idiopathic autoimmune disorder with substantial morbidity. The consequences of RA may be abated by appropriate anti-rheumatic therapy. Radiographic progression is a surrogate of poor long-term disease outcome. Bone erosions are a main component of radiographic progression measures. An important property of conventional pharmaceuticals is the inhibition of erosive disease progression. Towards this end, diagnostic imaging may serve a greater role in clinically managing RA pharmacotherapy.

This dissertation has been a source of knowledge advancement on the relative merits of x-ray and MRI for erosion detection in RA. The literature generally supports the stance that in geo-anatomically cross-referenced comparisons, MRI is superior at detecting bone erosions. Erosions are detectable on MRI prior to x-ray; however, not all MRI erosions evolve into x-ray erosions. The relative diagnostic test accuracy of the two modalities at the geo-anatomical level is dependent on disease duration. Other study-specific and general quality aspects of studies of diagnostic test accuracy need to be accounted for in assessing the relative merits of the two technologies. The assessment of erosions is subject to measurement error. The *"limits of agreement"* concept was applied at the unit of measurement of each scoring system to determine the SDD. Among the four participating radiologists, it was determined that only differences in the OMERACT RAMRIS erosion increments of 2 could be differentiated on the interval scale ranging from 0 to 10. Similarly, for evaluation of erosions on the vdHSS, only increments of 2 could be differentiated on the ordinal scale ranging from 0 to 5. These findings were

applied to the prospective assessment of the relative diagnostic test accuracy of x-ray and MRI. Overall, the inter-rater-reliability-adjusted relative diagnostic test accuracy of x-ray for erosions on MRI was characterized by the following properties: OR, 3.2 (95% CI: 1.5-6.1); Se, 0.19 (0.04); specificity, 0.93 (0.01); and accuracy, 0.79. At the joint level, examining MCP 2-5 joints, MRI detected 2.1-fold the erosions detected on x-ray. Comparing the ability of the modalities to detect erosive disease at the patient level, bilateral MRI of the MCP 2-5 joints detected a similar proportion of patients with erosive disease as did x-ray of the hands, wrists and feet ($p = 0.83$), with 12 of 57 exclusively detected on MRI and 10 exclusively detected on x-ray. Bilateral MRI scans of the MCP 2-5 joints detected a greater proportion of patients with erosive disease than did unilateral MRI of the same sites ($p<0.0001$). A comparison of unilateral MRI of the MCP 2-5 joints compared to x-ray of the hands, wrists, and feet favoured x-ray ($p=0.03$). Similar proportions of erosive disease were detected when unilateral MRI of the MCP 2-5 joints were compared to x-ray of either the hands and wrists ($p=0.12$), or the feet ($p=1.00$).

The study findings support the hypothesis that, technologically, MRI has a lower limit of detection for erosions than x-ray at the geo-anatomical site of evaluation. Clinically, the relative merits of the two modalities at detecting patients with erosive disease are dependent on the anatomy imaged. This work shed light on a number of study-specific and general methodological considerations that may affect the relative diagnostic test accuracy of the two modalities. Further, this research highlighted the importance of accounting for error at the unit of measurement when evaluating the reliability of a composite measure or aggregate score. Some measure-specific limitations

to the application of SDD-adjustment to the unit of measurement were identified. Nonetheless, the application of measurement and generalizability theory to the reliability measurement of these important rheumatology measures represents a substantial advance to the field.

7.7. References

1. O'Dell JR. Chapter 10. Rheumatoid arthritis: clinical aspects. In Koopman WJ, Boulware DW, Heudebert GR. Clinical Primer of Rheumatology. Philadelphia, Lippincott Williams and Wilkins, 2003. pp. 97-115.
2. Schett G. Erosive arthritis. Arthritis Res Ther 2007;9(Suppl 1):S2.
3. Smolen JS, Steiner G. Therapeutic strategies for rheumatoid arthritis. Nat Rev Drug Discov 2003;2:473-88.
4. Scott DL, Smith C, Kingsley G. Joint damage and disability in rheumatoid arthritis: an updated systematic review. Clin Exp Rheumatol 2003;21(5 Suppl 31):S20-7.
5. Scott DL. Radiological progression in established rheumatoid arthritis. J Rheumatol 2004;69(Suppl):55-65.
6. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, Lijmer JG, Moher D, Rennie D, de Vet HCW, for the STARD Group. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. Ann Intern Med 2003;138:40-4.
7. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, Moher D, Rennie D, de Vet HCW, Lijmer JG. The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. Clin Chem 2003;49:7-18.
8. Whiting P, Rutjes AWS, Reitsma JB, Bossuyt PMM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. BMC Med Res Methodol 2003;3:25.
9. Lassere M, McQueen F, Ostergaard M, Conaghan P, Shnier R, Peterfy C, Klarlund M, Bird P, O'Connor P, Stewart N, Emery P, Genant H, Edmonds J. OMERACT Rheumatoid Arthritis Magnetic Resonance Imaging Studies. Exercise 3: An international multicenter reliability study using the RA-MRI Score. J Rheumatol 2003;30:1366-75.
10. Landis R, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33(1):159-74.
11. Bruynesteyn K, Boers M, Kostense P, van der Linden S, van der Heijde D. Deciding on progression of joint damage in paired films of individual patients: smallest detectable difference or change. Ann Rheum Dis 2005;64:179-82.
12. He Q. Estimating the reliability of composite scores. Coventry, Qualifications and Curriculum Authority, 2010. Report no.: Ofqual/10/4703.
13. Robins et al. Handbook of Research Methods in Personality Psychology. New York: The Guilfod Press, 2009. p. 515.
14. Thompson B, Vacha-Haase T. Psychometrics is datametrics: the test is not reliable. Educ Psychol Meas 2000;60:174-95.

15. Kianifard F. Evaluation of Clinimetric Scales: Basic Principles and Methods. Statistician 1994;43:475-82.

16. Lukas C, Landewé R, Fatenejad S, van der Heijde D. Subtle changes in individual joints result in both positive and negative change scores in a patient: results from a clinical trial in patients with rheumatoid arthritis. Ann Rheum Dis 2009;68:1691-5.

17. Ejbjerg BJ, Vestergaard A, Jacobsen S, Thomsen HS, Ostergaard M. The smallest detectable difference and sensitivity to change of magnetic resonance imaging and radiographic scoring of structural joint damage in rheumatoid arthritis finger, wrist, and toe joints: a comparison of the OMERACT Rheumatoid Arthritis Magnetic Resonance Imaging Score applied to different joint combinaions and the Sharp/van der Heijde radiographic score. Arthritis Rheum 2005;52:2300-6.

18. Ejbjerg B, Narvestad E, Rostrup E, Szkudlarek M, Jacobsen S, Thomsen HS, Ostergaard M. Magnetic resonance imaging of wrist and finger joints in healthy subjects occasionally shows changes resembling erosions and synovitis as seen in rheumatoid arthritis. Arthritis Rheum 2004;50:1097-106.

19. Olech E, Crues III JV, Yocum DE, Merrill JT. Bone marrow edema is the most specific finding for rheumatoid arthritis (RA) on noncontrast magnetic resonance imaging of the hands and wrists: a comparison of patients with RA and healthy controls. J Rheumatol 2010;37:265-74.

20. Xie X, Webber CE, Adachi JD, O'Neill J, Inglis D, Bobba RS, Wu H. Quantitative, small bore, 1 Tesla, magnetic resonance imaging of the hands of patients with rheumatoid arthritis. Clin Exp Rheumatol 2008;26:860-5.

21. Tonolli-Serabian I, Poet JL, Dufour M, Carasset S, Mattei JP, Roux H. Magnetic resonance imaging of the writst in rheumatoid arthritis: comparison with other inflammatory joint diseases and control subjects. Clin Rheumatol 1996;15:137-42.

22. Olech E, Freeston JE, Conaghan PG, Hensor EMA, Emery P, Yocum D. Using extremity magnetic resonance imaging to assess and monitor early rheumatoid arthritis: the optimal joint combination to be scanned in clinical practice. J Rheumatol 2008;35:580-3.

23. van der Heijde DMFM, van Leeuwen M, van Riel PLCM, van de Putte LBA. Radiographic progression on radiographs of hand and feet during the first 3 years of rheumatoid arthritis measured according to Sharp's method (van der Heijde modification). J Rheumatol 1995;22:1792-6.

24. Aletaha D, Neogi T, Silman AJ, Funovits J, Felson DT, Bingham III CO, Birnbaum NS, Burmester GR, Bykerk VP, Cohen MD, Combe B, Costenbader KH, Dougados M, Emery P, Ferraccioli G, Hazes JMW, Hobbs K, Huizinga TWJ, Kavanaugh A, Kay J, Kvien TK, Laing T, Pincus T, Smoeln JS, Stanislawska-Biernat E, Symmons D, Tak PP, Upchurch KS, Vencovsky J, Wolfe F, Hawker G. 2010 Rheumatoid Arthritis Classification Criteria: An American College of Rheumatology/European

League Against Rheumatism collaborative initiative. Arthritis Rheum 2010;62:2569-81.

25. Arnett FC, Edworthy SM, Bloch DA, McShane DJ, Fries JF, Cooper NS, Healey LA, Kaplan SR, Liang MH, Luthra HS, Medsger Jr. TA, Mitchell DA, Neustadt DH, Pinals RS, Schaller JG, Sharp JT, Wilder RL, Hunder GG. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. Arthritis Rheum 1988;31:315-24.
26. Ropes MW for A committee of the American Rheumatism Association. Diagnostic criteria for rheumatoid arthritis 1958 revision. Ann Rheum Dis 1959;18:49-53.
27. Dohn UM, Ejbjerg BJ, Court-Payen M, Hasselquist M, Narvestad E, Szkudlarek M, Moller JM, Thomsen HS, Ostergaard M. Are bone erosions detected by magnetic resonance imaging and ultrasonography true erosions? A comparison with computed tomography in rheumatoid arthritis metacarpophalangeal joints. Arthritis Res Ther 2006;8:R110.
28. Duryea J, Jiang Y, Zakharevich M, Genant HK. Neural network based algorithm to quantify joint space width in joints of the hand for arthritis assessment. Med Phys 2000;27:1185-94.
29. Finckh A, de Pablo P, Katz JN, Neumann G, Lu Y, Wolfe F, Duryea J. Performance of an automated computer-based scoring method to assess joint space narrowing in rheumatoid arthritis: a longitudinal study. Arthritis Rheum 2006;54:1444-50.
30. Savnik A, Bliddal H, Nyengaard JR, Thomsen HS. MRI of the arthritic finger joints: synovial membrame volume determination, a manual vs a stereologic method. Eur Radiol 2002;12:94-8.
31. Ostergaard M, Hansen M, Stoltenberg M, Gideon P, Klaarlund M, Jensen KE, Lorenzen I. Magnetic resonance imaging-determined synovial membrame volume as a marker of disease activity and a predictor of progressive joint destruction in the wrists of patients with rheumatoid arthritis. Arthritis Rheum 1999;42:918-29.
32. Emond PD, Choi A, O'Neill J, Xie J, Adachi R, Gordon CL. The development of EERA: Software for assessing rheumatic joint erosions. Can Assoc Rad J 2009;60:63-8.
33. Emond PD, Inglis D, Choi A, Tricta J, Adachi JD, Gordon CL. Volume measurement of bone erosions in magnetic resonance images of patients with rheumatoid arthritis. Magn Reson Med 2012;67:814-23.
34. Dohn UM, Ejbjerg BJ, Hasselquist M, Narvestad E, Court-Paven M, Szkudlarek M, Moller J, Thomsen HS, Ostergaard M. Rheumatoid arthritis bone erosion volumes in CT and MRI: reliability and correlations with erosion scorFes on CT, MRI and radiography. Ann Rheum Dis 2007;66:1388-92.
35. Bird P, Joshua F, Lassere M, Shnier R, Edmonds J. Training and calibration improve inter-reader reliability of joint damage assessment using magnetic resonance

imaging scoring and computerized erosion volume measurement. J Rheumatol 2005;32:1452-8.

36. Bird P, Ejbjerg B, McQueen F, Ostergaard M, Lassere M, Edmonds J. OMERACT rheumatoid arthritis magnetic resonance imaging studies. Exercise 5: an international multicenter reliability study using computerized MRI erosion volume measurements. J Rheumatol 2003;30:1380-4.
37. OsiriX - DICOM Viewer. Available at: http://www.osirix-viewer.com/. Accessed 10 Oct 2010.
38. Streiner DL, Norman GR. Health measurements scales. A practical guide to their development and use. NewYork: Oxford University Press; 1995. pp. 128–43.
39. Lassere MN, van der Heijde D, Johnson KR, Boers M, Edmonds J. Reliability of measures of disease activity and disease damage in rheumatoid arthritis: implications for smallest detectable difference, minimal clinically important difference, and analysis of treatment effects in randomized controlled trials. J Rheumatol 2001;28:892-903.
40. Haavardsholm EA, Østergaard M, Ejbjerg BJ, Kvan NP, Uhlig TA, Lilleås FG, et al. Reliability and sensitivity to change of the OMERACT rheumatoid arthritis magnetic resonance imaging score in a multireader, longitudinal setting. Arthritis Rheum 2005;52: 3860–3867.
41. Conaghan P, Lassere M, Ostergaard M, Peterfy C, McQueen F, O'Connor P, Bird P, Ejbjerg B, Klarlund M, Shnier R, Genant H, Emery P, Edmonds J. OMERACT Rheumatoid Arthritis Magnetic Resonance Imaging Studies. Exercise 4: An international multicenter longitudinal study using the RA-MRI Score. J Rheumatol 2003;30:1376-9
42. Conaghan PG, Ejbjerg B, Lassere M, Bird P, Peterfy P, Emery P, McQueen F, Haavardsholm E, O'Connor P, Edmonds J, Genant H, Ostergaard M. A multicenter reliability study of extremity-magnetic resonance imaging in the longitudinal evaluation of rheumatoid arthritis. J Rheumatol 2007;34:857-8.
43. van der Heijde DM, van Riel PL, Nuver-Zwart IH, Gribnau FW, van de Putte LB. Effects of hydroxychloroquine and sulphasalazine on progression of joint damage in rheumatoid arthritis. Lancet 1989;1:1036-8.
44. Boers M, Verhoeven AC, Markusse HM, van de Laar MAFJ, Weshovens R, van Denderen JC, van Zeben D, Dilkmans BAC, Peeters AJ, Jacobs P, van den Brink HR, Schouten HJA, van der Heijde DMRM, Boonen A, van der Linden S. Randomised comparison of combined step-down prednisolone, methotrexate and sulphasalazine with sulphasalazine alone in early rheumatoid arthritis. Lancet 1999;350:309-18.
45. Klareskog L, van der Heijde D, de Jager JP, Gough A, Kalden J, Malaise M, Mola EM, Pavelka K, Sany J, Settas L, Wajdula J, Pedersen R, Fatenejad S, Sanda M, for

the TEMPO (Trial of Etanercept and Methotrexate with Radiographic Patient Outcomes) study investigators. Lancet 2004;363:675-81.

46. St Clair EW, van der Heijde DM, Smolen JS, Maini RN, Bathon JM, Emery P, Keystone E, Schiff M, Kalden JR, Wang B, Dewoody K, Weiss R, Baker D; Active-Controlled Study of Patients Receiving Infliximab for the Treatment of Rheumatoid Arthritis of Early Onset Study Group. Combination of infliximab and methotrexate therapy for early rheumatoid arthritis: a randomized, controlled trial. Arthritis Rheum 2004;50:3432-43.
47. Duer-Jensen A, Vestergaard A, Dohn UM, Ejbjerg B, Hetland ML, Albrecht-Beste E, Ostergaard M. Detection of rheumatoid arthritis bone erosions by two different dedicated extremity MRI units and conventional radiography. Ann Rheum Dis 2008;67:998-1003.
48. Dohn UM, Ejbjerg BJ, Hasselquist M, Narvestad E, Moller J, Thomsen HS, Ostergaard M. Detection of bone erosions in rheumatoid arthritis wrist joints with magnetic resonance imaging, computed tomography and radiography. Arthritis Res Ther 2008;10:R25.
49. Ejbjerg BJ, Narvestad E, Jacobsen S, Thomsen HS, Ostergaard M. Optimised, low cost, low field dedicated extremity MRI is highly specific and sensitive for synovitis and bone erosions in rheumatoid arthritis wrist and finger joints: comparison with conventional high field MRI and radiography. Ann Rheum Dis 2005;64:1280-7.
50. Klarlund M, Ostergaard M, Gideon P, Sorensen K, Jensen KE, Lorenzen I. Wrist and finger joint MR imaging in rheumatoid arthritis. Acta Radiologica 1999;40:400-9.
51. Ostergaard M, Hansen M, Stoltenberg M, Jensen KE, Szkudlarek M, Pedersen-Zbinden B, Lorenzen I. New radiographic bone erosions in the wrists of patients with rheumatoid arthritis are detectable with magnetic resonance imaging a median of two years earlier. Arthritis Rheum 2003;48:2128-31.
52. McQueen FM. The MRI View of Synovitis and Tenosynovitis in Inflammatory Arthritis. Ann NY Acad Sci 2009;1154:21-34.
53. Klarlund M, Ostergaard M, Jensen KE, Lysgard Madsen J, Skjodt H, Lorenzen I, and the TIRA group. Magnetic resonance imaging, radiography, and scintigraphy of the finger joints: one year follow up of patients with early arthritis. Ann Rheum Dis 2000;59:521-8.
54. Lindegaard H, Vallo J, Horslev-Petersen K, Junker P, Ostergaard M. Low field dedicated magnetic resonance imaging in untreated rheumatoid arthritis of recent onset. Ann Rheum Dis 2001;60:770-6.
55. Szkudlarek M, Klarlund M, Narvestad E, Court-Payen M, Strandberg C, Jensen KE, Thomsen HS, Ostergaard M. Ultrasonography of the metacarpophalangeal and proximal interphalangeal joints in rheumatoid arthritis: a comparison with magnetic

resonance imaging, conventional radiography and clinical examination. Arthritis Res Ther 2006;8:R52.

56. Altman DG, Bland JM. Diagnostic tests 1: sensitivity and specificity. BMJ 1994;308:1552.
57. Altman DG, Bland JM. Diagnostic tests 2: predictive values. BMJ 1994;309:102.
58. Altman DG, Bland JM. Diagnostic tests 3: receiver operating characteristic plots. BMJ 1994;309:188.
59. Deeks JJ, Altman DG. Diagnostic tests 4: likelihood ratios. BMJ 2004;329:168-9.
60. Glas AS, Lijmer JG, Prins MH, Bonsel GJ, Bossuyt PMM. The diagnostic odds ratio: a single indicator of test performance. J Clin Epidemiol 2003;56:1129-35.
61. Schunemann HJ, Oxman AD, Brozek J, Glasziou P, Jaeschke R, Vist GE, Williams Jr. JW, Kunz R, Craig J, Montori VM, Bossuyt P, Guyatt G. GRADE: grading quality of evidence and strength of recommendations for diagnostic tests and strategies. BMJ 2008;336:1106-10.
62. Schunemann HJ, Oxman AD, Brozek J, Glazziou P, Bossuyt P, Chang S, Muti P, Jaeschke R, Guyatt GH. GRADE: assessing the quality of evidence for diagnostic recommendations. EBM 2008;13:162-3.
63. Norman GR, Streiner DL. Biostatistics: The Bare Essentials. Second Edition. Hamilton, B.C. Decker Inc., 2000.

8. Bibliography

Alasaarela E, Suramo I, Tervonen O, Lähde S, Takalo R, Hakala M. Evaluation of humeral head erosions in rheumatoid arthritis: a comparison of ultrasonography, magnetic resonance imaging, computed tomography and plain radiography. Br J Rheumatol 1998;37:1152–6.

Aletaha D, Funovits J, Smolen JS. Physical disability in rheumatoid arthritis is associated with cartilage damage rather than bone destruction. Ann Rheum Dis 2011:70:733-9.

Aletaha D, Neogi T, Silman AJ, Funovits J, Felson DT, Bingham CO, 3rd, Birnbaum NS, Burmester GR, Bykerk VP, Cohen MD, Combe B, Costenbader KH, Dougados M, Emery P, Ferraccioli G, Hazes JM, Hobbs K, Huizinga TW, Kavanaugh A, Kay J, Kvien TK, Laing T, Mease P, Ménard HA, Moreland LW, Naden RL, Pincus T, Smolen JS, Stanislawska-Biernat E, Symmons D, Tak PP, Upchurch KS, Vencovský J, Wolfe F, Hawker G. 2010 Rheumatoid arthritis classification criteria: An American College of Rheumatology/European League Against Rheumatism Collaborative Initiative. Arthritis Rheum 2010;62:2569-81.

Aletaha D, Smolen J. The Simplified Disease Activity Index (SDAI) and the Clinical Disease Activity Index (CDAI): a review of their usefulness and validity in rheumatoid arthritis. Clin Exp Rheumatol 2005;23(Suppl):S100–8.

Altman DG, Bland JM. Diagnostic tests 1: sensitivity and specificity. BMJ 1994;308:1552.

Altman DG, Bland JM. Diagnostic tests 2: predictive values. BMJ 1994;309:102.

Altman DG, Bland JM. Diagnostic tests 3: receiver operating characteristic plots. BMJ 1994;309:188.

American College of Rheumatology Subcommittee on Rheumatoid Arthritis Guidelines. Guidelines for the management of rheumatoid arthritis: 2002 Update. Arthritis Rheum 2002;46:328-46.

Arnett FC, Edworthy SM, Bloch DA, McShane DJ, Fries JF, Cooper NS, Healey LA, Kaplan SR, Liang MH, Luthra HS, Medsger Jr. TA, Mitchell DA, Neustadt DH, Pinals RS, Schaller JG, Sharp JT, Wilder RL, Hunder GG. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. Arthritis Rheum 1988;31:315-24.

Backhaus M, Burmester GR, Sandrock D, Loreck D, Hess D, Scholz A, Blind S, Hamm B, Bollow M. Prospective two year follow up study comparing novel and conventional imaging procedures with arthritic finger joints. Ann Rheum Dis 2002;61:895-905.

Backhaus M, Kamradt T, Sandrock D, Loreck D, Fritz J, Wolf KJ, Raber H, Hamm B, Burmester G-R. Bollow M. Arthritis of the finger joints: a comprehensive approach

comparing conventional radiography, scintigraphy, ultrasound, and contrast-enhanced magnetic resonance imaging. Arthritis Rheum 1999;42:1232-45.

Bathon J, Robles M, Ximenes AC, Nayiager S, Wollenhaupt J, Durez P, Gomez-Reino J, Grassi W, Haraoui B, Shergy W, Park S-H, Genant H, Peterfy C, Becker J-C, Covucci A, Moniz Reed D, Helfrick R, Westhovens R. Sustained disease remission and inhibition of radiographic progression in methotrexate-naïve patients with rheumatoid arthritis and poor prognostic factors treated with abatacept: 2-year outcomes. Ann Rheum Dis 2011;70:1949-56.

Begg CB. Biases in the assessment of diagnostic tests. Stat Med 1987;6:411-23

Bennett PH, Burch TA. New York symposium on population studies in the rheumatic diseases: new diagnostic criteria. Bull Rheum Dis 1967;17:453-8.

Berbaum KS, Franken EA Jr, Dorfman DD, Barloon T, Ell SR, Lu CH, Smith W, Abu-Yousef MM. Tentative diagnoses facilitate the detection of diverse lesions in chest radiographs. Invest Radiol 1986;21:532-9.

Bevington PR, Robinson DK. Data reduction and error analysis of the physical sciences. 3rd edition. Boston, McGraw-Hill, 2003.

Bird P, Ejbjerg B, McQueen F, Ostergaard M, Lassere M, Edmonds J. OMERACT rheumatoid arthritis magnetic resonance imaging studies. Exercise 5: an international multicenter reliability study using computerized MRI erosion volume measurements. J Rheumatol 2003;30:1380-4.

Bird P, Joshua F, Lassere M, Shnier R, Edmonds J. Training and calibration improve inter-reader reliability of joint damage assessment using magnetic resonance imaging scoring and computerized erosion volume measurement. J Rheumatol 2005;32:1452-8.

Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;1:307-10.

Bleeker SE, Moll HA, Steyerberg EW. External validation is necessary in prediction research: A clinical example. J Clin Epidemiol 2003;56:826–832.

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;22:1109-26.

Boers M, Verhoeven AC, Markusse HM, van de Laar MAFJ, Weshovens R, van Denderen JC, van Zeben D, Dilkmans BAC, Peeters AJ, Jacobs P, van den Brink HR, Schouten HJA, van der Heijde DMRM, Boonen A, van der Linden S. Randomised comparison of combined step-down prednisolone, methotrexate and sulphasalazine with sulphasalazine alone in early rheumatoid arthritis. Lancet 1999;350:309-18.

Boers M. Understanding the window of opportunity concept in early rheumatoid arthritis. Arthritis Rheum 2003;48:1771-4.

Boini S, Guillemin F. Radiographic scoring methods as outcome measures in rheumatoid arthritis: properties and advantages. Ann Rheum Dis 2001;60:817-27.

Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, Moher D, Renie D, de Vet HCW, Lijmer JG. The STARD statement of reporting studies of diagnostic accuracy: explanation and elaboration. Clin Chem 2003;49:7-18.

Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glazziou PP, Irwig LM, Lijmer JG, Moher D, Rennie D, de Vet HCW, for the STARD Group. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. Ann Intern Med 2003;138:40-4

Boyesen P, Haavardsholm EA, Ostergaard M, van der Heijde D, Sesseng S, Kvien TK. MRI in early rheumatoid arthritis: synovitis and bone marrow edema are independent predictors of subsequent radiographic progression. Ann Rheum Dis 2011;70:428-33.

Boyle MH, Furlong W, Feeny D, Torrance GW, Hatcher J. Reliability of the Health Utilities Index – Mark III used in the 1991 cycle 6 canadian general social survey health questionnaire. Qual Life Res 1995 4:249-57.

Bracken MB. On stratification, minimization and protection against types 1 and 2 error. J Clin Epidemiol 2001;54:104-5.

Brook A, Corbett M. Radiographic changes in early rheumatoid disease. Ann Rheum Dis 1977;36:71-3.

Brown AK, Conaghan PG, Karim Z, Quinn MA, Ikeda K, Peterfy CG, Hensor E, Wakefield RJ, O'Connor PJ, Emery P. An explanation for the apparent dissociation between clinical remission and continued structural deterioration in rheumatoid arthritis. Arthritis Rheum 2008;58:2958-67.

Bruynesteyn K, Boers M, Kostense P, van der Linden S, van der Heijde D. Deciding on progression of joint damage in paired films of individual patients: smallest detectable difference or change. Ann Rheum Dis 2005;64:179-82.

Bruynesteyn K, van der Heijde D, Boers M, van der Linden S, Lassere M, van der Vleuten C. The Sharp/van der Heijde method out-performed the Larsen/Scott method on the individual patient level in assessing radiographs in early rheumatoid arthritis. J Clin Epidemiol 2004;57:502-12.

Centers for Disease Control and Prevention (CDC). Prevalence of arthritis--United States, 1997. Morb Mortal Wkly Rep. 2001;50:334-6.

Centers for Disease Control and Prevention. National and state medical expenditures and lost earnings attributable to arthritis and other rheumatic conditions - United States, 2003. Morb Mortal Wkly Rep 2007:56:4-7.

Cheung PP, Dougados M, Gossec L. Reliability and ultrasonography to detect synovitis in rheumatoid arthritis: a systematic literature review of 35 studies (1415 patients). Arthritis Care Res 2010;62:323-34.

CML Healthcare Inc. Mississauga, ON. Per test results received in 2010.

Coelho FM, Pinho V, Amaral FA, Sachs D, Costa VV, Rodrigues DH, Vieira AT, Silva TA, Souza DG, Bertini R, Teixeira AL, Teixeira MM. The chemokinereceptors CXCR1/ CXCR2 modulate antigen-induced arthritis by regulating adhesion of neutrophils to the synovial microvasculature. Arthritis Rheum 2008;58:2329–37.

Combe B, Landewe R, Lukas C, Bolosiu HD, Breedveld F, Dougados M, Emery P, Ferraccioli G, Hazes JMW, Klareskog L, Machold K, Martin-Mola E, Nielsen H, Silman A, Smolen J, Yazici H. EULAR recommendations for the management of early arthritis: report of a task force of the European Standing Committee for International Clinical Studies Including Therapeutics (ESCISIT). Ann Rheum Dis 2007;66:34-45.

Conaghan P, Bird P, Ejbjerg B, O'Connor P, Peterfy C, McQueen F, Lassere M, Emery P, Shnier R, Edmonds J, Østergaard M. The EULAR–OMERACT rheumatoid arthritis MRI reference image atlas: the metacarpophalangeal joints. Ann Rheum Dis 2005;64(Suppl 1):i11–i21.

Conaghan P, Edmonds J, Emery P, Genant H, Gibbon W, Klarlund M, Lassere M, McGonagle D, McQueen F, O'Connor P. Peterfy C, Shnier R, Stewart N, Ostergaard M. Magnetic resonance imaging in rheumatoid arthritis: summary of OMERACT activities, current status, and plans. J Rheumatol 2001;28:1158-62.

Conaghan P, Lassere M, Ostergaard M, Peterfy C, McQueen F, O'Connor P, Bird P, Ejbjerg B, Klarlund M, Shnier R, Genant H, Emery P, Edmonds J. OMERACT Rheumatoid Arthritis Magnetic Resonance Imaging Studies. Exercise 4: An international multicenter longitudinal study using the RA-MRI Score. J Rheumatol 2003;30:1376-9.

Conaghan PG, Ejbjerg B, Lassere M, Bird P, Peterfy C, Emery P, McQueen F, Haavardsholm E, O'Connor, P, Edmonds J, Genant H, Ostergaard M. A multicenter reliability study of extremity-magnetic resonance imaging in the longitudinal evaluation of rheumatoid arthritis. J Rheumatol 2007;34:857-8.

Conaghan PG, O'Connor P, McGonagle D, Astin P, Wakefield RJ, Gibbon WW, Quinn M, Karim Z, Green MJ, Proudman S, Isaacs J, Emery P. Elucidation of the relationship between synovitis and bone damage: a randomized magnetic resonance imaging study of individual joints in patients with early rheumatoid arthritis. Arthritis Rheum 2003;48:64–71.

Crues JV, Shellock FG, Dardashti S, James TW, Troum OM. Identification of wrist and metacarpophalangeal joint erosions using a portable magnetici resonance imagines system compared to conventional radiographs. J Rheumatol 2004;31:676-85.

Cyteval C, Miquel A, Hoa D, Daures JP, Mariette X, Combe B. Rheumatoid arthritis of the hand: monitoring with a simplified MR imaging scoring method--preliminary assessment. Radiology 2010;256:863-9.

Deeks JJ, Altman DG. Diagnostic tests 4: likelihood ratios. BMJ 2004;329:168-9.

Delfaut EM, Beltran J, Johnson G, Rousseau J, Marchandise X, Cotten A. Fat suppression in MR imaging: techniques and pitfalls. Radiographics 1999;19:373-82.

Dmitrienko A, Chuang-Stein C, D'Agostino R. Pharmaceutical Statistics Using SAS®. SAS® Press Series. Cary, The SAS Institute, 2007. pp. 228-33.

Dohn UM, Ejbjerg BJ, Court-Payen M, Hasselquist M, Narvestad E, Szkudlarek M, Moller JM, Thomsen HS, Ostergaard M. Are bone erosions detected by magnetic resonance imaging and ultrasonography true erosions? A comparison with computed tomography in rheumatoid arthritis metacarpophalangeal joints. Arthritis Res Ther 2006;8:R110.

Dohn UM, Ejbjerg BJ, Hasselquist M, Narvestad E, Court-Paven M, Szkudlarek M, Moller J, Thomsen HS, Ostergaard M. Rheumatoid arthritis bone erosion volumes in CT and MRI: reliability and correlations with erosion scorFes on CT, MRI and radiography. Ann Rheum Dis 2007;66:1388-92.

Dohn UM, Ejbjerg BJ, Hasselquist M, Narvestad E, Moller J, Thomsen HS, Ostergaard M. Detection of bone erosions in rheumatoid arthritis wrist joints with magnetic resonance imaging, computed tomography and radiography. Arthritis Res Ther 2008;10:R25.

Dore RK, Clements PJ, Fox RI, Furst DE, Kaplan H, Kitridou RC. Guidelines for Rheumatology Referral. Atlanta, American College of Rheumatology, 1996.

Doubilet P, Herman PG. Interepretation of radiographs: effect of clinical history. Am J Roentgenol 1981;137:1055-8.

Duer A, Ostergaard M, Horslev-Petersen K, Vallo J. Magnetic resonance imaging and bone scintigraphy in the differential diagnosis of unclassified arthritis. Ann Rheum Dis 2008;67:48-51.

Duer-Jensen A, Vestergaard A, Dohn UM, Ejbjerg B, Hetland ML, Albrecht-Beste E, Ostergaard M. Detection of rheumatoid arthritis bone erosions by two different dedicated extremity MRI units and conventional radiography. Ann Rheum Dis 2008;67:998-1003.

Dugowson CE, Nelson JL, Koepsell TD. Evaluation of the 1987 revised criteria for rheumatoid arthritis in a cohort of newly diagnosed female patients. Arthritis Rheum 1990;33:1042-1046.

Duryea J, Jiang Y, Zakharevich M, Genant HK. Neural network based algorithm to quantify joint space width in joints of the hand for arthritis assessment. Med Phys 2000;27:1185-94.

Egsmose C, Lund B, Borg G, Pettersson H, Berg E, Brodin U, Trang L. Patients with rheumatoid arthritis benefit from early 2nd line therapy: 5 year followup of a prospective double blind placebo controlled study. J Rheumatol 1995;22:2208-13.

Ejbjerg B, Narvestad E, Rostrup E, Szkudlarek M, Jacobsen S, Thomsen HS, Ostergaard M. Magnetic resonance imaging of wrist and finger joints in healthy subjects

occasionally shows changes resembling erosions and synovitis as seen in rheumatoid arthritis. Arthritis Rheum 2004;50:1097-106.

Ejbjerg BJ, Narvestad E, Jacobsen S, Thomsen HS, Ostergaard M. Optimised, low cost, low field dedicated extremity MRI is highly specific and sensitive for synovitis and bone erosions in rheumatoid arthritis wrist and finger joints: comparison with conventional high field MRI and radiography. Ann Rheum Dis 2005;64:1280-7.

Ejbjerg BJ, Vestergaard A, Jacobsen S, Thomsen HS, Ostergaard M. The smallest detectable difference and sensitivity to change of magnetic resonance imaging and radiographic scoring of structural joint damage in rheumatoid arthritis finger, wrist, and toe joints: a comparison of the OMERACT Rheumatoid Arthritis Magnetic Resonance Imaging Score applied to different joint combinaions and the Sharp/van der Heijde radiographic score. Arthritis Rheum 2005;52:2300-6.

Emery P, Breedveld F, van der Heijde D, Ferraccioli G, Dougados M, Robertson D, Pedersen R, Koenig AS, Freundlich B; Combination of Methotrexate and Etanercept in Early Rheumatoid Arthritis Trial Group. Two-year clinical and radiographic results with combination etanercept-methotrexate therapy versus monotherapy in early rheumatoid arthritis: a two-year, double-blind, randomized study. Arthritis Rheum. 2010;62:674-82. Erratum in: Arthritis Rheum. 2010;62:3005.

Emery P, Breedveld FC, Dougados M, Kalden JR, Schiff MH, Smolen JS. Early referral recommendation for newly diagnosed rheumatoid arthritis: evidence based development of a clinical guide. Ann Rheum Dis 2002;61:290-7.

Emery P, Genovese MC, van Vollenhoven R, Sharp JT, Patra K, Sasso EH. Less radiographic progression with adalimumab plus methotrexate versus methotrexate monotherapy across the spectrum of clinical response in early rheumatoid arthritis. J Rheumatol 2009;36:1429-41.

Emery P. Prognosis in inflammatory arthritis: the value of HLA genotyping and oncological analogy. The Dunlop Dotteridge Lecture. J Rheumatol 1997;24:1436-42.

Emond PD, Choi A, O'Neill J, Xie J, Adachi R, Gordon CL. The development of EERA: Software for assessing rheumatic joint erosions. Can Assoc Rad J 2009;60:63-8.

Emond PD, Inglis D, Choi A, Tricta J, Adachi JD, Gordon CL. Volume measurement of bone erosions in magnetic resonance images of patients with rheumatoid arthritis. Magn Reson Med 2012;67:814-23.

Eyles JL, Hickey MJ, Norman MU, Croker BA, Roberts AW, Drake SF, James WG, Metcalf D, Campbell IK, Wicks IP. A key role for G-CSF-induced neutrophil production and trafficking during inflammatory arthritis. Blood 2008;112:5193–201.

Finckh A, de Pablo P, Katz JN, Neumann G, Lu Y, Wolfe F, Duryea J. Performance of an automated computer-based scoring method to assess joint space narrowing in rheumatoid arthritis: a longitudinal study. Arthritis Rheum 2006;54:1444-50.

Forslind K, Johanson A, Larsson EM, Svensson B. Magnetic resonance imaging of the fifth metatarsophalangeal joint compared with conventional radiography in patients with early rheumatoid arthritis. Scand J Rheumatol 2003;32:131-7.

Forslind K, Larsson EM, Johansson A, Svensson B. Detection of joint pathology by magnetic resonance imaging in patients with early rheumatoid arthritis. Br J Rheumatol 1997;36:683-8.

Freeston JE, Conaghan PG, Dass S, Vital E, Hensor EMA, Stewart S, Emery P. Does extremity-MRI improve erosion detection in severely damaged joints? A study of long-standing rheumatoid arthritis using three imaging modalities. Ann Rheum Dis 2007;66:1538-40.

Fries JF, Spitz PW, Young DY. The dimensions of health outcomes: the health assessment questionnaire, disability and pain scales. J Rheumatol 1982;9:789-93.

Fries JF, Williams CA, Morfeld D, Singh G, Sibley J. Reduction in long-term disability in patients with rheumatoid arthritis by disease-modifying antirheumatic drug-based treatment strategies. Arthritis Rheum 1996;39:616-22.

Gamma-Dynacare Medical Laboratories, Brampton, ON. Per test results received in 2010.

Gaylis NB, Needell SD, Rudensky D. Comparison of in-office magnetic resonance imaging versus conventional radiography in detecting changes in erosions after one year of infliximab therapy in patients with rheumatoid arthritis. Mod Rheumatol 2007;17:273-8.

Genant HK. Methods of assessing radiographic change in rheumatoid arthritis. Am J Med 1983;75:35–47.

Glas AS, Lijmer JG, Prins MH, Bonsel GJ, Bossuyt PMM. The diagnostic odds ratio: a single indicator of test performance. J Clin Epidemiol 2003;56:1129-35.

Grant EP, Picarella D, Burwell T, Delaney T, Croci A, Avitahl N, Humbles AA, Gutierrez-Ramos J-C, Briskin M, Gerard C, Coyle AJ. Essential role for the C5a receptor in regulating the effector phase of synovial infiltration and joint destruction in experimental arthritis. J Exp Med 2002;196:1461–71.

Graudal Jurik AG, de Carvalho A, Graudal HK. Radiographic progression in rheumatoid arthritis: a long-term prospective study of 109 patients. Arthritis Rheum 1998;41:1470-80.

Green MJ, Marzo-Ortega H, McGonagle D, Wakefield R, Proudman S, Conaghan P, Gooi J, Emery P. Persistence of mild early in¯ammatory arthritis: the importance of disease duration, rheumatoid factor and the shared epitope. Arthritis Rheum 1999;42:2184-8.

Grespan R, Fukada SY, Lemos HP, Vieira SM, Napimoga MH, Teixeira MM, Fraser AR, Liew FY, McInnes IB, Cunha FQ. CXCR2-specific chemokines mediate leukotriene

B4-dependent recruitment of neutrophils to inflamed joints in mice with antigen-induced arthritis. Arthritis Rheum 2008;58:2030–40.

Guillemin F, Billot L, Boini S, Gerard N, Ødegaard S, Kvien TK. Reproducibility and sensitivity to change of 5 methods for scoring hand radiographic damage in patients with rheumatoid arthritis. J Rheumatol 2005;32:778-86.

Haavardsholm EA, Ostergaard M, Ejbjerg BJ, Kvan NP, Uhlig TA, Lilleås FG, Kvien TK. Reliability and sensitivity to change of the OMERACT rheumatoid arthritis magnetic resonance imaging score in a multireader, longitudinal setting. Arthritis Rheum 2005;52:3860-7.

Hamer RM. Sample 25031: Compute six intraclass correlation measures. Chapel Hill: Robert M. Hamer, 1990. Available: http://support.sas.com/kb/25/031.html. Accessed 5 Apr 2011.

Harrison B and Symmons D. Early inflammatory polyarthritis: results from the Norfolk Arthritis Register with a review of the literature. II. Outcome at three years. Rheumatology 2000;39:939-49.

He Q. Estimating the reliability of composite scores. Coventry, Qualifications and Curriculum Authority, 2010. Report no.: Ofqual/10/4703.

Holbrook AM (Chair) for Ontario Musculoskeletal Therapy Review Panel. Ontario Treatment Guidelines for Osteoarthritis, Rheumatoid Arthritis, and Acute Musculoskeletal Injury. Guideline. Toronto, Queen's Printer of Ontario, 2000.

Jansen AL, van der Horst-Bruinsma I, van Schaardenburg D, van de Stadt RJ, de Koning MH, Dijkmans BA. Rheumatoid factor and antibodies to cyclic citrullinated Peptide differentiate rheumatoid arthritis from undifferentiated polyarthritis in patients with early arthritis. J Rheumatol 2002;29:2074–6.

Jansen LM, van Schaardenburg D, van der Horst-Bruinsma I, van der Stadt RJ, de Koning MH, Dijkmans BA. The predictive value of anti-cyclic citrullinated peptide antibodies in early arthritis. J Rheumatol 2003;30:1691-5.

Jawaid WB, Crosbie D, Sholton J, Reid DM, Stewart A. Use of digital radiogrammetry in the assessment of joint damage in rheumatoid arthritis. Ann Rheum Dis 2006;65:459-64.

Jimenez-Boj E, Redlich K, Turk B, Hanslik-Schnabel B, Wanivenhaus A, Chott A, Smolen JS, Schett G. Interaction between synovial inflammatory tissue and bone marrow in rheumatoid arthritis. J Immunol 2005;175:2579-88.

Kameda H, Kanbe K, Sato E, Ueki Y, Saito K, Nagaoka S, Hidaka T, Atsumi T, Tsukano M, Kasama T, Shiozawa S, Tanaka Y, Yamanaka H, Takeuchi T. Continuation of methotrexate resulted in better clinical and radiographic outcomes than discontinuation upon starting etanercept in patients with rheumatoid arthritis: 52-week results from the JESMR study. J Rheumatol 2011;38:1585-92.

Kaye JJ, Nance EPJ, Callahan LF, Carroll FE, Winfield AC, Earthman WJ, Phillips KA, Fuchs HA, Pincus T. Observer variation in quantitative assessment of rheumatoid arthritis. Part II. A simplified scoring system. Invest Radiol 1987;22:41–6.

Kellgren JH. Diagnostic criteria for population studies. Bull Rheum Dis 1962;13:291-2.

Kelsey JL, Whittemore AS, Evans AS, Thompson WD. Methods in Observational Epidemiology. 2nd edition. New York, Oxford University Press, 1996. p. 15.

Kianifard F. Evaluation of Clinimetric Scales: Basic Principles and Methods. Statistician 1994;43:475-82.

Kirwan JR. The relationship between synovitis and erosions in rheumatoid arthritis. Br J Rheumatol 1997;36:225–8.

Klareskog L, van der Heijde D, de Jager JP, Gough A, Kalden J, Malaise M, Mola EM, Pavelka K, Sany J, Settas L, Wajdula J, Pedersen R, Fatenejad S, Sanda M, for the TEMPO (Trial of Etanercept and Methotrexate with Radiographic Patient Outcomes) study investigators. Lancet 2004;363:675-81.

Klarlund M, Ostergaard M, Jensen KE, Lysgard JM, Skjodt H, Lorenzen I, TIRA group. Magnetic resonance imaging, radiography and scintigraphy of the finger joints: one year of patients with early arthritis. Ann Rheum Dis 2000;59:521-8.

Klarlund M, Ostergaard M, Gideon P, Sorensen K, Jensen KE, Lorenzen I. Wrist and finger joint MR imaging in rheumatoid arthritis. Acta Radiologica 1999;40:400-9.

Klarlund M, Ostergaard M, Jensen KE, Lysgard Madsen J, Skjodt H, Lorenzen I, and the TIRA group. Magnetic resonance imaging, radiography, and scintigraphy of the finger joints: one year follow up of patients with early arthritis. Ann Rheum Dis 2000;59:521-8.

Kvien TK. Epidemiology and burden of illness of rheumatoid arthritis. Pharmacoeconomics 2004;22(2 Suppl 1):1-12.

Landewe R, van de Heijde D. Is radiographic progression a realistic outcome measure in clinical trials with early inflammatory arthritis? Clin Exp Rheumatol 2003;21(5 Suppl 31):S37-41.

Landewe RBM, Strand V, Conaghan PG, van der Heijde D. Damage and progression on radiographs in individual joints: data from pivotal randomized controlled trials. J Rheumatol 2011;38:2018-22.

Landis R, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159-74.

Larsen A, Dale K, Eek M. Radiographic evaluation of rheumatoid arthritis and related conditions by reference films. Acta Radiol Diagn 1977;18:481–91.

Lassere M, Boers M, van der Heijde D, Boonen A, Edmonds J, Saudan A, Verhoeven AC. Smallest detectable difference in radiological progression. J Rheumatol 1999;26:731-9.

Lassere M, McQueen F, Ostergaard M, Conaghan P, Shnier R, Peterfy C, Klarlund M, Bird P, O'Connor P, Stewart N, Emery P, Genant H, Edmonds J. OMERACT Rheumatoid Arthritis Magnetic Resonance Imaging Studies. Exercise 3: An international multicenter reliability study using the RA-MRI Score. J Rheumatol 2003;30:1366-75.

Lassere M, van der Heijde D, Johnson KR, Boers M, Edmonds J. Reliability of measures of disease activity and disease damage in rheumatoid arthritis: implications for smallest detectable difference, minimal clinically important difference, and analysis of treatment effects in randomized controlled trials. J Rheumatol 2001;28:892-903.

Lassere MN, Bird P. Measurements of rheumatoid arthritis disease activity and damage using magnetic resonance imaging. Truth and discrimination: does MRI make the grade? J Rheumatol 2001;28:1151-7.

Lassere MN, van der Heijde D, Johnson K, Bruynesteyn K, Molenaar E, Boonen A, Verhoeven A, Emery P, Boers M. Robustness and generalizability of smallest detectable difference in radiological progression. J Rheumatol 2001;28:911-3.

Lenth RV. Some practical guidelines for effective sample-size determination. Am Stat 2001;55:187-93.

Lifelabs® Medical Laboratory Services, Toronto, ON. Per test results received in 2010.

Lindegaard H, Vallo J, Horslev-Petersen K, Junker P, Ostergaard M. Low field dedicated magnetic resonance imaging in untreated rheumatoid arthritis of recent onset. Ann Rheum Dis 2001;60:770-6.

Lineker S, Badley E, Charles C, Hart L, Streiner D. Defining morning stiffness in rheumatoid arthritis. J Rheumatol 1999;26(5):1052-7.

Lukas C, Landewé R, Fatenejad S, van der Heijde D. Subtle changes in individual joints result in both positive and negative change scores in a patient: results from a clinical trial in patients with rheumatoid arthritis. Ann Rheum Dis 2009;68:1691-5.

Lukas C, van der Heijde D, Fatenajad S, Landewe R. Repair of erosions occurs almost exclusively in damaged joints without swelling. Ann Rheum Dis 2010;69:851-5.

Machold KP, Nell VP, Stamm TA, Eberl G, Steiner G, Smolen JS. The Austrian Early Arthritis Registry. Clin Exp Rheumatol 2003;21(5 Suppl 31):S113-7.

Maetzel A, Li LC, Pencharz J, Tomlinson G, Bombardier C, Community Hypertension and Arthritits Project Study Team. The economic burdent associated with osteoarthritis, rheumatoid arthritis, and hypertensions. Ann Rheum Dis 2004;63:395-401.

Mason JH, Anderson JJ, Meenan RF, Haralson KM, Lewis-Stevens D, Kaine JL. The rapid assessment of disease activity in rheeumatology (RADAR) questionnaire. Validity and sensitivity to change of a patient self-report measure of joint count and clinical status. Arthritis Rheum 1992;35:156-62.

McGonagle D, Conaghan PG, O'Connor P, Gibbon W, Green M, Wakefield R, Ridgway J, Emery P. The relationship between synovitis and bone changes in early untreated rheumatoid arthritis: a controlled magnetic resonance imaging study. Arthritis Rheum 1999;42:1706–11.

McQueen F, Lassere M, Edmonds J, Conaghan P, Peterfy C, Bird P, O'Connor P, Ejbjerg B, Klarlund M, Stewart N, Emery P, Shnier R, Genant H, Ostergaard M. OMERACT rheumatoid arthritis magnetic resonance imaging studies. Summary of OMERACT 6 MR imaging module. J Rheumatol 2003;30:1387-92.

McQueen F, Østergaard M, Peterfy C, Lassere M, Ejbjerg B, Bird P, et al. Pitfalls in scoring MR images of rheumatoid arthritis wrist and metacarpophalangeal joints. Ann Rheum Dis 2005;64(Suppl I):i48–i55.

McQueen F, Østergaard M, Peterfy C, Lassere M, Ejbjerg B, Bird P, O'Connor P, Genant H, Shnier R, Emery P, Edmonds J, Conaghan P. Pitfalls in scoring MR images of rheumatoid arthritis wrist and metacarpophalangeal joints. Ann Rheum Dis 2005;64(Suppl 1):i48–i55.

McQueen FM, Benton N, Crabbe J, Robinson E, Yeoman S, McLean L, Stewart N. What is the fate of erosions in early rheumatoid arthritis? Tracking individual lesions using x rays and magnetic resonance imaging over the first two years of disease. Ann Rheum Dis 2001;60:859-68.

McQueen FM, Benton N, Perry D, Crabbe J, Robinson E, Yeoman S, McLean L, Stewart N. Bone edema scored on magnetic resonance imaging scans of the dominant carpus at presentation predicts radiographic joint damage of the hands and feet six years later in patients with rheumatoid arthritis. Arthritis Rheum 2003;48:1814-27.

McQueen FM, Stewart N, Crabbe J, Robinson E, Yeoman S, Tan PL, McLean L. Magnetic resonance imaging of the wrist in early rheumatoid arthritis reveals progression of erosions despite clinical improvement. Ann Rheum Dis 1999;58:156–63.

McQueen FM, Stewart N, Crabbe J, Robinson E, Yeoman S, Tan PLJ, McLean L. Magnetic resonance imaging of the wrist in early rheumatoid arthritis reveals a high prevalence of erosions at four months after symptom onset. Ann Rheum Dis 1998;57:350-6.

McQueen FM, Stewart N, Crabbe J, Robinson E, Yeoman S, Tan PLJ, McLean L. Magnetic resonance imaging of the wrist in early rheumatoid arthritis reveals progression of erosions despite clinical improvement. Ann Rheum Dis 1999;58:156-63.

McQueen FM. The MRI View of Synovitis and Tenosynovitis in Inflammatory Arthritis. Ann NY Acad Sci 2009;1154:21-34.

Mohr W, Westerhellweg H, Wessinghage D. Polymorphonuclear granulocytes in rheumatic tissue destruction. III. an electron microscopic study of pmns at the pannus–cartilage junction in rheumatoid arthritis. Ann Rheum Dis 1981;40:396–9.

Mottonen T, Hannonen P, Korpela M, Nissilä M, Kautiainen H, Ilonen J, Laasonen L, Kaipiainen-Seppänen O, Franzen P, Helve T, Koski J, Gripenberg-Gahmberg M, Myllykangas-Luosujärvi R, Leirisalo-Repo M; FIN-RACo Trial Group. FINnish Rheumatoid Arthritis Combination 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;46:894-8.

Muller-Ladner U, Ospelt C, Gay S, Distler O, Pap T. Cells of the synovium in rheumatoid arthritis. Synovial fibroblasts. Arthritis Res Ther 2007;9:223.

Naraghi AM, White LM, Patel C, Tomlinson G, Keystone EC. Comparison of 1.0-T extremity MR and 1.5-T conventional high-field-strength MR in patients with rheumatoid arthritis. Radiology 2009;251:829-37.

Nell VP, Machold KP, Stamm TA, Eberl G, Heinzl H, Uffmann M, Smolen JS, Steiner G. Autoantibody profiling as early diagnostic and prognostic tool for rheumatoid arthritis. Ann Rheum Dis 2005;64:1731-6.

Neumann E, Lefevre S, Zimmerman B, Gay S, Muller-Ladner U. Rheumatoid arthritis progression mediated by activated synovial fibroblasts. Trends Mol Med 2010;16:458-68.

Nishimoto N. Interleukin-6 as a therapeutic target in candidate inflammatory diseases. Clin Pharmacol Ther. 2010;87:483-7.

Nishimura K, Sugiyama D, Kogata Y, Tsuji G, Nakazawa T, Kawano S, Saigo K, Morinobu A, Koshiba M, Kuntz KM, Kamae I, Kumagai S. Meta-analysis: diagnostic accuracy of anti-cyclic citrullinated peptide antibody and rheumatoid factor for rheumatoid arthritis. Ann Intern Med 2007;146:797–808.

Norman GR, Streiner DL. Biostatistics: The Bare Essentials. Second Edition. Hamilton, B.C. Decker Inc., 2000.

Noss EH, Brenner MB. The role and therapeutic implications of fibroblast-like synoviocytes in inflammation and cartilage erosion in rheumatoid arthritis. Immunol Rev 2008;223:252-70.

O'Dell JR. Chapter 10. Rheumatoid arthritis: clinical aspects. In Koopman WJ, Boulware DW, Heudebert GR. Clinical Primer of Rheumatology. Philadelphia, Lippincott Williams and Wilkins, 2003. pp. 97-115.

Olech E, Crues III JV, Yocum DE, Merrill JT. Bone marrow edema is the most specific finding for rheumatoid arthritis (RA) on noncontrast magnetic resonance imaging of the hands and wrists: a comparison of patients with RA and healthy controls. J Rheumatol 2010;37:265-74.

Olech E, Freeston JE, Conaghan PG, Hensor EMA, Emery P, Yocum D. Using extremity magnetic resonance imaging to assess and monitor early rheumatoid arthritis: the optimal joint combination to be scanned in clinical practice. J Rheumatol 2008;35:580-3.

Olga Kuznetsova and Alex Dmitrienko, personal communication, March 1, 2011: Dr. Olga Kuznetsova, Merck & Co., Inc., P.O. Box 2000, RY33-404, Rahway, NJ 07065-0900, Phone: (732) 594-5094, Fax: (732) 594-6075. Email: olga_kuznetsova@merck.com.

OsiriX - DICOM Viewer. Available at: http://www.osirix-viewer.com/. Accessed 10 Oct 2010.

Østergaard M, Conaghan PG, O'Connor P, Szkudlarek M, Klarlund M, Emery P, Peterfy C, Genant H, McQueen FM, Bird P, Lassere M, Ejbjerg B. Reducing invasiveness, duration, and cost of magnetic resonance imaging in rheumatoid arthritis by omitting intravenous contrast injection - Does it change the assessment of inflammatory and destructive joint changes by the OMERACT RAMRIS? J Rheumatol 2009;36:1806-10. Erratum in: J Rheumatol. 2010;37:2198.

Østergaard M, Edmonds J, McQueen F, Peterfy C, Lassere M, Ejbjerg B, Bird P, Emery P, Genant H, Conaghan P. An introduction to the EULAR–OMERACT rheumatoid arthritis MRI reference image atlas. Ann Rheum Dis 2005; 64(Suppl I):i3–i7.

Ostergaard M, Hansen M, Stoltenberg M, Gideon P, Klaarlund M, Jensen KE, Lorenzen I. Magnetic resonance imaging-determined synovial membrame volume as a marker of disease activity and a predictor of progressive joint destruction in the wrists of patients with rheumatoid arthritis. Arthritis Rheum 1999;42:918-29.

Ostergaard M, Hansen M, Stoltenberg M, Jensen KE, Szkudlarek M, Pedersen-Zbinden B, Lorenzen I. New radiographic bone erosions in the wrists of patients with rheumatoid arthritis are detectable with magnetic resonance imaging a median of two years earlier. Arthritis Rheum 2003;48:2128-31.

Østergaard M, Pedersen SJ, Dohn UM. Imaging in rheumatoid arthritis – status and recent advances for magnetic resonance imaging, ultrasonography, computed tomography and conventional radiography. Best Pract Res Clin Rheumatol 2008; 22:1019–44.

Ostergaard M, Peterfy C, Conaghan P, McQueen F, Bird P, Ejbjerg B, Schnier R, O'Connor P, Klarlund M, Emery P, Genant H, Lassere M, Edmonds J. OMERACT Rheumatoid Arthritis Magnetic Resonance Imaging Studies. Core set of MRI acquisitions, joint pathology definitions, and the OMERACT RA-MRI scoring system. J Rheumatol 2003;30:1385-6.

Pham T, Gossec L, Fautrel B, Combe B, Flipo RM, Goupille P, Le Loët X, Mariette X, Puéchal X, Wendling D, Schaeverbeke T, Sibilia J, Sany J, Dougados M. Physical examination and laboratory tests in the management of patients with rheumatoid arthritis: development of recommendations for clinical practice based on published evidence and expert opinion. Joint Bone Spine 2005;72:222-8.

Pierre-Jerome C, Bekkelund SI, Mellgren SI, Torbergsen T, Husby G, Nordstrøm R. The rheumatoid wrist: bilateral MR analysis of the distribution of rheumatoid lesions in axial plan in a female population. Clin Rheumatol 1997;1:80–6.

Pincus T, Bergman MJ, Yazici Y, Hines P, Raghupathi K, Maclean R. An index of only patient-reported outcome measures, routine assessment of patient index data 3 (RAPID3), in two abatacept clinical trials: similar results to disease activity score (DAS28) and other RAPID indices that include physician-reported measures. Rheumatology 2008;47:345-9.

Pincus T, Summey JA, Soraci SA Jr, Wallston KA, Hummon NP. Assessment of tpatient satisfaction in activities of daily living using a modified Standford Health Assessment Questionnaire. Arthritis Rheum 1983;26:1346-53.

Plant MJ, Jones PW, Saklatvala J, Ollier WE, Dawes PT. Patterns of radiological progression in early rheumatoid arthritis: results of an 8 year prospective study. J Rheumatol 1998;25:417–26.

Pocock SJ, Simon R. Sequential traetment assignment with balancing for prognostic factors in the controlled clinical trial. Biometrics 1975;31:103-15.

Pogue JM. Testing for treatment heterogeneity between the individual outcomes within a composite outcome. Open Access Dissertations and Theses. 2012. Paper 6796. http://digitalcommons.mcmaster.ca/opendissertations/6796. Accessed: 2012-Apr-03.

Prevoo ML, van 't Hof MA, Kuper HH, van Leeuwen MA, van de Putte LB, van Riel PL. Modified disease activity scores that include twentyeight–joint counts: development and validation in a prospective longitudinal study of patients with rheumatoid arthritis. Arthritis Rheum 1995;38:44–8.

Public Health Agency of Canada. Life with Arthritis in Canada: A personal and public health challenge. Ottawa, Public Health Agency of Canada, 2010.

Public Health Agency of Canada. Population Health Economics. Ottawa, Public Health Agency of Canada, 2010.

Quinn MA, Emery P. Window of opportunity in early rheumatoid arthritis: possibility of altering the disease process with early intervention. Clin Exp Rheumatol 2003;21(5 Suppl 31):S154-7.

Quinn MA, Green MJ, Conaghan P, Emery P. How do you diagnose rheumatoid arthritis early? Best Pract Res Clin Rheumatol 2001;15:49-66.

Rau R, Wassenberg S, Herborn G, Stucki G, Gebler A. A new method of scoring radiographic change in rheumatoid arthritis. J Rheumatol 1998;25:2094-107.

Robins et al. Handbook of Research Methods in Personality Psychology. New York: The Guilfod Press, 2009. p. 515.

Ropes MW, Bennett GA, Cobb S, Jacox R, Jessar RA. Revision of diagnostic criteria for rheumatoid arthritis. Bull Rheum Dis 1958;9:175-6.

Royston P, Altman DG, Sauerbrei W. Dichotomizing continuous predictors in multiple regression: a bad idea. Stat Med 2006;25:127-41.

Rydahl C, Thomsen HS, Marckmann P. High prevalence of nephrogenic systemic fibrosis in chronic renal failure patients exposed to gadodiamide, a gadolinium-containing magnetic resonance contrast agent. Invest Radiol 2008;43:141-4.

Saag KG, Teng GG, Patkar NM, Anuntiyo J, Finney C, Curtis JR, Paulus HE, Mudano A, Pisu M, Elkins-Melton M, Outman R, Allison JJ, Suarez Almazor M, Bridges SL Jr, Chatham WW, Hochberg M, MacLean C, Mikuls T, Moreland LW, O'Dell J, Turkiewicz AM, Furst DE, American College of Rheumatology. American College of Rheumatology 2008 recommendations for the use of nonbiologic and biologic disease-modifying antirheumatic drugs in rheumatoid arthritis. Arthritis Rheum 2008;59:762-84.

Sackett DL. Bias in analytic research. J Chronic Dis 1979;32:51-63.

Saraux A, Berthelot JM, Chales G, et al. Value of laboratory tests in early prediction of rheumatoid arthritis. Arthritis Rheum 2002;47:155–65.

Savnik A, Bliddal H, Nyengaard JR, Thomsen HS. MRI of the arthritic finger joints: synovial membrame volume determination, a manual vs a stereologic method. Eur Radiol 2002;12:94-8.

Savnik A, Malmskov H, Thomsen HS, Bretlau T, Graff LB, Nielsen H, Danneskiold-Samsoe B, Boesen J, Bliddal H. MRI of the arthritic small joints: comparison of extremity MRI (0.2 T) vs high-field MRI (1.5 T). Eur Radiol 2001;11:1030-8.

Savnik A, Malmskov H, Thomsen HS, Graff LB, Nielsen H, Danneskiold-Samsøe B, Boesen J, Bliddal H. Magnetic resonance imaging of the wrist and finger joints in patients with inflammatory joint diseases. J Rheumatol 2001;28:2193–200.

Savnik A, Malmskov H, Thomsen HS, Graff LB, Nielsen H, Danneskiold-Samsøe B, Boesen J, Bliddal H. MRI of the wrist and finger joints in inflammatory joint diseases at 1-year interval: MRI features to predict bone erosions. Eur Radiol 2002;12:1203-10.

Schedule of Benefits. Physician services under the Health Insurance Act (September 1, 2011). Toronto, Ontario Ministry of Health and Long Term Care, 2011.

Scheel AK, Hermann K-GA, Ohrndorf S, Werner C, Schirmer C, Detert J, Bollow M, Hamm B, Muller GA, Burmester GR, Backhaus M. Prospective 7 year follow up study comparing radiography, ultrasonography, and magnetic resonance imaging in rheumatoid arthritis finger joints. Ann Rheum Dis 2006;65:595-600.

Schett G. Bone Marrow Edema. Ann N Y Acad Sci 2009;1154:35-40.

Schett G. Erosive arthritis. Arthritis Res Ther 2007;9(Suppl 1):S2.

Schoels M, Knevel R, Aletaha D, Bijlsma JWJ, Breedveld FC, Boumpas DT, Burmester G, Combe B, Cutolo M, Dougados M, Emery P, van der Heijde D, Huizinga TWJ, Kalden J, Keystone EC, Kvien TK, Martin-Mola E, Montecucco C, de Wit M, Smolen JS. Evidence for treating rheumatoid arthritis to target: results of a systematic literature search. Ann Rheum Dis 2010;69:638-43.

Schunemann HJ, Oxman AD, Brozek J, Glasziou P, Jaeschke R, Vist GE, Williams Jr. JW, Kunz R, Craig J, Montori VM, Bossuyt P, Guyatt G. GRADE: grading quality of evidence and strength of recommendations for diagnostic tests and strategies. BMJ 2008;336:1106-10.

Schunemann HJ, Oxman AD, Brozek J, Glazziou P, Bossuyt P, Chang S, Muti P, Jaeschke R, Guyatt GH. GRADE: assessing the quality of evidence for diagnostic recommendations. EBM 2008;13:162-3.

Scott DL, Pugner K, Kaarela K, Doyle DV, Woolf A, Holmes J, Hieke K. The links between joint damage and disability in rheumatoid arthritis. Rheumatology 2000;39:122-32.

Scott DL, Smith C, Kingsley G. Joint damage and disability in rheumatoid arthritis: an updated systematic review. Clin Exp Rheumatol 2003;21(5 Suppl 31):S20-7.

Scott DL, Smith C, Kingsley G. What are the consequences of early rheumatoid arthritis for the individual? Best Pract Res Clin Rheumatol 2005;19:117-36.

Scott DL, Symmons DP, Coulton BL, Popert AJ. Long-term outcome of treating rheumatoid arthritis: results after 20 years. Lancet 1987;1:1108-10.

Scott DL. Radiological progression in established rheumatoid arthritis. J Rheumatol 2004;69(Suppl):55-65.

Segami N, Miyamaru M, Nishimura M, Suzuki T, Kaneyama K, Murakami K-I. Does joint effusion on T2 magnetic resonance images reflect synovitis? Part 2. Comparison of contrentration levels of preinflammatory cytokines and total preotein in synovial fluid of the temporomandibular joint with internal dergangements and osteoarthrosis. Oral Surg Oral Med Oral Pathol Oral Radiol Endod 2002;94:515-21.

Segami N, Nishimura M, Kaneyama K, Miyamaru M, Sato J, Murakami K-I. Does joint effusion on T2 magnetic resonance images reflect synovitis? Comparison of arthroscopic findings in internal derangements of the temporomandibular joint. Oral Surg Oral Med Oral Pathol Oral Radiol Endod 2001;92:341-5.

Segami N, Suzuki T, Sato J, Miyamaru M, Nishimura M, Yoshimura H. Does joint effusion on T2 magnetic resonance images reflect synovitis? Part 3. Comparison of histologic findings of arthroscopically obtained synovium in internal derangements of the temporomandibular joint. Oral Surg Oral Med Oral Pathol Oral Radiol Endod 2003;95:761-6.

Sharp JT, Lidsky MD, Collins LC, Moreland J. Methods of scoring the progression of radiologic changes in rheumatoid arthritis: correlation of radiologic, clinical and laboratory abnormalities. Arthritis Rheum 1971;14:706-20.

Sharp JT, Wolfe F, Lassere M, Boers M, van der Heijde D, Larsen A, Paulus H, Rau R, Strand V. Variability of precision in scoring radiographic abnormalities in rheumatoid arthritis by experienced readers. J Rheumatol 2004;31:1062-72.

Sharp JT, Young DY, Bluhm GB, Brook A, Brower AC, Corbett M, Decker JL, Genant HK, Gofton JP, Goodman N, Larsen A, Lidsky MD, Pussila P, Weinstein AS, Barbara N. Weissman BN. How many joints in the hands and wrists should be included in a score of radiologic abnormalities used to assess rheumatoid arthritis? Arthritis Rheum 1985;28:1326-35.

Shellock FG and Kanal E. Safety of magnetic resonance imaging contrast agents. J Magn Reson Imaging 1999;10:477–484.

Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull 1979;86:420-8.

Smolen JS, Steiner G. Therapeutic strategies for rheumatoid arthritis. Nat Rev Drug Discov 2003;2:473-88.

Society for Clinical Data Management. Good clinical data management practices, version 4.0. Milwaukee (Wisconsin): Society for Clinical Data Management, 2005.

Sokka T. Radiographic scoring in rheumatoid arthritis. A short introduction to the methods. Bull NYU Hosp Joint Dis 2008;66:166-8.

St Clair EW, van der Heijde DM, Smolen JS, Maini RN, Bathon JM, Emery P, Keystone E, Schiff M, Kalden JR, Wang B, Dewoody K, Weiss R, Baker D; Active-Controlled Study of Patients Receiving Infliximab for the Treatment of Rheumatoid Arthritis of Early Onset Study Group. Combination of infliximab and methotrexate therapy for early rheumatoid arthritis: a randomized, controlled trial. Arthritis Rheum 2004;50:3432-43.

Stemler SE. A comparison of consensus, consistency, and measurement approaches to estimating interrater reliability. Practical Assessment, Research and Evaluation. 2004;9(4). Accessed April 19, 2012, http://PAREonline.net/getvn.asp?v=9&n=4.

Stockman A, Tait BD, Wolfe R, Brand CA, Rowley MJ, Varney MD, Buchbinder R, Muirden KD. Clinical, laboratory and genetic markers associated with erosions and remission in patients with early inflammatory arthritis: a prospective cohort study. Rheumatol Int 2006;26:500-9.

Streiner DL, Norman GR. Health measurements scales. A practical guide to their development and use. NewYork: Oxford University Press; 1995. pp. 128–43.

Streiner DL, Norman GR. PDQ Epidemiology. 2nd edition. Hamilton, BC Decker Inc, 1998. p. 60.

Stucki G, Liang MH, Stucki S, Bruhlmann P, Michel BA. A self-administered rheumatoid arthritis disease activity index (RADAI) for epidemiologic research. Psychometric properties and correlation with parameters of disease activity. Arthritis Rheum 1995;38:795-8.

Suzuki T, Ito S, Handa S, Kose K, Okamato Y, Minami M, Sugihara M, Horikoshi M, Tsuboi H, Hayashi T, Goto D, Matsumoto I, Sumida T. New low-field extremity

MRI, compacTscan: comparison with whole-body 1.5 T conventional MRI. Mod Rheumatol 2010;20:331-6.

Symmons DP, Bankhead CR, Harrison BJ, Brennan P, Barrett EM, Scott DG, Silman AJ. Blood transfusion, smoking, and obesity as risk factors for the development of rheumatoid arthritis: results from a primary care-based incident case-control study in Norfolk, England. Arthritis Rheum 1997;40:1955-61.

Symmons DP. Epidemiology of rheumatoid arthritis: determinants of onset, persistence and outcome. Best Pract Res Clin Rheumatol 2002;16:707-22.

Symmons DPM. What is rheumatoid arthritis? Br Med Bull 1995;51:243-8.

Szkudlarek M, Klarlund M, Narvestad E, Court-Payen M, Strandberg C, Jensen KE, Thomsen HS, Ostergaard M. Ultrasonography of the metacarpophalangeal and proximal interphalangeal joints in rheumatoid arthritis: a comparison with magnetic resonance imaging, conventional radiography and clinical examination. Arthritis Res Ther 2006;8:R52.

Tak PP, Kalden JR. Advances in rheumatology: new targeted therapeutics. Arthritis Res Ther 2011;13(Suppl 1):S5.

Tavares R. UNCOVER: Understanding Near-term Care Of Very Early Rheumatoid arthritis in a national, multicentre, randomly selected, historical inception cohort of newly diagnosed disease. Toronto, University of Toronto, 2007.

Tavares R, Pope JE, Tremblay J-C, Thorne C, Bykerk VP, Lazovskis J, Blocka KLN, Bell MJ, Lacaille D, Hitchon CA, Fitzgerald AA, Fidler WK, Bookman AAM, Henderson JM, Mosher DP, Sholter DE, Khraishi M, Haraoui B, Chen H, Li X, Laupacis A, Boire G, Tomlinson G, Bombardier C. Early management of newly diagnosed rheumatoid arthritis by Canadian rheumatologists: a national, multicenter, retrospective cohort. J Rheumatol 2011;38:2342-5.

Tavares R, Wells GA, Scarf J, Cagaanan R, Bykerk VP, Bell MJ. Classification of rheumatologic opinion on early inflammatory arthritis: harmonization of a heterogeneous standard. Arthritis Care Res 2010;62:1407-14.

Taves DR. Minimization: a new method of assigning patients to treatment and control groups. Clin Pharmacol Ther 1974;15:443-53.

The EuroQol Group. EuroQol – A new facility for the measurement of health-related quality of life. Health Policy 1990;16:199-208.

The HERA Study Group. A randomized trial of hydroxychloroquine in early rheumatoid arthritis: the HERA Study. Am J Med 1995;98:156-68.

Therneau TM. How many stratification factors are "too many" to use in a randomization plan? Control Clin Trials 1993;14:98-108.

Thompson B, Vacha-Haase T. Psychometrics is datametrics: the test is not reliable. Educ Psychol Meas 2000;60:174-95.

Tonolli-Serabian I, Poet JL, Dufour M, Carasset S, Mattei JP, Roux H. Magnetic resonance imaging of the writst in rheumatoid arthritis: comparison with other inflammatory joint diseases and control subjects. Clin Rheumatol 1996;15:137-42.

United States Department of Health and Human Services, Food and Drug Administration. Guidance for industry: clinical development programs for drugs, devices and biological products for the treatment of rheumatoid arthritis. URL: www.fda.gov. In Paulus HE, Oh M, Sharp JT, Gold RH, Wong WK, Park GS, Bulpitt KJ, for the Western Consortium of Practicing Rheumatologists. Classifying structural joint damage in rheumatoid arthritis as progressive or nonprogressive using composite definition of joint radiographic change. A preliminary proposal. Arthritis Rheum 2004;50:1083-96.

van Aken J, van Dongen H, le Cessie S, Allaart CF, Breedveld FC, Huizinga TW. Comparison of long term outcome of patients with rheumatoid arthritis presenting with undifferentiated arthritis or with rheumatoid arthritis: an observational cohort study. Ann Rheum Dis 2006;65:20-5.

van der Heide A, Jacobs JW, Bijlsma JW, Heurkens AH, van Booma-Frankfort C, van der Veen MJ, Haanen HC, Hofman DM, van Albada-Kuipers GA, ter Borg EJ, Brus HL, Dinant HJ, Kruize AA, Schenk Y. The effectiveness of early treatment with "second-line" antirheumatic drugs. A randomized, controlled trial. Ann Intern Med 1996;124:699-707.

van der Heijde D, Lassere M, Edmonds J, Kirwan J, Strand V, Boers M. Minimal clinically important difference in plain films in RA: group discussions, conclusions, and recommendations. OMERACT Imaging Task Force. J Rheumatol 2001;28:914–7.

van der Heijde D. How to read radiographs according to the Sharp/van der Heijde method. J Rheumatol 1999;26:743-5.

van der Heijde D. How to read radiographs according to the Sharp/van der Heijde method. J Rheumatol 2000;27:261-3.

van der Heijde DM, van Leeuwen MA, van Riel PL, van de Putte LB. Radiographic progression on radiographs of hands and feet during the first 3 years of rheumatoid arthritis measured according to Sharp's method (van der Heijde modification). J Rheumatol 1995;22:1792-6.

van der Heijde DM, van Riel PL, Nuver-Zwart IH, Gribnau FW, van de Putte LB. Effects of hydroxychloroquine and sulphasalazine on progression of joint damage in rheumatoid arthritis. Lancet 1989;1:1036-8.

van der Heijde DM. Joint erosions and patients with early rheumatoid arthritis. Br J Rheumatol 1995;34(Suppl 2):74–8.

van der Heijde DM. Personal correspondence. 2010-Jun-24.

van der Heijde DMFM, van Leeuwen M, van Riel PLCM, van de Putte LBA. Radiographic progression on radiographs of hand and feet during the first 3 years of rheumatoid arthritis measured according to Sharp's method (van der Heijde modification). J Rheumatol 1995;22:1792-6.

van der Helm-van Mil AH, Detert J, le Cessie S, Filer A, Bastian H, Burmester GR, Huizinga TW, Raza K. Validation of a prediction rule for disease outcome in patients with recent onset undifferentiated arthritis: moving towards individualized treatment decision-making. Arthritis Rheum 2008;58:2241-7.

van der Helm-van Mil AH, le Cessie S, van Dongen H, Breedveld FC, Toes RE, Huizinga TW. A prediction rule for disease outcome in patients with recent-onset undifferentiated arthritis: how to guide individual treatment decisions. Arthritis Rheum 2007;56:433-40.

van Gaalen FA, Linn-Rasker SP, van Venrooij WJ, de Jong BA, Breedveld FC, Verweij CL, Toes RE, Huizinga TW. Autoantibodies to cyclic citrullinated peptides predict progression to rheumatoid arthritis in patients with undifferentiated arthritis: a prospective cohort study. Arthritis Rheum 2004;50:709-15.

van Venrooij WJ, Hazes JM, Visser H. Anticitrullinated protein/peptide antibody and its role in the diagnosis and prognosis of early rheumatoid arthritis. Neth J Med 2002;60:383-8.

Visser H, le Cessie S, Vos K, Breedveld FC, Hazes JMW. How to diagnose rheumatoid arthritis early: a prediction model for persistent (erosive) arthritis. Arthritis Rheum 2002;46:357-65.

Wakefield RJ, Gibbon WW, Conaghan PG, O'Connor P, McGonagle D, Pease C, Green MJ, Veale DJ, Isaacs JD, Emery P. The value of sonography in the detection of bone erosions in patients with rheumatoid arthritis: a comparison with conventional radiography. Arthritis Rheum 2000;43:2762–70.

Ward MM, Leigh JP, Fries JF. Progression of functional disability in patients with rheumatoid arthritis. Associations with rheumatology subspecialty care. Arch Intern Med 1993;153:2229-37.

Wener MH. Chapter 7: Diagnostic Laboratory Tests and Imaging. In Robbins L, Burckhardt CS, Hannan MT, deHoratius RJ (editors). Clinical Care in the Rheumatic Diseases. 2nd edition. Atlanta, Association of Rheumatology Health Professionals, 2001.

Whiting P, Rutjes AWS, Reitsma JB, Bossuyt PMM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. BMC Med Res Methodol 2003;3:25.

Wipke BT, Allen PM. Essential role of neutrophils in the initiation and progression of a murine model of rheumatoid arthritis. J Immunol 2001;167:1601–8.

Wittkowski H, Foell D, af Klint E, De Rycke L, De Keyser F, Frosch M, Ulfgren AK, Roth J. Effects of intra-articular corticosteroids and anti-TNF therapy on neutrophil activation in rheumatoid arthritis. Ann Rheum Dis 2007;66:1020–5.

Wright HL, Moots RJ, Bucknall RC, Edwards SW. Neutrophil function in inflammation and inflammatory diseases. Rheumatology 2010;49:1618-31.

Xie X, Webber CE, Adachi JD, O'Neill J, Inglis D, Bobba RS, Wu H. Quantitative, small bore, 1 Tesla, magnetic resonance imaging of the hands of patients with rheumatoid arthritis. Clin Exp Rheumatol 2008;26:860-5.

9. Appendix A: Study protocol

The concepts investigated in this dissertation were incorporated within the design of a double blind randomized controlled trial to compare the effect of biannual peripheral MRI, x-ray and standard of care on pharmacotherapeutic augmentation in early IA and RA over two years. The studies investigated in Chapters 5 and 6 were nested within this trial. The study protocol is detailed in this appendix.

9.1. Study title

A Double Blind Randomized Controlled Trial to Compare the Effect of Biannual Peripheral Magnetic Resonance Imaging, Radiography and Standard of Care on Pharmacotherapeutic Augmentation in Inflammatory Arthritis over Two Years

9.2. Introduction

The preceding chapters illustrated the relationship between erosion detection on magnetic resonance imaging (MRI) and x-ray in rheumatoid arthritis (RA). In the current chapter, the findings were applied prospectively to design a study to test the affect of specific MRI and x-ray diagnostic imaging interventions on pharmacotherapeutic augmentation in inflammatory arthritis (IA) over two years.

Inflammatory arthritis affects approximately 3% of the adult population (1). It is comprised of the following specific diagnoses: ankylosing spondylitis, psoriatic arthritis, reactive arthritis, RA, inflammatory osteoarthritis and undifferentiated inflammatory arthritis. Of these forms, RA is most common with a prevalence ranging from 0.5-2.0% (2-4) and a commonly accepted figure of 1%.

Early in the disease course, IA results in pain, joints swelling and stiffness, fatigue, and functional disability (3,5,6). In the long-term, the majority of IA patients suffer permanent erosive joint damage, increased risks for cardiovascular and systemic comorbidities, and a higher mortality rate (7-9). Joint destruction begins early in the disease course. Radiographic studies have demonstrated that 10-26% of RA patients develop irreversible erosive joint damage within three months from symptom onset (10,11). Within two years from symptom onset, up to 75-90% of patients develop irreversible joint damage (12,13). Various studies have demonstrated that magnetic resonance is more sensitive than radiography at detecting early articular changes leading to and including erosive joint damage (14-16).

Radiographic erosive disease is the classic hallmark of longstanding destructive RA (17,18). The second feature of note in the measurement of radiographic disease progression is joint space narrowing (19,20). Although erosions have been linked to physical deformity and long-term functional declines classically (21), recent data suggest that joint space narrowing may have a greater relative effect on these outcomes in disease subsets (22).

In contrast, the principal pathophysiological features of interest on MRI include synovitis, bone marrow edema, and erosions (23,24). Synovitis is a precursor to bone marrow edema (25). Synovitis and edema predict erosions (23,26,27). Cross-sectionally, erosions on MRI do not correlate well with x-ray erosions (28,29). Erosions on MRI represent bone lesions at an earlier stage of development below the lower limit of

detection of radiography. With time, MRI lesions increase in size to be later detected with radiography (30,31,32).

Limiting disease progression on diagnostic imaging remains an important clinical and regulatory outcome (33). Early pharmacologic treatment of IA with disease modifying anti-rheumatic drugs (DMARD) and biologics has proven to reduce joint damage and disability (34-7). Treatment-to-target strategies have been demonstrated to improve clinical outcomes in patients (38). Yet, recent data demonstrated that erosive disease progression is inadequately controlled despite clinical disease control (16). These data suggest that pharmacotherapeutic treatment decision-making based on clinical and laboratory measures may be supplemented by diagnostic imaging in treatment-to-target algorithms. Despite these findings, diagnostic imaging is infrequently and unsystematically applied to routine disease management for RA (39). Regular disease monitoring with sensitive modalities may help to augment therapy and further delay disease progression.

The primary objective of the study is to investigate the relative value of MRI or x-ray over standard of care as tools to guide DMARD treatment decision-making by rheumatologists for the care of IA. Specifically, the primary objective is to compare the effect biannual monitoring of disease progression with 1) 1.0T extremity MRI of the 2nd to 5th MCPs of the worst-effected hand at baseline, 2) x-rays of both hands and wrists, and 3) standard of care on the frequency of augmenting DMARD treatment decision-making in IA over two years.

A specific set of secondary objectives were defined for the study *a priori*. The effect of imaging-based changes in pharmacological treatment augmentation on changes in diagnostic imaging evidence of disease progression, composite measures of disease activity, function, and quality of life will be determined. The difference in the frequency of disease progression changes between the two active interventions will be determined. To determine effect of study participation bias on the standard of care frequency diagnostic imaging usage in disease management of the study population, the frequency of MRI and x-ray monitoring between participants allocated to standard of care and a matched historical control of patients fulfilling the eligibility criteria over the course of the two years immediately preceding the study will be determined. Health resource utilization (e.g. number of physician visits, hospitalizations) and productivity losses for this population and across study groups will be investigated. Blinded, biannual in-term assessments of these outcomes will also be investigated.

9.3. Methods

9.3.1. Study design

The proposed study is a three-group, double-blind, randomized, controlled trial to compare the effect of 1.0T extremity MRI of the 2nd to 5th MCPs of the worst-effected hand, x-ray of both hands using lateral, anteroposterior, and oblique views, and standard of care on the frequency of DMARD treatment escalation over two years follow-up (**Figure 1**). Intervention allocation will be concealed to study participants by having them undergo MRI and radiographic interventions at all timepoints: one-third will have positive or negative reports of disease progression reported back to the rheumatologist

based on MRI scores; one-third will have radiograph-based results reported; and, one-third will always have negative reports returned to the rheumatologist. All image sets will be anonymized per assessment and scored independently by a trained radiologist using the vdHSS for x-ray (40-2) and a non-contrast modification (43) of the RAMRIS system for MRI (32,44,45).

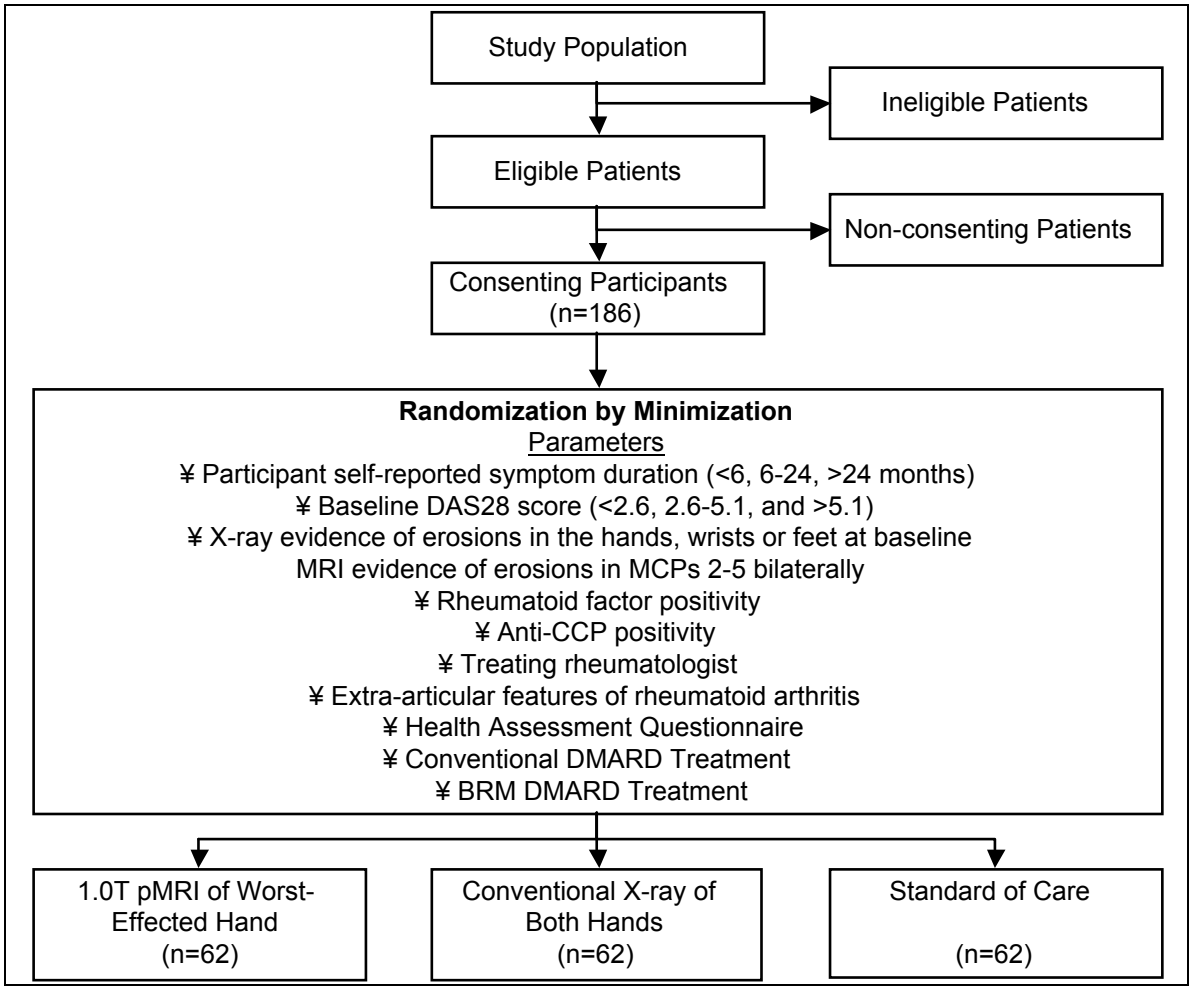


Figure 1. Study Participant Flow Diagram. DAS28 = 28-joint disease activity score; MRI = magnetic resonance imaging; MCP = metacarpophalangeal joint; Anti-CCP = anti-citrullinated cyclic peptide antibodies; DMARD = disease modifying anti-rheumatic drug; BRM = bioloigic response modifier.

Smallest detectable differences in imaging scores will be used to determine disease progression. Disease progression will be determined by an analyst blinded to the intervention allocation. Intervention allocation will be concealed from treating rheumatologists by reporting disease progression as changes greater than the SDD with follow-up disclosure of neither images, nor imaging features.

Participants will be assigned to the three intervention using minimization (46,47). A literature review was used to determine prognostic factors upon which intervention allocation was based. The selection of the anatomy to compare across study interventions was determined using a pragmatic approach taking into account the anatomy most commonly scanned in follow-up in this population (39), and considerations to promote patient compliance: minimize radiation exposure for x-ray, MRI scan time, efficiency (least anatomy required to visualize greatest change in disease progression) (14).

9.3.2. Population

Adults with IA will be recruited for the study. In this study, IA is defined as the fulfillment of the Emery *et al*. (2002) early referral to rheumatology recommendation for RA (48): at least 3 swollen joints, identified by a positive squeeze test; positive squeeze test for either the MCP or metatarsophalangeal (MTP) joints; or, at least 30 minutes of self-reported morning stiffness. These criteria have been found to be strongly associated with rheumatologic opinion on early IA and strongly associated with rheumatologic opinion the initiate DMARD therapy (49). Patients who fulfilled any of the early referral to rheumatology recommendations for RA, who were at least 18 years of age and had a

minimum of six weeks of self-reported symptom duration at enrolment were included in the study.

Study participants will be recruited from the practices of participating rheumatologists at McMaster University, Hamilton, ON, Canada. The study investigators prescreen the patients for eligibility and refer them to study staff for the provision of informed consent and enrolment. Patients fulfilling the following criteria were excluded from the study: a medical history of juvenile idiopathic or inflammatory arthritis; a rheumatologist confirmed primary clinical diagnosis of viral arthritis, a spondyloarthropathy (ankylosing spondylitis, psoriatic arthritis, reactive arthritis, or inflammatory bowel disease); at the investigator's discretion a concomitant condition with medical priority over IA; a concomitant condition that contraindicates treatment with DMARD (not including sulfa allergy or medically, controlled non-terminal liver disease); a psychological deficit or diminished capacity to provide independent, informed consent; a planned event that would prevent scheduled study follow-up; and MRI-specific safety precautions (a surgical history or scheduled procedure involving the introduction of a metallic foreign body, such as joint replacement, stent, pace-maker, neurostimulator, or other metal or electronic implant that could not be verified as 1.0T MRI-safe; an irremovable metallic foreign body with magnetic properties, or known to interfere with MRI; and, a current or historical chronic or high exposure to iron materials).

9.3.3. Interventions

Study participants will attend study visits at weeks 0, 13, 26, 39, 52, 78, and 104 (**Table 1**). All Participants will have MRI and x-ray scans at biannual study visits from weeks 0 to 104. This method will blind participants to treatment allocation. Clinical assessments, prognosis, and response to pharmacologic treatment will be assessed quarterly during the first year and biannually to year two. The increased frequency in the first year is to pickup any subtle escalations in pharmacological treatment. Study participants will be accrued over one year and followed for two.

9.3.3.1. Diagnostic imaging

At week 0 (baseline), all study participants will have an MRI conducted for the 2nd to 5th MCP joints of both hands. All participants will also have x-rays of the hands, wrists and feet conducted at baseline. Participants will be randomized into one of the following three diagnostic imaging modality monitoring intervention arms:

6) 1.0T pMRI of the 2nd to 5th MCP of the worst-affected hand[5]
7) Conventional x-ray of both hands and wrists
8) Standard of care intervention

MRI and x-ray will be conducted within one month of the study visits at weeks 0, 26, 52, 78, and 104. Additional MRI or radiography requisitions may be ordered by the treating rheumatologist per standard of care. If an MRI or x-ray is ordered as part of clinical care, the last study MRI or x-ray conducted will be sent to the treating rheumatologist. Other

[5] Cumulative swollen and tender joint count at baseline will determine the worst-affected hand imaged at weeks 26, 52, and 78.

diagnostic imaging requisitions are unrestricted. When requested, archived radiology reports will be sent to the treating rheumatologist per standard of care. Reports for all diagnostic imaging requisitions ordered by the treating rheumatologist will be blinded to the date of requisition in order to conceal diagnostic imaging modality allocation.

1.1.1.1.1. Non-contrast OMERACT RAMRIS

The RAMRIS is an internationally recognized system for the measuring RA disease changes on MRI (32,45,50-2). The system is comprised of the following components: a definition of the MRI sequences, planes of view, contrast agent requirements, and signal descriptions for each feature (32); scoring system (50); image atlas (45); guidance on pitfalls with the system (51); and, characterizations and validation of the system (52). In the current study, a non-contrast modification of the RAMRIS system will be used to follow disease progression on MRI.

Erosions and edema are measured as described on RAMRIS. Both features are measured per bone included. Erosions are scored 0 to 10 on a decile scale while edema is scored 0 to 3 on a tertile scale. The non-contrast modification pertains to the measurement of synovitis. In the current method, synovitis is detected on MRI without a contrast agent using Fast Spin Echo (FSE), T2-weighted (T2w), fat-saturated (+FS) sequences (43,53-57). The biological plausibility linking the T2w +FS signal and inflammation exists (53-5). Its application for synovitis detection is clinically accepted (43,56). Preliminary comparisons of the contrast and non-contrast performance using the RAMRIS scale indicate the two MRI sequences are scored similarly (56). This latter

study resolved good to excellent concordance between the two methods despite a bias towards demonstrating a difference (56). The original RAMRIS scale for synovitis is unaltered. Consistent with the RAMRIS system, pitfalls in scoring synovitis without contrast are noted (51). Further, a high-quality print of the RAMRIS image atlas was provided to the study radiologists for refence and an analogous image atlas for non-contrast synovitis scoring is under development by our group for application in the study. Using this non-contrast modification of the RAMRIS, the SDD will be determined from the first nine participants entered in the study. Changes in SDD will be monitored over time and adjusted to account for changes in scoring reliability.

1.1.1.1.2. van der Heijde-modified Sharp score

The vdHSS score measures erosions and joint space narrowing on non-continuous, ordinal scales (40-42). In this study, radiologists scoring the diagnostic imaging will be blinded to the identity of the participant, treatment allocation, and radiographs from previous time points. They will be provided high-quality printouts of vdHSS scoring examples to guide scoring and promote consistency over time (42,57). Given the independent evaluation of single radiographs by each rater, the SDD is the appropriate measure for determining real changes vdHSS and will be determined from anonymized, quadruplite scoring of the initial nine study participants.

1.1.1.1.3. Diagnostic image scoring

A designated radiologist will score each MRI and x-ray using the modified RAMRIS and vdHSS methods, respectively. The radiologist score for each diagnostic

imaging timepoint will be compared to the baseline score, or to the score from the previous timepoint on which disease progression was detected. For week 0 MRIs, the modified RAMRIS of the worst-affected will be used as the reference score to determine changes. Baseline total tender and swollen joint count will determine the worst-effected hand. If joint counts are equal bilaterally, the participants self-reported dominant hand will be used to determine disease progression. For x-ray, the vdHSS for both hands and wrists will be compared biannually to determine evidence of disease progression. The radiologist will be blinded to the diagnostic imaging modality allocation of each participant.

Changes in modified RAMRIS and vdHSS scores of at least the SDD will be sent to the treating rheumatologist in the form of blinded diagnostic imaging reports for participants allocated to the MRI and x-ray treatment groups, respectively.

The study-specific SDD will be determined using baseline data for the first nine participants enrolled in the study. For the first nine participants, a SDD value of 3 for RAMRIS (52) and 4 for vdHSS (58) will be used. Upon detection of an increase in RAMRIS or vdHSS of at least the SDD, a standardized, blinded report will be sent to the rheumatologist indicating disease progression for the patient:

> "Disease progression was detected on <date of current MRI/x-ray> relative to <date of previous MRI/x-ray>. This disease progression consists of evolving synovitis, and/or edema, and/or joint space narrowing, and/or erosive damage."

Diagnostic imaging reports sent to the rheumatologist will eliminate references to imaging modality, including reference to either RAMRIS or vdHSS, or imaging features specific to either modality. These considerations will help maintain the study blind.

1.1.1.1.4. Standard of care

All study participants will receive the standard of care for disease management. They will attend clinical assessments quarterly for year one and biannually for year two. A minimum of three months between two time points is required to assess treatment escalation. This window provides sufficient time for slow-acting DMARD to take effect. Pharmacologic treatment will be monitored and escalated over these clinical visits per rheumatologist-dependent standard of care. Pharmacologic management is based on monitoring a combination of clinical, laboratory and diagnostic imaging assessments. This standard of care varies by rheumatologist. As a result, treating rheumatologist is included as a parameter in the minimization treatment allocation scheme. At the treating rheumatologist's medical discretion, intra-articular or systemic corticosteroids may be used as bridge therapy, or per their standard of care. All administered corticosteroid therapy will be documented (parenteral/enteral; concentration; dose; duration; number of courses).

Study rheumatologists will receive baseline MRI and x-ray reports for all study participants. Upon requisition of MRI or x-ray diagnostic imaging, the images and radiology report for the last study-derived diagnostic imaging test will be made available to the treating rheumatologist per standard institutional practice, regardless of treatment

allocation. All radiography or MRI requisitions will be placed through the central study coordinator, thereby eliminating the conduct of redundant diagnostic imaging. Upon requisition, the treating rheumatologist will receive the report associated with the last scheduled study diagnostic imaging intervention. Any additional x-ray or MRI ordered beyond study requirements (additional time points; other joints) and their results will be delivered to the rheumatologist and recorded in the case report form. If no progression occurs between diagnostic imaging intervals, or if the participant is allocated to the standard of care intervention arm, negative diagnostic imaging reports will be reported to the treating rheumatologist.

A number of measures are captured in the study (**Table 1**). These include, but are not limited to, those required to assess the primary and secondary outcomes. The modified RAMRIS and vdHSS for all participants will be recorded centrally. Study investigators will be blinded imaging data. Measures required for the minimization allocation scheme and others commonly assessed as part of standard of care for IA patients will be captured. Much of the data acquired will be collected at patient study visits. Other data will be collected from the treating rheumatologist's clinical chart.

A chart audit will be conducted for the three years preceding study onset. The chart audit will be used to determine the frequency of diagnostic imaging requisition and treatment escalation in historically eligible patients. This chart audit will be conducted for all participating rheumatologists. Measurement of this historical control will provide a measure of contamination bias in the study; to determine if obtaining disease

progression reports for some participants increases the rate of diagnostic image requisition for others (59).

Table 1. Schedule of assessments

Assessment	**Week**						
	0	13	26	39	52	78	104
Informed Consent	X	X	X	X	X	X	X
OHIP Number (each visit)	X	-	-	-	-	-	-
Demographics							
Date of Birth	X	-	-	-	-	-	-
Gender	X	-	-	-	-	-	-
Race/Ethnicity	X	-	-	-	-	-	-
Dominant Hand	X	-	-	-	-	-	-
Household income (socioeconomic status)	X	-	-	-	-	-	-
Pack-years smoked	X	-	-	-	-	-	-
Medical History & Comorbidities							
Surgeries	X	X	X	X	X	X	X
Conditions	X	X	X	X	X	X	X
Medications	X	X	X	X	X	X	X
Clinical Assessment	X	X	X	X	X	X	X
Symptom Duration	X	-	-	-	-	-	-
Morning Stiffness Duration	X	X	X	X	X	X	X
Swollen Joint Count	X	X	X	X	X	X	X
Tender Joint Count	X	X	X	X	X	X	X
Extra-Articular Features	X	X	X	X	X	X	X
Fatigue	X	X	X	X	X	X	X
Pain – Visual Analog Scale	X	X	X	X	X	X	X
Global Physician Impression of Disease Progression	X	X	X	X	X	X	X
Global Patient Impression of Disease Progression	X	X	X	X	X	X	X
Diagnostic Imaging							
Magnetic Resonance Imaging – MCPs 2-5							
Worst-Effected Hand	-	-	X	-	X	X	-
Both Hands	X	-	-	-	-	-	X
Radiography							
Both Hands	X	-	X	-	X	X	X
Both Wrists	X	-	X	-	X	X	X
Both Feet	X	-	-	-	-	-	X
Laboratory Tests							
Rheumatoid Factor	X	O	O	O	O	O	O
Anti-Citrullinated Cyclic Peptide Antibodies(level)	X	-	-	-	-	-	-
Erythrocyte Sedimentation Rate	X	X	X	X	X	X	X
C-Reactive Protein	X	X	X	X	X	X	X
Other	X	X	X	X	X	X	X
Concomitant Medications	X	X	X	X	X	X	X
Adverse Events	X	X	X	X	X	X	X
Health-Related Quality of Life							
Health Assessment Questionnaire	X	-	X	-	X	X	X
Health Utility Index Mark 3 and EQ-5D	X	-	X	-	X	X	X
Health Resource Utilization	X	-	X	-	X	X	X
Work and Leisure Time Productivity	X	-	X	-	X	X	X

X = Assessment conducted; O = Assessment conducted only if preceding assessment negative.

OHIP = Ontario Health Insurance Plan; MCP = metacarpophalangeal joint; EQ-5D = EuroQOL general quality of life instrument.

9.3.4. Outcomes

9.3.4.1. Primary outcome

The primary outcome of the study is the frequency of DMARD augmentations over two years. All changes in DMARD and/or biologic treatment over the course of the study will be recorded. The primary endpoint definition is a change in DMARD agent or increase in dose that occurs on the visit following the date that the disease progression report is delivered to the treating physician. The secondary endpoint definition is a DMARD agent or dose change regardless of the timing of the delivery of the disease progression report.

9.3.4.2. Secondary outcomes

Secondary outcomes include changes in diagnostic imaging (changes in vdHSS, non-contrast modified RAMRIS, number of erosions, and proportion of patients with erosions), disease activity (DAS28 (60), CDAI (61), SDAI (61), RADAI (62), RADAR (63), RAPID1-5 (64)), function (HAQ (65), MDHAQ (66)), quality of life (HUI-III (67), EQ-5D (68), EQ-5D VAS) (68)). Health resource utilization and productivity losses will be determined as adapted from previous work in this population (69). Healthcare resource utilization will be supplemented by linking patients to provincial healthcare databases.

9.3.5. Randomization

Minimization (i.e. baseline adaptive randomization) will be used to allocate participants to one of the three study interventions. Allocation by minimization is recommended for small sample size clinical trials in diseases with many known prognostic markers (46). The minimization procedure can accommodate 10-20 variables without sacrificing statistical power (70-1). Parameters for minimization specific to this study include the following:

1) Participant self-reported symptom duration (<6, 6-24, >24 months)
2) Baseline DAS28 score (<2.6, 2.6-3.2, 3.2-5.1, and >5.1)
3) MRI or x-ray evidence of erosions from the hands, wrists or feet at baseline
4) Rheumatoid factor positivity (positive; negative; not available)
5) Anti-CCP positivity (positive; negative; not available)
6) Extra-articular features of rheumatoid arthritis
7) Health Assessment Questionnaire (HAQ) (<1, ≥1)
8) Conventional DMARD treatment
9) BRM DMARD treatment
10) Treating rheumatologist

With the exception of the treating rheumatologist, these variables are of prognostic value or otherwise informative for RA patient treatment (72). Treating rheumatologist was included as a minimization parameter to balance the effect of varying pharmacological disease management standards of care by rheumatologists across the three study groups.

Minimization was conducted using a SAS macro (**Figure 2**). The generic two-group SAS macro for minimization (73) was adapted into the three group-algorithm courtesy of Dr. Olga Kuznetsova (74).

```
* Macro that executes k-arm minimization;
* A random vector of covariates is generated for testing purposes;
* In a real trial, the vector of covariates of the patient in generation GEN should be read from the real one-observation dataset;
options nomprint;
%macro mink(n=,              /*the number of patients to be allocated*/
            k=3,             /*number of treatment arms*/
            nfact=2,         /*number of factors minimization baalnces upon*/
            lev1=2,          /*number of levels of teh first factor*/
            lev2=2,          /*number of levels of the second factor, add more macr variables Levi if NFACT>2*/
            pnon=0.1,        /*probability to allocate non-preferred treatment*/
            mseed=1971,      /*random seed*/
            outdata=         /*the dataset to store the complete allocation log*/);
*0. generate and store the dataset with the random string of a uniformly distributed variable;
data uni;       do gen=1 to &n;          ranu=ranuni(&mseed);       output;       end;          run;
*proc print data=uni;*run;
*1. generate the sequence of covariates for testing purposes modify code for  allocation in a real trial;
data covs;      n=&n;        do i=1 to n;    gen=i;        %do f=1 %to &nfact;        rand&f=ranuni(3171);
            %do lev=1 %to &&lev&f;    c&f._&lev=(ceil(rand&f*&&lev&f)=&lev);
            %end;        %end;        output;        end;          run;
proc print data=covs;run;
* 2. initiate the treatment totals;
data ttotals;    * marginal totals for each treatment;
%do i=1 %to &k;                      %do f=1 %to &nfact;        %do lev=1 %to &&lev&f;
            t&i._&f._&lev=0;                                    *treatment I total for level LEV of the factor F;
            %end;        %end;        %end;        run;
* initiate dataset to store all allocation records with complete info;
data store;gen=0;run;
proc print data=ttotals;run;
* 3. assign the treatments;
%do gen=1 %to &n;
data assign;    merge covs(where=(gen=&gen) /*covariates of the AN=gen are merged in, edit for real trial*/
            keep=gen     %do f=1 %to &nfact;        %do lev=1 %to &&lev&f;                    c&f._&lev
            %end;        %end; )
ttotals uni(where=(gen=&gen));                                                                 *merge in uniform random variable ranu;
            k=&k;        %do i=1 %to &k;
*calculate the imbalance function assuming that the patient is assigned TRT I;
* imbalance is the sum of the marginal ranges of the number of treatments across the factor levels of the patient in generation &gen;
            %do f=1 %to &nfact;        %do lev=1 %to &&lev&f;                    if c&f._&lev=1 then do;      %do j=1 %to &k;
            %if &j=&i %then %do;       if&i._&f._&j=t&j._&f._&lev+1;
                                                    *treatment I total in the new patient level of Factor F assuming I is allocated;
%end;       %else %do;    if&i._&f._&j=t&j._&f._&lev;
                                                                                    *other treatments total in the new patient level of Factor F;
%end;       %end;         end;         %end;
*now all treatment totals if.. assuming I is assigned are derived ;
*derive the range of trts within the factor F if TRT I is assigned;
            rg&i._&f=range(of if&i._&f._1-if&i._&f._&k);
*derive sum of ranges across all factors - total imbalance for group I;
            totrg&i=sum(of rg&i._1-rg&i._&nfact);      %end;        %end;
*derive minimum of total imbalances across k groups;
            mi=min(of totrg1-totrg&k);
* derive w - number of treatment groups that have total imbalance equal to mi;
            w=0;         %do i=1 %to &k;          if totrg&i=mi then w=w+1;   %end;
if w=k then do;                                  *if all groups have equal total imbalance, they are allocated in equal ratio ;
            %do i=1 %to &k;          p&i=1/k;    %end;        end;          else do;
*if W groups have total imbalance equal to minimum total imbalance MI (preferred groups), each of them are allocated with probability (1-&pnon)/w.  The remaining (K-W) non-
preferred groups are allocated with probability &pnon/(k-w) each;
            %do i=1 %to &k;          if totrg&i=mi then p&i=(1-&pnon)/w;        else p&i=&pnon/(k-w);       %end;        end;
*choose treatment TRT out of 1-K with probabilities p1-pk using uniform random variable RANU from the merged in dataset UNI;
if ranu le p1 then trt=1;else    %do i=2 %to &k;          if ranu le sum(of p1-p&i) then trt=&i;else    %end;;
*update the treatment totals with NEWTOTALS ;
            %do i=1 %to &k;          if &i=trt then do;%do f=1 %to &nfact;       %do lev=1 %to &&lev&f;
                        if c&f._&lev=1 then t&i._&f._&lev=t&i._&f._&lev+1;
*assigned treatment totals are increased by 1 for levels of the Subject GEN;
            %end;%end;end;%end;
*derive the treatment totals;
            %do i=1 %to &k;          tot&i=sum(of t&i._1_1-t&i._1_&lev1);       %end;        run;
*proc print data=assign;*title "gen=&gen"; *run;
* 4. update the dataset ttotals;
data ttotals;                                                                  * marginal totals for each treatment;
            set assign;
            keep         %do i=1 %to &k;          %do f=1 %to &nfact;        %do lev=1 %to &&lev&f;                    t&i._&f._&lev
                         %end;%end;%end;;         run;
*proc print data=ttotals;*title "gen=&gen"; *run;
*5. append ASSIGN to randomiztaion log;
data store;set store assign;if gen>0;run;
*proc print data=store;*title "gen=&gen"; *run;
%end;                                                                                          *end of a single allocation sequence generation;
data &outdata;set store;run;                                   *complete allocation log;
proc print data=&outdata;     title 'randomization log';       run;
%mend mink; %mink(n=30, k=3,nfact=2, lev1=2, lev2=2, pnon=0.1,mseed=1971, outdata=test1);
```

Figure 2. Generic SAS Macro for three-group Pocock and Simon-based (47) minimization intervention treatment allocation (74).

9.3.6. Blinding

Personnel involved with study participant follow-up, care, outcome assessments, or statistical analysis will be blinded to the intervention allocation scheme. Upon the determination of the eligibility, consent, and value of the minimization parameters for the participant, the enrolling study personnel will place a telephone call to the central study coordinator for the randomization of the participant. The central study coordinator will assign an intervention allocation number to the participant, record this number on the randomization scheme and report the number to the caller. The caller will record the randomization number in the participant's case report form.

Study radiologists will be blinded to the identity and intervention allocation of each participant. The MRI and x-ray images will be anonymized such that MRI or x-ray images for the same participant and time point, or images of the same modality for the same participant over time cannot be compared. All radiologist assessments will be reported back to the unblinded central study coordinator. The central coordinator will then process the results of the radiologists' assessments based on the intervention allocation of the participants. In the event of disease progression (i.e. a positive change in disease progression greater than the SDD), blinded results will be reported back to the study rheumatologist.

9.3.7. Allocation concealment

All study participants and the investigative team will be blinded to the intervention allocation scheme. Access to the treatment allocation scheme will be limited

to the central unblinded study coordinator or designate. This coordinator will not be involved with the clinical assessment of study subjects or intervention implementation. Statistical analysis will be blinded to treatment allocation.

9.3.8. Statistical analysis

9.3.8.1. Samples size

A sample size of 186 study participants was estimated from a one-way analysis of variance (ANOVA) statistical test to determine a difference in the rate of pharmacological treatment escalation over two years between three groups with intermediate dispersion of data (75). This parametric test is sufficiently robust to yield sample size estimates for non-parametric data (75). Two-year radiographic progression rates (76) were taken together with one-year MRI and radiographic progression data (77) to estimate testing a between 36% MRI progression of the 2nd to 5th MCP joints of a single hand, 25% radiographic progression of bilateral hands and wrists, and no progression for the standard of care group. Although some patients may progress several-fold over the clinically relevant threshold, progression was conservatively assumed to result in a single pharmacological treatment escalation per affected participant, since a change in therapy is expected to result in a DMARD intervention with a greater protective effect. Despite the liberal inclusion criteria used here, with the current setting characterized by patients with significantly progressed symptom duration, spontaneous remission was estimated at 5%. Missing data was estimated to be 15% and 5% annual attrition was estimated. Given these estimates, a sample size of 62

participants per intervention allocation group is required to determine differences in the mean frequency of treatment escalation with 90% power at a 5% level of significance (75,78). A total of 186 early IA and RA patients will be recruited.

9.3.8.2. Analysis

Differences in the primary and secondary outcomes between the three groups will be determined using non-parametric tests. Pharmacotherapy escalation will be compared using the Cochran-Mantel-Haenszel test. Multi-group differences in continuous or interval data will be determined using the Kruskal-Wallis test. Two groups differences will be tested using the Wilcoxon Rank Sum (Mann-Whitney) test. Correction for multiple comparisons will be conducted using the Tukey method.

The primary analysis population will be intent-to-treat. As-treated and per-protocol study populations will also be reported. Complete and Markov chain Monte Carlo multiple imputed data will be reported. All statistical analyses will be conducted using SAS/STAT version 9.2.

9.3.8.3. Data management

Data management and monitoring will be using iDataFax® (Clinical DataFax Systems Inc., Hamilton, ON).

9.3.8.4. Ethics

The study was conducted in accordance with the principles of the Declaration of Helsinki and governed by the St. Joseph's Healthcare Hamilton Research Ethics Board.

9.4. Discussion

The primary objective of the study is to investigate the relative value of MRI or x-ray over standard of care as tools to guide DMARD pharmacotherapeutic decision-making by rheumatologists for the care of early IA and RA. It is hypothesized that disease progression determined by MRI or x-ray will result in an increased frequency of pharmacologic treatment escalation. Recent evidence suggests that a proportion of RA patients persist with diagnostic imaging evidence of disease progression while in clinical remission (16). Due to the increased Se of MRI to detect disease progression relative to x-ray (14-16), participants allocated to this intervention may have a greater frequency of treatment escalations.

The potential impact of this study is multifold. First, it will determine if there is utility in implementing diagnostic-imaging guided pharmacotherapeutic clinical decision-making at the RA population level. The clinical trial literature is replete with trials demonstrating less than 50% of the population progressing on diagnostic imaging over two-years (79,80). Yet, predicting erosive progression is challenging (81-2). In the current study, a health economic evaluation was integrated within and the design and will contribute to the discussion on the benefits and value of this approach in a sample comprised of early IA and RA.

Second, the relative merits of MRI and radiography as prognosticators in IA will be determined. The increased Se of MRI for erosions that arises from the direct paired-bone comparisons is evident (28,29) but must be tampered by the increased false positive rate for MRI erosions in healthy controls (43,79-81). When the anatomies commonly

imaged in practice are compared, the relative merits are less clear (82-3). The probability of erosion is uniform neither by patient nor joint. Similar to the results reported in Chapter 6, erosive changes were noted in many joints of patients with active disease while others were spared (84). These findings set the expectation that the greater number of bones and joints imaged on radiography may negate the increased sensitivity of MRI in fewer joints. The advantages to imaging a greater number of joints need to be tampered by the increased demand on patients, cost, and data noise (data variability).

Further, the features measured on radiography and MRI include more than just erosions and they differ by modality. In RAMRIS, erosion score accounts for more than 69% of the total potential score (44). The other MRI features measured are more sensitive to change than erosion and account for a smaller percentage of the total potential score: synovitis, the most sensitive to change, accounts for up to 10% of the total RAMRIS; edema accounts for up to 21% of the total score (44). In contrast, the vdHSS measures erosions and joint space narrowing. The former vdHSS feature accounts 57% of the score for bilateral hands and wrists (42). Other factors expected to impact the relative merits of MRI and x-ray as a prognostic tool in IA include the stage of disease, factors affecting image resolution (e.g. magnet strength to bore diameter ratio for MRI; number of views for radiography). Measurement considerations including number or raters involved, the qualification and experience of the rater, number of readings and raters per image set, independent versus paired image evaluations, blinding the chronological order of image acquisitions, and current anti-rheumatic therapy are expected to affect the MRI and x-ray interventions equally. Taken together, there is

sufficient evidence to argue the position of clinical equipoise on the relative prognostic merits of MRI and radiography.

Similarly, it may be argued that treatment decision-making based on disease activity overlaps with imaging changes and that the latter is redundant. The evidence that 19% of RA patients in clinical remission progress on diagnostic imaging over one year (16) may be limited to a small proportion of patients in clinical practice with quiescent disease. For the remainder with active disease, treatment decision-making may be dictated by those disease activity findings. It has been demonstrated that clinical use of disease activity measures is limited (39). In this setting, where the use of intensive clinical management strategies based on formal assessments of disease activity is limited, the potential impact of diagnostic imaging guided care may be magnified.

This study represents an initial investigation into the merits of MRI and x-ray for guiding anti-rheumatic pharmacotherapeutic decision-making for early IA and RA. The work will contribute to the knowledge of how diagnostic imaging may be leveraged to optimize rheumatologic care. Pending findings, future work will investigate the optimization of this intervention strategy, including but not limited to identifying population subsets at greatest risk, adjusting the interval over which imaging is conducted, testing in conjunction with DMARD therapy-controlled, disease-activity based, intensive clinical management strategies.

9.5. Summary

Disease-modifying antirheumatic drugs have varying efficacies in limiting disease progression depending on treatment combination and dose. The revision of DMARD

therapy to maintain low disease activity has grown to become a major clinical management strategy. Yet, many patients continue to progress on diagnostic imaging despite the control of disease activity. The objective of this study was to determine if biannual monitoring of diagnostic imaging, using MRI or x-ray, had any additional impact over standard of care on anti-rheumatic pharmacotherapeutic decision-making over two year for early IA and RA. A two-year, double blind, randomized controlled trial of 186 IA patients will be undertaken. All patients will have an MRI and x-ray conducted biannually. Participants will be allocated into one of three groups of equal size (n=62) with differential diagnostic evidence of disease progression reported to the treating rheumatologist: MRI evidence of disease progression based on the RAMRIS of the 2^{nd}-5^{th} MCPs of the worst-affected at baseline, vdHSS of the hands and wrists, or no diagnostic imaging reported. Intervention allocation will be conducted by minimization. Participants, treating physicians, clinical assessors, radiologists, and analysts will be blinded to intervention allocation. The study will evaluate the relative merits of MRI and x-ray as pharmacotherapeutic decision-making aids for the care of IA. These findings will enable rheumatologists to better leverage MRI and x-ray for the care of this population.

9.6. References

1. Boers M. Understanding the window of opportunity concept in early rheumatoid arthritis. Arthritis Rheum 2003;48(7):1771-4.
2. O'Dell JR. Rheumatoid Arthritis: Clinical Aspects In: William J. Koopman, Dennis W. Boulware, Gustavo Heudebert, editors. Clinical Primer of Rheumatology. Philadelphia: Lippincott Williams and Wilkins; 2003; p. 97-115.
3. Kvien TK. Epidemiology and burden of illness of rheumatoid arthritis. Pharmacoeconomics 2004;22(2 Suppl 1):1-12.
4. Centers for Disease Control and Prevention (CDC). Prevalence of arthritis--United States, 1997. MMWR Morb Mortal Wkly Rep. 2001;50(17):334-6.
5. Ward MM, Leigh JP, Fries JF. Progression of functional disability in patients with rheumatoid arthritis. Associations with rheumatology subspecialty care. Arch Intern Med 1993;153(19):2229-37.
6. Scott DL, Smith C, Kingsley G. What are the consequences of early rheumatoid arthritis for the individual? Best Pract Res Clin Rheumatol 2005;19(1):117-36.
7. Fries JF, Williams CA, Morfeld D, Singh G, Sibley J. Reduction in long-term disability in patients with rheumatoid arthritis by disease-modifying antirheumatic drug-based treatment strategies. Arthritis Rheum 1996;39(4):616-22.
8. Scott DL, Symmons DP, Coulton BL, Popert AJ. Long-term outcome of treating rheumatoid arthritis: results after 20 years. Lancet 1987;1(8542):1108-10.
9. van Aken J, van Dongen H, le Cessie S, Allaart CF, Breedveld FC, Huizinga TW. Comparison of long term outcome of patients with rheumatoid arthritis presenting with undifferentiated arthritis or with rheumatoid arthritis: an observational cohort study. Ann Rheum Dis 2006 Jan;65(1):20-5.
10. Harrison B and Symmons D. Early inflammatory polyarthritis: results from the Norfolk Arthritis Register with a review of the literature. II. Outcome at three years. Rheumatology 2000;39(9):939-49.
11. Machold KP, Nell VP, Stamm TA, Eberl G, Steiner G, Smolen JS. The Austrian Early Arthritis Registry. Clin Exp Rheumatol 2003;21(5 Suppl 31):S113-7.
12. van der Heijde DM, van Leeuwen MA, van Riel PL, van de Putte LB. Radiographic progression on radiographs of hands and feet during the first 3 years of rheumatoid arthritis measured according to Sharp's method (van der Heijde modification). J Rheumatol 1995;22(9):1792-6.
13. Brook A, Corbett M. Radiographic changes in early rheumatoid disease. Ann Rheum Dis 1977;36(1):71-3.
14. Olech E, Freeston JE, Conaghan PG, Hensor EM, Emery P, Yocum D. Using extremity magnetic resonance imaging to assess and monitor early rheumatoid

arthritis: the optimal joint combination to be scanned in clinical practice. J Rheumatol 2008;35(4):580-3.

15. McQueen FM, Benton N, Crabbe J, Robinson E, Yeoman S, McLean L, Stewart N. What is the fate of erosions in early rheumatoid arthritis? Tracking individual lesions using x rays and magnetic resonance imaging over the first two years of disease. Ann Rheum Dis 2001;60(9):859-68.
16. Brown AK, Conaghan PG, Karim Z, Quinn MA, Ikeda K, Peterfy CG, Hensor E, Wakefield RJ, O'Connor PJ, Emery P. An explanation for the apparent dissociation between clinical remission and continued structural deterioration in rheumatoid arthritis. Arthritis Rheum 2008;58(10):2958-67.
17. Ropes MW for a Committee of the American Rheumatism Association. Diagnostic criteria for rheumatoid arthritis: 1958 Revision. Ann Rheum Dis 1959;18:49-53.
18. Arnett FC, Edworthy SM, Bloch DA, McShane DJ, Fries JF, Cooper NS, Healey LA, Kaplan SR, Liang MH, Luthra HS, *et al*. The American Rheumatism Association 1987 Revised Criteria for the Classification of Rheumatoid Arthritis. Arthritis Rheum 1988;31(5):315-24.
19. Sharp JT, Lipsky MD, Collins LS, Moreland J. Method of scoring the progression of radiologic changes in rheumatoid arthritis. Arthritis Rheum 1971;14:706-20.
20. Larsen A, Dale K, Eek M. Radiographic evaluation of rheumatoid arthritis and related conditions by reference films. Acta Radiol Diagn 1977;18:481–91.
21. Scott DL, Smith C, Kingsley G. Joint damage and disability in rheumatoid arthritis: an updated systematic review. Clin Exp Rheumatol 2003;21(5 Suppl 31):S20-7.
22. Aletaha D, Funovits J, Smolen JS. Physical disability in rheumatoid arthritis is associated with cartilate damage rather than bone destruction. Ann Rheum Dis 2011:70:733-9.
23. McGonagle D, Conaghan PG, O'Connor P, Gibbon W, Green M, Wakefield R, Ridgway J, Emery P. The relationship between synovitis and bone changes in early untreated rheumatoid arthritis: a controlled magnetic resonance imaging study. Arthritis Rheum 1999;42(8):1706-11.
24. McQueen FM, Benton N, Perry D, Crabbe J, Robinson E, Yeoman S, McLean L, Stewart N. Bone edema scored on magnetic resonance imaging scans of the dominant carpus at presentation predicts radiographic joint damage of the hands and feet six years later in patients with rheumatoid arthritis. Arthritis Rheum 2003;48(7):1814-27.
25. Jiminez-Boj E, Redlich K, Turk B, Hanslik-Schnabel B, Wanivenhaus A, Chott A, Smolen JS, Schett G. Interaction between synovial inflammatory tissue and bone marrow in rheumatoid arthritis. J Immunol 2005;175:2579-2588.
26. Conaghan PG, O'Connor P, McGonagle D, Astin P, Wakefield RJ, Gibbon WW, Quinn M, Karim Z, Green MJ, Proudman S, Isaacs J, Emery P. Elucidation of the

relationship between synovitis and bone damage. Ann Rheum Dis 2003; 48(1): 64-71.

27. Boyesen P, Haavardsholm EA, Ostergaard M, van der Heijde D, Sesseng S, Kvien TK. MRI in early rheumatoid arthritis: synovitis and bone marrow edema are independent predictors of subsequent radiographic progression. Ann Rheum Dis 2011;70(3):428-33.
28. Backhaus M, Burmester GR, Sandrock D, Loreck D, Hess D, Scholz A, Blind S, Hamm B, Bollow M. Prospective two year follow up study comparing novel and conventional imaging procedures with arthritic finger joints. Ann Rheum Dis 2002;61:895-905.
29. Scheel AK, Hermann K-GA, Ohrndorf S, Werner C, Schirmer C, Detert J, Bollow M, Hamm B, Muller GA, Burmester GR, Backhaus M. Prospective 7 year follow up study comparing radiography, ultrasonography, and magnetic resonance imaging in rheumatoid arthritis finger joints. Ann Rheum Dis 2006;65:595-600.
30. McQueen FM, Stewart N, Crabbe J, Robinson E, Yeoman S, Tan PLJ, McLean L. Magnetic resonance imaging of the wrist in early rheumatoid arthritis reveals progression of erosions despite clinical improvement. Ann Rheum Dis 1999;58:156-63.
31. Conaghan P, Edmonds J, Emery P, Genant H, Gibbon W, Klarlund M, Lassere M, McGonagle D, McQueen F, O'Connor P. Peterfy C, Shnier R, Stewart N, Ostergaard M. Magnetic resonance imaging in rheumatoid arthritis: summary of OMERACT activities, current status, and plans. J Rheumatol 2001;28(5):1158-62.
32. Ostergaard M, Peterfy C, Conaghan P, McQueen F, Bird P, Ejbjerg B, Shnier R, O'Connor P, Klarlund M, Emery P, Genant H, Lassere M, Edmonds J. OMERACT rheumatoid arthritis imaging studies. Core set of MRI acquisitions, joint pathology definitions, and the OMERACT RA-MRI Scoring System. J Rheumatol 2003;30:1385-6.
33. Emery P, Breedveld F, van der Heijde D, Ferraccioli G, Dougados M, Robertson D, Pedersen R, Koenig AS, Freundlich B; Combination of Methotrexate and Etanercept in Early Rheumatoid Arthritis Trial Group. Two-year clinical and radiographic results with combination etanercept-methotrexate therapy versus monotherapy in early rheumatoid arthritis: a two-year, double-blind, randomized study. Arthritis Rheum. 2010;62(3):674-82. Erratum in: Arthritis Rheum. 2010 Oct;62(10):3005.
34. The HERA Study Group. A randomized trial of hydroxychloroquine in early rheumatoid arthritis: the HERA Study. Am J Med 1995;98(2):156-68.
35. Egsmose C, Lund B, Borg G, Pettersson H, Berg E, Brodin U, Trang L. Patients with rheumatoid arthritis benefit from early 2nd line therapy: 5 year followup of a prospective double blind placebo controlled study. J Rheumatol 1995;22(12):2208-13.

36. van der Heide A, Jacobs JW, Bijlsma JW, Heurkens AH, van Booma-Frankfort C, van der Veen MJ, Haanen HC, Hofman DM, van Albada-Kuipers GA, ter Borg EJ, Brus HL, Dinant HJ, Kruize AA, Schenk Y. The effectiveness of early treatment with "second-line" antirheumatic drugs. A randomized, controlled trial. Ann Intern Med 1996;124(8):699-707.

37. Mottonen T, Hannonen P, Korpela M, Nissilä M, Kautiainen H, Ilonen J, Laasonen L, Kaipiainen-Seppänen O, Franzen P, Helve T, Koski J, Gripenberg-Gahmberg M, Myllykangas-Luosujärvi R, Leirisalo-Repo M; FIN-RACo Trial Group. FINnish Rheumatoid Arthritis Combination 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;46(4):894-8.

38. Schoels M, Knevel R, Aletaha D, Bijlsma JWJ, Breedveld FC, Boumpas DT, Burmester G, Combe B, Cutolo M, Dougados M, Emery P, van der Heijde D, Huizinga TWJ, Kalden J, Keystone EC, Kvien TK, Martin-Mola E, Montecucco C, de Wit M, Smolen JS. Evidence for treating rheumatoid arthritis to target: results of a systematic literature search. Ann Rheum Dis 2010;69:638-43.

39. Tavares R, Pope JE, Tremblay J-C, Thorne C, Bykerk VP, Lazovskis J, Blocka KLN, Bell MJ, Lacaille D, Hitchon CA, Fitzgerald AA, Fidler WK, Bookman AAM, Henderson JM, Mosher DP, Sholter DE, Khraishi M, Haraoui B, Chen H, Li X, Laupacis A, Boire G, Tomlinson G, Bombardier C. Early management of newly diagnosed rheumatoid arthritis by Canadian rheumatologists: a national, multicenter, retrospective cohort. J Rheumatol 2011;38:2342-5.

40. van der Heijde DM, van Riel PL, Nuver-Zwart IH, Gribnau FW, van de Putte LB. Effects of hydroxychloroquine and sulphasalazine on progression of joint damage in rheumatoid arthritis. Lancet 1989;1(8646):1036-8.

41. van der Heijde D. How to read radiographs according to the Sharp/van der Heijde method. J Rheumatol 1999;26:743-5.

42. van der Heijde D. How to read radiographs according to the Sharp/van der Heijde method. J Rheumatol 2000;27:261-3.

43. Xie X, Webber CE, Adachi JD, O'Neill J, Inglis D, Bobba RS, Wu H. Quantitative, small bore, 1 Tesla, magnetic resonance imaging of the hands of patients with rheumatoid arthritis. Clin Exp Rheumatol 2008;26(5):860-5.

44. Østergaard M, Edmonds J, McQueen F, Peterfy C, Lassere M, Ejbjerg B, et al. An introduction to the EULAR–OMERACT rheumatoid arthritis MRI reference image atlas. Ann Rheum Dis 2005; 64(Suppl I):i3–i7.

45. Conaghan P, Bird P, Ejbjerg B, O'Connor P, Peterfy C, McQueen F, et al. The EULAR–OMERACT rheumatoid arthritis MRI reference image atlas: the metacarpophalangeal joints. Ann Rheum Dis 2005;64(Suppl I):i11–i21.

46. Taves DR. Minimization: a new method of assigning patients to treatment and control groups. Clin Pharmacol Ther 1974;15(5):443-53.

47. Pocock SJ, Simon R. Sequential traetment assignment with balancing for prognostic factors in the controlled clinical trial. Biometrics 1975 ;31 :103-15.

48. Emery P, Breedveld FC, Dougados M, Kalden JR, Schiff MH, Smolen JS. Early referral recommendation for newly diagnosed rheumatoid arthritis: evidence based development of a clinical guide. Ann Rheum Dis 2002;61(4):290-7.

49. Tavares R, Wells GA, Scarf J, Cagaanan R, Bykerk VP, Bell MJ. Classification of rheumatologic opinion on early inflammatory arthritis: harmonization of a heterogeneous standard. Arthritis Care Res 2010;62(10):1407-14.

50. Østergaard M, Edmonds J, McQueen F, Peterfy C, Lassere M, Ejbjerg B, Bird P, Emery P, Genant H, Conaghan P. An introduction to the EULAR–OMERACT rheumatoid arthritis MRI reference image atlas. Ann Rheum Dis 2005; 64(Suppl I):i3–i7.

51. McQueen F, Østergaard M, Peterfy C, Lassere M, Ejbjerg B, Bird P, O'Connor P, Genant H, Shnier R, Emery P, Edmonds J, Conaghan P. Pitfalls in scoring MR images of rheumatoid arthritis wrist and metacarpophalangeal joints. Ann Rheum Dis 2005;64(Suppl I):i48–i55.

52. Haavardsholm EA, Ostergaard M, Ejbjerg BJ, Kvan NP, Uhlig TA, Lilleås FG, Kvien TK. Reliability and sensitivity to change of the OMERACT rheumatoid arthritis magnetic resonance imaging score in a multireader, longitudinal setting. Arthritis Rheum 2005;52(12):3860-7.

53. Segami N, Nishimura M, Kaneyama K, Miyamaru M, Sato J, Murakami K-I. Does joint effusion on T2 magnetic resonance images reflect synovitis? Comparison of arthroscopic findings in internal derangements of the temporomandibular joint. Oral Surg Oral Med Oral Pathol Oral Radiol Endod 2001;92:341-5.

54. Segami N, Miyamaru M, Nishimura M, Suzuki T, Kaneyama K, Murakami K-I. Does joint effusion on T2 magnetic resonance images reflect synovitis? Part 2. Comparison of contrentration levels of preinflammatory cytokines and total preotein in synovial fluid of the temporomandibular joint with internal dergangements and osteoarthrosis. Oral Surg Oral Med Oral Pathol Oral Radiol Endod 2002;94:515-21.

55. Segami N, Suzuki T, Sato J, Miyamaru M, Nishimura M, Yoshimura H. Does joint effusion on T2 magnetic resonance images reflect synovitis? Part 3. Comparison of histologic findings of arthroscopically obtained synovium in internal derangements of the temporomandibular joint. Oral Surg Oral Med Oral Pathol Oral Radiol Endod 2003;95:761-6.

56. Østergaard M, Conaghan PG, O'Connor P, Szkudlarek M, Klarlund M, Emery P, Peterfy C, Genant H, McQueen FM, Bird P, Lassere M, Ejbjerg B. Reducing invasiveness, duration, and cost of magnetic resonance imaging in rheumatoid arthritis by omitting intravenous contrast injection - Does it change the assessment of inflammatory and destructive joint changes by the OMERACT RAMRIS? J Rheumatol 2009;36(8):1806-10. Erratum in: J Rheumatol. 2010;37(10):2198.

57. van der Heijde DM. Personal correspondence. 2010-Jun-24.

58. Landewe R, van de Heijde D. Is radiographic progression a realistic outcome measure in clinical trials with early inflammatory arthritis? Clin Exp Rheumatol 2003;21(5 Suppl 31):S37-41.

59. Sackett DL. Bias in analytic research. J Chronic Dis 1979;32:51-63.

60. Prevoo ML, van 't Hof MA, Kuper HH, van Leeuwen MA, van de Putte LB, van Riel PL. Modified disease activity scores that include twentyeight–joint counts: development and validation in a prospective longitudinal study of patients with rheumatoid arthritis. Arthritis Rheum 1995;38:44–8.

61. Aletaha D, Smolen J. The Simplified Disease Activity Index (SDAI) and the Clinical Disease Activity Index (CDAI): a review of their usefulness and validity in rheumatoid arthritis. Clin Exp Rheumatol 2005;23 Suppl:S100–8.

62. Stucki G, Liang MH, Stucki S, Bruhlmann P, Michel BA. A self-administered rheumatoid arthritis disease activity index (RADAI) for epidemiologic research. Psychometric properties and correlation with parameters of disease activity. Arthritis Rheum 1995;38(6):795-8.

63. Mason JH, Anderson JJ, Meenan RF, Haralson KM, Lewis-Stevens D, Kaine JL. The rapid assessment of disease activity in rheeumatology (RADAR) questionnaire. Validity and sensitivity to change of a patient self-report measure of joint count and clinical status. Arthritis Rheum 1992;35(2):156-62.

64. Pincus T, Bergman MJ, Yazici Y, Hines P, Raghupathi K, Maclean R. An index of only patient-reported outcome measures, routine assessment of patient index data 3 (RAPID3), in two abatacept clinical trials: similar results to disease activity score (DAS28) and other RAPID indices that include physician-reported measures. Rheumatology 2008;47(3):345-9.

65. Fries JF, Spitz PW, Young DY. The dimensions of health outcomes: the health assessment questionnaire, disability and pain scales. J Rheumatol 1982;9(5):789-93.

66. Pincus T, Summey JA, Soraci SA Jr, Wallston KA, Hummon NP. Assessment of patient satisfaction in activities of daily living using a modified Standford Health Assessment Questionnaire. Arthritis Rheum 1983;26(11):1346-53.

67. Boyle MH, Furlong W, Feeny D, Torrance GW, Hatcher J. Reliability of the Health Utilities Index – Mark III used in the 1991 cycle 6 canadian general social survey health questionnaire. Qual Life Res 1995 4(3):249-57.

68. The EuroQol Group. EuroQol – A new facility for the measurement of health-related quality of life. Health Policy 1990;16(3):199-208.

69. Maetzel A, Li LC, Pencharz J, Tomlinson G, Bombardier C, Community Hypertension and Arthritits Project Study Team. The economic burdent associated with osteoarthritis, rheumatoid arthritis, and hypertensions. Ann Rheum Dis 2004;63(4):395-401.

70. Therneau TM. How many stratification factors are "too many" to use in a randomization plan? Control Clin Trials 1993;14(2):98-108.

71. Bracken MB. On stratification, minimization and protection against types 1 and 2 error. J Clin Epidemiol 2001;54(1):104-5.

72. Saag KG, Teng GG, Patkar NM, Anuntiyo J, Finney C, Curtis JR, Paulus HE, Mudano A, Pisu M, Elkins-Melton M, Outman R, Allison JJ, Suarez Almazor M, Bridges SL Jr, Chatham WW, Hochberg M, MacLean C, Mikuls T, Moreland LW, O'Dell J, Turkiewicz AM, Furst DE, American College of Rheumatology. American College of Rheumatology 2008 recommendations for the use of nonbiologic and biologic disease-modifying antirheumatic drugs in rheumatoid arthritis. Arthritis Rheum 2008;59(6):762-84.

73. Dmitrienko A, Chuang-Stein C, D'Agostino R. Pharmaceutical Statistics Using SAS®. SAS® Press Series. Cary, The SAS Institute, 2007. pp. 228-33.

74. Olga Kuznetsova and Alex Dmitrienko, personal communication, March 1, 2011: Dr. Olga Kuznetsova, Merck & Co., Inc., P.O. Box 2000, RY33-404, Rahway, NJ 07065-0900, Phone: (732) 594-5094, Fax: (732) 594-6075. Email: olga_kuznetsova@merck.com.

75. Norman, GR, Streiner, DL. Biostatistics: the bare essentials. 2nd ed. Lewiston, NY: B.C. Decker; 2000.

76. Emery P, Genovese MC, van Vollenhoven R, Sharp JT, Patra K, Sasso EH. Less radiographic progression with adalimumab plus methotrexate versus methotrexate monotherapy across the spectrum of clinical response in early rheumatoid arthritis. J Rheumatol 2009;36:1429-41.

77. Ejbjerg J, Vestergaard A, Jacobsen S, Thomsen HS, Østergaard M. The smallest detectable difference and sensitivity to change of magnetic resonance imaging and radiographic scoring of structural joint damage in rheumatoid arthritis finger, wrist, and toe joints: a comparison of the OMERACT rheumatoid arthritis magnetic resonance imaging score applied to different joint combinations and the Sharp/van der Heijde radiographic score. Arthritis Rheum 2005;52(8):2300-6.

78. Lenth RV. Some practical guidelines for effective sample-size determination. Am Stat 2001;55(3):187-93.

79. Ejbjerg B, Narvestad E, Rostrup E, Szkudlarek M, Jacobsen S, Thomsen HS, Ostergaard M. Magnetic resonance imaging of wrist and finger joints in healthy subjects occasionally shows changes resembling erosions and synovitis as seen in rheumatoid arthritis. Arthritis Rheum 2004;50(40):1097-106.

80. Olech E, Crues III JV, Yocum DE, Merrill JT. Bone marrow edema is the most specific finding for rheumatoid arthritis (RA) on noncontrast magnetic resonance imaging of the hands and wrists: a comparison of patients with RA and healthy controls. J Rheumatol 2010;37:265-74.

81. Tonolli-Serabian I, Poet JL, Dufour M, Carasset S, Mattei JP, Roux H. Magnetic resonance imaging of the writst in rheumatoid arthritis: comparison with other inflammatory joint diseases and control subjects. Clin Rheumatol 1996;15:137-42.

82. Olech E, Freeston JE, Conaghan PG, Hensor EMA, Emery P, Yocum D. Using extremity magnetic resonance imaging to assess and monitor early rheumatoid arthritis: the optimal joint combination to be scanned in clinical practice. J Rheumatol 2008;35(4):580-3.

83. Ejbjerg BJ, Vestergaard A, Jacobsen S, Thomsen HS, Ostergaard M. The smallest detectable difference and sensitivity to change of magnetic resonance imaging and radiographic scoring of structural joint damage in rheumatoid arthritis finger, wrist, and toe joints: a comparison of the OMERACT Rheumatoid Arthritis Magnetic Resonance Imaging Score applied to different joint combinations and the Sharp/van der Heijde radiographic score. Arthritis Rheum 2005;52(8):2300-6.

84. Landewe RBM, Strand V, Conaghan PG, van der Heijde D. Damage and progression on radiographs in individual joints: data from pivotal randomized controlled trials. J Rheumatol 2011;38(9):2018-22.

10. Appendix B: Systematic review search strategy

The literature search strategy for the systematic review described in Chapter 4 is summarized below.

10.1. Participants

Rheumatoid arthritis was the population of interest, however, early in the disease course, RA often presents with insidious, non-specific symptoms. Other inflammatory arthritides often exist in the differential diagnosis. For this reason, the population search string included generalized inflammatory arthritis terms. The following terms were used in population search string: exp Arthritis, Rheumatoid/ or exp Arthritis, Psoriatic/ or exp Spondylitis, Ankylosing/ or exp Arthritis, Reactive/ or exp Arthritis/ or (arthr$ or oligoarthritis or polyarthritis or monoarthritis or ankylosing spondylitis or rheum$).mp..

10.2. Index tests

With knowledge that erosions detected on MRI are more prevalent than those detected on x-ray, the review was structure to include x-ray as the index test or PICO "intervention" and MRI as the reference standard or "comparator". The search terms included, exp Radiography/ or exp X-Rays/ or radiography.mp. or x-ray.mp..

10.3. Target condition

The target condition, or PICO "outcome", was bone erosion. The following search terms were used detect citations pertaining to bone erosions. The search terms were designed to capture both erosion and the anatomical location: (exp "Bone and Bones"/ or (bone or erosi$ or damage or destructi$).mp.) and (exp Joints/ or exp Tarsal

Joints/ or exp Carpometacarpal Joints/ or exp Hand Joints/ or exp Carpal Joints/ or exp Foot Joints/ or exp Sternocostal Joints/ or joint$.mp.).

10.4. Reference standard

The reference standard, or PICO "comparator", search string included the following terms: exp Magnetic Resonance Imaging/ exp Magnetic Resonance Spectroscopy/ or (magnetic resonance imaging or nuclear magnetic resonance).mp..

11. Appendix C: MRI and x-ray technical specifications

The characteristics of the MR and x-ray image acquisitions were summarized below. The MRI technical specifications were summarized in **Table C1**.

Table C1. ONI 1.0T OrthOne MRI scanner (GE Healthcare, Wilmington, MA) study sequences.

Feature	Erosion	Erosion	Edema	Synovitis
Sequence Type	3D gradient echo	Fast spin echo	Fast spin echo	Fast spin echo
Orientation	coronal	axial	coronal	Axial
Repetition time (TR)	60	470	4000	2500
Echo time (TE)	6.6	15.1	15.1	40
Fat saturation	no	no	no	yes
Inversion recovery	no	no	yes	yes
Thickness, mm	1	2	2	2
Interslice gap, mm	0	0	0	0
Number of slices	40	16	16	18
Field of view, mm	140	110	110	110
Frequency	280	256	256	256
Phase	140	192	192	160
Minimum TE	yes	yes	yes	no
Number of excitations	1	2	1	2
Frequency direction	H/F	L/R	H/F	L/R
Flip angle	60	90	90	90
Bandwidth	50	35	35	25
Echo train	1	2	8	4
Number of echoes	1	1	1	1

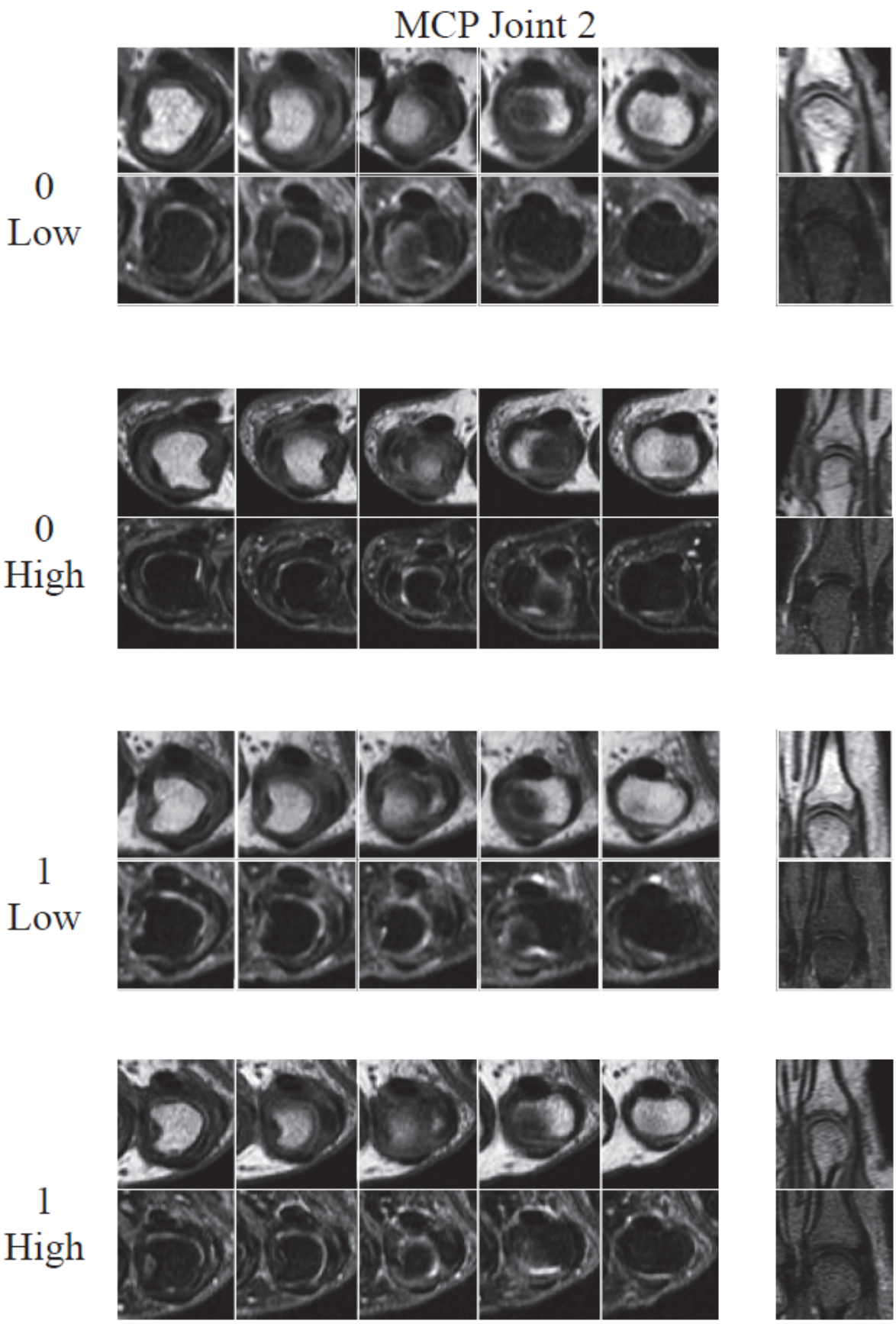
The two following radiography systems were used to acquire the x-ray images for Chapters 5 and 6. The first generator was a Canon CXDI, CTM Sedecal A-132 with a voltage of 60-4 kV and current of 100 mA. PA and oblique projections of the hands and wrists used 60 kV, and feet used 64 kV. All hand and wrist acquisitions, and PA and oblique projections output 3.2 mA at 32 ms; lateral feet projections output 5 mA at 50 ms. The second generator was a Villa Apollo G100 RFA, CTM Dunlee DU303 using Konica Nano Regius Model 110 software. The current output ranged from 2.5-3.2 mA at 25-32 ms. The voltage of the unit was 55-60 kV with a current of 100 mA. For both

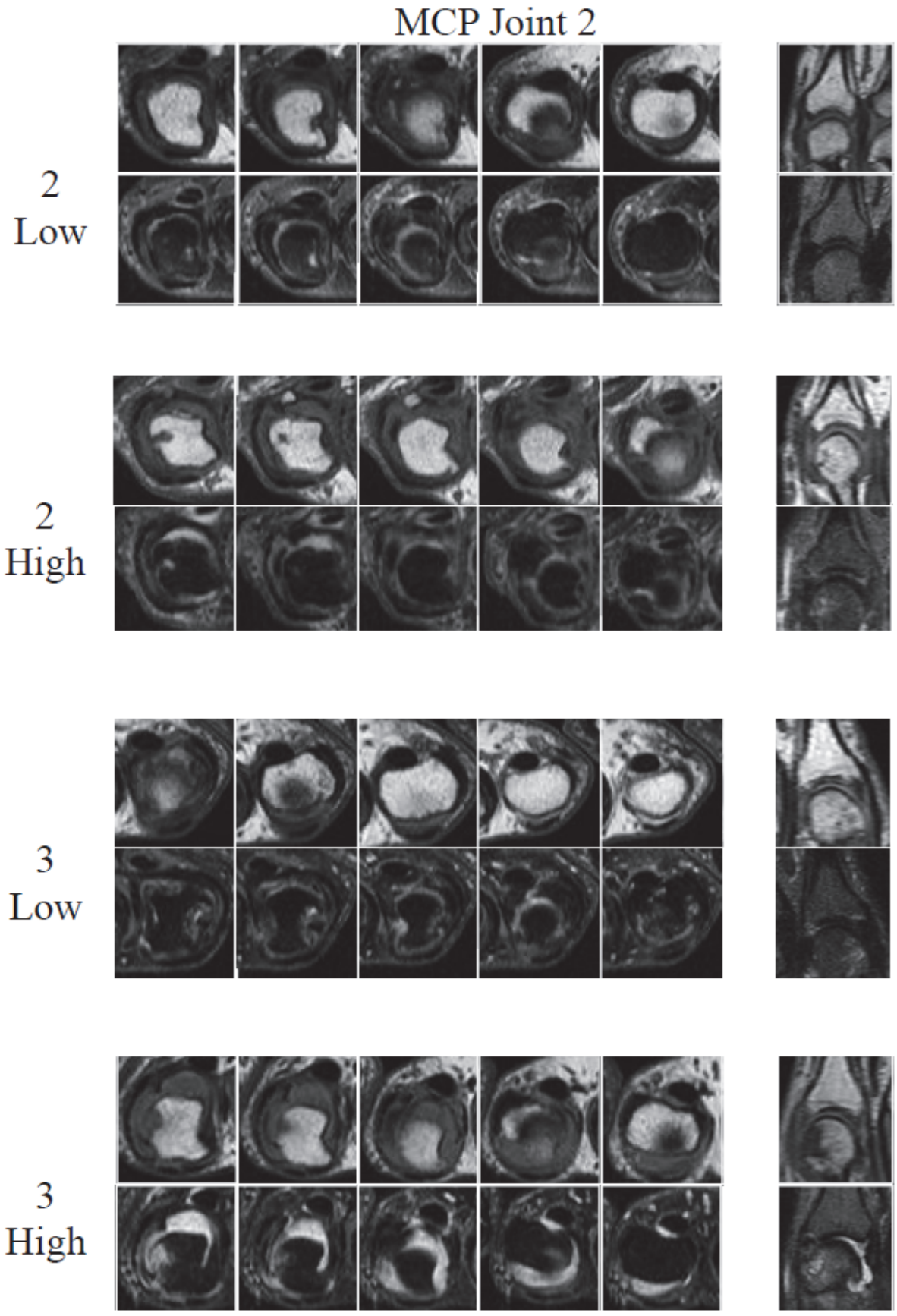
systems, the focus film distance (FFD) was 40” (approx. 100 cm) and had a focal spot of 0.6mm.

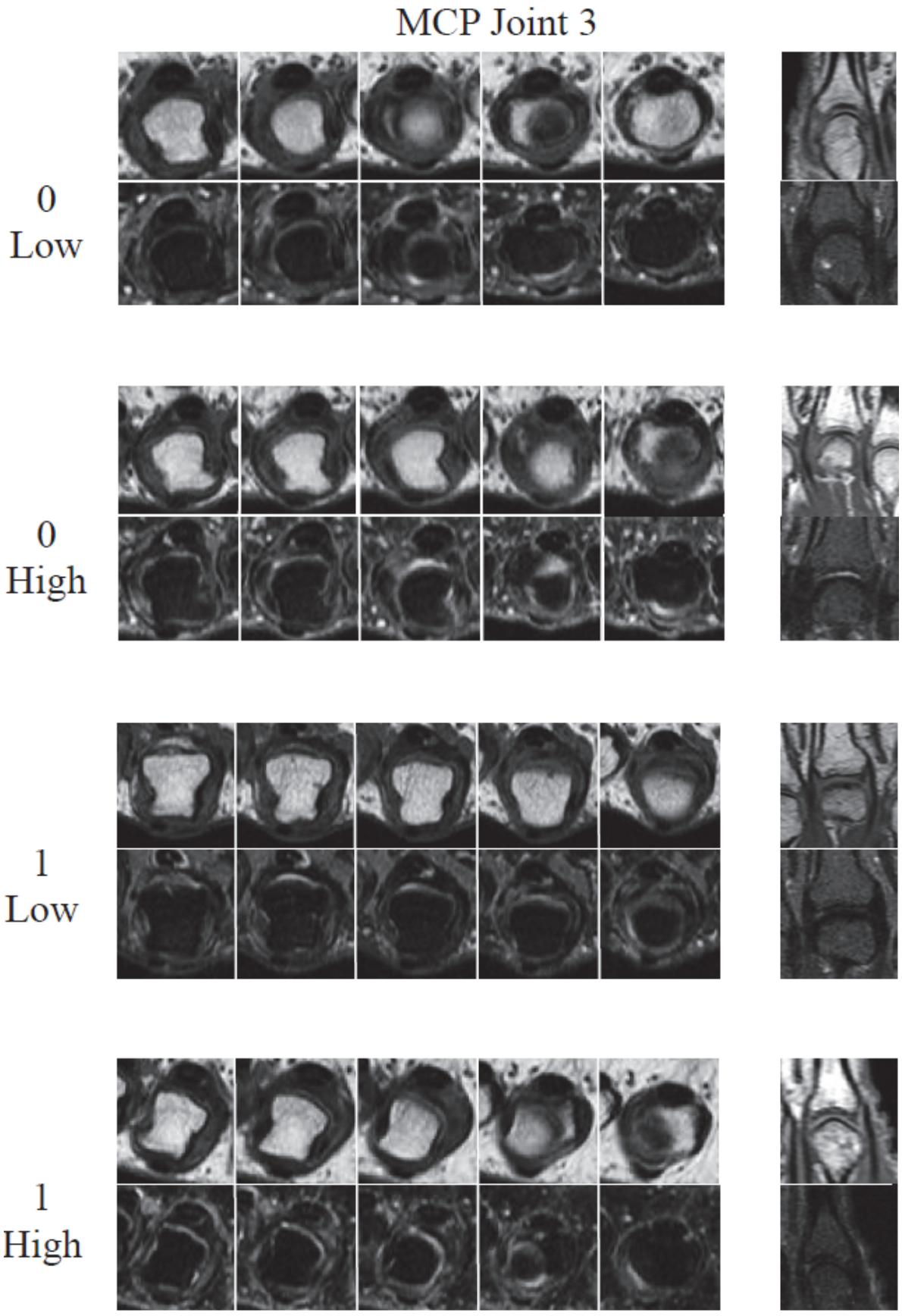
12. Appendix D: Contrast-independent MRI synovitis atlas

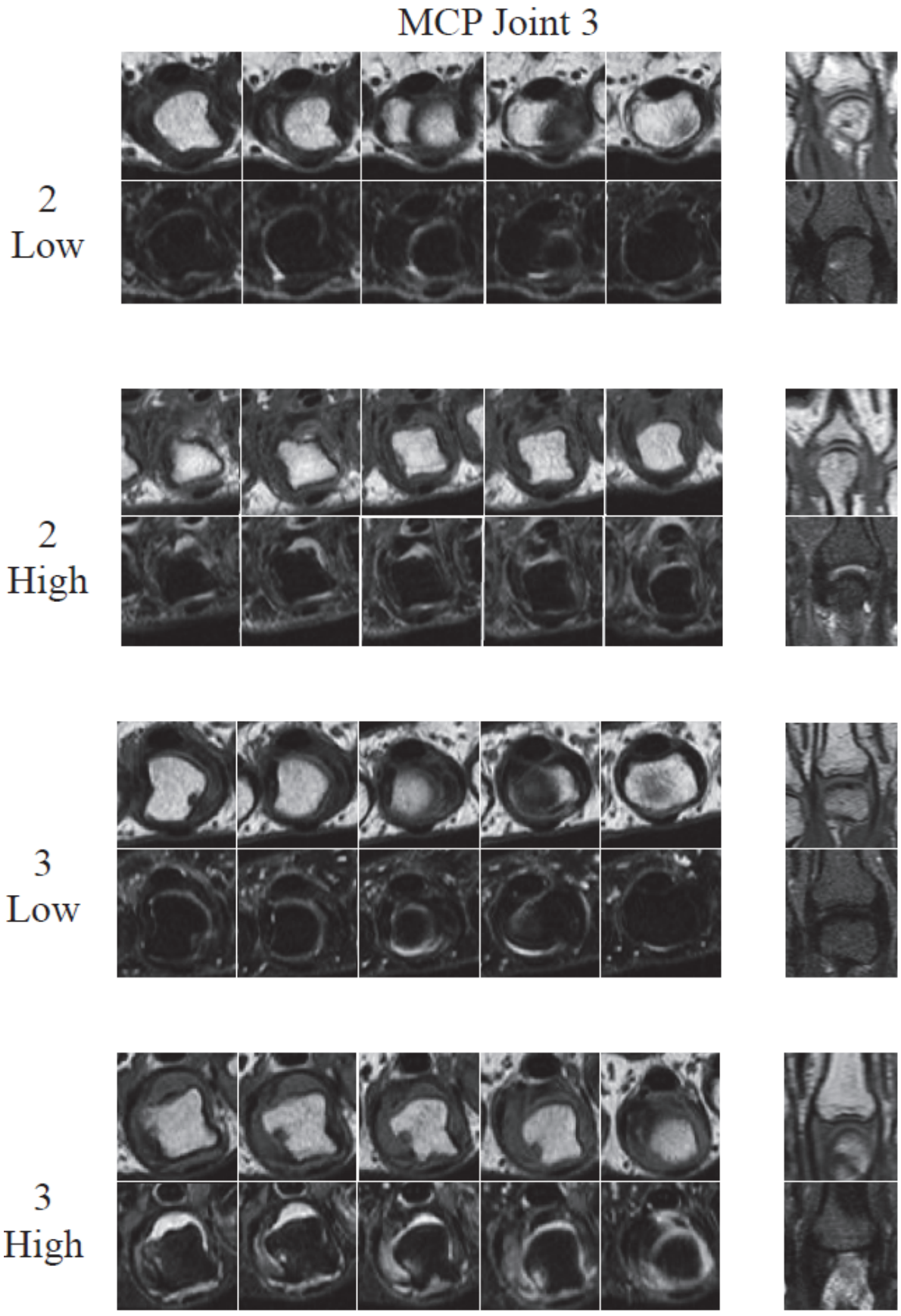
Synovitis grades 0-3 are presented with examples from the high and low end for scores for 1 and 2. Score 0 is presented only with the high end, as anything below it can be considered a 0, and score 3 is presented only with the low end, as anything above it can be considered a 3.

Grade						
	Slice 1 Axial T1w -FS	Slice 2 Axial T1w -FS	Slice 3 Axial T1w -FS	Slice 4 Axial T1w -FS	Slice 5 Axial T1w -FS	Coronal 3DGE-FS
	Slice 1 Axial T2w +FS	Slice 2 Axial T2w +FS	Slice 3 Axial T2w +FS	Slice 4 Axial T2w +FS	Slice 5 Axial T2w +FS	Coronal STIR


MCP Joint 2
0
Low
0
High
1
Low
1
High


MCP Joint 2
2
Low
2
High
3
Low
3
High


MCP Joint 3
0
Low
0
High
1
Low
1
High


MCP Joint 3
2
Low
2
High
3
Low
3
High

13. Appendix E: Contrast-independent MRI synovitis atlas grant notice

CANADIAN ARTHRITIS NETWORK | LE RÉSEAU CANADIEN DE L'ARTHRITE

November 22, 2011

Dr. Jonathan Adachi
McMaster University
501-25 Charlton Avenue East
Hamilton, ON L8N 1Y2

RE: Canadian Arthritis Network 2011 Rapid Impact Platform Program

Dear Dr. Adachi,

Thank you again your submission to the Canadian Arthritis Network (CAN) 2011 Rapid Impact Platform Program (RIPP). We are very pleased to announce that your project entitled "*A non-invasive alternative for evaluating rheumatoid arthritis inflammation on MRI: An OMERACT RAMRIS supplement*" has been funded at $60,000 and your project has been assigned the following project code: 11-01-RIPP-01-Adachi. A list of all funded grants will be posted on CAN's web site shortly.

Given the strategic nature of the RIPP program, all funded initiatives will also be assigned a Scientific Director and CAN Office representative to assist with the development and implementation of the proposed platform. Please include these individuals in your planning and decision-making activities for the platform. Your project has been assigned Dr. Claire Bombardier and Mr. Johnathan Riley.

Please complete the enclosed Fund Distribution form, and return it to the CAN office as soon as possible so that we can process payment of your award. The form must be signed by yourself (as PI) and all Co-PIs and faxed to (416) 586-8395 to the attention of Stefanie Cara or emailed to scara@mtsinai.on.ca.

All Investigators receiving funds must complete the enclosed Grant/Award Administrator Information form and return it by fax to (416) 586-8395 to the attention of Stefanie Cara or emailed to scara@mtsinai.on.ca. Please be advised that payments will be made quarterly (beginning soon after the return of the signed Fund Distribution Form) and sent to the Grant Administrator at the PI's/Co-PIs' institutions. A copy of this payment letter will also be sent to your attention.

Please be aware that this award may be modified should funding for CAN be modified or terminated by the Network of Centres of Excellence.

We would like to remind you that copies of all publications, abstracts and presentations arising from CAN-funded projects should be forwarded to Ms. Stefanie Cara, Manager of Research and Training Programs (scara@mtsinai.on.ca) at least 40 days prior to submission. We do not need final drafts. A first draft is sufficient to inform the CAN office of

522 University Avenue, Suite 1002 • Toronto, Ontario M5G 1W7 • Canada | 522, avenue University, bureau 1002 • Toronto (Ontario) M5G 1W7 • Canada
telephone/telephone 416-586-4770 **fax/télécopieur** 416-586-8395 **www.arthritisnetwork.ca**

the subject matter of the proposed disclosure and will not delay the publication process. In the event that any intellectual property protection needs to be filed on the disclosure, CAN will work with you and your institution to ensure that your ability to publish is not delayed.

In addition, please remember that, as a condition of Network funding, CAN must be acknowledged in all publications, presentations, abstracts and posters reporting the results of your CAN-funded project. Should you need a copy of the CAN logo for presentation purposes, please download one from the members' area of the CAN web site (www.arthritisnetwork.ca).

If you have any questions or concerns regarding the outcome of this review process, please forward these in writing to Ms. Stefanie Cara by email at scara@mtsinai.on.ca or by mail to:

522 University Avenue, Suite 1002
Toronto, Ontario M5G 1W7
Canada

Kindest regards,
CANADIAN ARTHRITIS NETWORK

Stefanie Cara
Manager of Research and Training Programs, Canadian Arthritis Network
Encl.